

Mirostław Koziarski

**RETRODIGITALIZACJA
SŁOWNIKA WARSZAWSKIEGO.
CELE, METODY, NARZĘDZIA**

Praca doktorska napisana
na Wydziale Neofilologii
Uniwersytetu im. Adama Mickiewicza w Poznaniu
pod kierunkiem prof. Mirosława Górnego

Poznań, 2022

Spis treści

Streszczenie	10
Abstract	11
Podziękowania	12
Motto	13
0. Wstęp	15
0.1. Cele pracy.....	15
0.2. Hipotezy badawcze.....	16
0.2.1. Hipoteza I – technologia a postać słownika.....	16
0.2.2. Hipoteza II – technologia a funkcje słowników.....	16
0.3. Układ pracy.....	16
1. Stan badań	19
1.1. Zakres dziedzinowy.....	19
1.1.1. Językoznawstwo.....	19
1.1.2. Leksykografia.....	21
1.1.3. Humanistyka cyfrowa.....	26
1.1.4. Nauka o informacji.....	30
1.2. Przegląd literatury.....	31
1.3. Luka badawcza.....	32
2. Cechy słownika	34
2.1. Cechy słownika a cechy języka.....	34
2.2. Cechy słownika a cechy nadawcy.....	36
2.3. Cechy słownika a cechy odbiorcy.....	37
2.4. Cechy słownika a cechy medium.....	40
2.5. Cechy słownika a jego funkcje.....	41
3. Słownik warszawski – charakterystyka	48
3.1. Autorzy.....	48
3.1.1. Jan Aleksander Karłowicz.....	49
3.1.2. Adam Antoni Kryński.....	50
3.1.3. Władysław Marcin Niedźwiedzki.....	51
3.1.4. Pozostali.....	51
3.2. Materiały pomocnicze.....	53
3.2.1. Trzy piramidy.....	53
3.2.2. Prace dotyczące słownika.....	53

3.2.3.	Korespondencja	55
3.3.	Założenia	55
3.4.	Powstawanie	62
3.5.	Suplement.....	63
3.6.	Materiał źródłowy	64
3.7.	Analiza treści.....	65
3.8.	Recepcja.....	69
3.9.	Wartość SW	73
4.	Struktura SW.....	78
4.1.	Megastruktura SW.....	78
4.2.	Makrostruktura SW	79
4.3.	Mikrostruktura SW	82
4.3.1.	Metodologia analizy artykułu hasłowego	82
4.3.2.	Segmentacja artykułu hasłowego.....	83
4.3.3.	Segment inicjalny	86
4.3.4.	Segment centralny	95
4.3.5.	Segment terminalny.....	123
4.3.6.	Hasła zagnieżdżone	126
5.	Skrócenia.....	128
5.1.	Różnice pomiędzy tomami	128
5.2.	Klasyfikacja skrótów	130
5.2.1.	Terminologia	131
5.2.2.	Poziomy analizy skrótów.....	131
5.2.3.	Kategoryzacja funkcjonalna.....	132
5.3.	Uzupełnienie źródeł	152
5.4.	Reguły tworzenia skrótów.....	156
5.4.1.	Skrócenia ortograficzne i ortofonetyczne	156
5.4.2.	Dyferencjacja nazw autorów.....	157
5.4.3.	Pozostałe osobliwości	159
5.5.	Standaryzacja zapisu.....	160
5.5.1.	Uzgodnienie treści <i>Spisu</i> i haseł.....	161
5.5.2.	Dezambiguacja skrótów.....	162
5.5.3.	Warianty skrótów	162
5.5.4.	Zapis treści rozwinięć.....	163
5.5.5.	Inne uwagi	165
6.	Metodologia retrodigitalizacyjna	166
6.1.	Przegląd podejść	166

6.2.	Przyjęta metodologia	168
6.2.1.	Wybór źródła	169
6.2.2.	Analiza źródła	171
6.2.3.	Wybór narzędzi	171
6.2.4.	Pozyskanie kopii cyfrowej	172
6.2.5.	Przygotowanie do digitalizacji	172
6.2.6.	Optyczne rozpoznanie tekstu (OCR).....	173
6.2.7.	Korekta rezultatów OCR	173
6.2.8.	Format tekstu cyfrowego	174
6.2.9.	Analiza składniowa (parsowanie).....	174
6.2.10.	Kontrola danych wyjściowych	175
6.2.11.	Indeksowanie	175
6.2.12.	Generowanie danych pochodnych	175
6.2.13.	Prezentacja i interpretacja danych	176
6.2.14.	Archiwizacja i utylizacja	176
7.	Narzędzia	178
7.1.	Środowisko programistyczne	178
7.1.1.	Języki programowania Python i Ruby	178
7.2.	Standard Unikonod (Unicode).....	179
7.3.	Język znaczników XML.....	182
7.4.	Technologia DjVu	184
7.5.	Oprogramowanie i język LaTeX.....	185
7.6.	Graficzna obróbka haseł.....	185
7.7.	Parsowanie struktury hasłowej.....	186
7.8.	Optyczne rozpoznawanie znaków.....	186
7.9.	Metoda wprowadzania danych.....	187
7.10.	Edycja tekstu.....	190
7.11.	Porównywanie tekstu	190
7.12.	Wyrażenia regularne	191
8.	Ekstrakcja kart hasłowych.....	193
8.1.	Wprowadzenie	193
8.2.	Cyfrowe przetwarzanie obrazów.....	194
8.3.	Symulacja metody manualnej	196
8.4.	Wybór metody.....	198
8.4.1.	Koszt wdrożenia	198
8.4.2.	Nakład na przetworzenie pojedynczej strony.....	198
8.4.3.	Podatność na błędy.....	199

8.4.4.	Skalowalność	199
8.4.5.	Jakość obrazu wyjściowego	199
8.4.6.	Wnioski	200
8.5.	Implementacja metody automatycznej	200
8.5.1.	Strategia programistyczna	200
8.5.2.	Wybór źródła	201
8.5.3.	Środowisko programistyczne	202
8.5.4.	Dobre praktyki	203
8.6.	Działanie programu IndexCardSplitter	204
8.6.1.	Etap pierwszy	204
8.6.2.	Etap drugi	204
8.6.3.	Etap trzeci	205
8.6.4.	Etap czwarty	213
8.6.5.	Etap piąty	217
8.6.6.	Etap szósty	219
8.6.7.	Znane problemy	223
8.6.8.	Wydajność	224
8.7.	Rezultaty	225
8.8.	Alternatywne rozwiązania	226
9.	Konwersja warstwy graficznej na tekstową	227
9.1.	Przebieg procesu OCR	228
9.1.1.	Przetwarzanie wstępne	228
9.1.2.	Rozpoznawanie znaków	229
9.1.3.	Przetwarzanie końcowe	229
9.2.	Rozpoznawanie tekstu a SW	229
10.	Przetwarzanie danych SW	232
10.1.	Standardy w dziedzinie	234
10.2.	Dane wejściowe	235
10.2.1.	Dane podstawowe	236
10.2.2.	Dane pomocnicze	237
10.2.3.	Dane predefiniowane	239
10.3.	Analiza składniowa	242
10.3.1.	Podział na litery	242
10.3.2.	Wczytanie i korekcja danych	243
10.3.3.	Deklaracje wyrażeń regularnych	243
10.3.4.	Wykorzystanie skrótów	247
10.3.5.	Supersegmentacja artykułu hasłowego	248

10.3.6.	Parsowanie segmentu inicjalnego	249
10.3.7.	Parsowanie segmentu centralnego	252
10.3.8.	Parsowanie segmentu terminalnego.....	256
10.3.9.	Wykorzystanie znaków pomocniczych.....	257
10.3.10.	Zagregowane związki wyrazowe	257
10.4.	Generowanie danych pochodnych.....	259
10.4.1.	Indeks artykułów hasłowych	259
10.4.2.	Indeks stałych połączeń wyrazowych	260
10.4.3.	Indeks form wyrazowych	260
10.4.4.	Korpus zdań i wyrażenia przykładowych	260
10.4.5.	Wykaz znaków w systemie Unikod.....	261
10.4.6.	Fotokorpus haseł	261
10.4.7.	Informacja gramatyczna	261
10.4.8.	Słowniczek walencyjny.....	262
10.4.9.	Słowniczki przekładowe	262
10.4.10.	Matryca siatek hasłowych SW i wybranych słowników	262
10.4.11.	Lista podmian w treści.....	264
10.4.12.	Skonsolidowany spis skrótów.....	265
10.4.13.	Dane liczbowe	265
10.4.14.	Słowniczek par derywacyjnych.....	265
10.4.15.	Słowotwórstwo rzeczowników żeńskich (feminatywów)	266
10.5.	Dane wyjściowe.....	266
10.5.1.	Format i prezentacja danych.....	266
10.5.2.	Wyświetlanie znaków	266
10.6.	Podsumowanie procesu.....	268
10.7.	Niezaimplementowane moduły.....	271
10.7.1.	Klasyfikacja informacji etymologicznej.....	271
10.7.2.	Generowanie tabel fleksyjnych	271
10.7.3.	Słownik sprawdzania pisowni	271
10.7.4.	Słownik gniazd derywacyjnych	272
10.7.5.	Lista słoworodów	272
10.7.6.	Mapa relacji semantycznych.....	272
10.7.7.	Detekcja martwych odniesień.....	272
10.7.8.	Makieta SW on-line	272
10.8.	Potencjalne usprawnienia.....	273
10.8.1.	Ekstrakcja haseł ukrytych.....	273
10.8.2.	Detekcja glos łacińskich	273

10.8.3.	Detekcja zakresu w wyrażeniach	274
10.8.4.	Kwalifikatory opcjonalne	274
10.8.5.	Pozostałe osobliwości hasłowe.....	274
10.8.6.	Usprawnione parsowanie mikrostruktury.....	275
10.8.7.	Rezygnacja z podziału na litery	275
10.8.8.	Walidacja i DTD.....	276
11.	Ewolucja słowników polszczyzny	277
11.1.1.	Okres przedstownikowy	279
11.1.2.	Wybrane dawniejsze słowniki drukowane	279
11.1.3.	Słowniki narodowe przed SW	282
11.1.4.	Słowniki narodowe po SW	284
11.1.5.	Słowniki nowego milenium	286
11.1.6.	Miejsce SW wśród słowników polszczyzny	289
12.	Praktyczna aplikacja wersji cyfrowej.....	293
12.1.	Słownik elektroniczny	293
12.1.1.	Układ i prezentacja	296
12.1.2.	Nowe ścieżki dostępu.....	299
12.1.3.	Hipertekstowość i integracja zasobów	300
12.1.4.	Słownik jako korpus	302
12.2.	Przetwarzanie komputerowe	303
12.2.1.	Słownik sprawdzania pisowni i OCR	305
12.2.2.	Autosuplement.....	306
12.2.3.	Analiza składniowa i morfologiczna.....	307
12.2.4.	Klasyfikacja maszynowa	309
12.3.	Wartość badawcza.....	309
12.3.1.	Porównanie siatek hasłowych.....	310
12.3.2.	Test hipotezy lingwochronologiczycznej.....	311
12.3.3.	Badania metaleksykograficzne	313
12.3.4.	Pozostałe obszary	314
13.	Falsyfikacja tez badawczych.....	315
13.1.	Weryfikacja hipotezy I	317
13.2.	Weryfikacja hipotezy II.....	318
13.3.	Podsumowanie	320
14.	Prezentacja danych	321
14.1.	Aneks A: Indeks artykułów hasłowych.....	321
14.2.	Aneks B: Indeks stałych połączeń wyrazowych.....	322
14.3.	Aneks C: Indeks form wyrazowych.....	322

14.4. Aneks D: Korpus zdań i wyrażen przykładowych	323
14.5. Aneks E: Wykaz znaków w systemie <i>Unikod</i>	323
14.6. Aneks F: Fotokorpus haset <i>SW</i>	323
14.7. Aneks G: Słowniczek gramatyczny	324
14.8. Aneks H: Klasyfikacja paradygmatów gramatycznych.....	325
14.9. Aneks I: Słowniczek walencyjny.....	325
14.10. Aneks J: Słowniczeki przekładowe	325
14.11. Aneks K: Matryca siatek hastowych <i>SW</i> i wybranych słowników	326
14.12. Aneks L: Lista podmian w treści.....	326
14.13. Aneks M: Skonsolidowany spis skrótów.....	327
14.14. Aneks N: Dane liczbowe	327
14.15. Aneks O: Słowniczek par derywacyjnych.....	328
14.16. Aneks P: Słowotwórstwo rzeczowników żeńskich (feminatywów).....	328
15. Opis pracy	329
Bibliografia	331
Literatura cytowana	332
Literatura uzupełniająca.....	354
Wykaz użytych skrótów.....	376
Wykaz tabel.....	380
Wykaz listingów	381
Wykaz ilustracji	382
Aneksy.....	384
Aneks A-1: Indeks a fronte artykułów hastowych	385
Aneks A-2: Indeks a tergo artykułów hastowych	420
Aneks B: Indeks stałych połączeń wyrazowych.....	455
Aneks C-1: Indeks a fronte grafoform wyrazowych	463
Aneks C-2: Indeks a tergo grafoform wyrazowych	667
Aneks D: Korpus zdań i wyrażen przykładowych	871
Aneks E: Wykaz znaków w systemie <i>Unikod</i>	962
Aneks F: Fotokorpus haset <i>SW</i>	969
Aneks G: Słowniczek gramatyczny	1791
Aneks H: Klasyfikacja paradygmatów gramatycznych	2011
Aneks I: Słowniczek walencyjny	2020
Aneks J: Słowniczeki przekładowe	2022
Aneks K: Matryca siatek hastowych <i>SW</i> i wybranych słowników	2081
Aneks L: Lista podmian w treści <i>SW</i>	3381
Aneks M: Skonsolidowany spis skrótów <i>SW</i>	3412

Aneks N: Dane liczbowe <i>SW</i>	3634
Aneks O: Słowniczek par derywacyjnych.....	3691
Aneks P: Słowotwórstwo rzeczowników żeńskich (feminatywów).....	3698

Streszczenie

Słownikowi języka polskiego autorstwa Karłowicza, Kryńskiego i Niedźwiedzkiego, znanego także jako *Słownik warszawski*, poświęcono do tej pory stosunkowo niewiele uwagi, mimo że należy on do wielkich dzieł leksykograficznych polszczyzny. Jednocześnie jest on – ze względu na rejestrowanie jednostek z różnych rejestrów i epok języka polskiego – kluczowym zasobem dla badań diachronicznych, który jednak do tej pory nie został w pełni wykorzystany. W ramach niniejszej tezy – umiejscowionej na skrzyżowaniu lingwistyki stosowanej, leksykografii, humanistyki cyfrowej oraz nauki o informacji – zostanie podjęta próba wyekstrahowania danych zawartych w *Słowniku...*, co posłuży także jako punkt wyjścia do uogólnienia dotychczasowych trendów w leksykografii cyfrowej i zaproponowania nowej metodyki w tym obszarze.

W tym celu najpierw od strony teoretycznej, w oparciu o literaturę jak i analizę własną, omówiona zostanie ewolucja postaci słowników i kolejnych etapów zmiany ich postaci fizycznej i organizacji, następnie opisany zostanie proces ucyfrowienia fragmentu *Słownika warszawskiego* przeprowadzony na potrzeby niniejszej tezy, aż wreszcie zostanie podjęta próba podtrzymania tez badawczych, które brzmią: (1) Rozwój technologii jest głównym czynnikiem zmian postaci fizycznej i organizacji słowników, oraz (2) Pojawienie się nowych funkcji słowników jest pochodną zmian technologicznych.

Powyższe tezy zostaną poddane testowi empirycznemu, w ramach którego fragment oryginalnej postaci materiału badawczego zostanie skontrastowany z jego cyfrową reprezentacją wygenerowaną przez autora, co pozwoli zweryfikować, czy i w jakim stopniu digitalizacja wpływa na sposoby prezentacji danych leksykograficznych, oraz czy poszerza ona zakres funkcji, jakie dane te mogą pełnić.

Na końcu zaprezentowane zostaną wyniki procesu ucyfrowienia materiału badawczego i przedstawione zostaną uogólnione kierunki zmian postaci i funkcji słowników cyfrowych, a także zostanie podjęta próba ekstrapolacji przyszłych przemian w obszarze leksykografii elektronicznej.

Abstract

The *Dictionary of the Polish Language* by Karłowicz, Kryński and Niedźwiedzki, also known as the *Varsovian Dictionary*, has received relatively little attention so far, despite the fact that it belongs to the great lexicographic works of the Polish language. At the same time, it is—due to the registration of items from different registers and epochs of Polish—a key resource for diachronic research, of which, however, the full potential has not been used so far. As part of this thesis—located at the intersection of applied linguistics, lexicography, digital humanities and information science—an attempt will be made to extract the data contained in the *Dictionary...*, which will also serve as a point of departure to generalize the current trends in digital lexicography and propose a new methodology in this area.

For this purpose, first from the theoretical side, based on the literature and own analysis, the evolution of the form of dictionaries and subsequent stages of changing their physical form and organization will be discussed, then the process of digitization of a fragment of the *Varsovian Dictionary* carried out for the purposes of this thesis will be described, and finally, an attempt will be made to maintain the research theses, which state that: (1) The development of technology is the main factor in changes in the physical form and organization of dictionaries, and (2) The emergence of new functions of dictionaries is a derivative of technological changes.

The above-mentioned theses will be subjected to an empirical test, which entails contrasting a fragment of the original form of the research material with its digital representation generated by the author in order to verify whether and to what extent digitization influences the methods of presenting lexicographic data and whether it extends its the range of functions that these data can perform.

Finally, the results of the process of digitization of the research material will be presented and generalized trends of changes in the form and functions of digital dictionaries will be presented, and an attempt will be made to extrapolate future changes in the field of electronic lexicography.

Podziękowania

Autor pragnie na wstępie podziękować wszystkim osobom, które miały swój nieodzowny wkład w powstanie niniejszej tezy doktorskiej, czyli przede wszystkim:

- Prof. **Mirosławowi Górnemu** – za olbrzymi wkład merytoryczny i pomoc w uformowaniu ostatecznego kształtu pracy;
- Prof. **Piotrowi Wierzchoniowi** – za wszechstronne pokierowanie mojego rozwoju naukowego i wkład merytoryczny w pracę doktorską;
- Dr. **Robertowi Wołoszowi** – za pomoc w zdefiniowaniu tematu tezy doktorskiej i udostępnione materiały;
- Mgr. inż. **Pawłowi Zakrzewskiemu** – za całokształt wieloletniego wsparcia technicznego w rozwiązywaniu zawłości programistycznych;
- Dr. **Magdalenie Majdak** – za udostępnione materiały;
- Dr. **Ewie Ulitzkiej** – za udostępnione materiały;
- **Najbliższym**, a zwłaszcza **Ukochanej** – za wsparcie w trudnym i długim okresie powstawania tejże pracy.

Motto

W 1632 roku Joseph Justinian Scalinger wydał *Thesaurus linguae arabicae*, w którym zamieścił wiersz zatytułowany „In Lexicorum compilatores, inscriptum Lexico Arabico a se collecto, in Batavis”, poświęcony leksykografom. Wersja oryginalna (łacińska) brzmi następująco:

*Si quem dura manet sententia indicis olim,
Damnatum aerumnis suppliciisque caput:
Hunc neque fabrili lassent ergastula massa,
Nec rigidas vexent fossa metalla manus.
Lexica contexit, nam caetera quid moror? omnes
Poenarum facies hic labor unus habet.*

Powyższy tekst został przetłumaczony na język angielski przez Barry’ego Baldwina¹ na dwa sposoby: dosłowny i poetycki, które przytoczone są poniżej:

Tłumaczenie dosłowne:

*If the harsh sentence of the judge in the past pursues anyone,
His head condemned to miseries and sufferings,
May the prisons not weary him with heavy manual toil
Nor the digging of minerals pain his stiffened hands
Let him start a Dictionary: why linger amongst the rest?
This labour contains all forms of punishments in one*

Tłumaczenie poetyckie:

*If harassed by the former sentence of a judge,
Condemned to miseries & sufferings like a drudge,
May prisons not weary him with manual toil,
Let not mining his work-stiffened hands besoil.
A Dictionary then? Let other things be gone
This job contains all punishments in one*

¹ *The Latin & Greek poems of Samuel Johnson: text, translation, and commentary*, London, Duckworth, 1995.

Słowa te zgrabnie ilustrują trud i złożoność pracy słownikarza; piszący te słowa, choć bada słowniki miał je redagować, odnosi treść wiersza w pewnym stopniu także do siebie, dlatego też pozwolił sobie podjąć próbę przetłumaczenia² ich na język polski:

Tłumaczenie na język polski:

*Gdy sędziego wyroki człowieka pogrązą,
A ciemięga, trud, boleść nad głową mu ciężką
Miejsce jego nie w więziennej pracy pocie,
Ni to w kopalni, by grzebał w ziemi i błocie.
Niechże słownik napisze! Znajdź to tak głęboki,
Że mieści w sobie wszelkie możliwe wyroki.*

² Autor pragnie tutaj podziękować Urszuli Dwojak za pomoc w opracowaniu ostatecznej wersji tłumaczenia.

0. Wstęp

Spośród wszystkich słowników okresu, z jakimi piszący te słowa miał styczność, jest jedno, które wyłania się jako unikatowe na wielu płaszczyznach – który odznacza się zarówno wyjątkowym bogactwem treści, złożonością artykułów hasłowych, ciekawą historią, jak i nowatorskimi (względem epoki) rozwiązań leksykograficznych.

Dziełu, jakim jest tzw. *Słownik warszawski*, czyli *Słownik języka polskiego* Karłowicza, Kryńskiego i Niedźwiedzkiego (zwany dalej *SW*), nie poświęcono wystarczającej uwagi, bo choć pojawiły się liczne publikacje na jego temat, to pozostaje on najsłabiej opracowanym spośród wielkich przedwojennych słowników polszczyzny. Dlatego też drugim filarem pracy będzie analiza formalna dzieła, która następnie posłuży do stworzenia makiety rozwiązania programistycznego, które pozwoli na opracowanie fragmentu dzieła z założeniem, że będzie ono mogło być wykorzystane do opracowania ogółu treści słownika.

Oba te filary – ramy teoretyczne i egzemplifikacja przy pomocy autorskiego rozwiązania inżynierskiego – posłużą do uzyskania pewnej abstrakcji, a konkretnie teoretycznej relacji między technologią, formą, funkcją a treścią słownika. Choć relacje te są badane od kilku dekad, to uwydatnienie skali ich wzajemnego wpływu było dotychczas niedostateczne – wykorzystanie metod automatycznych, jak i zakres egzemplifikacji zalet formy cyfrowej uważa się za największe novum pracy; jej wartością poznawczą będzie także skonstruowanie ram teoretycznych do wyznaczenia granicy treści i formy w słownikach.

Teza ta pozostaje jednak osadzona w ramach lingwistyki stosowanej i komputerowej, ponieważ aspekt programistyczny przedstawionych przez autora tych słów, choć prominentny, jest silnie oparty o teorie językoznawcze i jest narzędziem, a nie przedmiotem prowadzonych tu badań.

0.1. Cele pracy

Głównym celem pracy jest ustalenie, opisanie i wyjaśnienie wybranych rezultatów retrodigitalizacji w dziedzinie leksykografii.

Aby tego dokonać, zdecydowano się na samodzielne dokonanie pełnego procesu retrodigitalizacji fragmentu konkretnego dzieła słownikowego. Wybór padł na *Słownik warszawski*.

Digitalizacji poddano tylko fragment słownika z uwagi na skalę przedsięwzięcia. Digitalizacja całego słownika przekroczyłaby możliwości techniczne i finansowe doktoranta, a równocześnie jej rezultaty w kontekście głównego celu pracy nie odbiegałyby istotnie od rezultatów digitalizacji części słownika, którą można określić swoisty model pozwalający na symulację ucyfrowienia całego dzieła.

Zatem takie podejście do pracy narzuciło pewną grupę celów podrzędnych. Były to:

1. Wybór i opracowanie (fragmentu) danych zawartych w *SW* – pod kątem ich reprezentatywności.
2. Opracowania wyspecjalizowanych narzędzi do retrodigitalizacji – ze względu na brak ogólnodostępnych rozwiązań, które łatwo można by wykorzystać.

3. Wykazanie, że poszczególne etapy procesu mogą zostać zautomatyzowane – zapewniając tym samym większą wydajność przetwarzania.
4. Wykazanie, że proponowane rozwiązanie jest skalowalne, co oznacza że może być wykorzystane do retrodigitalizacji całego słownika.

Realizacja powyższych celów pozwoliła podjąć próbę realizacji celów głównych:

1. Ustalenie efektów, jakie daje wykorzystanie digitalizacji w leksykografii.
2. Opisanie tychże efektów.
3. Wyjaśnienie mechanizmu uzyskiwania wspomnianych efektów.

0.2. Hipotezy badawcze

0.2.1. Hipoteza I – technologia a postać słownika

Pierwsza hipoteza postawiona w niniejszej pracy brzmi: **Rozwój technologii jest głównym czynnikiem zmian postaci fizycznej i organizacji słowników.** Rozpatrywana będzie relacja między rodzajem danych a sposobem ich prezentacji. Eksplorowane będą także sposoby parametryzacji danych słownikowych (takich jak: hipertekstowość, sposób organizacji danych, struktura dostępu, kompresja danych, możliwość przetwarzania automatycznego, wyłapywanie błędów), a także potencjalne sposoby ich prezentacji.

Inne czynniki wpływające na formę leksykonów, takie jak komercjalizacja, polityka edukacyjna, przemiany społeczne itp. potraktowano jako poboczne. Założenie to jest zgodne z treścią powyższej hipotezy.

0.2.2. Hipoteza II – technologia a funkcje słowników

Druga hipoteza postawiona w niniejszej pracy brzmi: **Pojawienie się nowych funkcji słowników jest pochodną zmian technologicznych.**

Pod uwagę brano również hipotetyczne funkcje słowników. Podobnie jak w przypadku hipotezy I nie analizowano tu roli innych czynników wpływających na funkcje leksykonów. Do nich można zaliczyć: potrzeby i świadomość użytkowników, rodzaj zawartych danych i formę ich prezentacji, przekonania ideologiczne i wiedzę redaktorów oraz stan leksykografii i językoznawstwa. Powodem takiego podejścia była konieczność ograniczenia badań do sfery technicznej. Uwzględnienie wspomnianych czynników wiązałoby się z ogromnym nakładem prac badawczych.

0.3. Układ pracy

Niniejsza teza doktorska będzie miała budowę trójdzielną, co odzwierciedla strukturę procesu badawczego, który został przeprowadzony.

Część pierwsza poświęcona będzie ramom teoretycznym.

W pierwszej kolejności zostaną przedstawione podstawy badawcze, obejmujące językoznawstwo i jego dziedziny, leksykografię, humanistykę cyfrową i retrodigitalizację, w których osadzone będą badania.

W dalszej kolejności słowniki zostaną scharakteryzowane jako gatunek, z próbą ich zdefiniowania, klasyfikacji typologicznej, opisanie ich treści, struktury, cech oraz kryteriów opisu.

Następnie zostanie przedstawiony rys historyczny słowników aż do epoki współczesnej *Słownikowi warszawskiemu*, z charakterystyką wybranych dzieł oraz ogólnych tendencji na czterech płaszczyznach: leksykografii na świecie, leksykografii w Europie, leksykografii na Słowiańszczyźnie, i w końcu leksykografii w Polsce.

W kontraście do rysu historycznego zaprezentowane zostaną współczesne rozwiązania w dziedzinie leksykografii elektronicznej, potencjalne kierunki jej rozwoju, a także zastosowanie słowników w ramach językoznawstwa komputerowego.

Część druga poświęcona będzie *Słownikowi warszawskiemu* i procesowi jego ucyfrowienia.

Najpierw zostanie przedstawiona charakterystyka dzieła: jego geneza, autorzy, proces redakcji, recepcja oraz miejsce w historii leksykografii oraz narodu polskiego.

Następnie szczegółowo omówiona zostanie organizacja słownika: kryteria analizy, jego mega-, makro- i mikrostruktura oraz omówienie cech wyróżnionych jednostek na poszczególnych poziomach układu hierarchicznego.

Osobno potraktowane będą skrócenia, począwszy od rozbieżności między tomami, przez ich klasyfikację, proces ich uzupełnienia, próbę rekonstrukcji reguł ich tworzenia, aż po zaproponowanie standaryzacji zapisu.

Kolejno sformułowana zostanie metodologia retrodigitalizacyjna w oparciu o literaturę oraz wcześniej doświadczenia autora tych słów, ale ze szczególnym uwzględnieniem wyzwań związanych z charakterystyką *SW* i należytą adaptacją metod.

Dalej zostaną zaprezentowane narzędzia, które wykorzystane zostaną do opracowania cyfrowej wersji *Słownika warszawskiego* – uwzględniając programy komercyjne, programy o otwartym kodzie źródłowym, relewantne systemy i standardy w środowisku komputerowym, oraz rozwiązania opracowane na potrzeby niniejszej tezy.

Następnie przedstawiony będzie warsztat inżynieryjno-programistyczny pracy, egzemplifikujący metodologię retrodigitalizacyjną:

Pierwszym elementem inżynieryjnym będzie opis procesu ekstrakcji kart hasłowych z zeskanowanych stron *Słownika...* przy pomocy autorskiego rozwiązania programistycznego.

Drugim elementem inżynieryjnym będzie opis procesu konwersji warstwy graficznej na tekstową i wyzwania związane z uzyskaniem danych o wysokiej wierności względem oryginału.

Trzecim elementem inżynieryjnym będzie opis działania analizatora składniowego opracowanego specjalnie dla *SW* wraz z krokami, które muszą być powzięte względem poszczególnych elementów, a także metody przetwarzania danych pierwotnych w celu uzyskania danych wtórnych (pochodnych).

Część trzecia poświęcona będzie podsumowaniu rezultatów i na ich podstawie wyabstrahowaniu teoretycznej relacji między technologią, formą, funkcją a treścią słownika.

Najpierw zostanie zaproponowane podsumowanie ewolucji słowników polszczyzny oraz pochylenie się nad teorią i rezultatem podjętego w tej pracy eksperymentu polegającego na separacji formy od treści i próby wykorzystania nowopowstałych danych na nowe sposoby.

Następnie wymienione zostaną wybrane korzyści i zastosowania uzyskanej cyfrowej wersji treści *Słownika warszawskiego*: powstanie nowych narzędzi, wygenerowanie nowych rodzajów danych, a także ułatwienie dostępu do treści i wykorzystania jej do celów badawczych.

Na tej podstawie podjęta zostanie próba falsyfikacji hipotez badawczych i podsumowanie.

Wreszcie na zwieńczenie niniejszej pracy zostaną zaprezentowane rezultaty przetwarzania danych i formy ich prezentacji.

1. Stan badań

1.1. Zakres dziedzinowy

Obszar badawczy niniejszej pracy obejmuje cztery zasadnicze dziedziny nauki: (1) językoznawstwo (a dokładniej: lingwistykę stosowaną), (2) leksykografię, (3) humanistykę cyfrową, oraz (4) naukę o informacji (informatologię). Dziedziny te, choć fundamentalnie odrębne, są ze sobą w kontekście prowadzonych tu badań ściśle powiązane:

- **językoznawstwo i leksykografia** dotyczą układu słowników językowych – językoznawstwo w dużym stopniu opiera się o leksykografię, leksykografia czerpie natomiast z językoznawstwa metodologię;
- **leksykografia i humanistyka cyfrowa** dotyczą zarządzanie informacjami – zwłaszcza ich organizacją i archiwizacją w formie czytelnej dla człowieka;
- **humanistyka cyfrowa i nauka o informacji** dotyczą pozyskiwania, przechowywania i manipulacji danych w formie cyfrowej.

Poniżej przedstawiona jest ogólna charakterystyka wymienionych wyżej dyscyplin.

1.1.1. Językoznawstwo

Najprościej mówiąc, językoznawstwo to nauka o języku we wszelkich jego aspektach – aby zatem móc przybliżyć charakterystykę dziedziny, konieczne jest ustalenie, czym jest sam język.

Bañczerowski & Pogonowski & Zgółka (1982: 29) piszą, że „[j]ęzyk we wszystkich praktycznie ujęciach zarówno lingwistycznych, jak i logicznych, rozumiany jest jako system znaków.” Znaki te uporządkowane są zawsze w ramach pewnego systemu – o tym pisze Piotrowski (2010: 52):

System języka istnieje jedynie hipotetycznie: to zbiór hipotez, jakie badacze wysuwają, by wyjaśnić dostrzegane regularności występujące w badanych tekstach, każda zaś teoria językoznawcza to zapis tego rodzaju hipotez (...)

Jedną z głównych trudności w badaniu tego fenomenu jest fakt, że język „na równi z kulturą jest uważany za najważniejszy wytwór ludzkiej umysłowości” (Rudnicka 2016: 52; za: Anusiewicz 1994: 3). Oznacza to, że realizacja języka zależy od jednostki, ale system ten jest negocjowany między członkami pewnej zbiorowości. Język związany jest też ze stylem myślowym, co jest jego najważniejszym społecznie aspektem (Doroszewski 1966: 433), co negocjację tę utrudnia.

Co więcej, badanie języka jest nierozdzielnie powiązane z badaniem fragmentów świata pozajęzykowego – Doroszewski (1966: 5–6) zauważa, że choć świat idei (lub słów) i świat rzeczy (zmysłów) często bywają sobie przeciwstawiane, to w praktyce te pierwsze są znakami reprezentującymi fragmenty rzeczywistości, a więc z nimi nierozdzielnie związane. Čermák (2011: 53) myśl tę pogłębia pisząc:

Obecně ovšem platí, že mluvčí lépe pracuje s informací, která je jazykově kódovatelná, tj. pojmenovatelná jasně, jednoduše (psychologové tu mluví o kompaktnosti) a bezpříznakově než naopak.

Na ogól jednak mówiący lepiej pracuje z informacją, która jest możliwa do zakodowania w języku, tj. wyraźnie, prosto (psychologowie mówią o zwięzłości) i bezobjawowo niż na odwrót.

Doroszewski (1966: 364) pisze też o społecznym aspekcie języka:

[J]eżeli język jest – a jest to niesporne – jedną z form ludzkiego działania, to i praca nad językiem jest pracą nad tym, co czynią ludzie, nad pewnym typem faktów społecznych.”

Čermák (2011: 14) idzie dalej wskazując, jak istotna, a wręcz centralna dla społeczeństwa jest rola języka:

Postavení jazyka je pro fungování společnosti centrální a jedinečné a jazykem je podmíněna i hlavní část lidské civilizace a její kultury; stejně tak je jazyk nezbytný i pro většinu oborů lidské činnosti.

Pozycja języka jest centralna i unikalna dla funkcjonowania społeczeństwa, a główna część ludzkiej cywilizacji i jej kultury jest również uwarunkowana przez język; podobnie język jest niezbędny w większości dziedzin ludzkiej działalności.

Głównym powodem tak istotnej roli języka w funkcjonowaniu społecznym jest fakt, że stanowi on *de facto* medium, poprzez które zachodzi ciągła socjalizacja jednostki – proces, który trwa całe jej życie:

Pero tanto la objetivación como la reflexión acerca de la lengua [es] (...) resultado de un proceso de socialización que comienza con el nacimiento de cada ser humano y no se detiene nunca, durante el cual la lengua se enseña (y se aprende) primero, y después se educa o se cultiva. (Lara 1990: 134)

Jednak cel, jakim jest uprzedmiotowienie języka, jest (...) wynikiem procesu socjalizacji, który zaczyna się wraz z narodzinami każdego istnienia ludzkiego i nigdy się nie zatrzymuje, a podczas którego języka najpierw się naucza (i uczy), a później się w nim kształci i dba o niego.

Język jest często rozpatrywany w kontekście negocjacji pomiędzy światem zewnętrznym (obiektywnym) a światem wewnętrznym (subiektywnym), między jednostką a zbiorowością. Jest jednocześnie produktem kultury, ale i jej medium – wynikiem selekcji relewantnym w danym kręgu informacji o środowisku, które tworzą językowy obraz świata, ale i sposobem jej kodowania przy pomocy symboli – choć fragmentarycznym, skompresowanym i subiektywnym, to skutecznym.

Jeśli język jest fenomenem kulturowym, to czyni to językoznawstwo nauką o kulturze (Piotrowski 1994: 10), a raczej pewnych jej produktów. Bańcerowski & Pogonowski & Zgółka (1982: 21) definiują ogólne zadanie dziedziny w sposób następujący:

Przedmiotem i zadaniem lingwistyki jest więc rekonstrukcja, opis w postaci odpowiedniej teorii tej formy świadomości społecznej, jaką stanowi język.

Jednostki języka, na których operuje się najczęściej – przynajmniej w europejskiej tradycji językoznawczej – stanowią wyrazy. Poszczególnymi wyrazami, jak i ogółem słownictwa zajmuje się dział językoznawstwa, jakim jest leksykologia. O roli wyrazów w języku pisze Doroszewski (1966: 107):

Teoretyczna praca nad językiem musi się, choćby pośrednio, wiązać z pracą nad powiększaniem sprawności w posługiwaniu się tymi narzędziami przekazywania myśli i uczuć, jakimi są wyrazy.

W ramach leksykologii można badać różne aspekty słownictwa – według Adamo (1998: 75) dziedziną ta zajmuje się jednak przede wszystkim:

1. znaczeniem wyrazów i ich relacjami;
 2. zmianami historycznymi w formie wyrazów;
 3. regułami morfologicznymi, którym podlegają;
- a także, co szczególnie istotne z punktu widzenia niniejszej tezy:
4. opisem jednostek na potrzeby leksykograficzne.

Ze względu na swój szeroki zakres leksykologia dzieli się na dalsze obszary badawcze – Kania & Tokarski (1984: 129) wymieniają cztery najważniejsze z nich:

Jako podstawowe działy leksykologii wymienia się zazwyczaj:

- 1) etymologię – naukę o pochodzeniu i pierwotnym znaczeniu wyrazów,
- 2) semantykę – naukę o znaczeniu jednostek leksykalnych (w badaniach współczesnych punkt ciężkości przesunął się z wyrazów na zdania),
- 3) frazeologię – naukę o związkach wyrazowych,
- 4) onomazjologię – naukę o słownictwie, przyjmującą klasyfikację desygnatów za podstawę klasyfikacji nazw.

Prezentacją informacji wskazanych wyżej kategorii leksykalnych zajmuje się jednak odrębna dziedzina – leksykografia.

1.1.2. Leksykografia

Leksykografia lub słownikarstwo to dziedzina, która – najzwężej to ujmując – zajmuje się słownikami. O ile trudno opisowi takiemu odmówić prawdziwości, to jest on mało przydatny, jeśli nie określić najpierw, czym są słowniki i w jaki sposób można je badać.

Malkiel (1968: 278) słusznie zauważa, że *słownik* to wyraz, który „dość luźno odnosi się do wszelkich prac źródłowych uporządkowanych według słów lub nazw”. Fuertes-Olivera (2018: 1) ujął to natomiast w sposób następujący:

Lexicography is the science concerned with the theory and practice of dictionaries, that is, dictionaries, encyclopaedias, lexica, glossaries, vocabularies, terminological knowledge bases, and other information tools covering areas of knowledge and its corresponding language—in other words, reference and information tools dealing with “things”, “facts”, and “language”.

Leksykografia to nauka zajmująca się teorią i praktyką słowników, tj. słowników, encyklopedii, leksykonów, glosariuszy, zbiorów słownictwa, terminologicznych baz

■ wiedzy i innych narzędzi informacyjnych obejmujących dziedziny wiedzy i odpowiadający im język – innymi słowy odniesienie i informacje narzędzia zajmujące się „rzeczami”, „faktami” i „językiem”.

Z tego względu słowniki można przypisać do tego samego gatunku, co książki telefoniczne albo przewodniki (Kammerer & Wiegand 1998).

O nieuchwytności definicyjnej tego gatunku tekstu pisze również Bańko (2020: 15):

Leksykografia jest dziedziną o otwartych granicach, w których mieści się tak wielka różnorodność słowników, że próba uchwycenia ich wspólnych cech prowadzi do określeń o treści zaskakująco ubogiej, a mimo to dopuszczających wyjątki.

Mimo tego Tarpowi (2018a: 21) udało się zaproponować trzy zasadnicze cechy leksykografii jako dziedziny, choć te odznaczają się wysokim poziomem ogólności:

1. leksykografia nie jest poddziedziną językoznawstwa, ponieważ nie każdy słownik wymaga wiedzy lingwistycznej;
2. leksykografia nie należy do żadnej innej dziedziny, bo żadna inna dziedzina nie jest obecna w każdym słowniku;
3. leksykografia powstaje w wyniku współpracy interdyscyplinarnej.

Pierwsza z wymienionych wyżej cech pozwala potraktować leksykografię językową wspólnie z leksykografią rzeczową, która zajmuje się słownikami encyklopedycznymi, które są poświęcone miejscom, autorom lub wydarzeniom. Niniejsza praca skupia się jednak wyłącznie na leksykografii językowej i odrzuca słowniki rzeczy z zakresu rozważań.

W przypadku dzieł leksykograficznych, którym poświęcona jest ta teza doktorska, najistotniejsze jest rejestrowanie zmian w obszarze leksyki. Wawrzyńczyk (2015: 7) tak oto o tym wymiarze języka pisze:

Leksyka to artefakty. Przyroda nie daje człowiekowi słów, on je wytwarza, wszystkie pokolenia ludzkie są ich producentami, zachowują je w swym dorobku kultury duchowej i materialnej.

Zasadne jest zainteresowanie ilością, głębsza refleksja, czy dany język ma bogate słownictwo.

Nie wszystkie słowa jednak w ten sam sposób odzwierciedlają procesy leksykalne zachodzące w języku – niektóre z nich są trwalsze od innych, jeśli nie liczyć systemowych zmian w fonetyce i grafii, przez co z kolei trudno jest je wykorzystać w badaniach porównawczych (Górski et al. 2019: 82). Z perspektywy leksykologii historycznej nie tylko istnienie i forma słów jest interesująca, ale także zakres ich występowanie – interesujące są m.in. jednostki niepoświadczane w języku ogólnym (Borejszo 2013: 61).

Górski et al. (2019: 82) zwraca jednak uwagę na ograniczenia operowania w skali pojedynczych słów:

Trudno śledzić zmiany pojedynczych słów (...) Znacznie sensowniej będzie przyjąć hipotezę, że za ewolucją systemu językowego stoi bardzo duża liczba ledwie zauważalnych zmian leksykalnych, być może da się zaobserwować *en masse*.

Rozwiązania masowe są jednak praktycznie wykonalne dopiero od niedawna – umożliwiły to dopiero komputery, a ich pierwsze użycie do tego typu celów miało w Polsce miejsce

w 1978. Od tamtej pory wiele się zmieniło – z ostatnich dokonań warto wspomnieć o osiągnięciach Piotra Wierzchoń w obszarze masowej datacji i redatacji dziesiątek tysięcy wyrazów, w co piszący te słowa miał swój wkład. Aby to osiągnąć Wierzchoń sięgnął do tekstów źródłowych, wykorzystując przy tym repozytoria tekstów cyfrowych.

Dlatego konieczne jest uwzględnienie treści słowników historycznych, które często podają informację o chronologii jednostek (Seco 1987: 187), a także rejestrują wyrazy dialektalne oraz wszystkie te, które trudno z różnych względów znaleźć w tekście. Podhajecka (2018: 134–135) wyjaśnia, że informacja historyczna może być w słowniku reprezentowana poprzez:

1. odniesienia etymologiczne;
2. kwalifikatory chronologiczne;
3. informację encyklopedyczną i kulturową w ramach definicji;
4. diachroniczne formy wyrazowe;
5. chronologicznie uporządkowane cytaty.

W przypadku słowników dawniejszych *de facto* cała informacja jest w pewnym sensie historyczna – o ile nie traktuje się diachronii jako informacji wyłącznie eksplicytnie historycznej – ale wówczas potrzebny jest materiał referencyjny, na przykład nowszy słownik tego samego języka. Słowniki tego rodzaju są cennym źródłem w badaniu pochodzenia frazeologizmów, a także rejestrują zachodzące w nich zmiany i pozwalają wykryć „zapomniane” jednostki czy osobliwości syntaktyczne (Dziamska-Lenart 2013: 453).

Z drugiej strony słowniki mogą też być źródłem dezinformacji, a do uzyskania pełnego i właściwego obrazu polskiej leksyki i jej ciągłości konieczne jest badanie materiału źródłowego (Wawrzyńczyk 2010: 3). Alvar (1976: 44) pisze o tym problemie następującymi słowami:

Los diccionarios antiguos (...) [no deberían utilizarse] como textos informativos de primer orden ya que sus contenidos pueden responder a la conciencia lingüística del autor y no a la situación real.

Stare słowniki (...) [nie powinny być używane] jako teksty informacyjne pierwszego rzędu, gdyż ich treść może odpowiadać świadomości językowej autora, a nie rzeczywistej sytuacji.

Poza tym słowniki nie wszystko zdążyły odnotować – istotnym problemem w kontekście kompletności opisu leksykograficznego są m.in. efemerydy leksykalne, czyli jednostki o stosunkowo dużej liczbie poświadczeń, z których niemal wszystkiego pochodzą z bardzo wąskiego okresu, są zaś dużo rzadsze poza nim (Graliński & Wierzchoń 2017). Efemerydy takie są więc używane w skali, którą – zwłaszcza w okresie przedkorpusowym – bardzo łatwo przeoczyć. Problemów badań leksykograficznych nad przeszłością języków jest znacznie więcej, które opisuje m.in. Wanicowa (2009) w kontekście staropolszczyzny.

Kluczowym problemem jest jednak nie tylko sam fakt rejestracji jednostek w słowniku, ale także odpowiednia notacja informacji stylistycznej. Aby informacja stylistyczna mogła się w słowniku znaleźć, musi najpierw zostać zidentyfikowana poprzez analizę leksykologiczną, na co zwraca uwagę Hemming (1968: 818):

One of the first tasks of the lexicologist (...) is to distinguish the various ‘registers’, to transform the unclassified and heterogeneous body of extant writing into something more orderly and more manageable.

Jednym z pierwszych zadań leksykologa (...) jest rozróżnienie różnych „rejestrów”, przekształcenie niesklasyfikowanego i niejednorodnego zbioru zachowanego pisma w coś bardziej uporządkowanego i łatwiejszego w zarządzaniu.

To jednak nie zamyka pocztu problemów z opisem leksykograficznym. Jak pisał Quirk (1986: 5), innym z problemów tej dyscypliny jest fakt, że zbyt mało uwagi poświęcano rozróżnieniu semantyki, słowotwórstwa i gramatyki. Można rozważyć zatem relację na przykład między leksykografią a leksykologią – obie dziedziny zajmują się przecież słownictwem:

Leksykografia jest rejestrowaniem i objaśnianiem wyrazów, pojęta zaś ogólniej jako praca nad badaniem historii wyrazów obejmuje także ich gromadzenie, a więc sporządzanie ich zbiorów, czyli słowników. (Doroszewski 1954: 5)

Zakres leksykologii jest z reguły postrzegany jako węższy niż słownikarstwa (Bielińska 2019: 119). Przeciwniej opinii jest jednak Miodunka (1989: 5), według którego różnica pomiędzy leksykologią a leksykografią „[s]prowadzała się (...) do przedstawienia leksykologii jako nauki o znaczeniu i funkcjonowaniu wyrazów, leksykografii zaś – wyłącznie jako leksykologii stosowanej”. Obie dziedziny są więc ze sobą ściśle powiązane.

Z leksykografią silnie związane jest także pojęcie metaleksykografii. Lew (2007: 212) wskazuje, że jest to „podstawa teoretyczna dla praktyki leksykograficznej”. Termin ten można więc uznać w zasadniczo jako tożsamy z leksykografią teoretyczną. Jednak, jak wspomniano, trudno w słownikarstwie oddzielić teorię od praktyki. Istotę problemu uchwycił w tej kwestii Tarp (2018a: 27)

It is frequently claimed that lexicography is the discipline that deals with the theory and practice of dictionaries. The statement is not false, but neither is it complete as it does not cover the whole discipline which, in conformity with the concept of science introduced earlier, also comprises other elements than theory and practice. Something similar can be said about the frequent distinction between lexicography as a practice and metalexigraphy as a research area: Which component parts does metalexigraphy contain? What is the subject matter of metalexigraphical research? What is meant by lexicographical practice? And how is the overall field of lexicography delimited in this respect?

Często twierdzi się, że leksykografia jest dyscypliną zajmującą się teorią i praktyką słowników. Stwierdzenie to nie jest fałszywe, ale też nie jest zupełne, gdyż nie obejmuje całej dyscypliny, która zgodnie z wprowadzoną wcześniej koncepcją nauki obejmuje także elementy inne niż teoria i praktyka. Coś podobnego można powiedzieć o częstym rozróżnianiu między leksykografią jako praktyką a metaleksykografią jako obszarem badawczym: Jakie części składowe zawiera metaleksykografia? Jaki jest przedmiot badań metaleksykograficznych? Co oznacza praktyka leksykograficzna? A jak pod tym względem jest wytyczona cała dziedzina leksykografii?

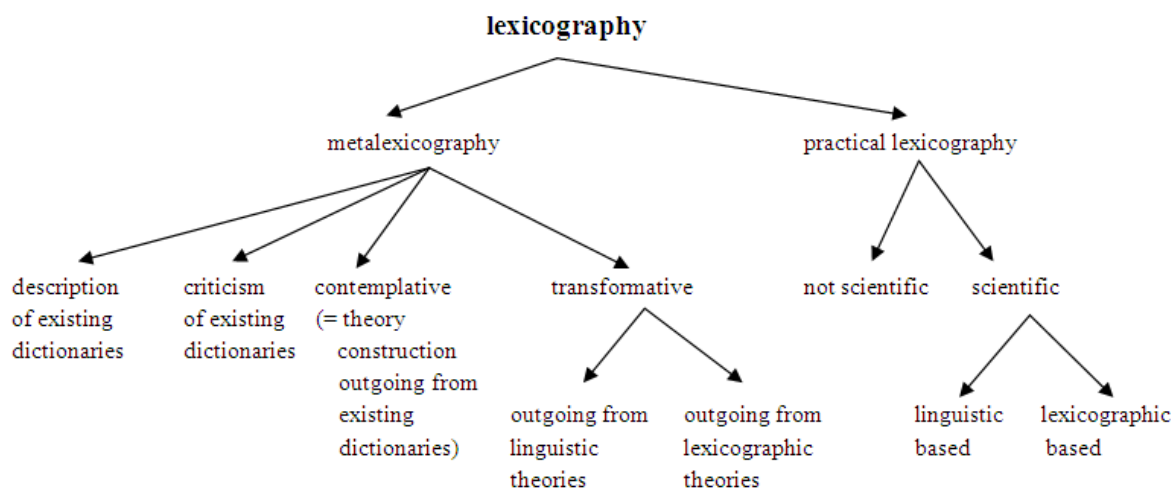
Badania w tym obszarze to nie lada wyzwanie także z innego powodu – Podhajecka (2009: 158–162) zauważa, że trudno jest je prowadzić ze względu na zróżnicowaną jakość materiałów dotyczących poszczególnych dzieł i zagadnień.

Żeby z zagadnieniami tymi móc skutecznie pracować, konieczna jest ich odpowiednia systematyka. Tę proponuje Hausmann (1986: 101ff), wyliczając następujące zadania metaleksykografii:

1. teoria leksykografii;
 - (a) teoria tekstów leksykograficznych;
 - (b) typologia słowników;
 - (c) kolekcja i przetwarzanie danych;
 - (d) teoria organizacji pracy leksykograficznej;
 - (e) teoria celu prac leksykograficznych.
2. krytyka słowników;
3. badanie korzystania ze słowników;
4. badanie statusu i marketingu słowników;
5. historia leksykografii.

Bergenholtz & Gouws (2012: 40) proponują następujący diagram, który przedstawia działy i zadania leksykografii praktycznej oraz teoretycznej (metaleksykografii):

Ilustracja 1. Leksykografia praktyczna i teoretyczna (metaleksykografia).



Jest jednak jeszcze jeden bardzo ważny aspekt metaleksykografii, który należałoby dodać, a mianowicie układ słowników. Jak zauważa Gouws (2018b: 226):

Dictionary structures can be regarded as one of the focal areas of metalexicographic research of the last three decades.

Struktury słownikowe można uznać za jeden z głównych obszarów badań metaleksykograficznych ostatnich trzech dekad.

Aspekt ten uwzględnia także u Hartmanna (2018: 767–768), który wymienia następujące aspekty praktyczne i teoretyczne kompilacji słowników:

(...) „badania słownikowe”, najpierw w ramach trzech głównych aspektów procesu kompilacji, w zakresie:

- zbierania dowodów (praca terenowa w celu stworzenia korpusów danych),
- redagowania wpisów (czynności opisowe w celu zdefiniowania i zilustrowania konkretnych słów/fraz), oraz

- formatowania tekstu (staranie się o przedstawienie materiału w sposób kompleksowy),
- a następnie, w ramach pięciu lub sześciu aspektów ich obiektywnej analizy, pod względem:
 - kontynuacji tradycji leksykograficznych (słownik historii),
 - klasyfikowania ich gatunków referencyjnych (typologia słownikowa),
 - oceny ich cech (krytyka słownikowa),
 - określenia ich części składowych (struktura słownikowa),
 - przestrzeganie konsultowania ich treści (korzystanie ze słownika), oraz
 - zastosowanie wsparcia informatycznego (technologia informatyczna).

Budowa słowników jest bardzo istotna, zwłaszcza z punktu widzenia niniejszej pracy, która temu ich aspektowi poświęca wiele uwagi. Struktura leksykonu silnie wpływa na to, jak się z niego korzysta, co determinuje również jego kształt już na etapie planowania – relacja między strukturą a użytkowaniem słownika więc jest dwustronna. Bielińska (2019: 133) pisze w sposób następujący o tej kategorii badań metaleksykograficznych:

Badanie użycia słowników to najmłodszy dział metaleksykografii, który ostatnio dzięki coraz lepszym możliwościom pozyskiwania danych szybko się rozwija. W centrum zainteresowań badaczy znajdują się następujące kwestie: Kto używa słownika, w jakich sytuacjach, z jakiego powodu, w jakim celu, z jakim rezultatem? Czy użytkownik posiada odpowiednie umiejętności? Jakie kompetencje są najważniejsze? Jakie potrzeby mają użytkownicy? Jak przebiegają konsultacje słownikowe różnego rodzaju?

Na wspomnienie zasługuje jeszcze jedno zadanie metaleksykografii – badanie aspektu kulturowego słowników. Rudnicka (2016: 50) zadanie to opisuje w sposób następujący:

Umiejętność osadzenia danego dzieła w kontekście kulturowym czy powiązania jego formy i treści z czynnikami istotnymi dla procesu jego tworzenia to równie ważne aspekty rzetelnej metaleksykografii.

Ten ów kontekst kultury również zostanie podjęty w niniejszej tezie w odniesieniu do *Słownika warszawskiego*, i obejmie zarówno to, jak ówczesna kultura została odzwierciedlona w słowniku, jak i jego wpływ, jaki miało jego wydanie na stan kultury w Polsce.

1.1.3. Humanistyka cyfrowa

Według polskiej Wikipedii humanistyka cyfrowa (ang. *digital humanities*, w skrócie DH) to „dział badań humanistycznych zajmujący się człowiekiem i jego wytworami w przestrzeni wirtualnej”. Definicja ta jednak jest bardzo wąska – trzy znacznie trafniejsze próby opisu dziedziny przytacza Maryl (2017: ak. 5):

[H]umanistyka cyfrowa [to termin] oznaczający nowe podejścia naukowe i jednostki instytucjonalne, służące zespołowym, transdyscyplinarnym i zaawansowanym cyfrowo badaniom, szkolnictwu i publikacjom. Humanistyka cyfrowa to (...) szeroki

wachlarz konwergentnych praktyk, służących eksploracji uniwersum, w którym druk nie jest już głównym medium. (Burdick et al. 2012: 127; tłumaczenie za: Maryl 2017)

Projekty z zakresu humanistyki cyfrowej są z definicji oparte na współpracy: łączą techników, bibliotekarzy, przedstawicieli nauk społecznych, artystów, architektów, badaczy komunikacji i informatyków we wspólnym ujmowaniu i rozwiązywaniu problemów, które często cechuje duży wpływ³, zaangażowanie społeczne, szeroki zakres i długotrwałość. Jednocześnie humanistyka cyfrowa to kontynuacja i rozszerzenie tradycyjnego obszaru humanistyki, nie substytucja czy negacja badań humanistycznych. (Presner 2010: 3; tłumaczenie za: Maryl 2017)

[H]umanistyka cyfrowa stanowi pole w luźnym sensie. Nie suponuję tutaj dobrze zdefiniowanego i ograniczonego pola akademickiego poza tradycyjnymi dyscyplinami humanistycznymi, tylko raczej inkluzywne pojęcie, które pozwoli nam mówić o różnych rodzajach inicjatyw i czynności na przecięciu humanistyki i informatyki lub cyfrowości. (Svensson 2010: ak. 12; tłumaczenie za: Maryl 2017)

Powyższe opisy traktują zagadnienie bardzo szeroko, przez co są bardzo nieostre. Ucińska (2017: 125) zauważa, iż humanistyka cyfrowa nadal nie posiada jednoznacznej definicji, ani nawet jednolitych ram metodologicznych, natomiast w *DH Manifesto 2.0* (2009: ak. 9) autorzy stwierdzają, że to „nie jednolita dziedzina, ale kolekcja zbieżnych praktyk”, które łączy koncentracja na cyfrowym aspekcie danych; w podobnym duchu pisze dziedzinę tę podsumowuje Maryl (2017: ak. 6):

Kluczowe jest tu uznanie humanistyki cyfrowej nie tyle za konkretną dyscyplinę czy metodę, ile zespół różnych praktyk badawczych, wykorzystujących cyfrową (pierwotnie bądź wtórnie) formę przedmiotu badań.

Nie ma więc mowy o humanistyce cyfrowej bez materiałów źródłowych w formie digitalnej. Te mogą mieć formę cyfrową pierwotnie (ang. *born digital*), ale także wtórnie, poprzez ucyfrowienie – czyli digitalizację (ang. *digitization*). Warto też wspomnieć, że sam proces transformacji źródeł analogowych na cyfrowe również jest włączany w obszar *digital humanities*.

Nie inaczej jest w przypadku leksykografii cyfrowej. Jest to dyscyplina, która leży na przecięciu humanistyki cyfrowej i językoznawstwa komputerowego, co otwiera wiele możliwości badawczych; z jakiegoś powodu dziedziny te jednak mało się komunikują (Piotrowski 2012: 6), mimo szerokiej gamy zastosowań, które przetwarzanie języka naturalnego ma w DH (Piotrowski 2012: 7–10).

Na silny związek językoznawstwa komputerowego i leksykografii cyfrowej wskazuje choćby liczba wariantów tego ostatniego pojęcia – Hartmann & James (2002) wymieniają następujące terminy: *computational lexicography* (idem 2002: 26), zwana także *computer-assisted lexicography*, *computer-based lexicography*, *computer-enhanced lexicography* (idem),

³ W oryginale został zastosowany angielski termin *high impact*, który ma nieco inny odcień znaczeniowy.

machine-aided lexicography, *machine lexicography* (idem: 90), *automatic lexicography* (idem: 11) oraz *electronic lexicography* (idem: 48).

Związek ten wynika z faktu, iż leksykografia to dziedzina poświęcona przechowaniu informacji (a raczej dość specyficznym jej wycinkiem), zarządzania nią i tworzeniem odpowiednich struktur dostępowych, które pozwalają ją wyekstrahować. Zgrabnie istotę problemu sformułował Tarp (2009: 45):

The common nature of all lexicographically relevant needs is that they are needs for information which may subsequently be used for a huge variety of purposes.

Wspólną naturą wszystkich istotnych leksykograficznie potrzeb jest to, że są to potrzeby informacji, które mogą być następnie wykorzystane do bardzo różnych celów.

Współczesna leksykografia silnie bazuje na rezultatach pracy innych działów humanistyki cyfrowej, zwłaszcza technikom eksploracji tekstu (ang. *text mining*), które wykorzystują publikacje tekstowe o postaci numerycznej. O zagadnieniu tym pisze Piotrowski (2012: 25):

Digital text is an essential prerequisite for natural language processing. However, books on NLP rarely discuss how to acquire texts, since electronic texts in modern languages are easy to obtain today, whether as raw texts or in the form of corpora.

Tekst cyfrowy jest niezbędnym warunkiem przetwarzania języka naturalnego. Jednak książki o NLP rzadko omawiają sposób pozyskiwania tekstów, ponieważ teksty elektroniczne we współczesnych językach są dziś łatwo dostępne, czy to w postaci surowych tekstów, czy w formie korpusów.

Tekstem tego typu mogą być słowniki, także dawniejsze – wówczas mowa o **retrodigitalizacji**. Termin ten związany jest z pojęciem **dygitalizacji** lub **digitalizacji** (ang. *digitizing* oraz *digitization*), określanego także mianem **cyfryzacji** czy **ucyfrowienia**. W kontekście bibliotekarstwa i niniejszej pracy wyrazy te oznaczają proces wprowadzenia do pamięci komputera w postaci cyfrowej dokumentu, którego oryginał miał formę analogową. Innymi słowy digitalizacja obejmuje wprowadzanie danych z nośników fizycznych na nośniki elektroniczne dla potrzeby humanistyki cyfrowej i dziedzin pokrewnych.

W przypadku tekstu proces ten zachodzi z wykorzystaniem sensorów optycznych (zwykle skanerów), które umożliwiają wprowadzenia danych pisanych lub drukowanych bezpośrednio do pamięci komputera; alternatywą jest ręczne przepisywanie treści, czego jednak tradycyjnie nie określa się mianem digitalizacji.

Choć samo pojęcie digitalizacji i jego warianty mają pochodzenie techniczne, to dzisiaj funkcjonuje w powszechnym użyciu i można je przypisać do słownictwa współnoodmianowego⁴. Warto tu odnieść się do słów Majewskiej (2020: 25), która porusza temat zapisu wyrazu **dygitalizacja/digitalizacja** i wskazuje, że ten pierwszy wariant pisowni (*dygitalizacja*) ma formę zgodniejszą ze współczesnymi regułami rządzącymi w polszczyźnie, na co przytacza stosowne argumenty. W niniejszej pracy jednak stosuje się wariant z zapisem miękkiem (*digitalizacja*) ze względu na jego częstotliwość i stosunkowo niedawne zapożyczenie z angielszczyzny, co uzasadnia utrzymanie anglosaskiej formy ortograficznej.

⁴ Mowa tutaj o rozumieniu tego pojęcia w interpretacji Markowskiego (1992:10ff).

Retrodigitalizacją natomiast – jak sugeruje to przedrostek *retro-* – nazywa się specjalny przypadek digitalizacji, która dotyczy dokumentów z epok dawniejszych; te mogą być praktycznie dowolne, muszą poprzedzać epokę współczesną. Oczywiście to, jak definiowana jest epoka współczesna i te uprzednie względem niej, to odrębna, acz w dużym stopniu drugorzędna kwestia.

Za ogólne cele digitalizacji uważa się:

1. archiwizację danych;
2. ochronę treści oryginału przed zniszczeniem;
3. upowszechnienie zasobu.

Majewska (2020: 51) wymienia natomiast cztery cele digitalizacji bibliotecznej:

1. uchronienie cennej książki przed zniszczeniem wskutek jej użytkowania;
2. udostępnianie elektronicznej kopii egzemplarza zabytkowej książki;
3. ułatwienie dostępu do zbiorów dzięki wykorzystaniu przestrzeni cyfrowej;
4. upowszechnienie dziedzictwa kulturowego.

Ogólny proces digitalizacji oraz zagadnienia związane z bibliotekami cyfrowymi opisują Bień & Szafran (2010). Nie można jednak zapominać, że procesy cyfryzacji indywidualnych dzieł się różnią, co jest pewnym wyzwaniem:

Z braku ujednoczenia procesu cyfryzacji wzmiankowanych słowników wynika konieczność ich „ręcznego”, wyzutego z automatyzacji, przeszukiwania i zestawiania – zatem sekwencji czynności niezwykle żmudnych, wymagających precyzji, dużego nakładu sił oraz czasu. (Kaszewski 2015: 96)

Wyzwania retrodigitalizacji są tu większe, ponieważ publikacje na papierze nie tylko ulegają degradacji wynikającej z upływu czasu, ale także zostały wykonane mniej dokładnymi technikami niż współczesne – wszystko to utrudnia dokładny odczyt treści w nich zawartych, zwłaszcza przez metody automatycznego rozpoznawania znaków.

W przypadku dzieł językoznawczych pojawiają się dodatkowe wyzwania, o których piszą Bouda & Cysouw (2012: 16):

The real challenge of linguistic retrodigitization is thus not the digital encoding as such, but the continuing update, refinement, and interpretation of the digital products.

Prawdziwym wyzwaniem retrodigitalizacji językoznawczej jest nie kodowanie cyfrowe jako takie, ale ciągłe aktualizowanie, udoskonalanie, i interpretowanie produktów cyfrowych.

Jednym z rodzajów takich obiektów o charakterze językowym są słowniki elektroniczne. Żmigrodzki (2008: 98–123) opisuje ich dwa rodzaje: słowniki **prymarnie elektroniczne** (niepoddane digitalizacji, powstałe już w formie cyfrowej) i słowniki **wtórne elektroniczne** (które są rezultatem procesu digitalizacji formy analogowej).

Retrodigitalizacja leksykograficzna stanowi dosyć szczególną podklasę retrodigitalizacji językoznawczej, ponieważ – jak zauważa Bień (2010b: 132) – w szerokim znaczeniu jest ona połączeniem dwóch procesów:

W przypadku takich „urodzonych cyfrowo” (ang. *digitally born*) tekstów lepiej mówić o ich komputeryzacji, to znaczy o przekształceniu do postaci bardziej wygodnej dla

użytkowników. Kiedy mówi się o dygitalizacji dziedzictwa kulturowego, do którego należą w szczególności słowniki, należy mieć na myśli zarówno dygitalizację właściwą, jak i komputeryzację.

W Polsce ukończonych zostało już kilka projektów w ramach tej dziedziny. Bień (2006a) opisuje pierwsze digitalizacje słowników, ich archiwizację przy pomocy technologii DjVu oraz indeksację ich treści, a także projekty pierwszych słowników polszczyzny *born digital*, natomiast bardziej aktualnego przeglądu słowników dawnych w formie cyfrowej dokonali Pałka & Kwaśnicka-Janowicz (2017: 58–85).

Retrodigitalizacja leksykograficzna posiada swoje specyficzne problemy, jak nietypowy układ typograficzny charakteryzujący słowniki. Jednocześnie pozwala ona na znacznie głębsze przekształcenie treści niż w przypadku digitalizacji większości obiektów niesłownikowych – wynika to m.in. z tego, że słowniki same służą do indeksacji wiedzy. Proces ucyfrowienia pozwala też wykryć błędy w słownikach – nie tylko w samej treści, ale i strukturze, czego przykład przytacza Bień (2015: 5):

[Z]akłócenia porządku alfabetycznego haseł, co mogło mieć trojaki przyczyny: błąd rozpoznania hasła, błędne potraktowanie podhasła jako hasła oraz – bardzo rzadkie – pomyłka redaktorów.

Oczywiście poza korektą słownika digitalizacja pozwala także na uzyskanie treści o większym stopniu przeszukiwalności, a także na głębszą manipulację informacjami w nim zawartymi dla celów praktycznych oraz badawczych, zwłaszcza przy pomocy parsowania. Słowniki rzadko mają zupełnie sztywną i zamkniętą strukturę, którą łatwo jednoznacznie sparsować, co przysparza problemów z głęboką digitalizacją, która wymaga do rozwiązania poszczególnych problemów pewnej dozy kreatywności.

Warto się jednak zastanowić, czymże jest właściwie retrodigitalizacja leksykograficzna. Usytuowana jest ona na przecięciu humanistyki cyfrowej i leksykografii, i jest zaaplikowaniem metod językoznawstwa komputerowego do starych źródeł, a właściwie dosyć szczególnego ich podzbioru, jakim są słowniki.

Ze względu na tematykę niniejszej tezy piszący te słowa wyodrębnił ten jakże wyspecjalizowany obszar, ponieważ posiada on specyficzną problematykę niespotykaną w takiej konfiguracji w dziedzinach pokrewnych, trudno jednak na chwilę obecną mówić o teorii retrodigitalizacji leksykograficznej czy nawet o koherentnym ujęciu jej założeń. Autor tych słów liczy jednak, że praca ta przyczyni się do wyodrębnienia dyrektyw metodologicznych, które pozwolą na uformowanie silniejszych podstaw tej dziedziny.

1.1.4. Nauka o informacji

Zarządzanie i przetwarzanie szeroko pojętej informacji jest przedmiotem badań dyscypliny teoretycznej, jaką jest nauka o informacji, czyli informatologia. Istnieje także powiązana dyscyplina inżynierska, która zajmuje się realizacją założeń nauki o informacji – jest nią informatyka.

W niniejszej pracy zagadnienia ogólne teorii informacji poruszane są tylko marginalnie, ponieważ do opisu specjalnego przypadku organizacji danych w materiale badawczym najlepiej służy leksykografia. Do praktycznego operowania na wspomnianych danych będą

wykorzystane rozwiązania informatyczne, a zagadnienia z tego obszaru wykraczają poza zakres słownikarstwa. Informatyka traktowana będzie tu więc instrumentalnie.

1.2. Przegląd literatury

Ze względu na interdyscyplinarną naturę prowadzonych tu badań, trudno jest dokonać wyczerpującego przeglądu obecnego stanu wiedzy – o ile można tu przytoczyć literaturę poświęconą relewantnym problemom z poszczególnych dziedzin i poddziedzin, to nadrzędny cel i *novum* tejże pracy stanowi właśnie połączenie dyscyplin w takiej, a nie innej konfiguracji, które składają się na retrodigitalizację leksykograficzną.

Samej dziedzinie – zwłaszcza w kontekście słowników – poświęcone są jednak bardzo nieliczne prace naukowe, a publikacji użytkowych jest niewiele więcej, co będzie dokładniej opisane w **rozdziale 6.1**.

W związku z tym piszący te słowa ograniczy się w tym rozdziale wyróżnienia prac pod względem konceptualnym najbardziej zbliżonych do niniejszej tezy doktorskiej – są to publikacje:

1. **Joanny Bilińskiej** (2013), poświęcona *Słownikowi języka polskiego* autorstwa Samuela Bogumiła Lindego, zwanemu *Słownikiem Lindego*;
2. **Magdaleny Majdak** (2018a), poświęcona *Słownikowi warszawskiemu*, oraz
3. **Małgorzaty Majewskiej** (2020), poświęcona *Słownikowi języka polskiego* pod redakcją Aleksandra Zdanowicza, wydanemu przez Maurycego Orgelbranda, zwanemu *Słownikiem wileńskim*.

Wszystkie trzy wymienione dzieła poświęcone są opisowi poszczególnych słowników narodowych polszczyzny i przedstawiają ich kompleksową charakterystykę: strukturę, treść i proces redakcji, a także inne relewantne informacje. Prac w innych językach o podobnej tematyce jest mało, a te znalezione niewiele wnoszą ani od strony teoretycznej, ani praktycznej (por. np. Belyaev et al. 2021).

W swojej pracy Majewska (2020: 18) postawiła sobie dwa cele: (1) poszerzenie i weryfikację wiedzy o konkretnym słowniku języka polskiego (tu: o *Słowniku wileńskim*), oraz (2) przedstawienie problematyki digitalizacyjno-bazodanowej tegoż dzieła; podobny zamysł przyświecały także Bilińskiej.

Podobna koncepcja realizowana jest również w niniejszej tezie doktorskiej, a wymienione prace zostały wykorzystane jako materiał referencyjny, choć w ograniczonym zakresie. Dzieje się tak, ponieważ w przeciwieństwie do pozostałych prac autorowi tych słów przyświecały dwa częściowo odmienne cele:

1. Przedstawienie opisu struktury słownika nie w sposób wrywkowy, ale możliwie kompleksowy i pełny, tak, aby był on wystarczający na potrzeby inżynierii językowej⁵;
2. Uogólnienie wyników analizy materiału badawczego na całą dyscyplinę i zaproponowanie standardowej metodyki w przypadkach analogicznych.

⁵ Inżynieria językowa jest tu rozumiana jak u Piotrowskiego (2001: 29–30)

Należy wymienić tutaj również pracę Kaszewskiego (2015), która dokonuje wyczerpującego przeglądu najważniejszych historycznych dzieł słownikarskich polszczyzny.

W sposób odrębny została potraktowana tutaj także literatura informatyczna, ponieważ propagacja wiedzy inżynierskiej odbywa się w tej dziedzinie innymi kanałami. Poza tym, jak ujął to Kamil J. Dudek⁶:

Faktem, który w 2017 zdumiewa jeszcze mocniej, niż to, że dalej istnieją gazety jest fakt, że w dalszym ciągu drukowana jest literatura informatyczna. W dodatku tłumaczona! Co jeszcze bardziej oddala treść od punktu w czasie, w którym była względnie świeża.

W związku z powyższym, a także wspomnianym wcześniej instrumentalnej (użytkowej) roli tejże dyscypliny w niniejszej pracy, literatura z tego zakresu będzie przytoczona tylko tam, gdzie jest to niezbędne.

1.3. Luka badawcza

Na podstawie zgromadzonej literatury – a w dużym stopniu także jej braku – autor tych słów zidentyfikował lukę badawczą w postaci trzech problemów, które postara się w tej pracy należyście zaadresować.

Pierwszym problemem badawczym jest brak *Słownika warszawskiego* w formie ubazdanowionego zasobu elektronicznego. W chwili pisania tych słów nadal nie istnieje ustrukturyzowana cyfrowa reprezentacja treści *Słownika warszawskiego*. Zasób tak byłby komplementarny względem tego, co w kontekście digitalizacji zasobów leksykograficznych Bień (2008a) określa mianem „The comprehensive lexicographic description of Polish” (‘Całościowego opisu leksykograficznego języka polskiego’), czyli sześć projektów słownikarskich:

1. *Słownik staropolski*;
2. *Słownik polszczyzny XVI wieku*;
3. *Słownik języka polskiego XVI i pierwszej połowy XVIII wieku*;
4. *Słownik języka polskiego PAN*;
5. *Wielki słownik języka polskiego*;
6. *Słownik geograficzny Królestwa Polskiego i innych krajów słowiańskich*.

Główne cele opracowania cyfrowej wersji SW są następujące:

1. Publikacja słownika w formie elektronicznej.
2. Stworzenie narzędzie do przetwarzania języka naturalnego.
3. Opracowanie nowych, do tej pory trudno dostępnych danych.

Cel pierwszy ma charakter zarówno terminalny (dostarcza narzędzia użytkownikom i umożliwia powszechny dostęp do informacji), jak i instrumentalny (prowadzi do opracowania narzędzi NLP). Cel drugi ma charakter wyłącznie instrumentalny, ponieważ prowadzi do opracowania danych będących przedmiotem celu trzeciego. Cel trzeci natomiast ma charakter przede wszystkim terminalny poprzez dostarczenie wiedzy o języku.

⁶ Opublikowane pod pseudonimem @wielkipiec na portalu DobreProgramy: <https://www.dobreprogramy.pl/@wielkipiec/poczytaj-mi-wielkipiecu-biblioteczka-programisty,blog,79847>.

Drugim problemem badawczym jest brak dostatecznie głębokiej digitalizacji dla słowników historycznych. Dotychczasowe digitalizacje takie są ograniczone i nie wykorzystują w pełni potencjału medium elektronicznego, co wymaga odłączenia struktury danych od ich reprezentacji, a to z kolei wiąże się z ich zaawansowanym przetworzeniem przy pomocy rozwiązań informatycznych. Sytuacja tak w istotny sposób ogranicza potencjał badawczy już opracowanych materiałów, a podejście to wytycza niepożądany precedens dla przyszłych przedsięwzięć.

Opracowanie treści *SW* – nawet fragmentarycznie, jak ma to miejsce w niniejszej pracy – ma więc wymiar nie tylko praktyczny, ale może także posłużyć jako makieta dla analogicznych przedsięwzięć.

Wreszcie trzecim problemem badawczym jest brak literatury, a więc i dostępnej metodyki retrodigitalizacji leksykograficznej, która pozwoliłaby na ucyfrowienie kolejnych dzieł w sposób usystematyzowany, głębszy i wydajniejszy niż dotychczas. Niniejsza teza doktorska powstał także w odpowiedzi na ten brak i zostanie tu podjęta próba sformułowania takiej metodologii w oparciu o opracowanie materiału badawczego.

2. Cechy słownika

Choć zróżnicowanie wewnątrz gatunku tekstu, jakim są słowniki, jest bardzo duże, a kryteriów, na podstawie których różnice te można wyodrębnić, jest bardzo wiele, to w rozdziale tym eksplorowany będzie ogólny związek między cechami poszczególnych dzieł a różnymi aspektami teoretycznymi i praktycznymi, których dotyczy praca leksykografa.

2.1. Cechy słownika a cechy języka

Jako że omawiane tu słowniki opisują system językowy, ich forma będzie od niego przynajmniej w pewnym stopniu zależna. Jest to zresztą stan pożądany: leksykografia powinna być silnie oparta o język i pozostawać w relacji wzajemnej zależności z gramatyką (Lombard 1994: 207–208). Oznacza to jednak, że leksykograf dziedziczy problemy lingwisty.

Wyjaśniona wcześniej trudność z rozgraniczeniem słowników i encyklopedii jest przejawem trudności z odseparowaniem wiedzy o języku od wiedzy o świecie. Według Gleasona (1967: 101) słownik to „[T]he meeting place of all the systems, linguistic and non-linguistic which bear relevantly on speech behavior”. Oznacza to, że leksykograf musi uwzględnić szeroko pojęte badania socjolingwistyczne i arbitralną naturę języka. Ta ostatnia objawia się nie tylko na poziomie cech systemowych, ale także na wewnętrznym zróżnicowaniu leksyki poszczególnych języków. Dlatego słowniki bywają opisywane jako repozytoria osobliwości (Kenstowicz & Kisseberth: 1979), które opisując język jako taki (Mochiwa 2007: 117).

Wynika to z faktu, że język ma wielu autorów, z których większość się nie zna i może mieć ze sobą stosunkowo niewiele wspólnego. Fakt ten Murray (1900) dostrzegł już przeszło przed stuleciem:

[T]he English Dictionary, like the English Constitution, is the creation of no one man, and of no one age; it is a growth that has slowly developed itself adown the ages.

The English Dictionary, podobnie jak angielska konstytucja, nie jest dziełem jednego człowieka ani jednego wieku; jest to wzrost, który przez wieki rozwijał się powoli.

Słownik nie jest więc obrazem samego języka, ale i kultury, ponieważ – jak ujmuje to Landau (2001: 10): „Many words, too, are *culture specific* (or *culture bound*, as some linguists prefer)” (wiele wyrazów jest właściwych danej kulturze). Piotrowski (1993: 572) zauważa, że „słownik nie jest neutralnym opisem słownictwa”, ale zależy od kultury, bo ta wpływa na wybór i sposób opisu:

Opis słownictwa od strony znaczeniowej jest jednocześnie opisem kultury, ponieważ słownictwo można traktować jako zbiór znaków odnoszących się do tych aspektów kultury, które dana społeczność traktuje jako stabilne i ważne dla jej funkcjonowania. Wyrazy kodują te doświadczenia lub obiekty, które są intersubiektywnie postrzegane jako ważne dla członków danej kultury. (Piotrowski 1993: 572)

Słowniki są zatem nośnikami norm obecnych w kulturze, którą rozumie się jako „zbiór asercji o człowieku i o społeczeństwie” (Dubois & Dubois 1971: 99).

Kolejnym aspektem związanym z postacią języka, która ma wpływ na formę jego reprezentacji w słowniku, jest jego zapis, co zostało zasugerowane podczas opisu typologii słowników: w przypadku języków zapisywanych alfabetem lub sylabariuszem sprawdza się porządek ortograficzny, ale rozwiązania analogiczne mogą nie być optymalne czy nawet możliwe w aplikacji w przypadku pozostałych języków. Tak dzieje się w przypadku języków semickich, w których – ze względu na derywacyjną i fleksyjną funkcję samogłosek – w słownikach stosuje się organizację gniazdową.

Dużo trudniej jest zorganizować leksykony dla języków opartych o ideograficzny system zapisu. Te zazwyczaj stosują metodę kluczy, czyli elementów znaków – samodzielnych lub niesamodzielnych – które wchodzi w skład znaków złożonych. Tak więc w słowniku chińskim wszystkie znaki zawierające klucz 口 kǒu ('usta') będą występowały w słowniku obok siebie – np.:

- 听 tīng ('słyszeć');
- 告 gào ('poinformować');
- 回 huì ('odpowiedzieć'),
- 叫 jiào ('nazywać; wzywać');
- 唱 chàng ('śpiewać');
- 哈 hā (dźwięk śmiechu: 'ha ha');
- 吗 mā (partykuła pytajna wprowadzająca pytania tak-nie: 'czy').

Taka metoda opracowywania słowników podług formy graficznej znaków-wyrazów określana jest mianem metody grafologicznej. W Chinach jednak już za panowania dynastii Qīng (1644–1911) leksykografowie zaczęli eksperymentować z opracowywaniem leksykonów metodą fonetyczną, co miałoby przydatność ze względu na nieoczywistą odpowiedniość i częściową nieprzewidywalność czytań znaków. To jednak było trudne do zrealizowania ze względu na trudność zapisu tej informacji używając systemu ideograficznego, dlatego poszukiwano się rozwiązaniami w postaci książek i tablic rymów, zwanych kolejno 韻書 yùnsū i dēngyùn[tú] 等韻【圖】; eksperymentowano także z innymi metodami, ale bez większych sukcesów (Saarela 2014: 188–194).

Na zapisie problemy leksykografii z językiem się jednak nie kończą. Jak zauważa Mochiwa (2007: 115) „[A] dictionary reveals the non-correspondence between form and meaning in language” ('Słownik obnaża nieodpowiedniość między formą a znaczeniem w języku'). Na dodatek pracę słownikarzy utrudnia fakt, że leksykografia jest przeważnie w tyle za językoznawstwem i nie korzysta z najnowszej wiedzy – jednocześnie wiedza im dostępna często stoi w sprzeczności z ograniczeniami, jakie nakłada metodologia redakcyjna słowników (Lombard 1994: 211).

Opis leksykograficzny dodatkowo komplikuje fakt, że jest on dokonany przy pomocy języka naturalnego – w przypadku słowników jednojęzycznych ten sam język jest więc nie tylko obiektem, ale i narzędziem do opisu. Lara (2001: 153–162) izoluje cztery problemy z opisem leksykograficznym dokonany przy pomocy języka naturalnego:

1. trudno jest zdefiniować podstawy metajęzykoznawcze do opisu;
2. żaden język naturalny nie jest wystarczający, bo jako system nie jest zdolny do opisanego w całości innego systemu;
3. język nie jest system zamkniętym;
4. język i koncepty nie są rozłączne wbrew klasycznej tradycji filozoficznej.

Język (rozumiany jako sposób tworzenia wypowiedzi) stosowany do opisu słownika nie jest jednak tym samym co język opisywany i może charakteryzować się pewną dewiacją względem naturalnej składni języka (Gouws 2018a: 49). Ten więc będzie zależał także od tradycji słownikarskiej i założeń metodologicznych przyjętych przez redaktorów danego dzieła.

2.2. Cechy słownika a cechy nadawcy

Na kształt słownika wpływ ma nie tylko stan języka, ale także to, jak jest on stosowany i prezentowany przez osobę go redagującą. Słowniki w sposób nieunikniony odnotowują przekonania użytkowników języka w danej epoce (Loiseau 2013: 155) – dzieje się to za pośrednictwem leksykografów, którzy nieświadomie przenoszą wartości społeczeństwa, w tym i swoje uprzedzenia (Landau 2001: 420). Oznacza to, że istotna jest „jednostkowa świadomość leksykografa” (Kaszewski 2014: 11).

Słowniki ogólne często jednak posiadają wielu autorów, co sprawia, że wiele indywidualnych idiosynkrazji jest odfiltrowywanych przez redaktorskie gremium. Fakt ten jest także nie bez znaczenia w dla odbiorców, ponieważ:

Slovník si bez uvedeného kolektivu autorů a hlavní redakce tak odebírá auru důvěryhodnosti, které by se každý slovník měl těšit. (Čermáková 2013: 196)

Nie przytaczając zespołu autorów i redakcji głównej słownik odbiera sobie w ten sposób aurę wiarygodności, którą każdy słownik powinien się cieszyć.

To, co trafi do słownika i jak zostanie zaprezentowane, nie zawsze jest wyborem słownikarzy – świadomym czy nieświadomym: leksykografia jest osadzona w danej kulturze, którą jednocześnie odzwierciedla (Żmigrodzki 2009: 119–123), podlega ona także naciskom politycznym (Żmigrodzki 2009: 116–118). Redaktor słownika musi się więc liczyć z tym, co stanowi temat tabu, z kontekstem ideologicznym wyrazów, a także aspektami prawnymi swoich poczynań. Wszystko to jednak może stać w opozycji do oczekiwań odbiorcy:

This attempt to provide a dictionary containing only ‘good’ words reflects a conflict between the views of the public and the views of dictionary editors as to what a dictionary is supposed to be. (Gates 1992: 265)

Owa próba dostarczenia słownika zawierającego tylko „dobre” słowa odzwierciedla konflikt między poglądami opinii publicznej a poglądami redaktorów słownika na temat tego, czym powinien być słownik.

Można więc posunąć się do konstatacji, że forma słownika jest przynajmniej po części rezultatem postawy jego autora względem odbiorcy. Crystal (1986: 72) opisuje tę relację, twierdząc, że leksykografia zajmuje się (wyidealizowaną) dynamiką na linii twórca słownika – użytkownik słownika, i zwraca uwagę na ciekawą okoliczność: twórcy słowników nie są użytkownikami swoich dzieł, a użytkownicy z reguły nie kompilują słowników, tylko z nich korzystają.

Relacja ta jednak może nie być postrzegana jako dwustronna, ale także jako trójstronna: w układzie tym jako odrębna grupa pojawiają się także językoznawcy. Lombard (1994) pisze o problematyce relacji między leksykografami, lingwistami i użytkownikami słowników: leksykografowie wprowadzają bowiem rozwiązania, które zadowolają ich samych i lingwistów, ale wprowadzają zwykłego użytkownika w konsternację; czasem pogodzenie tych dwóch celów jest jednak niemożliwe (Lombard 1994: 205).

2.3. Cechy słownika a cechy odbiorcy

Jak wspomniano wyżej, słownik wyraża niełatwą relację między nadawcą a odbiorcą – bez odbiorcy dzieło takie nie ma racji bytu. Opis leksykograficzny musi więc odzwierciedlać potrzeby użytkowników i nie może być od nich odizolowany (Argebo 2018: 61). Oznacza to, że również i odbiorca będzie miał wpływ na kształt leksykonu, choć nie wprost – reprezentować go musi sam leksykograf, z jednej strony musi uwzględniając kompetencje docelowego odbiorcy, a z drugiej – jego oczekiwania.

W kontekście kompetencji Bielska-Brodziak & Tobor (2007: 33) stwierdzają: „Słownik jest narzędziem i jak każde narzędzie powinien być odpowiednio używany”. Zgodnie z tym odbiorcom słownika mogą być stawiane różne wymagania (Zgółkowska 1994: 363–364). W przypadku słowników ogólnych warto jednak wziąć pod uwagę stan świadomości użytkowników:

Regardless of age, dictionary use is – to a large extent – dependent on experience, and it is surprising to see how many ‘adults’ are apparently unable to use a dictionary properly or optimally. (Nielsen 1990: 58)

Bez względu na wiek, korzystanie ze słownika jest w dużej mierze zależne od doświadczenia i zaskakujące jest to, jak wielu „dorosłych” najwyraźniej nie potrafi posługiwać się słownikiem właściwie lub optymalnie.

W związku z niskimi umiejętnościami w posługiwaniu się słownikiem użytkownicy mogą potrzebować przeszkolenia w tym zakresie (Whitcut 1986: 115). Powstaje pytanie, czy powinien zajmować się tym sam słownik – przedtem jednak trzeba zdecydować, jak przekonać użytkownika o potrzebie nabycia takich kompetencji:

While dictionary-using skills must be learnt for dictionaries to appropriately respond to the community needs, users must be made to realise the importance of dictionaries. (Chebanne 2010: 98)

O ile umiejętność korzystania ze słowników jest konieczna, aby te mogły odpowiadać na potrzeby społeczności, to należy uświadomić użytkowników, jak ważne są słowniki.

Odpowiedzi na powyższe pytania nie są odizolowane jednak od oczekiwań i potrzeb użytkowników – ich oczekiwania względem słowników różnią się w poszczególnych krajach i okresach (Crystal 1986: 81), a także w różnych częściach społeczeństwa. U podstawy jednak korzystanie ze słownika składa się z trzech niezmiennych etapów (Corbin 2013: 24):

1. wyszukania artykułu hasłowego;
2. wyszukania w artykule hasłowym pożądanej informacji;

3. zrozumienia (zinterpretowania) informacji weń zawartej.

Relevantne umiejętności odbiorcy mogą dotyczyć jednak nie tylko korzystania z samej struktury dostępowej, ale także jego kompetencji językowej – słowniki monolingwalne wymagają opanowania języka docelowego przynajmniej w pewnym stopniu, ponieważ ten pełni funkcję metajęzykową. Wyjaśnia to Bergenholtz (2005: 130) w kontekście wydanego w 1999 r. *DanskOrdbogen*:

Dieses Wörterbuch ist als Textproduktionswörterbuch für Benutzer konzipiert, die dänisch als Muttersprache haben. Es ist somit kein Rezeptionswörterbuch, das dem Benutzer auch vollkommen unbekannte Wörter erklären muss. Der Benutzer kennt das Wort oder hat zumindest irgendeine Vorstellung davon, worum es bei dem betreffenden Wort geht.

Słownik ten został zaprojektowany jako słownik do produkcji tekstu dla rodzimych użytkowników języka duńskiego. Nie jest więc słownikiem stworzonym do odbioru, który musi wyjaśniać użytkownikowi także zupełnie nieznanne słowa. Użytkownik zna słowo lub przynajmniej ma pojęcie, o co chodzi w jego znaczeniu.

Z tego względu parametry takie jak gęstość informacji będzie różna w zależności od przewidywanego użytkownika – względem rodzimych użytkowników zakłada się pewną kompetencję, której nie muszą posiadać użytkownicy uczący się języka (Mochiwa 2007: 115). W przypadku słowników wykładowych (definicyjnych) za grupę docelową przyjmuje się użytkowników języka z wykształceniem średnim oraz ich uzusy językowe, co jest punktem wyjścia do rozpatrywania ich postaw względem języka i rzeczywistości (Filipec 1995: 22). Poziom wykształcenia i świadomości językowej wpływa nie tylko na odbiór treści, ale także na sposób użycia leksykonu. Już ponad 170 lat temu Johnson (1747: 5) zauważył, że:

[T]he unlearned much oftner consult their dićtionaries, for the meaning of words, than for their structures or formations (...).

[Ludzie] niewykształceni znacznie częściej sięgają do słowników po znaczenie wyrazów, niż po ich struktury czy formację (...).

Użycie słownika nie musi mieć charakteru poszukiwania konkretnej jednostki języka – jak wspomniano, może to niekiedy też być czytanie go w całości, jednak częściej ludzie słowniki kartkują – wówczas oczekują losowego natrafienia na terminy ciekawe lub zabawne (Hanks 2010a: 589).

Bergenholtz (2005: 125) zwraca także uwagę na dwa powszechne – choć nie uniwersalne – wśród użytkowników słowników przekonania:

Zwei Grundhaltungen haben viele potentielle Wörterbuchbenutzer gemeinsam:

- Hin Wörterbuch führt die Wörter in alphabetischer Reihenfolge an und erklärt ihre Bedeutung.
- Wenn es im Wörterbuch steht, ist es richtig.

Wielu potencjalnych użytkowników słowników dzieli dwa podstawowe założenia:

- Słownik wymienia słowa w kolejności alfabetycznej i wyjaśnia ich znaczenie.
- Jeśli coś jest w słowniku, to jest to poprawne.

To pierwsze przekonanie wynika z doświadczenia w danym kręgu kulturowym i dotyczy czysto technicznego aspektu słowników. Przekonanie drugie jest jednak bardziej złożone i posiada poważniejsze implikacje. Niektórzy użytkownicy pokładają w słownikach bardzo duże zaufanie, o czym pisze Gates (1992: 266): „If a word is in the dictionary, they believe, that means it is all right to use it anywhere, anytime” (‘Jeśli wyraz jest w słowniku, jak wierzą, to oznacza to, że można go stosować zawsze i wszędzie’).

To często zbyt wielkie zaufanie użytkowników do leksykografów lub ich niezrozumienie wynika z braku kompetencji oraz niechęci poznania złożonych reguł rządzących językiem, przez co sami pozbawiają się możliwości wybrania innego źródła wiedzy (Bańko 2020: 48). Taka postawa wpływa na społeczną percepcję leksykonów:

Niewielka wiedza na temat słowników i bezkrytyczny stosunek do nich tyleż jednak obniżają ich status, co świadczą o ich prestiżu w oczach odbiorców. (Bańko 2020: 67)

Nie jest jednak dobrą strategią rozpatrywać kompetencje i oczekiwania użytkowników jako homogenicznej grupy, o czym pisze Čermák (2011: 107):

Ke strategii tvorby slovníku patří výchozí rozhodnutí o typu uživatele a jeho potřebách, rozsahu slovníku, volba hesel v podobě hesláře (včetně např. zkratek) apod.

Strategia tworzenia słownika obejmuje wstępną decyzję o typie użytkownika i jego potrzebach, zakresie słownika, doborze haseł w formie hasłownika (zawierającego np. skróty) itp.

Należy pamiętać, że ze słownika zawsze korzysta konkretny człowiek, co do którego zakłada się pewną wiedzę (Ряполова 2018: 387–388); wiedza ta będzie zróżnicowana między poszczególnymi grupami oraz w obrębie danej grupy docelowej. Jednym ze skrajnych przypadków jest zderzenie wiedzy popularnej z informacjami specjalistycznymi. Zgusta (1971: 113) zwraca uwagę na fakt, że „the folk classification of plants and similar subsystems of the extralinguistic world is often different from the scientific one” (‘ludowa klasyfikacja roślin i podobnych podsystemów świata pozajęzykowego często różni się od naukowej’).

Oczekiwania fachowców są przeważnie bardziej szczegółowe i trudniejsze do spełnienia, czego przykład podaje Bergenholtz (2005: 127) w odniesieniu do definiowania jednostek leksykalnych powiązanych z wirusami:

3. Falsche Bedeutungsangaben aus der Sicht von Fachleuten

Mit der Frage, ob Viren leben oder nicht, wird allerdings ein echtes Problem angesprochen: Das Verständnis der allgemeinen Bevölkerung ist manchmal nicht identisch mit dem der Fachleute. Das ist in Wirklichkeit ein interessanteres Problem als die der einfachen Fehlerlexikographie.

3. Błędne wskazanie znaczenia z punktu widzenia ekspertów

Jednak pytanie, czy wirusy są żywe, czy nie, odnosi się do prawdziwego problemu: rozumienie opinii publicznej czasami różni się od rozumienia profesjonalistów. W rzeczywistości jest to bardziej interesujący problem niż prosta leksykografia błędów.

Oznacza to, że to, co może być poprawne w rozumieniu potocznym, może nie spełniać oczekiwań fachowców, jednak Bergenholtz (2005: 127) uważa, że trudno o sytuację odwrotną:

Ob auch noch der fachlich unkorrekte Ausdruck Zellelement für einen Laien gut genug ist, kann bezweifelt werden.

Wątpliwe jest jednak, czy technicznie niepoprawna jednostka terminologiczna jest wystarczająco dobra dla laika.

Poszczególne grupy specjalistów będą miały różne oczekiwania. Gruszczyński (1999) na przykład pisze o dziennikarzach jako specjalnym przypadku odbiorców słownika, w których pracy zasadniczym narzędziem jest język.

Ciekawym zagadnieniem jest także wykorzystywanie leksykonów w orzecznictwie, gdzie widać ich wysoki prestiż, ale także gdzie darzy się je zbyt wielkim zaufaniem (traktując ich normatywność jako rozstrzygającą w kontekście prawnym) i stosuje nie wybierając ich w spójny sposób. Widać przy tym „wyraźny dystans prawników do słowników elektronicznych niemających wydania papierowego” (Czelakowska & Kubicka & Klubińska 2016: 53), czego głównym powodem jest niska permanentność tego medium – choć informacji nie można zniszczyć, to może ona łatwo zniknąć z Internetu.

Zasadniczo przyjmuje się, że słownik do zaspokojenia potrzeb użytkownika musi zawierać „dokładne rodzaje danych – nie więcej i nie mniej” (Tarp 2008: 83). Część z tych danych będzie znajdowała się w treściach zewnętrznych słownika – jednak badania wykazują, że mało kto czyta wstępy do słowników (Berglund 2014). Dlatego bardzo ważne są decyzje leksykografa, który musi przewidzieć zachowania odbiorców – według Tarpa (2008: 54–55) istnieje dziewięć kryteriów sformułowanych jako pytania, które musi sobie zadać leksykograf:

Jaki jest ojczysty język użytkowników?

W jakim stopniu władają oni swoim językiem ojczystym?

W jakim stopniu opanowali dany język obcy?

W jakim stopniu władają danym językiem specjalistycznym w języku ojczystym?

W jakim stopniu opanowali dany język specjalistyczny w języku obcym?

Jak wiele doświadczenia przekładowego posiadają?

Jak rozległa jest ich ogólna wiedza o kulturze?

Jak rozległa jest ich wiedza na temat kultury na danym obcym obszarze językowym?

Jak wiele wiedzą na dany temat lub o danej dziedzinie nauki?

Powyższe pytania skupiają się wyłącznie na treści, pomijając zupełnie jej podania. Odseparowanie formy prezentacji od zawieranej informacji pozwoliłoby jednak rozszerzyć zakres odbiorców słownika; dopasowanie układu graficznego czy ścieżki dostępu mogłoby całkowicie odmienić kontekst tejże informacji (Martinez 2013).

2.4. Cechy słownika a cechy medium

Jak zostało już w niniejszej tezie wspomniane, na poziomie abstrakcyjnym słownik jest sposobem organizacji danych o pewnej przyjętej formie i posiadający ograniczenia co do natury danych w nim zawartych. Jednak abstrakt taki jest niczym bez realizacji – tą jest proces kompilacji słownika, w ramach którego najpierw tworzy się strukturę teoretyczną, a następnie umieszcza

się w niej zebrane dane; rezultatem jest uzyskanie konkretnego dzieła. To, jak proces ten wygląda, jest odzwierciedlone w budowie i treści leksykonów – lub jak ujmują to Bielska-Brodziak & Tobor (2007: 29): „To, co znajduje się w słownikach, jest konsekwencją metody opracowywania słowników”.

Jednym z najważniejszych aspektów, jakie będzie terminowało medium, jest struktura dostępu: słowniki drukowane mają ją przeważnie tylko jedną, choć zdarzają się indeksy pozwalające na inny dostęp – to jednak zawsze odbywa się kosztem miejsca przeznaczonego na indeksowanie. Słowniki elektroniczne oferują znacznie więcej ścieżek dotarcia do danych, w dodatku bez ponoszenia niemal jakichkolwiek kosztów technicznych, jeśli tylko ich architektura jest dostosowana do medium. Niestety, wiele słowników nie wykorzystuje w pełni możliwości baz danych, oddzielając jedynie hasło od treści artykułu hasłowego – innymi słowy, są jedynie cyfrową reprezentacją nie tylko treści, ale i linearnego układu ze słownika tradycyjnego.

Ograniczenia medium będą miały też wpływ na stopień kondensacji informacji nie tylko w ramach artykułów hasłowych, ale także w treściach zewnętrznych, o czym dokładniej pisze Wiegand (1996). W słownikach drukowanych zaoszczędzenie miejsca przekłada się na mniejszy koszt ich wydania, a więc na bardzo istotny parametr w przypadku dzieł komercyjnych (i nie tylko). Rozwiązania elektroniczne ograniczeń takich nie mają, ponieważ tekst zajmuje niewiele przestrzeni dyskowej; skrócenia, jeśli stosowane, mają za zadanie przede wszystkim przyspieszyć czytanie. Ograniczenia w ilości danych dotyczą słowników elektronicznych w zasadzie tylko w przypadku multimediiów, takich jak obrazy, nagrania audio czy video, które wymagają znacznie więcej miejsca niż tekst.

Sama forma prezentacji danych również jest zależna od nośnika: w przypadku ograniczeń miejscowych można stosować mniejszą czcionkę i redukować światła między elementami strukturalnymi, co może istotnie zmniejszyć czytelność tekstu. Co więcej, raz obrana forma w druku jest silnie związana z treścią i niezmienna, podczas gdy słowniki elektroniczne pozwalają na odseparowanie treści od formy, dzięki czemu te same elementy mogą zostać zaprezentowane na różne sposoby – nie tylko od strony typograficznej, ale i strukturalnej.

2.5. Cechy słownika a jego funkcje

Słowniki należy traktować zasadniczo jako narzędzia (Tarp 2018b: 244). Jako takie najogólniej leksykony opisuje to Piotrowski (2001: 45):

Użytkownik sięga do słownika przy natknięciu się na jakiś problem. Powstaje dzięki temu możliwość rozwiązania takiego problemu na szerszą skalę za pomocą słownika (...).

Funkcje, jakie pełnią słowniki, można rozważać na różnych poziomach analizy – jednak w sposób sformalizowany zajmuje się tym tzw. teoria funkcji leksykograficznych (Tarp 2008: 33–124). Tarp (2008: 81) wyjaśnia, że „funkcja leksykograficzna” czy też „funkcja słownika” to jego:

[Z]dolność do zaspokojenia konkretnego rodzaju relewantnych leksykograficznie potrzeb, które mogą zaistnieć u konkretnego rodzaju użytkowników w konkretnej sytuacji pozasłownikowej.

Założenia, na jakich opiera się ta teoria, są następujące (Tarp 2008: 40–41):

1. słowniki są odpowiedzią na konkretne rodzaje potrzeb zarejestrowanych w danym społeczeństwie;
2. wyartykułowane oczekiwania i potrzeby użytkowników słownika są cieniem ich faktycznych, nieuświadomionych potrzeb;
3. przyjazność słownika dla użytkownika nie zależy tylko od łatwości i prędkości, z jaką wyszukiwana jest przez niego informacja, ponieważ zależy to od poprzedniego doświadczenia tegoż użytkownika.

W ramach teorii funkcji leksykograficznych uwzględnia się cztery rodzaje sytuacji, przy pomocy których można określić funkcje słownika (Tarp 2008: 43–53):

1. **sytuacje potencjalnego użytkownika** – w tym hipotetyczne sytuacje, w jakich może się znaleźć jego obiorca;
2. **sytuacje kognitywne** – rzeczy, wydarzenia i pojęcia, na które natrafia użytkownik; te dzielą się na systematyczne i sporadyczne;
3. **sytuacje komunikacyjne** – powtarzane akty w ramach danego kontekstu społecznego;
4. **sprawdzanie/korektę tekstu** oraz **ocenę tekstu** – ogół zabiegów dokonywanych między nadaniem tekstu przez jego autora a przed otrzymaniem go przez grupę docelową – czyli np. redakcja tekstu przed jego wydaniem i ocena przez nauczyciela w kontekście dydaktycznym.

Wyróżnione na tej podstawie funkcje słownika można pogrupować w dwie kategorie (Tarp 2008: 83–85):

1. **komunikatywne** – które dotyczą trzech typów czynności:
 - a. recepcji (rozumienia) tekstu (Leroyer 2018);
 - b. produkcji (wytwarzania) tekstu (Fuertes-Olivera & Bergenholtz 2018);
 - c. przekładu tekstu (Ciacomini 2018).
2. **kognitywne** – które dotyczą sytuacji, w jakich użytkownik znajduje się powtarzalnie (to zależy od rodzaju użytkownika).

Bergenholtz (2005: 131) podkreśla, że określenie głównych funkcji słownika jest na tyle ważne, iż wszelkie inne rozważania natury leksykograficznej są mu podrzędne:

Ohne genaue Überlegungen über den genuinen Zweck eines Wörterbuches, d.h. die Menge der vorgesehenen Wörterbuchfunktionen, sind Überlegungen wie die von Pitzek (2000) oder Wierzbicka (1993) und auch die meisten lexikologischen Vorschläge lexikographisch gesehen ohne Belang.

Bez dokładnego rozważenia rzeczywistego przeznaczenia słownika, tj. zestawu funkcji słownikowych, rozważania Pitzka (2000) czy Wierzbickiej (1993), a także większość propozycji leksykologicznych, są z leksykograficznego punktu widzenia nieistotne.

Funkcji szczegółowych słownika można wyróżnić jednak więcej, a jedno dzieło pełni ich nierzadko wiele naraz. Zgółkowa (1994: 363) wyraża na przykład opinię, że słowniki mają dwie zasadnicze funkcje:

1. słuzenie w formie „poradników” lub „podręczników”;
2. dokumentowanie stanu języka (jego leksyki i semantyki).

Cała systematyka funkcjonalna zaprezentowana wyżej przez Tarpa w pierwszym ze wskazanych przez Zgólkową zadań, które można opisać jako dydaktyczne i które jest nastawione na odbiorcę. Drugi wskazany przez badaczkę cel słowników natomiast jest nastawiony przede wszystkim na język. Choć w teorii obie te funkcje zdają się być w opozycji, to w praktyce słowniki ogólne (zwłaszcza narodowe) je w sobie łączą. Oczywiście słowniki innych rodzajów również mogą pełnić tę rolę.

Wartość dokumentacyjną słowników podkreśla Rudnicka (2016: 53):

Warto zatem myśleć o słowniku (...) jako o swoistym repozytorium doświadczenia zbiorowego danej społeczności (utrwalonego w języku) (...) Język dokumentuje pewien model świata, a słownik utrwała ten model w wersji językowej w schematycznej formie szeregu artykułów hasłowych (...).

W przypadku słowników narodowych można także mówić o związanej z powyższą funkcji integracyjnej (integrującej), o której pisze Hanks (2010c: 581):

A great monolingual dictionary (...) brings together speakers of a language, it has a socially integrative function, making explicit the basis of words and meanings and usage, which all uses of the language rely on.

Wielki⁷ słownik jednojęzyczny (...) zrzesza osoby mówiące językiem, pełni funkcję integracyjną w społeczeństwie, uwydatniając podstawy słów i znaczeń oraz użycia, na których opierają się wszystkie użycia języka.

Rolę integracyjną w kulturze mają także wszelkie słowniki historyczne (Hanks 2012b: 2), ponieważ demonstrowują ciągłość języka, który w europejskim kręgu kulturowym jest często podwaliną tożsamości etnicznej. W pewnym stopniu przyczyniają się do tego wszystkie słowniki – jak zauważają Kahane & Kahane (1992: 20): „Dictionaries tend to cultivate the heritage” (‘Słowniki przeważnie kultywują tradycję’).

Ta identyfikująco-integracyjna funkcja języka była kluczowa w Polsce dwudziestolecia międzywojennego ze względu na swoją wieloetniczność i próbę unifikacji trzech obszarów należących przedtem do odrębnych bytów państwowych (Woźniak 2019: 30).

O ile słownik ma funkcję społecznie integrującą, to tylko do pewnego stopnia (Hanks 2010a: 587) – jest on dziełem przede wszystkim pragmatycznym i służy do przetwarzania tekstu, czyli jego odczytania lub stworzenia (Piotrowski 2001: 44).

Pragmatyzm ten przejawia się przede wszystkim poprzez funkcję informatywną: przekazują one użytkownikom wiedzę na temat rzeczywistości zakodowaną przy pomocy tłumaczeń lub definicji. Sposób realizacji tej funkcji przez leksykony jest drugorzędny – jak konstatuje Bolinger (1965: 572): „Dictionaries do not exist to define, but to help people grasp meanings” (‘Słowniki nie istnieją po to, aby definiować, ale po to, aby pomóc ludziom zrozumieć znaczenia’). Istotne jest w nich jednak to, że jako narzędzia odpowiadają na zapotrzebowanie informacyjne, o czym piszą Dubois & Dubois (1971: 7) tymi słowami:

⁷ Wyraz angielski jest w tym kontekście użyty niejednoznacznie – może być też tu przetłumaczony jako „dobry”.

Les dictionnaires sont des objets manufacturés dont la production, importante dans les sociétés développées, répond à des exigences d'information et de communication.

Słowniki to wytworzone przedmioty, których produkcja, ważna w rozwiniętych społeczeństwach, spełnia wymagania odpowiadające na zapotrzebowania komunikacyjne.

Słowniki nie tylko prezentują nowe informacje, ale także standaryzują te już istniejące – albo, jak ujmuje to Bańko (2020: 49): „[S]łownik pełni funkcję objaśniającą, jak i normotwórczą”. W ramach funkcji normatywnej (normotwórczej) podaje on nie tylko poprawne i niepoprawne formy, ale jest także arbitrem znaczenia – rozstrzyga, co dany wyraz znaczy (Cotter & Damaso 2007: 6).

W swoim artykule Ceil (2003: 323) zestawia – a nawet przeciwstawia – charakterystyczną dla języka wariantywność z typową słownikom kodyfikacją. Ta ostatnia jest możliwa, ponieważ „dictionaries suggest authority, scholarship, and precision” (‘słowniki sugerują autorytet, naukowość i dokładność’). Jak zostało wcześniej w tej pracy wspomniane, stosunek danego dzieła wobec normy może być różny. Zadaniem leksykografa jest jednak przede wszystkim obserwować i rejestrować konwencje, a nie je zmieniać. Słownik ma co prawda funkcję kodyfikującą, ale:

Výklad významu kodifikuje významovou normu, ale distribuční oblast není neměnná a uzavřená. Výběrem dokladů a charakteristik lexikograf i prognózuje. (Filipec 1995: 41)

Objaśnienie znaczenia kodyfikuje normę znaczeniową, ale obszar dystrybucji nie jest niezmienny i zamknięty. Leksykograf wybierając przykłady i charakterystykę jednocześnie dokonuje prognozy.

Ten wybór jest istotny, ponieważ „słowniki oddziałują swoim autorytetem przez pokolenie i dłużej” (Filipec 1995: 42). Decyzje leksykografów mogą mieć poważne implikacje, zwłaszcza gdy wpływają na dopiero kształtujące się języki narodowe:

[L]exicographers will have to help in making decisions concerning the establishment of different national languages. There is no guaranteed prescription for such a decision (Zgusta 1971: 185)

Leksykografowie będą musieli pomóc w podejmowaniu decyzji dotyczących ustanowienia różnych języków narodowych. Na taką decyzję nie ma gwarancji przedawnienia.

Niektóre decyzje leksykografów mają jednakże konsekwencje wykraczające poza świat językowy i przeplatają się z zagadnieniami o wymiarze etyczno-prawnym (Landau 2001: 402). Analizę tego, jak słowniki stosowane są w polskim orzecznictwie, przeprowadzili w swojej pracy Bielska-Brodziak & Tobor (2007):

[W]ydano w naszym kraju ponad 700 orzeczeń, w których słownik był jednym z argumentów na rzecz podjętego rozstrzygnięcia. I mamy tu do czynienia z wyraźną tendencją wzrostową. (Bielska-Brodziak & Tobor 2007: 20)

Proponują oni w swojej pracy trzy typy orzeczeń sądów wyróżnionych ze względu na sposób, w jaki opierano się na słownikach:

1. orzeczenia, w których znaczenie słownikowe interpretowanych zwrotów nie spotkało się z uznaniem sądu;
2. orzeczenia, w których słownik stanowił tylko jeden z argumentów, najczęściej punkt wyjścia;
3. orzeczenia, w których znaczenie słownikowe stanowiło podstawę rozstrzygnięcia sprawy.

Zanim takie orzeczenie zostanie jednak wydane, organ interpretujący musi w pierwszej kolejności zdecydować, czy decyzje będą tylko na argumentach pozatekstowych, czy także tekstowych; w tym drugim przypadku konieczne trzeba też podjąć decyzję o tym, jaki rodzaj języka będzie brany pod uwagę: język potoczny, język prawny czy język aktów normatywnych (Bielska-Brodziak & Tobor 2007: 21).

Przytoczeni badacze krytykują jednak to, w jaki sposób leksykony są stosowane w orzecznictwie, wskazują na brak uzasadnienia do wyboru danego słownika, nieuwzględnianie „akontekstowości definicji słownikowych” czy implicytne i aprioryczne założenie o prymacie znaczenia językowego (Bielska-Brodziak & Tobor 2007: 27–30).

Zarówno funkcja normatywna, jak i informatywna są silnie związane z inną – funkcją pedagogiczną. Słowniki należą bowiem do dyskursu pedagogicznego (dydaktycznego), a ich zadaniem jest wypełnić odstęp między czytelnikami a normą lingwistyczną i kulturową, pełniąc rolę mediatora między językiem a zbiorowością, ponieważ jego materiałem źródłowym jest ogół wypowiedzi tejże zbiorowości; słownik można więc uznać za „instytucję społeczną” (Dubois & Dubois 1971: 10–11; 49–52). Dobrym przykładem dzieła o roli czysto dydaktycznej jest słownik ortograficzny, który komunikuje o poprawnych i niepoprawnych formach oraz ich zapisie. Podobną funkcję w założeniu pełnią słowniki wyrazów obcych, jednak Doroszewski (1966: 64) poczynił następującą obserwację:

Miało się czasem wrażenie, że zadaniem słownika ma być nie ułatwienie czytelnikowi wysłowienia się po polsku możliwie bez użycia wyrazów obcych, ale usprawnienie czytelnika do tak zwanego brylowania w rozmowach towarzyskich językową erudycją.

Kolejna funkcja słownika – perswazyjna – jest silnie związana z poprzednią. Przyjmuje się, że perswazyjność w słownikach to „przekonywanie do czegoś, wpływanie na postawę i poglądy odbiorcy” (Żmigrodzki 2009: 113). Definicje, niezależnie od rodzaju i strategii normatywnej, pełnią funkcję perswazyjną (Kucala 1985): próbują one przekonać, że w dany sposób się mówi albo powinno mówić; mogą to oczywiście czynić także inne elementy opisu słownikowego, na przykład kwalifikatory czy glosy.

Perswazja ta niekiedy napotyka jednak na opór użytkownika:

[S]e da por sentido que los diccionarios estipulan la técnica del habla; pero como esa técnica se ve interpretada por la imagen social de la lengua, toda estipulación lexicográfica contraria a la imagen compartida por los miembros de la sociedad tiene por efecto una reacción social en contra del diccionario. (Lara 1990: 137)

Przyjmuje się za pewnik, że słowniki określają technikę mowy; ale ponieważ ta technika jest interpretowana przez społeczny obraz języka, każde zastrzeżenie leksyko-graficzne sprzeczne z obrazem podzielanym przez członków społeczeństwa powoduje reakcję społeczną w opozycji do słownika.

Oczywiście perswazyjność ta może przyjmować różne postaci zależnie od założeń reaktorów słownika: w słowniku deskryptywnym będzie ona miała inny charakter niż w słowniku normotwórczym (na przykład ortograficznym), podobnie jak w przypadku słowników ogólnych i specjalistycznych (terminologicznych), czy jednojęzycznych i przekładowych (zorientowanych na dany język).

Do zagadnienia funkcji słownika można jednak podejść znacznie bardziej kompleksowo, co czyni Tarp (2008: 69–76) opisując potrzeby różnych typów użytkowników co do konkretnych rodzajów informacji zawartych w słowniku. Opis ten ma postać matrycy elementów słownikowych z informacją o ich przydatności do wykonywania poszczególnych zadań, których badacz uwzględnia cztery kategorie: odbiór, produkcję, sprawdzanie oraz ocenę (Tarp 2008: 75–76); w swej klasyfikacji dodatkowo wyróżnia funkcje pierwotne (**tabela 1**) oraz wtórne (**tabela 2**):

Tabela 1. Elementy słownika a jego funkcje pierwotne.

Rodzaj informacji	Odbiór	Produkcja	Sprawdzanie	Ocena
klasa wyrazu	X	✓	✓	✓
rodzaj gramatyczny	X	✓	✓	✓
warianty rodzajowe	X	X	✓	✓
zapis ortograficzny	X	✓	✓	✓
warianty ortograficzne	X	X	✓	✓
znaczenie lemm	✓	X	✓	✓
ograniczenia pragmatyczne i kulturowe	✓	✓	✓	✓
lokalne ograniczenia polityki językowej	X	✓	✓	X
odmiana	X	✓	✓	✓
warianty odmiany	X	X	✓	✓
nieregularne formy odmiany w funkcji lemm	✓	X	X	X
połączenia wyrazowe	X	✓	✓	✓
właściwości syntaktyczne	X	✓	✓	✓
kolokacje	X	✓	✓	✓
warianty kolokacji	X	X	✓	✓
idiomy	✓	✓	✓	✓
znaczenie idiomów	✓	X	✓	✓

Rodzaj informacji	Odbiór	Produkcja	Sprawdzanie	Ocena
przysłowia	✓	✗	✓	✓
znaczenia przysłów	✓	✗	✓	✓
warianty przysłów	✗	✓	✓	✓
standardowe wyrażenia	✗	✓	✓	✓
relacje semantyczne (synonimy, antonimy, hiponimy, hiperonimy)	✗	✓	✓	✓

Tabela 2. Elementy słownika a jego funkcje wtórne.

Rodzaj informacji	Odbiór	Produkcja	Sprawdzanie	Ocena
lemmy	✓	✓	✓	✓
połączenia wyrazowe w funkcji lemm	✓	✗	✓	✓
formy nieregularne w funkcji lemm	✓	✗	✓	✓
warianty ortograficzne w funkcji lemm	✓	✓	✓	✓
(błędy ortograficzne w formie lemm)	✓	✓	✓	✓
ortografia	✓	✓	✓	✓
klasa wyrazu	✓	✓	✓	✓
rodzaj gramatyczny	✓	✓	✓	✓
znaczenie lemm	✗	✓	✓	✓
znaczenie idiomów	✗	✓	✗	✗
znaczenie przysłów	✗	✓	✗	✗

Tak kompleksowe podejście jest przy tym bardzo generyczne i nie uwzględnia potrzeb szczegółowych użytkownika. Leksykograf musi zatem każdorazowo przy projektowaniu słownika brać pod uwagę ograniczenia funkcjonalne wynikające z cech słownika (rodzaju zawartych w nim informacji oraz struktury dostępu), a także – analogicznie – determinować własności leksykonu, które będą wynikać z jego przeznaczenia (jako że nie każdy rodzaj i układ informacji będzie w stanie daną funkcję realizować).

3. **Słownik warszawski – charakterystyka**

Przedmiotem opisu niniejszej pracy jest *Słownik języka polskiego* pod redakcją Jana Aleksandra Karłowicza, Adama Antoniego Kryńskiego i Władysława Marcina Niedźwiedzkiego, wydawany w latach 1900–1927 przy wsparciu Kasy im. Józefa Mianowskiego. Dzieło to składa się z ośmiu tomów i łącznie ma prawie 7 800 stron i jest szerzej znane pod nazwą *Słownik warszawski* – nazwy tej jako pierwszy użył najprawdopodobniej Ułaszyn:

Ze względu, iż znaczna większość współpracowników tego słownika są to warszawscy uczeni, że myśl ułożenia jego powstała w kółku językoznawców i literatów warszawskich, że ostatecznie miejscem druku i wydania tego słownika jest Warszawa – sądzę, że najodpowiedniejszą nazwą jego będzie: „Warszawski słownik języka polskiego” (...) (Ułaszyn 1900: 4)

Słownik warszawski (określany dalej *SW*) jest bardzo cennym zasobem, co będzie wykazane w dalszej części tego rozdziału, jednak mimo to nie posiada w literaturze językoznawczej zbyt wielu analiz ani omówień (Żmigrodzki 2009: 142). Na szczęście jego treść – nie licząc nigdy nieopublikowanego suplementu – została zachowana w całości, podobnie jak i część materiałów dotyczących jego powstawania.

Poniżej przybliżone zostaną sylwetki autorów, a także zostanie przedstawiona charakterystyka dzieła.

3.1. **Autorzy**

Redaktorami *SW* i zarazem jedynymi wskazanymi w samym dziele autorami są Jan Aleksander Karłowicz, Adam Antoni Kryński, Władysław Marcin Niedźwiedzki. Skład autorski *Słownika...* się jednak zmieniał: istotny udział w jego treść mieli Józef Peszke i Kazimierz Król, chociaż brakuje szczegółowych informacji o naturze ich wkładu w projekt (Majdak 2014: 93).

Do powstania *SW* przyczyniło się jednak znacznie więcej osób, a ich cele, które im przyświecały, były związane z odrodzeniem się narodu polskiego:

W parze z zaangażowaniem naukowym i społecznym szła odpowiedzialność wobec ojczyzny – służyli jej w sposób, który uznawali za słuszny: uczestnicząc w powstaniu styczniowym lub organizując pracę organiczną lub pracę u podstaw. Niektórzy byli usuwani przez rosyjskich wizytatorów, odmawiali wykładów po rosyjsku bądź rezygnowali z zajmowanych stanowisk. Cechowała ich ofiarność i żywa chęć pomocy naukowej, materialnej i osobistej. (Majdak 2008b: 241)

„Bezinteresowności” i „szlachetności” autorów i współpracowników dowodzi także niska cena *SW*, która pokrywała tylko niezbędne materiały i koszty wydania (Ułaszyn 1900: 4).

3.1.1. Jan Aleksander Karłowicz

Jan Aleksander Karłowicz (1836–1903) był inicjatorem przedsięwzięcia, jakim był *Słownik warszawski*. Kryński (1904: 164) pisał o jego roli w ten sposób: „K[arłowicz] był także początkodawcą, organizatorem i głównym redaktorem dzieła, mającego objąć cały słowozbiór mowy polskiej i wychodzącego obecnie p. n.”.

Karłowicz był wysoce płodnym uczonym – był autorem 3627 dzieł w 6435 tomach (Rak 2021a: 11). Jego twórczość objęła wiele dziedzin, w tym (Rak 2021a: 7f):

- historia dziejów;
- językoznawstwo;
- ludoznawstwo;
- muzyka;
- krytyka literacka;
- pedagogika;
- filozofia;
- publicystyka.

Badacz ten zajmował się także teorią i historią muzyki, przez krótki czas jej także uczył w konserwatorium muzycznym w Warszawie (Kryński 1903: 6–7). W 1866 r. obronił w Berlinie doktorat zatytułowany „De Boleslai Primi bello kiovensi” („O wyprawie kijowskiej Bolesława I-go”); rok później złożył pracę habilitacyjną pt. „Don Karlos, królewicz hiszpański, szkic biograficzno-historyczny” w Szkole Głównej w Warszawie, jednak nie udało mu się tam uzyskać katedry. Wówczas wrócił na Litwę, skąd pochodził, i poświęcił się wszechstronnym badaniom językoznawstwa, etnografii i mitologii (Kryński 1903: 6–24, 26). Brak formalnego związku z uniwersytetem nie był dla niego przeszkodą:

Na działalność naukową J. Karłowicza wpłynęły nie tylko jego wykształcenie uniwersyteckie, lektura opracowań i kontakty z innymi badaczami, ale także kondycja naszej nauki bezpośrednio powiązana z sytuacją historyczno-społeczną ziem polskich pod zaborami. (Rak 2021a: 15)

W wieku ok. 30 lat Karłowicz nauczył się litewskiego – trzymał bowiem na dworze dwóch Litwinów tylko w celu nauki języka (Rak 2021b). Jego zapał akademicki i przygotowanie językoznawcze były doceniane m.in. przez Kryńskiego, który tak pisał o *SWOKarł* (Kryński 1903: 25):

[P]raca (...) mająca rozjaśnić rodowody zawile tych wszystkich wyrazów obcych, które w mowie naszej uzyskały prawo obywatelstwa, ale uległy przytym tak znacznemu nieraz przekształceniu pierwotnych swych postaci, że odgadnienie ich dla najbieglejszego językoznawcy istną bywa zagadką. K. przystąpił do tej pracy uzbrojony znajomością wszystkich środków, jakimi rozporządza językoznawstwo dzisiejsze, i część jej znaczną wykonał.

Kryński (1903: 5) podsumowuje Karłowicza w ten sposób: „[J]eden z najznakomitszych uczonych polskich, znawca rzeczy ludowych, mitologii i językoznawca”. U Karłowicza cenił jednak nie tylko jego intelekt, ale i siłę charakteru: „Czczono w nim charakter prawy i umysł niepospolity człowieka, który w ciągu życia umiał cenić postęp, wiedzę i prawdę” (Kryński 1903: 26–27). To dzięki temu Karłowicz doprowadził do powstania *Słownika war-*

szawskiego, jego *opus magnum*. Jego sylwetka w umysłach ludzi nie była jednak jednoznaczna – jak zauważa Urbańczyk (1966: 55): „Jako językoznawca za życia przeceniany przez niefachowców, został po śmierci przez lingwistów zanadto zapomniany”.

Karłowicz doczekał publikacji tylko dwóch z sześciu tomów – po jego śmierci, która nastąpiła 14 maja 1903 r., udało się mimo wydać dzięki wkładowi innych autorów, o czym więcej pisze Rak (2021b: 34ff). Wpływ Karłowicza na polskie językoznawstwo i etnografię był jednak istotny, czego dowodem jest życiorys wydany tuż po jego śmierci: *Życie i prace Jana Karłowicza (1836–1903)*.

3.1.2. Adam Antoni Kryński

Adam Antoni Kryński (1844–1932) był badaczem, który mocno angażował się w tworzenie nauki polskiej i podwalin odrodzenia kulturalnego polski:

Wolał Kryński, jak wielu innych jego współtowarzyszów, ograniczyć zakres swojej działalności i skazać się na cięższe warunki naukowej pracy, niż się „przetapiać w obce bryły” i pomnażać obce kulturalne dobra. (Szober 1934: XI–XII)

Kryński najsilniej udzielał się w obszarze badań językoznawczych i etnograficznych. Najpierw jednak cztery lata studiował na wydziale matematyczno-fizycznym, dopiero potem zmienił kierunek studiów i wstąpił na wydział historyczno-filologiczny; rok studiował także w Lipsku (Szober 1934: VII–VIII).

Kryński był ważną osobą dla polskiej nauki. Założył on Seminarjum Filologii Słowiańskiej wraz z biblioteką na Uniwersytecie Lwowskim (Szober 1934: VIII). Od 1875 r. był członkiem Komisji językowej Akademii Umiejętności (Łopaciński 1901: 740). W roku 1884 założył także czasopismo „Prace Filologiczne”, pierwsze w kraju czasopismo językoznawcze (Szober 1934: IX).

Bardzo istotnym dziełem Kryńskiego była *Gramatyka języka polskiego* z 1897 r. (Szober 1934: XIII). Łopaciński (1901: 738) o gramatyce tej pisał w ten sposób:

[P]ierwszy raz pomyślano o tem, aby uczącej się młodzieży i szerszemu ogółowi uprzystępnąć wyniki wiedzy lingwistycznej w dziele, które treścią i sposobem wykładu do umysłu każdego czytelnika polskiego trafić powinno.

Kryński starał się uwzględnić w badaniach nad gramatyką także gwary i narzecza (Łopaciński 1901: 743). Pierwsza część książki, poświęcona „głosowni”, czyli fonetyce, była najmniej rozbudowana, co wynikało z faktu, iż badania w tym zakresie były wówczas w Polsce na bardzo wczesnym etapie (Łopaciński 1901: 743). Pomiął on także słowotwórstwo, składnię i nieodmienne części mowy w swojej pracy, ale jak zaznacza Łopaciński (1901: 742):

Nie o to jednak autorowi chodziło, aby książka jego ogromem podziw budziła, i nie z tego też punktu widzenia oceniać ją należy, ale z tego, czy i o ile wymaganiom naukowym odpowiada.

Kryński był jednak nie tylko gramatykiem, ale także encyklopedystą i leksykografem. Od 1890 r. udzielał się przy tworzeniu *Wielkiej Encyklopedji Powszechnej Ilustrowanej* (Szober 1934: IX). Doświadczenie tam zdobyte przydało mu się przy pracach nad *SW*, który współtworzył od 1898 r. (Szober 1934: IX).

To zapewne on miał wpływ na bezprecedensową ilość informacji gramatycznej w *Słowniku...* dzięki warsztatowi, jaki wniósł. Można też przypuszczać, że to on miał też wpływ na włączenie do hasłownika tak wielu form produktywnych i potencjonalizmów. Kryński był również proponentem zapisu zgodnego z wymową w swoich publikacjach, a nie archaizującego (Łopaciński 1901: 740) – to przełożyło się na pisowni jednostek poświadczonych wyłącznie ustnie. Oczywiście Kryński również opracowywał hasła do *SW* – były to przede wszystkim wyrazy staropolskie (Łopaciński 1901: 742).

W 1919 r. Kryńskiemu została nadana profesura honorowa na UW (Szober 1934: IX). U schyłku życia pracował nad Słownikiem Jana Kochanowskiego, którego nie zdołał ukończyć (Szober 1934: XIV). Był jednak autorem licznych prac, których bibliografię opracował Szober (1934: XVII–XXXIX).

3.1.3. Władysław Marcin Niedźwiedzki

Władysław Marcin Niedźwiedzki (1849–1930) był nauczycielem języka polskiego, zajmował się także językoznawstwem słowiańskim i etnografią tej części Europy (Szober 1930: 362–364). W 1891 r. Karłowicz i Kryński zaprosili go do współtworzenia *SW*.

Nad słownikiem pracował jednak tylko do tomu IV (1908), ponieważ po śmierci Karłowicza nie mógł dojść do porozumienia z Kryńskim w kwestiach merytorycznych i wydawniczych (Szober 1930: 363). Następnie, wraz z Antonim Krasnowolskim, opracował podręczny *Słownik staropolski* wydany w 1914 r. w 2 tomach przez Michała Arcta.

Niedźwiedzki miał według Szobera największy wkład w *SW*, ponieważ poświęcił się mu niemal absolutnie: „Oddany (...) całkowicie, oddalił się od ludzi i bieżącej literatury tak, że w ruchu naukowym ostatnich lat kilkudziesięciu nie brał zupełnie udziału” (Szober 1930: 364).

3.1.4. Pozostali

Zasadniczo trudno jest ustalić, kto w ogóle wchodził w skład redakcji wielkich przedsięwzięć leksykograficznych, nawet jeśli miał udział w zbieraniu materiału (Rossi & Čermák 2008: 145). W przypadku *SW* również nie było to łatwe. Nazwiska współpracowników *SW* miały się ukazać w ostatnim tomie, do tego wszak nie doszło.

Ułaszyn (1900: 4) wymienia jako współpracowników redakcji *SW* następujące osoby:

[P]rof J. Baudouin da Courtenay (słoworód wyrazów rodzimych i porównaczy spis pni), prof. Ad. Ant. Kryński (część staropolska), prof. J. Karłowicz (wyrazy gwarowe, zapożyczone i muzyczne), prof. Kalina, p. Niedźwiedzki, ś. p. prof. Łaguna i wielu innych.

Większość współtwórców była z Warszawy; przejawiali oni postawy społeczne i patriotyczne, za co często bywali prześladowani (Horodyska 1986: 158–159). Horodyskiej (1986: 151) udało się ustalić 30 nazwisk powiązanych z tym dziełem. Poza głównymi autorami badaczka wskazuje, że byli to (Horodyska 1989: 659):

- Jan Baudouin de Courtenay;
- Józef Bliziński;

- Józef Bojasiński;
- Florian Czepieliński;
- Samuel Dickstein;
- Karol Drymmer;
- Zygmunt Fiszer;
- Konstanty Górski;
- Edmund Jankowski;
- Antoni Kalina;
- Ludwik Korotyński;
- Henryk Kotłubaj;
- Stanisław Kramsztyk;
- Kazimierz Król;
- L. Leciejewski;
- Stosław Laguna;
- Hieronim Łopaciński;
- Jan Łoś;
- R. Łuszczkiewicz;
- Adam Mahrburg;
- Lucjan Malinowski;
- Leopold Meyet;
- Wacław Nałkowski;
- Jan Nitowski;
- Feliks Ochimowski;
- Józef Peszke;
- Antoni Semenowicz;
- Wacław Taczanowski;
- Jan Trejdosiewicz;
- Wawrykiewicz;
- Władysław Wiorogórski;
- Józef Wojno;
- August Wrześniowski;
- Władysław Wścieklica;
- Bronisław Znatowicz.

Wielu z nich nie było lingwistami, ale specjalistami z dziedzin, których wiedza była przydatna przy opracowywaniu słownictwa i terminologii w poszczególnych obszarach. Horodyska (1989: 661) pisze o nich tak:

[F]akt, że pomimo upływu lat i wielu przemian w nauce i w życiu narodu zachował w znacznym stopniu aktualność, wynika w jakimś stopniu z poziomu wiedzy i intelektu, jaki reprezentowali jego twórcy.

O wielu z wymienionych osób jest wiadomo stosunkowo niewiele. Szober (1930: 363) pisze np. o tym, że Wacław Taczanowski był filologiem i urzędnikiem Biblioteki Warszawskiej, który pomagał w opracowaniu SW, jednak tylko na początku, ponieważ zginął w 1904 r. podczas wojny rosyjsko-japońskiej. Wiadomo też, że Niedźwiedzki zaprosił do współtworzenia

tomów V i VI Kazimierza Króla. Dokładny zakres, w jakim poszczególne osoby były zaangażowane w tworzenie słownika, nie jest jednak znany.

3.2. Materiały pomocnicze

Słownik warszawski nie jest dziełem, które należy rozważać w izolacji – po pierwsze dlatego, że zostało ono oparte o wcześniejsze prace leksykograficzne Karłowicza, a po drugie dlatego, iż powstały materiały mające wspomóc proces redakcyjny, po trzecie natomiast zachowała się korespondencja, która nadaje dziełu temu szerszy kontekst. Materiały te są omówione poniżej.

3.2.1. Trzy piramidy

Ze względu na ich wartość dla polszczyzny, Erazm Majewski (1904: 24) trzy wielkie dzieła leksykograficzne Karłowicza – czyli *Słownik gwar polskich (SKarł)*⁸, *Słownik wyrazów obcego a mniej jasnego pochodzenia w języku polskim używanych (SWOKarł)* oraz *Słownik języka polskiego (SW)* – nazywa „trzema piramidami”, które dały podwaliny kształtowaniu się świadomości językowej Polaków (Rak 2021a: 8).

Karłowicz za rękopis „Słownik wyrazów cudzoziemskich, w języku polskim używanych, ze szczególnym względem na ich słoworód”, który stał się przyczynkiem dla jego *Słownika wyrazów obcych...*, otrzymał nagrodę konkursową im. Lindego (Karłowicz 1894–1905: wstęp).

Rak (2018) kartotekę nieukończonego słownika etnolingwistycznego, którą opracował Karłowicz, nazywa „czwartą piramidą”, a także opisuje jej historię oraz treść.

Należy zaznaczyć, że *Słownik warszawski* zawiera przy tym hasła z pozostałych piramid i większość zawartych w nich informacji – godnym uwagi przykładem jej redukcji jest nieuwzględnienie w *SW* informacji o pochodzeniu geograficznym wyrazów.

3.2.2. Prace dotyczące słownika

Sam *Słownik warszawski* nie posiada żadnego wstępu teoretycznego, objaśnień dotyczących procesu kompilacji czy wskazówek redaktorskich. Majdak (2006a: 231) wymienia jednak cztery źródła dostarczające wiedzy o powstawaniu *SW*:

1. *Przyczynki do projektu wielkiego słownika polskiego* z r. 1876, autorstwa Jana Karłowicza;
2. *Wskazówki dla współpracowników* z r. 1889;
3. *Zapowiedź* słownika;
4. Arkusz próbny.

Niestety nie wszystkie wskazane publikacje są dostępne:

Mimo wielu dyskusji nad kształtem słownika nie zamieszczono w nim wstępu zawierającego idee i reguły opracowania leksykograficznego. (...) Inne teksty, czyli

⁸ Więcej o *SKarł* pisze Karaś (1965).

Wskazówki dla współpracowników [1889] i *Zapowiedź* [1898], uważa się za zaginione. (Majdak 2014: 97)

Poniżej znajduje się podsumowanie informacji, które są znane w odniesieniu do wymienionych publikacji.

3.2.2.1. Przyczynki do projektu...

W literaturze można spotkać się z opiniami, że publikacja *Przyczynki do projektu wielkiego słownika polskiego* Jana Karłowicza jest wstępem do *Słownika warszawskiego*. Majdak (2008b: 242) przytacza silne argumenty wskazujące na to, że są to odrębne dzieła – kluczowy jest również fakt, że w *SW* ostatecznie metodologia była znacznie mniej rygorystyczna niż ta, którą Karłowicz planował zastosować w *Wielkim słowniku polskim* (Majdak 2006a: 231f; 2008b: 242).

Przyczynki... opiewają na 82 strony i mają formę zbioru notatek podejmujących różne zagadnienia związane z tworzeniem słownika.

3.2.2.2. Wskazówki dla współpracowników

Kryński (1904: 164) wymienia *Wskazówki dla współpracowników*, publikację „obrabiającą część główną słownika”, mającą zaledwie trzy strony. Do dzieła tego odnosi się także Majdak (2006a: 232):

Źródło drugie – *Wskazówki dla współpracowników* z roku 1889 – to opracowana przez Karłowicza instrukcja, „szczegółowy plan słownika wraz z metodą jego układu”. Powstała w wersji pisanej i drukowanej, niestety jak dotąd nie udało się jej odnaleźć.

3.2.2.3. Zapowiedź

Majdak (2006a: 232; za: Kryński 1904: 164) wskazuje także na istnienie *Zapowiedzi*, której treść również jest utracona:

Ostatni tekst to czterostronicowa *Zapowiedź* mającego się ukazać słownika: „uwiedamiam – jak pisał Kryński – o rozmiarach dzieła, współpracownikach, warunkach przedpłaty, oraz podaje na dwóch stronicach opracowanie wyrazów słownika od A do Ablegieriek”. Tej broszury również nie odnaleziono.

3.2.2.4. Arkusz próbny

Drugim zachowanym w całości tekstem źródłowym Arkusz próbny, o którym pisze Majdak (2006a: 232):

Kolejny dokument – Arkusz próbny – to kilkunastostronicowa publikacja zachowana do dziś w kilku egzemplarzach. Zawiera przedmowę, wykaz skrótów oraz próbkę samego słownika. Arkusz rozsyłano do znawców, do gazet w celu zebrania opinii na temat prezentowanej próbki dzieła. Jest to właściwie jedyne źródło wiedzy o zamierzeniach autorów.

Kryński (1904: 164) podaje nawet szczegółową budowę tego dzieła:

Arkusze ten zawierał przedmowę (str. 3—6), wykaz skrótów i wyrazy od B do Bakiera (str. 9—16).

Publikacja ta posiada hasła w formie praktycznie identycznej do tej, w jakiej opublikowane zostały pełnoprawne tomy.

3.2.3. Korespondencja

Istotnym, jeśli nie najważniejszym źródłem metawiedzy o *SW* jest korespondencja między jego redaktorami, a także między autorami a współpracownikami czy osobami niezwiązanymi bezpośrednio z przedsięwzięciem. Można się z niej wiele dowiedzieć o koncepcji słownika, procesie kompilacji, sposobie opracowania haseł, kształtowaniu się użytej pisowni, a także o samych redaktorach i współpracownikach.

Korespondencja ta nie będzie jednak tutaj ona analizowana, ponieważ temat ten został wyczerpująco podjęty przez Majdak (2006a).

3.3. Założenia

Trudno jest dokładnie i jednoznacznie określać, jakie założenia przyświecały autorom *Słownika warszawskiego*, ponieważ:

1. w samym dziele brak bezpośredniej informacji o założeniach redakcyjnych;
2. dokumentacja o słowniku jest fragmentaryczna i częściowo dotyczy odrębnych dzieł (np. *Przyczynki...*, które odnoszą się do niepowstałego *Wielkiego słownika polskiego*);
3. historia powstawania *SW* była burzliwa.

Istnieje jednak wiele wskazówek co do tego, jakie cele przyświecały autorom – te są zarówno wyrażone przez nich samych, a także ukryte w treści zasadniczej samego słownika.

Jedną z pierwszych rzeczy, co do których musi zostać przez redaktorów podjęta decyzja, jest wybór materiału źródłowego. Do niedawna podstawową formą zbierania materiału stanowiła ekscerpacja, czyli wypisywanie przykładów wyrazów napotkanych w tekstach, najczęściej na osobnych kartach (fiszkach), często wraz z kontekstem i innymi informacjami (Zgusta 1971: 225ff); metodę tę zastosowano także w przypadku omawianego dzieła. Istnieją dwa rodzaje ekscerpacji:

- **ekscerpacja totalna** – polega na zebraniu całego możliwego materiału, a następnie przystąpienie do jego selekcji, odrzucając karty, które nie zawierają nowych informacji. Jest to najczęstszy jej rodzaj w przypadku słowników ogólnych;
- **ekscerpacja częściowa lub specjalna** – charakteryzuje się przyjęciem kryteriów początkowych i preselekcją. Stosowana jest ona do pozostałych rodzajów słowników.

Ekscerpacja totalna nie była możliwa w przypadku *SW*, jednak o kryteriach preselekcji można wnioskować jedynie na podstawie użytych źródeł; nie wszystkie źródła zostały jednak wymienione, a te, które zostały, nie podają szczegółowej lokalizacji.

Nierzadko też źródłem słowników są inne słowniki. Jak zwięźle podsumował ten proces Reed (1962: 95): “borrowing — even plagiarism — is no sin to lexicographers” (‘pożyczanie – a nawet plagiat – nie jest grzechem dla leksykografów’). W przypadku czerpania z innych słowników nadal należy wykazać się podejściem krytycznym. Historia pokazuje jednak, że materiał leksykograficzny bywa powielany bezrefleksyjnie, co miało miejsce przed stuleciem, a co ma miejsce także i dziś. I *SW* czerpał wprost z poprzedników, tj. *SL* i *SWil*.

Kiedy materiał źródłowy jest dobrany i przygotowany, słownikarz stoi przed dylematem, jakie jednostki włączyć do słownika, a jakie odrzucić – kryteria te mogą być merytoryczne i związane jedynie z kwestiami językowymi, ale przeważnie wynikają także z ograniczeń wielkości słownika. Opis leksykograficzny jest dodatkowo trudny, bo wiąże się z opisem „możliwości wyrazowych”, a słownik to kolekcja „produktów językowych”, co czyni ze słownika „próbę empiryczną” języka jako fenomenu społecznego (Lara 2001: 149–151). Użytkownicy języka mogą tworzyć jednak jednostki przez leksykony niezarejestrowane – fenomen ten opisuje Haspelmath (2002: 39):

Dictionaries can record only the actual words, but at any time a speaker may use a possible (but non-actual) word, and, if it is picked up by other speakers, it may join the set of actual words (...).

Słowniki mogą odnotowywać tylko rzeczywiste słowa, jednak w każdej chwili mówiący może użyć możliwego (ale nierzeczywistego) słowa, a jeśli zostanie podchwycone przez innych mówiących, może dołączyć do zbioru rzeczywistych słów (...).

Słowniki zajmują się przyjętymi rezultatami produktywności będącej przejawem kompetencji derywacyjnej, którą definiuje Pena (1994–1995: 164)

Esta competencia derivativa puede definirse como un conjunto de reglas, es decir, de relaciones que unen de manera regular las propiedades formales y semánticas de una palabra derivada con las de la palabra base de derivación.

Tę kompetencję derywacyjną można zdefiniować jako zbiór reguł, to znaczy relacji, które regularnie łączą formalne i semantyczne właściwości słowa pochodnego z właściwościami słowa podstawowego.

Można zatem mówić o jednostkach (jeszcze) niezrealizowanych lub (jeszcze) niepoświadczonych, ale z punktu widzenia systemu możliwych i zgodnych z jego regułami; każda taka jednostka to tzw. forma potencjalna, czyli:

[F]orma fleksyjna (lub derywat), możliwa do utworzenia przy zastosowaniu ogólnie obowiązujących reguł fleksyjnych lub słowotwórczych, ale o małej (czasem niezwykle małej) zaświadczonej frekwencji tekstowej (Żmigrodzki 2008: 168)⁹.

Zgólkowa (1994: 367) zauważa, iż zachodzi także sytuacja odwrotna, to znaczy istnieją w systemie językowym wyrazy, które nie zostały zarejestrowane:

Czynne słownictwo współczesnego Polaka to hasła autentycznie używane (choć niekoniecznie przez wszystkich), a nie tylko potencjalnie obecne w źródłach leksykograficznych i ewentualnie biernie rozpoznawalne.

⁹ Podczas cytowania miejsce zamknięcia nawiasu zostało poprawione względem oryginału.

Wynika z tego, że podstawa materiałowa nie może być wyznacznikiem istnienia jednostki. Zgodna z tym twierdzeniem jest opinia Saloniego (1992), który pisze:

Empiryczne potwierdzenie istnienia danej jednostki w tekście (...) nie może być jedynym dowodem jej istnienia.

Żmigrodzki (2008: 71) sugeruje, że twory potencjalne powinny być odnotowywane, ale dobrze byłoby brak ich poświadczenia w jakiś sposób zaznaczać.

Istnieje także inny powód, dla którego formy potencjalne bywają pomijane, a jest nim przewidywalność znaczenia. Podejście takie krytykuje Azpiazu Torres (1999), która podejmując problematykę opisu przysłówków odprzymiotnikowych w języku hiszpańskim proponuje gwoźli eksplicytności informacji wypisywać wszystkie formy, w przeciwnym razie nie wiadomo, czy wynika to z faktu, że (1) twór jest regularny i nie posiada żadnej nieoczekiwanej zmiany w znaczeniu, czy (2) wyraz został pominięty przez zaniedbanie. Stosuje się to również do polskiego: odsłownik „robienie” jest wyłącznie nazwą czynności czasownika „robić”, podczas gdy wyraz „pranie” oznacza dodatkowo obiekt, który był (albo zostanie) poddany czynności opisywanej przez czasownik „prać”, co jest zjawiskiem nieprzewidywalnym w rozumieniu precyzyjnych oczekiwań w stosunku do systemu.

Nie wszystkie jednostki można jednak uwzględnić – ograniczenia nie płyną jednak wyłącznie z samego materiału źródłowego i kryteriów selekcyjnych *per se*, ale także z ograniczeń samego medium: z jednej strony słowniki powinny być bowiem dokładne i duże, ale z drugiej konieczna jest selekcja informacji, aby były używalne – uwzględnienie zbyt wielu jednostek może powodować szum informacyjny (Crystal 1986: 75).

Zgusta (1986: 142) wskazują na problem ze skróconymi słownikami – dobór słownictwa według samej tylko frekwencji doprowadza do sytuacji, w której trafiają do niego tylko wyrazy, które każdy zna, co obniża jego użyteczność przynajmniej w niektórych obszarach. Pod pewnymi względami ich przeciwieństwem są tezaury, o których pisze ten sam badacz:

A big dictionary which tries to be exhaustive is called a thesaurus. When we take into consideration what we know about the constant innovations caused in language by new applications of lexical units, we are not surprised that it is only the dead languages (including, of course, languages like Middle English) which is represented by a finite set of texts and which, therefore, can be studied, analyzed, and presented in its totality. (Zgusta 1971: 217)

Duży słownik, który stara się być wyczerpujący, nazywa się tezaurem. Gdy weźmiemy pod uwagę to, co wiemy o ciągłych innowacjach wprowadzanych w języku przez nowe zastosowania jednostek leksykalnych, nie dziwi nas, że tylko martwe języki (w tym oczywiście języki takie jak średnioangielski) są reprezentowane przez skończony zbiór tekstów, które w związku z tym mogą być badane, analizowane i prezentowane w całości.

W kontekście tego rodzaju słowników istotny staje się problem gęstości haseł, czyli negocjacji między kompletnością i mocą generatywną słownika (ta nie jest istotna dla języków martwych, gdyż w ich przypadku nowe teksty nie powstaną), która może być źródłem wspomnianego wcześniej szumu informacyjnego (Zgusta 1971: 243); ten szum zredukować mogą jednak strategie kompresji informacji, które zostały wspomniane wcześniej. Leksykograf zawsze musi znaleźć równowagę między zwięzłością a pełnością – jak ujmuje to Lombard

(1994: 210): „One of the great challenges confronting lexicographers is to give a comprehensive yet coherent description of language” (‘Jednym z wielkich wyzwań stojących przed leksykografami jest przedstawienie wyczerpującego, ale spójnego opisu języka’). W przypadku jednotomowego słownika należy rozważyć, co ma do niego wejść, a co nie (Filipec 1975: 228f).

W przypadku *Słownika warszawskiego* zdają się być odziedziczone po *Wielkim słowniku polskim* zamiary stworzenia „obszernego i dokładnego ile możliwości skarbcza mowy naszej”, który miał zawierać „całość zasobu mowy naszej” (Karłowicz 1985: 3; por. Majdak 2008b: 242). W zamierzeniu *SW* (a właściwie *Wielki słownik polski*) miał być „przeznaczony dla ogółu inteligencji, pisarzy, poetów, czytelników literatury folklorystycznej, dla pracowników nauki, językoznawców” (Horodyska 1986: 153). Bańko (2014: 96) twierdzi, że *SW* ten był „zaplanowany jako popularna publikacja referencyjna średnich rozmiarów”, w *Wielki słownik polski* miało zostać włączone całe zgromadzone słownictwo, a który miał być jednocześnie oparty o *SL* i *SWil*:

Staraliśmy się pomnożyć słowniki Lindego i Wileński całym zasobem opuszczonych tam, lub nowoprzybyłych wyrazów, czerpiąc ze znacznej ilości dzieł drukowanych, z dostarczonych nam przez licznych współpracowników notat rękopiśmiennych i z pamięci. (Karłowicz 1895: 4)

W praktyce jednak oba słowniki referencyjne nie były wystarczającą podstawą, więc autorzy czerpali też bezpośrednio z nowych źródeł. W rezultacie *SW* zawiera leksykę od polszczyzny średniowiecznej – choć w niepełnym wymiarze – aż po w. XIX, gdzie jest reprezentowana w pełnym zakresie stylistycznym Piotrowski (1993: 575).

Powstało jednak dzieło kompromisowe, średniej wielkości, ponieważ redaktorzy nie byli w stanie zawrzeć całego zgromadzonego materiału (Karłowicz 1985: 3). Niemniej zawierało ono jednostki z różnych epok i odmian stylistycznych polszczyzny, co odzwierciedlało założenia Karłowicza:

[Ż]aden język, ani sam przez się, ani w stosunku do innych, bliżej lub dalej z nim spokrewnionych, nie może być gruntownie zbadany bez ścisłego poznania go w czasie i przestrzeni, to jest w jego historycznym rozwoju i miejscowych objawach, czyli, mówią po prostu: bez jego dziejów i form prowincjonalnych, oraz nazw ludzi i miejsc. Historia języka ma już gotowe źródła drukowane i pisane, bo się zajmuje przeszłością mowy; ze zbiorów więc gwar ludowych pośrednio tylko i częściowo korzysta. Opis zaś języka dzisiejszego w jego przestrzeniowych odmianach żyje wyłącznie i jedynie z zasobów żywych, ze zbiorów rzeczy ludowych. (Karłowicz 1871: 4)

Jednocześnie w formie odrębnych zawartych było wiele jednostek regularnych, czysto produktywnych, czasem wręcz potencjonalizmów, takich jak odsłowniki, abstrakcyjne rzeczowniki żeńskie zakończone na *-ość*, zdrobnienia itd., a także wiele wariantów. Trudno jest stwierdzić, czy wynikało to z chęci mechanicznego zwiększenia objętości słownika, czy z czegoś innego – być może autorzy nie chcieli poprzez pominięcie danych form nieumyślnie zasugerować, że one nie funkcjonują, ani utrudnić ich wyszukania poprzez gniazdowanie. Nawigację w tak obszernym dziele ułatwia na szczęście bogaty system kwalifikatorów (Majdak 2014: 95).

Wśród jednostek, które były celowo opuszczane, znajdują się przede wszystkim informacje typowo encyklopedyczne (co jednak nie było czynione w pełni konsekwentnie):

[O]puszczamy imiona własne, mitologiczne i geograficzne, a także nazwy herbów i t. p. i wogóle takie, które właściwie należą nie do słownika, lecz do encyklopedji, lub do słowników specjalnych. (Karłowicz 1895: 5)

Jeśli chodzi o przeznaczenie, to *Wielki słownik polski* miał być dziełem „czysto naukowym” (Karłowicz 1876: 6). *SW* taki do końca nie jest – posiada zarówno cechy dzieła naukowego, jak i praktycznego (popularnego) (Majdak 2014: 95). Jednocześnie słownik ten stosuje podejście proskryptywne, które polega na empirycznym (deskryptywnym) opisie form istniejących (niezależnie od ich poprawności) przy jednoczesnym (preskryptywnym) różnicowaniu form zalecanych i odradzanych podle panujących standardów (Bergenholtz 2003: 77; Bergenholtz & Gouws 2010: 36). *SW* nie miał wartościować, tylko zebrać całość bogactwa polszczyzny – dopiero na tym materiale można by, w założeniu autorów, opracować słownik normatywny (wartościujący) mniejszych rozmiarów (Piotrowski 1993: 574).

To podejście do normy jest istotne nie tylko badawczo, ale i w kontekście społecznym. Бабаев (2013: 15) zauważa, że leksykografia normatywna odzwierciedla pewien etap rozwoju języka, i postuluje:

[В]ечное стремление литературного языка к избавлению от дублетности и вариантности, но, в тоже время, именно дублетность и вариантность являются стимулом к оптимизации языковой системы.

[O]dwieczne dążenie języka literackiego, aby pozbyć się dubletów i wariantów, ale jednocześnie dublety i warianty są bodźcem do optymalizacji systemu językowego.

Wspomniane wcześniej kategorie można pogrupować także na **wpływające na standard języka** (podejście preskryptywne i proskryptywne – mające charakter produktywny i wpływające na użycie) oraz **niewpływające** na niego (podejście deskryptywne – opis całego języka, bez narzucania norm). Zgusta (1971: 210) opisuje tę pierwszą kategorię jako *standard-descriptive dictionaries*, a tę drugą mianem *overall-descriptive* lub *informative dictionaries*, jednak w swojej późniejszej pracy (Zgusta 1989) rozszerza swoją klasyfikację i dzieli słowniki preskryptywne na podklasy:

1. słowniki, których celem jest stworzenie standardu pisanego (ang. *standard-creating dictionaries*):
 - a. tworzące nowy standard;
 - b. próbujące reanimować martwy język;
 - c. próbujące wywyższyć odmianę językową do rangi standardu.
2. słowniki starające się uwspółcześnić standard (ang. *modernizing dictionaries*);
3. słowniki starające się powstrzymać zmiany w standardzie języka, a nawet niektóre z nich cofnąć (ang. *antiquating dictionaries* lub *archaizing dictionaries*);
4. słowniki opisujące istniejące standard, objaśniające go (ang. *standard-descriptive dictionaries*).

Słownik warszawski zdaje się mieć w tym zakresie cechy dzieła (2) starającego się uwspółcześnić standard, (3) starającego się powstrzymać niektóre zmiany, ale przede wszystkim (3) opisać standard już istniejący.

W centrum uwagi redaktorów *SW* była także normatywność ortograficzna – Karłowicz napisał we wstępie do tomu I (1900) dzieła w ten sposób:

Wobec niejednostajności pisowni drukowanych dotąd przyczynków gwarowych, zachodziła trudność w wyborze jednolitej przepisowni (transkrypcji). Miałem do wyboru: albo powtarzać bez zmian pisownię przyczynków, albo też obmyślić jakąś jednostajną, normalną, powszechną. Wybrałem tę drugą drogę, mniemając, że oszczędzi ona *Słownikowi* nieuniknionej w przeciwnym razie pstrocizny, któraby była tylko zawadą dla czytelników.

Natomiast według Karasia (1961: 359) to dzięki szkole Malinowskiego:

[Z]aczęto doceniać postać fonetyczną wyrazów, w rozmaity zresztą sposób; oddaje się ją początkowo przy pomocy pisowni literackiej, później zaś wprowadza nowe odpowiednie znaki. Od tego też czasu możemy mówić o nowoczesnej dialektologii, a więc i o leksykografii gwarowej.

Przyjęta strategia zapisu była też rezultatem trwającego cztery lata (1892–1896) sporu o reguły ortograficzne, w rezultacie którego zostały przyjęte propozycje wysunięte przez Kryńskiego, a nie te forsowane przez Akademię Umiejętności. Nie wygrał on jednak na każdym polu: Kryński nie mógł Kopczyńskiemu – który był uważany za „prawodawcę w dziedzinie gramatyki i pisowni” – darować zasady pisowni końcówek *-ym*, *-em*, *-ymi*, *-emi* w zależności od końcówki mianownika; regułę tę zniesiono dopiero w 1936 r. (Doroszewski 1966: 308).

Z zasady Karłowicz unikał zapożyczeń i poniekąd starał się oczyścić polszczyznę (Łempicka 1958: 204–206). Pisał on tak:

Wyrazy i wyrażenia, brane żywcem z języków obcych, a uczuwane jako barbaryzmy, usuwamy ze słownika; takie zaś, które mają tylko pozory polskości, a nie powinny być używane, odznaczamy wykrzyknikiem. (Karłowicz 1895: 5)

Obszerniej zagadnienie to traktuje Rak (2021a: 20):

W XIX w. w trosce o polszczyznę i polskość terminologii naukowej unikano bezpośrednich zapożyczeń z języków obcych (wyjątkiem jest *folklor*), ustalając rodzimy zasób terminów albo dbając przynajmniej o ich polską formę, stąd ludoznawstwo i językoznawstwo jako kalki niemieckich *Volkskunde* i *Sprachkunde*. Mając do wyboru *dialekt* i *gwarę*, *dialektologię* i *gwaroznawstwo*, *lingwistykę* i *językoznawstwo*, *etnografię* i *ludoznawstwo*, które właściwie znaczyły to samo, Karłowicz częściej posługiwał się drugim elementem z tych par (słownik gwarowy równie dobrze mógł być zatytułowany *Słownik dialektów polskich*).

Podobne tendencje widać zresztą w wielu słownikach specjalistycznych z tego okresu, nad którymi pracował piszący te słowa. Brückner (1904: 170) podkreśla jednak, że Karłowicz nie przejawiał postawy absolutystycznej w kwestii pożyczek językowych:

Źle zrozumiany patryjotyzm – raczej szowinizm obłądny – uważa niby za ubliżenie godności własnego języka, skoro się w nim zbyt wiele obcych żywiołów wyławia. Podobna skrajność była Karłowiczowi zawsze obcą (...)

W *SW* realizowana była idea językoznawstwa, które nie ingeruje w naturalny rozwój języka (Doroszewski 1954: 44). Na zasadniczo nienormatywny charakter planowanego słownika wskazuje też rozumienie samej lingwistyki przez Karłowicza (1876: 6f)

Językoznawstwo, jako nauka, zapatruje się na język jako na fakt dokonany, jako na swój materiał nie zna i znać nie powinno różnicy wartościowej pomiędzy faktami, to jest nie ma prawa uważania jednego objawu językowego za błędny, drugiego za poprawny, tak jak np. botanika nie ma prawa decydowania o piękności lub szpetności kwiatów, o przyjemności lub wstrętności ich zapachu.

Porównanie językoznawstwa do botaniki – w opozycji do dbającego o praktyczny pożytek ogrodnictwa – cieszyło się popularnością wśród językoznawców w XIX i XX w. (Doroszewski 1954: 39). Karłowicz również realizował idee pozytywizmu i ewolucjonizmu (Okoniowa 2013: 246), wówczas nowoczesne, jednak które dzisiaj są krytykowane (Quirk 1986: 4).

Brückner (1904: 165) pisał, że Karłowicz nie chciał opisać ogólnych reguł języka, a studiować jego pojedyncze elementy (zwłaszcza słowa). Przy tworzeniu ich inwentarza – czyli *Słownika warszawskiego* – autorzy starali się też wykorzystać najnowocześniejsze dokonania leksykograficzne w Europie – część z nich był zaimplementowana przez samego Karłowicza przed rozpoczęciem prac nad swoim *opus magnum*:

Historia powstania SKarł jest bezpośrednio powiązana z planami leksykograficznymi Akademii Umiejętności. Jak wiadomo, wraz z utworzeniem tej instytucji pojawił się pomysł, by koordynowała ona (a dokładniej działająca w jej obrębie Komisja Językowa Wydziału Filologicznego) prace nad słownikiem języka polskiego (roboczo nazywanym Słownikiem Akademii Umiejętności), wykorzystującym rozwiązania leksykograficzne francuskie (DLF) oraz niemieckie (DWB) i znacznie poszerzonym w porównaniu ze słownikiem Samuela B. Lindego (SL) i Słownikiem wileńskim (SWil) (Rak 2021a: 23ff)

Definicje w *SW* mają jednak z reguły charakter opisowy, a nie synonimiczny, jak to w dużym stopniu ma miejsce w *DWB*, a także znacznie późniejszego *SSJČ*. Kolejność znaczeń podawana jest natomiast jak w *DLF*, czyli najpierw podawane jest znaczenie pierwotne wyrazu. Istotnym postulatem w *SW* było też podawanie jak najbardziej wyczerpujących informacji (Doroszewski 1954: 46).

Autorzy podawali też informację gramatyczną, zwłaszcza morfologiczną, która opisywała poszczególne leksemy lub ich wybrane znaczenia. Nie podawano jednak pełnego repertuaru form fleksyjnych, a jedynie oznaczano tylko nieoczywiste końcówki i przynależności do części mowy (Karłowicz 1895: 6).

Podobnie było w przypadku informacji o wymowie – wskazówki były podawane tylko w sytuacji, gdy polska ortografia jest niejednoznaczna dla rodzimego użytkownika języka polskiego.

O ile prowincjonalizmy stanowią integralną część *Słownika warszawskiego*, to niestety przełom metodologiczny w gwaroznawstwie, który nastąpił po roku 1901, nie przeniknął do niego. Poza tym traktowanie morfologii gwarowej w *SW* jest niepełne, a czasem wręcz błędne (Okoniowa 2013: 251–253).

Zasadniczo powyższe założenia udało się w *SW* zrealizować, choć nie zawsze w stopniu zadowalającym, co po części wynikało z dużych ambicji i problemów warsztatowych, a po części z trudnego przebiegu prac nad tymże dziełem.

3.4. Powstawanie

Zwięzłego podsumowania bogatej historii *SW*, w tym jego powstawania, dokonała Majdak (2008b). Autor teŝe pracy doktorskiej pozwoli sobie nakreślić jedynie najważniejsze kwestie, które odbiły się na ostatecznym kształcie dzieła.

Piotrowski (1993: 574) proces publikacji streszcza następująco:

Prace nad tym słownikiem rozpoczęto w 1889 r., a finansowała je Kasa im. Mianowskiego oraz prenumeratorzy. Zeszyty tomu I zaczęły wychodzić w 1898 r., cały tom I ukazał się w 1900 r., a tom ostatni, VIII, wydrukowano w 1927 r. (przedruk Warszawa 1952-1953).

Ze względu na sytuację polityczną, finansowanie przedsięwzięcia miało charakter oddolny. Należy tu wymienić trzy główne źródła finansowania, które miały różny udział na różnych etapach (Rak 2021a: 17):

1. założona w Warszawie w 1881 r. Kasa im. Mianowskiego, stanowiąca główne źródło finansowania;
2. wkład własny Karłowicza;
3. mecenat prywatny.

Jak zostało to wcześniej wspomniane, istotną część *SW* stanowi *SKart*:

Karłowicz przygotował słownik gwarowy samodzielnie we wszystkich aspektach – organizacyjnym, naukowym i finansowym. (Rak 2021a: 31)

Chociaż Karłowicz pisał, że w dziele tym udział mieli teŝ inni, to jednak miał na myśli tylko dostarczane przez nich materiały. Kartoteki pozostałych „piramid” również zostały wykorzystane podczas kompilacji *SW*.

Sam proces wydawniczy trwał niemal trzy dekady, podczas których wydanych zostało 48 zeszytów – pierwszy ukazał się w roku 1898. Zeszyty te razem składają się na 8 tomów. Oryginalne założenia przewidywały tylko 4 tomy, a prace były rozpisane na 5 lat (Majdak 2014: 98).

Dopóki żył Karłowicz, prace redakcyjne były prowadzone w jego mieszkaniu, gdzie teŝ przechowywana była kartoteka i zgromadzone materiały (Horodyska 1986: 164–165). Po jego śmierci mecenat naukowy przejmowały kolejne osoby, ale ostatecznie udało się doprowadzić słownik do końca. Niestety, jak pisze Majdak (2014: 98):

[Z] powodów finansowych i kadrowych (redaktorem był już wówczas tylko W. Niedźwiedzki) opis etymologiczny doprowadzono tylko do hasła **Szmerglować** w tomie VI.

Karłowicz miał także zamiar publikowania informacji o postępach w pracach nad *Słownikiem warszawskim*, o czym informuje Doroszewski (1954: 47):

Bardzo słuszna natomiast była myśl Karłowicza o niezbędności w pracy nad słownikiem specjalnego czasopisma sprawom tej pracy poświęconego. Takie czasopismo, dla którego Karłowicz proponował nazwę „Przyczynki do Słownika Polskiego”, ukazywałyby się nie w ściśle oznaczonych terminach, zeszyty zaś jego zawierałyby

bibliograficzne spisy materiałów, same materiały, spisy i objaśnienia wyrazów, pni i „rostków” (afiksów).

Strategia taka została zaimplementowana później przez Kancelář Słovníku jazyka českého – instytucję, której praca doprowadziła do powstania *Akademického slovníku*, a której spadkobiercą stał się Ustav pro jazyk český – instytucja regulująca język czeski (Filipec 1981: 225). W ramach swojej działalności publikowała ona dwie serie artykułów dostępnych w czasopiśmie *Naše řeč*: „Z kanceláře Słovníku jazyka českého” (1929–1932) oraz „Zpráva z kanceláře Słovníku jazyka českého” (1935, 1937–1942). W owych artykułach znalazły się informacje o postępach w pracach nad słownikiem (podsumowywano liczbę opracowanych fiszek i źródeł) oraz imionach i nazwiskach współpracowników, a także o poszukiwaniu dalszych ekscerptorów i kwotach, jako oferowano za pomoc w przedsięwzięciu.

Podczas II wojny światowej zaginął prawie cały nakład *SW*, co spowodowało, że wznowiono jego wydanie (w formie faksymilowej) wraz z *SL*, aby zapobiec utracie obu dzieł (Doroszewski 1954: 37).

3.5. Supplement

Redaktorzy *SW* po wydaniu pierwszych ośmiu tomów, obejmujących słownictwo zarejestrowane w okresie do I wojny światowej, zaczęli opracowywanie suplementu, który miał uzupełnić to dzieło o wcześniej pominięte hasła, jak i nowe słownictwo, które pojawiło się w polszczyźnie. Do suplementu zgromadzono 60 000 haseł, cała kartoteka zaginęła jednak podczas II wojny światowej i uważa się ją za zniszczoną (Piotrowski 1993: 577f).

Próby opracowania suplementu do *SW* podjął się jednak Doroszewski wraz ze swoimi studentami (Skarżyński & Smułkowa 2018: 485), prace te jednak przeobraziły się w odrębny projekt leksykograficzny, który doprowadził do powstania *SJPDor*. Zdaje się to być dobrą decyzją, jako że suplementy do słowników – zwłaszcza innych autorów – są w literaturze polonistycznej mocno krytykowane.

Lubaś (1994) pisząc o suplementach do słownika Szymczaka oraz Doroszewskiego zarzuca im, że popełniają te same błędy, co dzieła oryginalne, jednocześnie nie wnosząc prawie żadnego metodologicznego novum. Wtórują mu słowa Puzyriny (1994: 381–382) o słowniku Szymczaka:

W sumie otrzymujemy mozaikę, na którą składa się to, co nowe, i to, co stare, to, co całkiem obce, to, co specjalistyczne lub książkowe, i to, co potoczne, pospolite lub wulgarne.

Bańko (1995) uzupełnia, że publikacja „Polszczyzna, jaką znamy” Bogusławskiego nie jest suplementem do żadnego konkretnego dzieła, a jedynie „rejestruje jednostki leksykalne pominięte w większości słowników polskich”. W dopisku do tegoż artykułu Lubaś zaznacza, iż według niego tworzenie uzupełnienia do słownika innego autora nie może być zasadne.

Podobna krytyka najprawdopodobniej miałaby zastosowanie do suplementu do *SW*, zwłaszcza jeśli wziąć pod uwagę, jak fundamentalnie różne były koncepcje leksykograficzne Karłowicza i Doroszewskiego.

3.6. Materiał źródłowy

Materiał źródłowy, na podstawie którego powstał *SW*, był nie tylko obszerny, ale także zróżnicowany pod względem typów tekstów, które zostały wykorzystane.

Pierwszym rodzajem źródeł, które posłużyły do stworzenia *Słownika warszawskiego*, były poprzednie prace leksykograficzne Karłowicza lub ich kartoteki. Należy tu wspomnieć, że kartoteki *SKarl* i *SW* były częściowo pomieszane, ponieważ pochodzą z pudeł *Wielkiego słownika polskiego* (Rak 2021b). Fiszki te miały dużo bardziej szczegółowe informacje o lokalizacji terminów prowincjonalnych i gwarowych, niż zostało to odnotowane w *SW* (Rak 2021b).

Drugim rodzajem źródła są dwa słowniki referencyjne. Karłowicz dokonuje nawet fragmentarycznego porównania siatek hasłowych: tam, gdzie *SL* ma 228 haseł, *SWil* ma 321, a *SW* aż 570. Autorzy *SW* podważają też zasadność niektórych haseł w pozostałych wymienionych dziełach – głównie nazw własnych, których do swojego leksykonu nie włączyli (Karłowicz 1895: 4).

Niestety, *SW* przejął także wiele błędów *SL* (nie sprawdzono cytatów), opuszczono też tytuły i strony dzieł, z których pochodziły, co zaoszczędziło miejsce, ale obniżyło wartość naukową warsztatu: „Straty poniesionej przez opuszczenie źródła cytatów nie zastąpi ani ilość słów, ani wskazówki poprawnościowe” (Urbańczyk 1964: 24).

Drugim ważnym wykorzystanym dziełem był *SWil* Orgelbranda – niektóre definicje autorzy *SW* cytowali w całości (Okoniowa 2013: 246, 250).

Formalnie należy też uwzględnić tu dwa słowniczki Władysława Siarkowskiego, które zostały włączone w *Słownik gwar* Karłowicza, jednak ich udział jest niewielki: 355 i 162 wyrazy (Cygan 2020: 232, 234).

Nitsch pisał także o wykorzystaniu materiałów Passendorfera w suplemencie do *SW*, ten jednak nie jest tutaj omawiany, ponieważ jego treść jest zaginiona (Skarzyński & Smułkowa 2018: 314).

Trzecim źródłem były wreszcie materiały pierwotne, z których ekscerpcji. Na to, jak ważny bywa materiał źródłowy, może wskazywać poświęcenie autorom, z których czerpał Linde, osobnej publikacji (Hrabec & Peplowski 1963). Pełna ich lista znajduje się w **Aneksie M** do niniejszej pracy, zatytułowanym *Skonsolidowany spis skrótów SW*. Wiele z nich jednak było w oryginale niewymienionych – zwłaszcza rękopisów (głównie z XV w.), do których Linde nie miał dostępu (Urbańczyk 1964: 24).

Karłowicz (1871: 9ff) wymienia rodzaje także źródeł w tradycji oralnej, które są dla niego szczególnie ważne – obejmują: podania, zabobony, zamawiania, gusła, obchody, gry, pieśni, oracje, przysłowia, zagadki. O tych ostatnich pisze on (Karłowicz 1871: 12):

Zagadki ważne są nietylko jako łamigłówki lub dowcipy ludowe, lecz jako mieszczące czasem porównania i przenośnie czysto mitologicznej natury.

Część zebranego przez autorów *SW* materiału nie została w pełni wykorzystana, tzn. nie powstały odrębne hasła dla każdego zebranego wyrazu – według Wawrzyńczyka (2009a: 3–4) pozostawia to możliwość stworzenia suplementu „wewnętrznego”, czy też „autosuplementu”.

3.7. Analiza treści

Słownik warszawski posiada bogatą treść – zarówno ze względu na liczbę opisywanych jednostek, jak i bogactwo rodzajów informacji. Poszczególne typy wiadomości mają zazwyczaj specjalnie do tego przeznaczone rubryki – budowie hasłowej *SW* uwaga będzie jednak poświęcona w **rozdziale 4.3**.

Siatka pojęciowa *SW* reprezentują one wszystkie odmiany polszczyzny – myśl tę Ułaszyn (1900: 4) wyraził tymi słowami:

„Słownik” ten zawiera wyrazy polskie powszechnie używane w języku żywym i literackim, z szerokim uwzględnieniem wyrazów dawniejszych i gwarowych, jak również i wyrazy techniczne, czyli t. zw. terminologię.

Część z tych jednostek jest jednak kwestionowana – autorzy *SW* niektóre wyrazy brali z gwar innych języków, a nie odmian polszczyzny.

Niemal nigdy odrębnego artykułu hasłowego nie stanowią jednostki wieloelementowe (nie liczą czasowników zwrotnych z partykulą „się”); często – choć nie zawsze – są one wyróżnione na jeden z dwóch sposobów: albo są one gniazdowane jako podhasła (zwłaszcza w przypadku jednostek derywowanych od hasła głównego przy pomocy metody produktywnej, np. **Faraoński** → **Po faraońsku**), albo odznaczają się zmienioną typografią. Część jednostek ma jednak charakter ukryty (ang. *hidden entries*) i tylko analiza semantyczna pozwala je zidentyfikować.

Sam fakt wyróżniania jednostek wieloelementowych świadczy o wysokim poziomie warsztatu *SW*. Sam Karłowicz (1876: 16) pisze o nich w następujący sposób:

Skłonność tę wyrazów do tworzenia pewnych grup można nazwać obrazowo *chemicznym powinowactwem słów*, chociaż właściwie niema tu ani powinowactwa, ani chemizmu, lecz tylko mechaniczne zestawienie, uświęcone dawnością używania, nałogiem dziejowojęzykowym, ukutą wiekami foremką, w którą myśl przy swem objawianiu się chętniej wstępuje niż w formę nieużywaną, nową (...)

Hasłownik *SW* zawiera też wiele „mechanicznych” derywatów, przez co powiększone są tomy *SW* z literami, na które zaczynają się prefiksy czasownikowe (Doroszewski 1954: 20). Ponadto niektóre działy – np. medycyna – były opracowane znacznie wnikliwiej niż inne (Łempicka 1958: 204).

Jeśli chodzi o wewnętrzną strukturę artykułów hasłowych, to układ znaczeniowy w *SW* jest oparty o konkretne realizacje, ale traktuje przy tym znaczenia konkretne synonimicznie z obiektywnymi i realnymi, pomijając fakty podmiotowe, które na znaczenie również się składają (Doroszewski 1954: 50). Znaczenia uszeregowane jest w pierwszej kolejności logicznymi powiązaniem, w drugiej – frekwencyjnie (Piotrowski 1993: 575).

Słownik ten zawiera wszystkie typy definicji wskazane w klasyfikacji Doroszewskiego: realnoznaczeniowe, strukturalno-znaczeniowe, strukturalne, zakresowe, synonimiczne oraz gramatyczne, a także definicje mieszane. Są także podawane informacje peryferyjne względem znaczenia, zawarte w postaci kwalifikatorów, ograniczeń gramatycznych, antonimów, przykładów, objaśnień i glos.

Majdak (2007: 230) zauważa, że jest sporo zwłaszcza definicji strukturalno-znaczeniowych, które można rozpatrywać jako konkurencyjne względem odsyłaczy. Badaczka podkreśla także, że choć zarzucano to jako błąd w warsztacie, to definicje w formie ciągów synonimicznych były zamierzone (idem 2007: 224). Terminy naukowe są natomiast różnie, wręcz niekonsekwentnie traktowane w *Słowniku*: czasem zupełnie scjencyficznie i encyklopedycznie, kiedy indziej zupełnie nienaukowo i potocznie (Majdak 2007: 226–227).

Przy niektórych hasłach znajduje się także informacja o wymowie – dzieje się tak tylko wtedy, kiedy polska ortografia jest niejednoznaczna. Istnieją dwa zasadnicze sposoby podawania informacji fonetycznej:

1. przy pomocy alfabetu fonetycznego, który używa nierzadko abstrakcyjnych symboli do reprezentacji dźwięków – przykładem takiego alfabetu jest Międzynarodowy Alfabet Fonetyczny (ang. *International Phonetic Alphabet*, IPA) – jego odczytanie jest niełatwe i stosuje się je zwłaszcza dla osób używających danego języka jako obcego (Ilson 1986b: 55);
2. przy pomocy *respellingu* (zwanego także po angielsku *pronunciation key* ‘klucz wymowy’), czyli zapisu przy pomocy alfabetu tego samego języka w sposób ujednoznaczniiony i wraz z dodatkowymi oznaczeniami diakrytycznymi – rozwiązanie takie jest intuicyjne dla rodzimych użytkowników języka (Landau 2001: 118).

Tę pierwszą strategię coraz częściej stosuje się w słownikach przekładowych, ponieważ korzystanie z alfabetu innego języka wprowadza zamieszanie i może mieć zbyt ograniczony inwentarz znaków graficznych, aby oddać fonetykę języka docelowego. Ta druga strategia natomiast dominuje w słownikach jednojęzycznych ze względu na swoją prostotę i intuicyjność, jako że użytkownik nie musi się uczyć nowego systemu zapisu.

W przypadku języka polskiego informacja fonetyczna jest bardzo rzadko uwzględniana ze względu na dużą odpowiedniość systemową między zapisem a wymową. Informacja o wymowie w *SW* jest podana przy pomocy alfabetu polskiego, z tą różnicą, że w środek kombinacji liter, nie należy interpretować jako dwuznaków wstawiane są dywizy (np. c-h, r-z itd.). Nie jest podawana zmiana wymowy zgodna z regułami fonetyki polskiej, np. denazalizacji czy zjawiska udźwięcznienia i ubezdźwięcznienia itd.; klucz wymowy jest więc minimalistyczny. Wskazuje to na przeznaczenie słownika dla rodzimych użytkowników polszczyzny.

Słownik warszawski odznacza się także bardzo bogatym systemem kwalifikatorów, o czym pisze Walczak (1995: 90):

[R]ozbudowany system kwalifikatorów pozwolił twórcom słownika [warszawskiego] na wyodrębnienie z tego morza leksyki słownictwa objętego normą ówczesnego języka ogólnego i jego poszczególnych warstw.

Karłowicz (1876: 4–6) rozpatrywał słownictwo w trzech wymiarach pod względem:

1. kryterium **zasięgu**:
 - a. wyrazy **powszechne** lub **wszechpolskie** – odpowiadające słownictwu współnoodmianowemu;
 - b. wyrazy **mniej powszechne** – wszystkie pozostałe jednostki.

2. kryterium **chronologicznego**:
 - a. wyrazy **dawnopolskie** – pochodzące sprzed okresu współczesnego Karłowiczowi;
 - b. wyrazy **nowopolskie** – współczesne Karłowiczowi.
3. kryterium **zakresowego**:
 - a. wyrazy **czystopolskie nowsze** – czyli powszechne (wszechpolskie);
 - b. wyrazy **staropolskie**;
 - c. wyrazy **provincjonalne** – zarówno polskiego. jak i obcego pochodzenia;
 - d. wyrazy **obce, własne i odwłasnoimienne**;
 - e. wyrazy **techniczne** – niezależnie od pochodzenia
 - f. **pnie** – reprezentujące podstawę słowotwórczą, a nie samodzielne leksemy.

Kwalifikatory w samym leksykonie nie odzwierciedlają jednak do końca tego systemu, tzn. nie zawsze danej kategorii odpowiada konkretny kwalifikator lub ich grupa, nie są też w żaden sposób pogrupowane. Autorzy SW także praktycznie wcale nie używali kwalifikatorów do opis związków frazeologicznych.

SW stosuje także „nieliterową metodę kwalifikowania wyrazów”, czyli systemu oznaczeń (parakwalifikatorów), od której nowsze słowniki odeszły. Z powodu częstego używania kwalifikatorów symbolicznych słownik nazywano też niekiedy „znaczkowym” (Mariak 1991/1992: 146). Rozwiązanie to jest zbliżone do tego, które można spotkać w słowniku Adelunga. Czasem oznaczenia te są stosowane w dosyć zagadkowy i trudny do przewidzenia sposób.

Mariak (1991/1992: 146) zwraca uwagę na to, pod jakim względem nietypowy jest kwalifikator ×:

Obejmuje bowiem słownictwo przestarzałe, tzn. recesywne oraz słownictwo nowe, czyli progresywne, do którego należą: neologizmy rodzime, zapożyczenia, indywidualizmy terminologia specjalna z dziedzin nauki oraz nieliczne gwaryzmy.

Łempicka (1958) natomiast dokładnie analizuje wyrazy oznaczone w SW wykrzyknikiem ! i ubolewa nad niedostatecznym stosowaniem tego kwalifikatora:

Mimo głębokiego szacunku dla pomnikowej pracy autorów Słownika Warszawskiego, stanowiącego pod wieloma względami niezastąpione źródło, nie mogę się powstrzymać od westchnienia żalu, że tak mało stosowali wykrzykników tam, gdzie szło o dziwolągi postaci. (Łempicka 1958: 207)

Zastanawiające dla Łempickiej (1958: 203) jest też zaklasyfikowanie do „kategorii ujemnych” poniższych rodzajów jednostek leksykalnych:

- a) wyrazy należące do ścisłej terminologii naukowej, utworzone na podstawie elementów łacińskich i greckich,
- b) wyrazy przeniesione na grunt polski z języków europejskich w brzmieniu prawie nie zmienionym, a mające odpowiedniki polskie,
- c) zniekształcenia wyrazów obcych,

- d) ortograficzne i fonetyczne zniekształcenia wyrazów polskich,
- e) dziwaczne formacje składniowe.

Jednocześnie, jak zauważa badaczka, wiele form, które autorzy starali się wyrugować, zachowało się w polszczyźnie (Łempicka 1958: 207).

Należy tu napomknąć, że kwalifikatory w *Słowniku warszawskim* stanowią część znacznie większej grupy oznaczeń, która opisana jest w **rozdziale 5**. Karłowicz był niesłusznie krytykowany za brak wykazu użytych skrótów (Rak 2021a: 67):

Nitsch nie miał racji, zarzucając w recenzji Karłowiczowi to, że nie przygotował wykazu skrótów, mylił się też, przypisując tę pracę w całości Łosiowi (Nitsch 1911/1958: 197, za Nitschem tak podają E. Koniusz (2001: 12, 33) i H. Karaś (2011: 95)). Jak się bowiem okazuje, Karłowicz zestawiał część skrótów do SKarł i dodatkowo załączył do kartoteki słownika gwarowego trzydziestosześciostronicowy Klucz znaków i skrótów, który dotyczy także SMP. Do obydwu dokumentów miał dostęp Łoś i z obydwu korzystał, choć tego nie ujawnił. Rękopiśmienny Wykaz skrótów był dołączony do Przedmowy tymczasowej do „Słownika gwar polskich”.

SW wiele z tych skrótów odziedziczył, co tłumaczy wiele problemów ze spójnością; dodatkowo skróty te były w międzyczasie uzupełniane przez Łosia w *Słowniku gwarowym*.

Podstawa materiałowa dla SW była uzyskana metodą ekscerpacji totalnej, a znalezione wyrazy były ilustrowane przykładami autentycznymi i spreparowanymi. Jak pisze sam Karłowicz (1876: 12):

„a) na każdy odcień znaczenia wyrazu (...), b) na każdą epokę jego rozwoju, c) na każdą ciekawszą lub podwójną (resp. potrójną i t. d.) jego postać i formę t. j. odmiankę (waryjant).

I dalej (Karłowicz 1876: 14):

Byłbym za tem, aby każdy przykład pisał się bezwarunkowo pisownią pisma, z którego jest przytoczony, bez najmniejszej zmiany w głoskach, znakach pisarskich i t. d.

Niestety, przykłady nie są wystarczające ani pod względem ilościowym, ani jakościowym – opinię tę wyraża m.in. Żmigrodzki (2009: 142) pisząc: „Dokumentacja materiałowa w słowniku [warszawskim] jest słaba”.

Słownik warszawski podaje także informację o kluczowych (nieprzewidywalnych z perspektywy systemu) formach fleksyjnych, część mowy (jeśli nie jest to oczywiste), a z rzadka także o aspekcie i walencji. Informacja ta może dotyczyć zarówno całego hasła, jak i poszczególnych jego znaczeń. Niekiedy informacja o liczbie jest podawana w charakterze glosy, jeśli zmienia ona interpretację semantyczną w nieregularny sposób.

Omawiane dzieło jest też pierwszym słownikiem polszczyzny o zawierającym informacje etymologiczne, które dotyczą nie tylko zapożyczeń, ale także leksemów rodzimych:

Wielką zasługą redaktorów nowego słownika jest wprowadzenie do niego żywiołu etymologicznego czyli słoworodowego, przez podawanie pierwiastków, od których pochodzą odpowiednie wyrazy (...) (Ułaszyn 1900: 4)

Informacja ta w *SW* rzadko jest realizowana w pełni retrogresywnie, czyli do najdawniejszego źródła pożyczki lub poprzez prześledzenie całej drogi, jednak zazwyczaj zawiera kluczowe informacje, w tym komentarze słowotwórcze. Niespójnie traktowane są w słowniku „za-pożyczenia semantyczne, które polegają na utworzeniu nowego znaczenia dla jednostki leksykalnej na wzór takiego samego procesu w innym języku – przykładem jest polska „mysz”, która uzyskała nowe znaczenie pod wpływem ang. *computer mouse*, określenia na jedno z peryferiów komputerowych (Zgusta 1971: 180f). Karłowicz (1876: 39) przybliżyła to pojęcie w tenże sposób:

Należy także zwracać uwagę na to, co nazywam naśladowaniem wyrazów. Wyrazy naśladowane utworzyły się pod wpływem obcych języków: ponieważ cudzoziemiec, nazywając przedmiot jakiś, używał pewnej przenośni, więc i my, wprowadzając do naszego języka odpowiedni wyraz, nieraz chwytałyśmy się tej samej przenośni, to jest tłómaczyłyśmy tylko obcy wyraz, albo całkowicie, albo częściowo. Tak n. p. chociaż wyraz *drobnowid* jest czysto polski, lecz powstanie swoje zawdzięcza do-słownemu przekładowi greckiego *mikroskopu*.

W opisie etymologicznym także w sposób niejasny, niemal zamienny, stosuje się poję-cia: pień, osnowa i słoworód, mimo że słowa samego autora wynika, że nie są równoznaczne – por.:

- „z pnia rozwinięte osnowy polskie” (Karłowicz 1895: 5);
- „Mniejszą pewność słoworodu oznaczamy dodanym do wyrazu znakiem zapy-tania.” (Karłowicz 1895: 3) – mowa tu o pochodzeniu języka;
- „słoworód” użyty jako opis innego pochodzenia (Karłowicz 1895: 5).

Jest to niespójne z tym, że rdzenie – czyli „pierwiastki” lub „pnie” (ang. *root, radix*) – w słowniku określane są właśnie mianem słoworodów.

Niestety, zaprzestano podawania etymologii w tomie VI ze względu na sytuację w kraju:

Wskutek obecnych wypadków wojennych ustała możliwość komunikowania się z mieszkającym stale w Krakowie prof. J. Rozwadowskim, który wykazuje pochodzenie wyrazów, objaśnionych w Słowniku. Ażeby pomimo to nic przerywać wydawnic-twa, Słownik będzie wydawany tymczasowo (od wyrazu Szlifowanie) bez uwag ety-mologicznych do czasu, w którym stosunki normalne pozwolą podać wykaz opusz-czonych słoworodów oraz umieszczać je nadal w swoich miejscach.

Jest to wielką stratą, ponieważ *SW* stanowi przez to jeden z kilku niedokończonych słowników etymologicznych języka polskiego¹⁰.

3.8. Recepcja

Choć *SW* był niewątpliwie bardzo ważnym i przełomowym słownikiem – zarówno pod wzglę-dem praktycznym, jak i akademickim – jego recepcja w środowisku naukowym była mieszana. Wiąże się to z pojęciem obrazu słownika, które Hartmann & James (2002: 71–72) definiują w sposób następujący:

¹⁰ Por. prace Sławskiego i Bańkowskiego.

image of the dictionary

The public perception of DICTIONARIES and other reference works. Very little is known about this subject except that they evoke mixed feelings, from an appreciation of their value as cultural tools, which is associated with their AUTHORITY on etymology, spelling, pronunciation and definition, to a complete ignorance of their benefits and limitations.

obraz słownika

Publiczny odbiór SŁOWNIKÓW i innych publikacji źródłowych. Niewiele wiadomo na ten temat poza tym, że budzą one mieszane uczucia, od doceniania ich wartości jako narzędzi kulturowych, co wiąże się z ich AUTORYTETEM w zakresie etymologii, ortografii, wymowy i definicji, do całkowitej nieznajomości ich zalet i ograniczeń.

Wkład naukowy Karłowicza był marginalizowany, mimo że *SKart* był najlepszym słownikiem gwarowym na słowiańszczyźnie (Rak 2021a: 179ff); dopóki żył Nitsch, mało kto odważył się publikować o dziele tym pozytywne opinie (Rak 2021b); Nitsch (1911) poświęcił zresztą cały artykuł również krytyce *SW*. Choć o *SW* był oparty na *SL* i *SWil*, to w drugiej połowie XX w. zaczął być oceniany krytycznie, ponieważ sam miał być podstawą dla *SJPDor* (Rak 2021a: 179). W literaturze można spotkać wiele zarzutów *SW*, jak i pochwał za innowacje.

Głosy krytyczne dotyczyły wielu kwestii – najważniejsze problemy to (wraz z *kontrargumentacją*):

1. **dobór jednostek leksykalnych** – objawiało się to szerokim włączaniem rzadkich i wątpliwych leksemów przy jednoczesnym pomijaniu pewnych istotniejszych jednostek: „[*SW*] jest z pewnych względów za obszerny, a z drugiej strony wykazuje luki” (Szober 1928: 551). *Bańko (2014: 96) zauważa jednak, że z założenia miał być słownikiem popularnym, a nie zadowolającym językoznawców;*
2. **encyklopedyzm** – polegający na zbytnim opisie faktograficznym (poznawczym), wykraczającym poza opis językoznawczy dla niektórych jednostek. *Patrz: komentarz wyżej;*
3. **obszerność terminologii specjalistycznej** – wskazywano na jej zbyt bogatą reprezentację wśród haseł słownika (Korbut 1900: 5). *Ta jednak czyni z *SW* dzieło pełniejsze, co podnosi jego wartość praktyczną i naukową;*
4. **nadreprezentacja folkloru słownego** – czyli zbyt wiele leksyki niepoświadczonej w języku pisanim (Rak 2021a: 55). *To jednak świadczy o dużej wartości dzieła i unikatowości jego zasobu;*
5. **brak „rozsądnej normatywności”** – która objawiała się zbyt szerokim włączeniem leksemów, ale i zbyt ostrym oznaczaniem części z nich (Doroszewski 1952: 2);
6. **masowe włączanie derywatów** – Linde oraz autorzy *SWil* krytykowali nadużywanie słowotwórstwa przy tworzeniu artykułów hasłowych (Bańko 2001: 31; por. Majdak 2014: 95), które w *SW* było stosowane na szeroką skalę. Ważnym krytykiem jest tu Doroszewski (1966: 114), który zarzuca *SW* „brak rozsądnej selekcji, która powinna być oparta na kryteriach społecznych” tylko dodawanie form „dokombinowanych”, jako „pewne możliwości gramatyczne”. Badacz ten

wymienia tutaj m.in. łączenie „dowolnych prefiksów z dowolnymi tematami czasownikowymi i dotwarzanie do każdego czasownika odpowiadającej mu formy rzeczownikowej na -anie lub -enie” (Doroszewski, 1968: XIII). *Z punktu widzenia badań leksykologicznych uwzględnianie form nawet silnie generatywnych jest istotną informacją o stanie i rozwoju leksyki, zwłaszcza o mechanizmach słowotwórczych – por. Zarębski (2012), który zwraca uwagę na to, że „historyczny ogląd marginalizowanych, ale przecież wciąż obecnych w języku faktów słowotwórczych może uzupełniać naszą wiedzę o systemie słowotwórczym.”* (Zarębski 2012: 118);

7. **budowa definicji** – które pod wieloma względami były niedoskonałe metodologiczne i treściowo (Korbut 1895: 295–296). *Niemniej były one i tak bardziej rygorystyczne niż w wielu słownikach SW współczesnych i późniejszych;*
8. **nieadekwatna egzemplifikacja** – niektóre przykłady nieodpowiednio obrazują znaczenie opisane w treści definicji, jak to ma miejsce w przypadku hasła **Czas** (Doroszewski 1966: 115–116);
9. **brak „porządnej ekscerpcji”** – według Nitscha materiał źródłowy nie był odpowiednio opracowany (Skarżyński & Smułkowa 2018: 501);
10. **brak źródeł cytatów** – niektóre cytaty nie mają podanych źródeł, a podane źródła są „niedokładnie zlokalizowane” (Piotrowski 1993: 575): nie zawierają tytułu pracy ani numeru strony: „[N]iestety w monumentalnym Słowniku warszawskim – *ars citandi*, sztuka cytowania, często pozostawia wiele do życzenia.” (Doroszewski 1966: 320);
11. **problemy z cytowaniem** – fragmenty są przytaczane niedokładnie, a poza tym ten sam fragment bywa przytaczany w różnych formach (Rak 2021a: 55);
12. **zbytня drobiazgowość rubryk** – co prowadzi do zamieszania, nawet u samych redaktorów, czego dowodem są liczne błędy (Doroszewski 1954: 52f). *Drobiazgowość ta jednak ma wiele zastosowań przy wykorzystaniu metod opisanych w części inżynierskiej niniejszej pracy;*
13. **mieszanie kategorii leksemów nieliterackich** – według Nitscha w SW pomieszane są prowincjonalizmy (kategoria ta nie jest tam wyróżniona), regionalizmy oraz wyrazy gwarowe (Skarżyński & Smułkowa 2018: 467, 485, 486). *Wynikało to z dopiero kształtującej się w Polsce dyscypliny, jaką jest dialektologia;*
14. **brak lokalizacji geograficznej leksemów gwarowych** – informacja ta, zebrana na potrzeby *SKarł*, została w SW zupełnie pominięta (Rak 2021a: 55);
15. **niesymetryczność reprezentacji leksyki regionalnej** – w słowniku jest nieproporcjonalnie duży udział leksyki kresowej przy jednoczesnej małej reprezentacji słownictwa pochodzącego z obszarów takich jak Wielkopolska czy Śląsk. *Należy jednak mieć na uwadze, że słownik ogólnogwarowy najczęściej wykorzystuje dostępne dzieła leksykograficzne* (Rak 2021a: 55);
16. **wątpliwa wartość niektórych leksemów** – dotyczy to dialektyzmów oraz wyrazów o „podejrzanej polskości” (zwłaszcza pochodzenia rosyjskiego) (Rak 2021a: 55);
17. **niedokładna i błędna morfologia gwarowa** – podawane formy fleksyjne wyrazów często nie są zgodne ze stanem faktycznym (Rak 2021a: 55);
18. **stosowanie pni wyrazów** – kwestionowano ich rekonstrukcje i wartość naukową (Rusticus 1895: 2);

19. **opis etymologiczny** – który w przypadku niektórych jednostek był pod różnymi względami niesatysfakcjonujący. *Nie oznacza to jednak, że jest bez wartości – wiele informacji etymologicznej z SW zostało podtrzymanych w USJP, co oznacza, że jest ona solidna (Majdak 2008a: 224; Kwapien 2010: 223);*
20. **niekonsekwentna ortografia** – pisownia nie zawsze stosowana jest spójnie, zwłaszcza przypadku wyrazów kaszubskich (Rak 2021a: 55);
21. **brak jawnych zasad redakcyjnych** – co utrudnia interpretację niektórych informacji zawartych w słowniku. *Te jednak w dużym stopniu powstały, ale nie zostały opublikowane przez Łosia (Rak 2021a: 55).*

Doceniono jednak także wiele **pozytywnych cech SW** – przedmiotem pochwał były przede wszystkim:

1. **obecność leksyki dawnej i gwarowej** – co daje pełniejszy obraz polszczyzny (Brückner 1901: 98);
2. **objaśnienia etymologiczne** – które najczęściej były wystarczające dla szerokiej publiczności (Ułaszyn 1899: 273)
3. **uwzględnianie wyrazów bliskoznacznych** – co jest interesujące z perspektywy leksykologicznej, i co stanowiło podstawę do przyszłych prób standaryzacji języka polskiego;
4. **system objaśniania znaczeń** – ten był dużo dokładniejszy niż dziełach SW współczesnych, a także wielu wydanych po nim, zarówno w kraju, jak i za granicą (Nehring 1901: 161);
5. **obfitość materiału** – dzieło to posiada wiele informacji, w tym wiele przykładów użycia (Łoś 1900: 1358–1359);
6. **podanie informacji o wymowie** – co było innowacyjne w tej części Europy (Majdak 2008b: 244);
7. **podanie informacji etymologicznej** – był to pierwszy słownik ogólny podający etymologię wyrazów rodzimych (Walczak 2006: 422);
8. **typografię słownika** – która na owe czasy była nowoczesna i bardzo zbliżona do rozwiązań obecnych (Majdak 2008b: 244).

Niestety, jakość warsztatu leksykograficznego nie szła w parze z wielkością przedsięwzięcia. Wiele zarzutów było zasadnych, ale niektóre z nich są sprzeczne ze sobą lub się zdezaktualizowały, także w kontekście niniejszej tezy doktorskiej. Koniecznie należy też mieć na względzie, że zastosowana metodologia leksykograficzna kształtowała się równoległe do powstawania słownika, który był pod wieloma względami innowacyjny.

Istotne jest również to, że *SKarł* stanowi początki polskiej dialektologii, a niedociągnięcia w słowniku wypunktowane przez Nitscha często wynikają ze stanu ówczesnej wiedzy, a nie warsztatu leksykograficznego (Rak 2017: 263–264). Ma to oparcie w słowach Łosia (1904: 178), który traktując o niedostatku badań ludoznawczych pisze: „Dobrze przynajmniej, że mamy już słownik gwarowy, tak bogaty i tak świetnie wykonany”, który potem nazywa „kamieniem węgielnym językoznawstwa” (idem 1904: 180).

Na część zarzutów odpowiada niejako sam autor:

Naród spokojny o byt polityczny, może spokojnie też badać ludoznawstwo swoje ze strony czysto naukowej; ale plemiona, wyzute z samoistności, wystawione

na nieustanną groźbę pogńębienia i zatrąty, gdy się zastanawiają nad naturą i warunkami bytu swego skupienia narodowego, muszą do badań takich koniecznie mieszać pytania praktyczne: gdy bowiem mowa jest o właściwościach i odrębnościach narodowych, nasuwa się im bez ustanku pytanie: być albo nie być? Czy te zagrożone właściwości i odrębności utrzymają się, czy też skazane są na zagładę? Nie trzeba nawet się dziwić, jeżeli te pytania praktyczne więcej uwagi od teoretycznych zajmują: inaczej być nie może (Karłowicz 1906: 213–214).

Choć *SW* spotkał się umiarkowanym entuzjazmem szerszej publiczności, to z dzieła tego korzystano aż do 1968 r. (Majdak 2008b: 243). Słownik ten nie wystarcza współczesności Doroszewskiego, ponieważ oparty był na zdezaktualizowanej koncepcji – brakowało mu „rozsądnej normatywności” i „przemysłanej zasady selekcionowania materiału”, co było wynikiem unikania „jakiegokolwiek wartościowania faktów językowych” (Doroszewski 1966: 125–126).

Mimo wszystkich tych problemów Majdak (2008b: 243) wysoko ocenia ten słownik:

Ogólnie rzecz biorąc, zalety dzieła przyćmiewały jednak jego wady a ogrom opracowywanego materiału usprawiedliwiał braki.

Słowa Okoniowej (2013: 256), choć dotyczą *SKarł*, zdają się stosować również do *SW*:

Musiało minąć sto lat od wydania ostatniego tomu *Słownika gwar polskich* Jana Karłowicza i 108 lat od śmierci jego autora, żeby móc dostrzec wielkość uczonego oraz wartość jego dzieł.

Trudno się nie zgodzić też ze słowami, zwłaszcza biorąc pod uwagę ogrom pracy, jaki został włożony w to dzieło, przeszkody, z jakimi zmagali się jego redaktorzy, a mimo których powstał monumentalny i historyczny skarb polszczyzny.

3.9. Wartość *SW*

Słownik warszawski jest niewątpliwie jednym z najważniejszych dzieł leksykograficznych polszczyzny, nie tylko ze względu na objętość i przełomowe rozwiązania metodologiczne, ale także ze względu na udział, jaki miał w tworzeniu się polskiej państwowości. Z tego tytułu jest uważany za tzw. słownik narodowy – rodzaj dzieła leksykograficznego, który Hartmann & James (2002: 98) definiują w sposób następujący:

national dictionary

A cultural or historical dictionary of significance to the official language of a country, often financed by central government or an ACADEMY and compiled by a team of scholarly lexicographers.

słownik narodowy

Słownik kulturowy lub historyczny mający znaczenie dla języka urzędowego danego kraju, często finansowany przez rząd centralny lub AKADEMIE i opracowywany przez zespół uczonych leksykografów.

Słowniki narodowe są jednocześnie rodzajem słowników kulturowych (Hartmann & James 2002: 32):

cultural dictionary

The collective term for a range of REFERENCE WORKS which are both the result of cultural practice and agents of its promotion. During the early stages of a language community, NATIONAL DICTIONARIES can contribute to the development of a cultural identity, while DIALECT DICTIONARIES can document internal regional and social variety. General monolingual USAGE dictionaries can help to support a common linguistic STANDARD, thus benefiting education. ENCYCLOPEDIAS and similar reference works can help to raise the level of factual knowledge, and BILINGUAL DICTIONARIES can improve individual and public awareness of intercultural contact and contrast.

słownik kulturowy

Zbiorowe określenie na szereg DZIEŁ REFERENCYJNYCH będących zarówno rezultatem praktyki kulturowej, jak i pośrednikami jej promocji. We wczesnych stadiach wspólnoty językowej SŁOWNIKI NARODOWE mogą przyczynić się do rozwoju tożsamości kulturowej, podczas gdy SŁOWNIKI DIALEKTU mogą udokumentować wewnętrzną różnorodność regionalną i społeczną. Ogólne słowniki jednojęzyczne UŻYTKOWANIA mogą pomóc we wspieraniu wspólnego STANDARDU lingwistycznego, przynosząc w ten sposób korzyści edukacji. ENCYKLOPEDIAE i podobne opracowania mogą przyczynić się do podniesienia poziomu wiedzy faktograficznej, a SŁOWNIKI DWUJĘZYCZNE mogą poprawić indywidualną i publiczną świadomość kontaktów i kontrastów międzykulturowych.

Według Piotrowskiego (1993: 574) *SW* opierał się na idei słownika narodowego Passowa, na których opierały się też słowniki braci Grimm w Niemczech, Murraya w Anglii czy Littrego we Francji.

Kategoria słownika narodowego odnosi się jednak zasadniczo do jego roli w społeczeństwie, a nie jego charakteru; tu *SW* należy traktować za słownik polifunkcyjny. Matuszczyk (2006: 83–110) rozważa następujące koncepcje leksykograficzne, które realizuje *SL*: przekładową, porównawczą, panslawistyczną, etymologiczną, narodową, historyczną i normatywną. Analogicznie jest w przypadku *SW* – piszący te słowa uważa, że dzieło to powinno być rozpatrywane jako słownik:

- definicyjny;
- poprawnościowo-ortograficzny;
- wyrazów obcych;
- gramatyczny;
- historyczny;
- etymologiczny;
- gwarowy;
- potoczymów;
- specjalistyczny/terminologiczny.

A także, choć w mniejszym zakresie:

- frazeologiczny;
- przekładowy;
- walencyjny.

Tak bogate i wielowymiarowe dzieło miało nieodzowną wartość praktyczną. *SW* był stosowany przez Nitscha – ostrego krytyka działalności leksykograficznej Karłowicza – do rozstrzygnięcia problemów językowych (Skarżyński & Smułkowa 2018: 430). Badacz ten przyrównał nawet *SW* do słownika Webstera pod względem zakresu i funkcji (Skarżyński & Smułkowa 2018: 363). To, że słownik tego rodzaju jest potrzebny, sygnalizował zresztą Ułaszyn (1900: 4):

Że wydawnictwo „Słownika języka polskiego” jest rzeczą niezmiernie ważną, nie ulega chyba najmniejszej wątpliwości, a szczególnie dziś, kiedy w trzech dzielnicach dawnego państwa polskiego, szpecimy mowę naszą wyrazami cudzoziemskimi i dziwolągami gramatycznymi.

Słownik warszawski nie jest jednak wyłącznie produktem użytkowym, ale niesie ze sobą olbrzymią wartość badawczą. Wpłynął on na rozwój metodologii warsztatu naukowego, zwłaszcza polskiego językoznawstwa i leksykografii, a także zapoczątkował dyskusję nad normatywnością, miał swój udział w standaryzacji ortografii, zwiększył świadomość językową i pełnił istotną rolę dla ruchu narodowościowego.

Dzieło to jest też ważnym źródłem informacji historycznej w myśl koncepcji Doroszewskiego, który uważał słowniki za „dokumenty epok historycznych i prądów umysłowych” (Doroszewski 1954: 55).

Szczególnie bolesne jest tutaj utracenie suplementu. Wawrzyńczyk (2009b: 3) zwraca uwagę na to, że do dzisiaj słownictwo międzywojenne jest słabo znane:

Wbrew rozpowszechnionemu mniemaniu dziesięciolecia bezpośrednio powojenne nie wniosły aż tak wielu nowości (neologizmów, neosemantyzmów), jak by to miało wynikać z zestawienia danych, których dostarczył ponadtrzydziestoletni akademicki program badawczy tzw. nowego słownictwa polskiego.

Badacz ten (Wawrzyńczyk 2015: 9) zaznacza jednak, że nie można traktować *SW* jako pełnego obrazu leksyki i przestrzega (idem 2015: 11):

Leksykograficzne zauroczenie, koncentracja wyłącznie na przyroście słownictwa rejestrowanego w słownikach wypacza obraz rozwoju tych zasobów; nie przystaje on do rzeczywistości, rzeczywistości językowej, do faktografii opartej na tekstach, źródłach tekstowych (...).

I dalej:

Tymczasem użytkownicy słowników, także specjaliści językoznawcy, zapominają o tej okoliczności i na podstawie słowników ogólnych języka polskiego, które są przecież zawsze słownikami materiałowo cząstkowymi, konstruują sądy o słownictwie, o języku, o jego rozwoju jako całości.

Z punktu widzenia metaleksykografii *Słownik warszawski* wprowadza też wiele innowacji warsztatowych (Majdak 2014: 96).

Po pierwsze dzieło to obejmuje największy do dziś zbiór leksyki gwarowej – jest to istotne źródło informacji o odmianach regionalnych polszczyzny, choć niedoskonałe, bo nieposiadające dokładnej informacji geograficznej wyrazów. Gwary w Polsce są w trakcie postępującego zaniku, przez co waga zasobów takich jak *SW* jest duża. Leksykon ten wyprzedzał

epokę także dlatego, że oprócz izolowanych wyrazów uwzględniał związki frazeologiczne i dokładny opis gramatyczny (Piotrowski 1993: 574).

Po drugie podawanie w słowniku etymologii dla wielu jednostek, w tym dla leksyki rodzimej, uważane jest za pionierskie rozwiązanie w słowniku ogólnym (Walczak 2006: 422), istotne także ze względu na odradzanie się tożsamości narodowej Polaków. *SW* jest też kontynuacją pierwszego słownika etymologicznego, *SWOKarł*, który co prawda nie został ukończony, ale który poprzedzał słownik Brücknera, który stworzył pierwszy wyłącznie etymologiczny słownik polszczyzny (Rak 2021a: 10).

Po trzecie jednym z większych problemów, z jakimi spotykają się twórcy korpusów historycznych, jest brak znormalizowanej grafii. Jakkolwiek istnieją pytania, na które można odpowiedzieć tylko wtedy, gdy sięgnąć do oryginalnego zapisu, w większości wypadków badacze potrzebują ujednoczonego kształtu graficznego tekstu. Jednolita i uwspółcześiona grafia jest jednym z podstawowych warunków korzystania z dostępnych analizatorów morfologicznych (tagerów) (Górski et al. 2019: 31). *Słownik warszawski*, który posiada ujednoczony zapis (fonetyczny) tak obszernego zbioru leksyki, jest zatem wyjątkowo przydatny do badań nad rozwojem języka.

Po czwarte *SW* wprowadza rozbudowany i nowoczesny system kwalifikatorów, który ważny jest nie tylko dla językoznawców, ale i dla zwykłych użytkowników pogłębiających swoją wiedzę o języku. Rozwiązanie to wyznaczyło – poza stosowaniem parakwalifikatorów symbolicznych – standard w dziedzinie, który wykraczał poza granice kraju. Przeglądu kwalifikatorów dziedzinowych i stylistycznych, których zaczęto używać w czeszczyźnie szerzej dopiero po ukończeniu wydawania *SW*, dokonuje Lubomír (1955: 12ff). Kwalifikatory symboliczne były co prawda już w słowniku Jungmanna, ale na mniejszą skalę – autor ten użył znaku * dla leksemów przestarzałych, a dla „cudze, źle utworzone, lub podejrzane słowa” opatruje znakiem † (Lubomír 1955: 13).

Po piąte w *Słowniku warszawskim* osobne hasła stanowią feminytywa, czyli rzeczowniki żeńskie. Feminytywami w polszczyźnie zaczęto się interesować na początku XX w., kiedy to zaczęły się masowo pojawiać nowe w wyniku zmian społecznych w Polsce – cechowały się jednak brakiem symetryczności i regularności, a także „unieruchamianiem fleksji” (Kubiszyn-Mędrala 2007: 31–32). Decyzja o ich uwzględnieniu miała konsekwencje społeczne i ideologiczne, jednocześnie odzwierciedlając przemiany w rzeczywistości narodu polskiego. Powód, dla którego jest to ważne wyjaśnia, Krysiak (2016: 84):

Konsekwencją uniwersalności *maskulinum* i nacechowania rodzaju żeńskiego w systemie językowym jest przekonanie o prymarności form męskich i swoistej wtórności, podrzędności rzeczownikowych form żeńskich (które najczęściej są od form męskich derywowane).

Po szóste *SW* jest innowacyjny od strony typograficznej, także ze względu na rubryki, które wyróżnia strukturalnie. Wyznaczony przezeń standard w zasadzie w formie niezmienniczej panuje do dziś, co wskazuje na jego funkcjonalność. Ma to też nie lada znaczenia dla części inżynierskiej niniejszej pracy, w której struktura ta jest wykorzystana do opracowania głębokiej digitalizacji tegoż dzieła.

Powyższa lista nie jest jednak wyczerpująca – wartości badawczej *SW* jako materiału źródłowego lub referencyjnego dowodzą liczne prace przeprowadzone na jego podstawie.

Mańczak-Wohlfeld (1988) pisze o anglicyzmach poświadczonych w *Słowniku warszawskim* i niektórych błędach związanych z ich klasyfikacją jako takich, a także o wykorzystaniu zawartej w nim informacji gramatycznej do badań nad rozwojem języka polskiego. W podobny sposób Koniusz (2013) wykorzystywała prace Karłowicza do badań nad wzajemnym wpływem różnych języków na ziemiach polskich.

Badania słowników pozwalają na rekonstrukcję regionalizacji leksyki, czego dowodzi m.in. praca Piotrowskiej-Wojaczyk (2011: 12). Do opracowania leksyki gwarowej Kurzowa (1985: 127ff) korzysta ze źródeł takich jak *SKarł* i *SW*, a także *SWil* i *SL*, porównując uzyskaną siatkę hasłową z innymi słownikami gwarowymi (Kurzowa 1985: 127ff).

Badać można też poszczególne pola semantyczne – Piotrowicz (2004) skonstruowała słownik tematyczny poświęcony leksyce życia towarzyskiego, który oparty był na XX-wiecznych źródłach słownikowych, w tym *SW* (Piotrowicz 2004: 22–24)

Doroszewski z kolei wykorzystywał *SW* do swoich badań przymiotnikami złożonymi (Doroszewski 1966: 320), a także jako materiału referencyjnego do prac chronologizacyjnych (Doroszewski 1966: 59).

Wreszcie podczas badania przykładów w tymże dziele Majdak (2006b: 103ff) udało się wyodrębnić zmiany w zakresie ortografii, interpunkcji, morfologii, składni i leksyki w nowopolszczyźnie.

I w ramach niniejszej pracy podjęte zostaną próby wyizolowania różnego rodzaju informacji, które będą objaśnione w **rozdziale 10**, a ich rezultaty przedstawione w formie aneksów.

4. Struktura SW

W niniejszym rozdziale zostanie przedstawiona struktura *Słownika warszawskiego*, która jest bardzo bogata – bez analizy tejże struktury wszak nie można przetworzyć treści dzieła metodami automatycznymi. Wysokiej jakości opisu budowy SW podjęła się co prawda Majdak (2008b: 244), jednak w niniejszej tezie poziom analizy musi zostać pogłębiony i uwzględnąć perspektywę programistyczną; podobnie jest w przypadku prac poświęconym innym wielkim słownikom, takich jak Bilińskiej, której opis mega- (2013: 49–72), makro- (idem 2013: 73–95) oraz mikrostruktury (idem 2013: 97–127), choć pod wieloma względami zbliżony, nie jest wystarczający na potrzeby tejże rozprawy doktorskiej.

Zasadniczy opis struktury *Słownika warszawskiego* zawarty w tym rozdziale bazuje na powszechnie przyjętym podziale na mega-, makro- i mikrostrukturę leksykonu, natomiast wyróżnione elementy strukturalne artykułu hasłowego oparte są o charakterystykę samego źródła, jak i konieczność wyodrębnienia jednostek na potrzeby stworzenia analizatora składniowego, opisanego w rozdziale.

4.1. Megastruktura SW

Dominującą częścią megastruktury *Słownika warszawskiego* jest korpus słownika – dla przykładu w tomie I stanowi 955 z 968 stron, czyli 98,66% całej jego objętości. W korpusie strony numerowane są przy pomocy cyfr arabskich. Ostatnią stroną korpusu zwieńcza informacja w formacie: „Koniec tomu *n*-ego”, gdzie ostatnim wyrazem jest zawsze liczebnik porządkowy w dopełniaczu liczby pojedynczej rodzaju męskiego, którego wartość odpowiada aktualnemu tomowi.

Treści zewnętrzne obecne w SW to przede wszystkim treści początkowe. Numeracja w tej części słownika zapisana jest za pomocą cyfr rzymskich. Stałymi elementami treści początkowych we wszystkich tomach są karty tytułowe oraz spis skrótów. Pozostałe elementy nie występują we wszystkich tomach, ich nazwa lub kolejność jest zmienna, lub stanowią część treści końcowych; treści końcowe numeracji nie posiadają. Miejsce występowania tychże sekcji w poszczególnych tomach słownika zostało zestawione w **tabeli 3**.

Tabela 3. Porównanie treści zewnętrznych tomów SW.

Tytuł lub opis sekcji	Tom							
	I	II	III	IV	V	VI	VII	VIII
Karty tytułowe – pierwsza część	1	1	1	1	1	1	1	1
Karty tytułowe – druga część	2	3	3	3	3	4	2	3
[Nota od autorów]	–	–	–	–	–	2	–	–
„Dalsze skrócenia”	–	2	2	2	2	3	–	2
„Dalsze skrócenia” – ciąg dalszy	–	7	–	–	–	–	–	–

Tytuł lub opis sekcji	Tom							
	I	II	III	IV	V	VI	VII	VIII
„Spis skrótów”	3	4	4	4	4	5	3	4
KORPUS SŁOWNIKA	4	5	6	6	6	6	4	6
„Sprostowania”	5	6	5	5	5	–	–	–
„Opuszczone skrócenia”	6	–	–	–	–	–	–	–
„Opuszczone”	–	–	–	–	–	7	–	–
„Sprostowanie do tomu VIII”	–	–	–	–	–	–	–	5
„Cena słownika”	7	–	–	–	–	–	–	–

4.2. Makrostruktura SW

Makrostruktura *Słownika warszawskiego* nie odbiega formatem od porównywalnych słowników epoki ani od słowników współczesnych. Hasła zostały ułożone w porządku alfabetycznym z uwzględnieniem polskich liter, czyli od A do Ź. Dodatkowo część zasadnicza podzielona jest na sekcje poświęcone hasłom zaczynającym się na daną literę, która to jest podana jako nagłówek każdego tekstu; nowa sekcja może być rozpoczęta na nowej stronie, co widać na przykładzie strony 119 tomu II (**ilustracja 2**). W przypadku liter o złożonej funkcji ortograficznej zamiast haseł może także pojawić się informacja o ich stosowaniu zamiast haseł, co obrazuje **ilustracja 3**.

Same artykuły hasłowe natomiast umieszczone są w dwóch kolumnach, które oddzielone są pionową linią. Niektóre z nich, ze względu na swoje położenie i obszerność, mogą obejmować kilka kolumn lub stron. Jednemu artykułowi hasłowemu zawsze odpowiada jeden akapit. Wszystkie akapity zaczynają się od wcięcia tekstu i wytłuszczonego druku lemmy głównej. Niektóre artykuły hasłowe nie występują samodzielnie, ale zagnieżdżone są wewnątrz innych – co opisane zostało w **rozdziale 4.3** Mikrostruktura SW.

W celu ułatwienia nawigacji, lemmy pierwszego i ostatniego hasła na każdej stronie dodatkowo umieszczone są w nagłówku. Wszystkie strony posiadają też numerację w centralnej części stopki, poza stroną pierwszą i ostatnią korpusu słownika; numeracja nie jest kontynuowana między tomami.

IZOTERMA

Izoterma, *y*, częśc. w *lm. y*, **Linja izotermiczna** f. fiz. p. **Izotermiczny**. 2. *gie. linja, łącząca na kartach geograficznych punkty powierzchni ziemi, mające jednakową temperaturę roczną. Izotermi miesięczne = linje, łączące na kartach geograficznych miejsca o jednakowej średniej temperaturze miesięcznej.* <Now. z Gr. *isos*=równy + *thermos*=ciepły >

Izotermiczny przym. od **Izoterma**: **Linje izotermiczne** p. **Izotermz.** Fiz.: I. przebieg zjawiska = przy którym temperatura pozostaje staćca. Krzywa izotermiczna a. **Izoterma** = linja, okazująca zmiany objętości gazu pod wpływem różnego ciśnienia, jeżeli przytym temperatura pozostaje niezmienną.

Izoteryczny przym. od **Izotera**.

Izotropiczny p. **Jednorodny**. <Now. z Gr. *isos* równy + *trópos* = sposób >

[**Izbyon**, *u*, *lm. y*] p. **Ajzbyon**. <Nm. Eisenbahn >

Iż, **†Iże**, **†Eż**, **†Eze**, **†Hiż**, [**Jeż**, **Eż**, **Iże**] sp. f. a. **†Iżby że, jakoby**: Słyszał, iż mówiono, że bitwa przegrana. L. Żona mi utonęła, żona, iż tak rzekę, wpadła mi w rzekę. Mick. Tak do weipne żarciki uniał komponować, iżby je w kalendarzu można wydrukować. Mick. **†Apostoł Paweł św.** tak napisał, iż łakomstwo jest korzeń wszystkiego złego. Kazania r. 1555. 2. *ponieważ*: Klucznikiem siebie tytułował, iż ten urząd na zamku przed laty piastował. Mick. <I + Ż(e)>

†Iż z. p. **Iż.**

JA

× **Iżali** p. **Iża**. Wor.

× **Iżaliż** p. **Iża**. Wor.

Iżby f. a. [**Iżeby**], **†Iżeby** sp. *aby, ażeby, żeby* by. 2. **†p. Iż**: Przyganiono mu, I. był wyleganec. Ort. Nauceza, I. nasza służba w kościele była marna. Op. <I + Ż(e) + By>

[**Iżci**] p. **Iżci**.

I. Iże z. p. **Iż.**

2. **†Iże**, [**Iże**] sp. f. p. **Iż**: Ślubuję, **iże**, stanąwszy w Krakowie, powieszę pawęż na gospodzie. Sienk. Począł Zbyszkowi przymawiać, **iże** wasów nie ma. Sienk. Coś mi **ś. iże** wiedzie, **iżem** nie postuchał, **tatuleczku**, ciebie. Konik **żałobam** nadkrzyty, **iże** (= *ponieważ*) mój Janiczek na wojnie zabity. Piotr odpowiedział, I. Jan w tej istej dziedzinie ze mną mieszkał. Św. z Woc. 2. **†czy**: Pytaliśmy was, I. człowiek może dać swe imię. Ort. Chcąc on to od niej przewidzieć, I. narodził **ś. który** człowiek. Kaz. Gnieź. 3. [**I. Jeno**] = o mało co, *bezmata, omal, tylko co*. <I + Że>

[**Iżeby**], **†Iżeby** p. **Iżby**: Wojski przykazuje, I. byli gotowi na wojnę. S. Par.

[**Iże jeno**] p. **Iże**: Leć, stał się ptakiem, a **car-noksiężnik** jastrzębiem; **juz** go miał **iże jeno** zgłonić, a **tén** stał się **lisem**.

[**Iżeli**] p. **Iżeli**.

× **Iżówka**, f, *lm. i* p. **Iżówka**: Śmiało szukał **iżówek** po lesie, **idą** bezpieczne, bez **męzów** swych **kozy**. Nar.

†Iżto z. p. **Iż.**

†Iżtoć z. p. **Iż.**

J

I spółgłoska podniebienna a. średniojęzykowa; niegdys wywierała (podobnie jak samogłoski e, i) wpływ miękczący na spółgłoskę poprzedzającą twardą, zmieniając ją na odpowiednią podniebienną czyli miękką, która następnie przerodziła **ś.** w inną spółgłoskę. Tą drogą powstały między innymi z połączeń *sj, zj* ogólnie-słowiańskie *sz, ż*, a z połączeń *tj, dj* ogólnie-polskie *c, dz*, i f. p. Na początku wyrazów oraz na początku zgłosek, gdzie dziś piszemy *ż*, np. *inny, istota, moim, boi ś.*,... mamy właściwie połączenie głosek *jt, tj*. j. jinny, jistota, mojim, boji ś., i t. p. W najstarszych zabytkach (w. XIV – XV) spółgłoskę *j* wyrażano bądź literami *i* i *y*, bądź też z czeska literą *g* (jedno, moy, jest i gest = jest...), od wieku XVI w drukach używano w tym celu dwóch znaków, *i* i *y* (jest, móy, wojsko, Julian, historia); od wieku zaś XIX-go zamiast tak stosowanych liter *i, y* weszła w użycie litera *j* (jest, mój, wojsko, Julian, historia...).

J. w skrótach = a) *Jego, Jej*: J. K. M. = *Jego a. Jej Królewska Mość*. J. C. M. = *Jego a. Jej Cesarzowska Mość*. J. C. W. = *Jego a. Jej Cesarzowska Wysokość*. J. Ks. M. = *Jego a. Jej Książęca Mość*. J. E. = *Jego Eksceleńcja*. J. Em. = *Jego Emnienecja*. b) *Jaśnie*: J. W. = *Jaśnie Wielmożny*. J. J. W. W. = *Jaśnie Wielmożni*. J. O. = *Jaśn Oświecony*. J. J. O. O. = *Jaśnie Oświeceni*. c) *-gomość*:

JP. a. J. M. P. a. J. M. C. P. = *Jegomość Pan. J. Ks. (J. X.) = Jegomość Książęca*. d) *jać*: j. w. = *jak wyżej*; j. n. = *jak niżej*. e) *jest*: t. j. = *to jest*. f) *Jezus*: JHS. = *Jezus Chrystus*. Na krucyfikach: J. N. R. I. = *Jesus Nazarenus Rex Judaeorum* (= *Jezus Nazareński, król żydowski*). g) *ja*: j. n. pod. = *ja niżej podpisany*.

Ja, mnie, a. mię, *lm. my*, [2 pp. **mie, mię, me**, **mi**; 3 pp. **mie, mię, me, mę, my**; 4 pp. **mie, mię, me, mę, mi, ma, nia**; 6 pp. **mnąm, mnom, mą**; 7 pp. **mię**] f. z. osobowy pierwszej osoby, wyrażający osobę mówiącą: Ja mówię. **Któż** tam jest? — Ja. — **Co** za ja? — Ja. Zabł. Przyjaciel dobry jest drugi ja. Bud. **Tys** mnie jest drugi ja, a **zaś** ja z tobą **jedno** jest. Petr. **Twarzę** tak jest do mnie podobny, że trudno wyrazić; ja a ja (= *drugi ja*). Jak ty mnie, tak ja tobie. Prz. **Mną** **ś.** mieni, i wobec mnie za mnie **uchodzi**. Zabł. Ja w tym, **zdaj** to na mnie (= *moja w tym głowa*). Co mi po tym! Co mi tam **będziesz** **głowę** **zawraca!** **Oóż** mi z tego? Jak mi **ś.** masz? Szanujże mi **go!** Rusz mi **ś.** jeno! Niech mi tu nie **popasa!** A **będziesz**-ze mi tu **stał?** Dobrze mi **wyglądasz**. To mi **człowiek** **szczęśliwy**, co **pańskiej** **piewien** **obrony**. Ryb. Ja, któremu on tyle **szkodzi**, **staram** **ś.** mu w **każdej** **okazji** **usłużyć**. Przen.: **Mnie** w **Nowogródku**, **mnie** w **Mińsku** **czytuje** **młodź**. Mick. (= *moje dzieła*).

WZE

kiedy uderzy. Kn. (= nic dziwnego, że...). Wzdyć państwa zbytkiem gina. Opal. Przebóg, cóż czynisz, na myśl dawne wiodąc żale, wzdyć ś. to sama we łzach swych pogrążasz cale. Min. f) Wzdyby = toby' jednak, mozeły przecie: By ś. nadętości panów dobry przyjaciel sprzeciwił, wzdyby do uznania siebie przychodzili. Skar. g) zgoła, ze wszystkim; wogóle, żywnie: Któż W. z każdej strony uszczęśliwiony? Koryt. Bóstwo samo z siebie nigdy z żadnym stworzeniem nie mówiło, jeno zawždy przez jaki instrument: a. przez anioła, a. przez proroka, a. W. przez jaki pośrzodek do tego godny. Rej. Ja pewny, że przyzwolisz, nie pierwsza to proba, na to wszystko, cokolwiek W. mi ś. podobna. Min.

- [Wze] już.
- † Wzecz, wżegnien, wżęgl, † Wżegnąd, nied.
- † Wżęgać wpalic. L.
- † Wżęgać, a, ał p. Wżec. L.
- † Wżęganie; a, blm., czynność cz. Wżęgać.
- † Wżęgnąć, nie, nął p. Wżec. L.
- † Wżęgnięcie, a, blm., czynność cz. Wżęgnąć.
- [Wżenę, wżenie] p. Węgnąć.
- Wżerać się, a ś., ał ś. p. Wżeręć ś.: Mróz wżerał ś. ostremi kłami w ziemię. Reym. Wżerał ś. oczyma w tę twarz, by z niej wyczytałę straszną tajemnicę. Przybysz. Wojska republikańskie wżerały ś. w kamienną Pontebbę. Zer. Wżeranie się, a ś., blm., czynność cz. Wżerąć ś.

Y

- Wżęcie, a, blm., czynność cz. Wżąć.
- Wżęcie się, a ś., blm., czynność cz. Wżąć ś.
- Wżęty l. im. od Wżąć. 2. im. od Wżąć ś., który ś. wżął, znacząc zapuścił ś. w co.
- × Wżmiąć, weźmie, wżmiął p. Wżąć.
- Wżycie się, a ś., blm., czynność cz. Wżyć ś.: Dar wżycia ś. w nieskończoną mnogość obcych istnień.
- Wżyć się, yje ś., ył ś., nied. Wżywać ś. żywszy ś. wmięszac ś., wglębić ś., zespolić ś.: Dosia potrafiła zaraz W. ś. tutaj, poznać ludzi, zastosować do nich. Krasz. Tak ś. wżywa w szalony wir życia paryskiego.. Przen.: Wżył ś. już był potrochu w te formy, które tam są praktykowane. Kaczk. (= przywykł do nich, otrząsał ś., oswoił ś. z niemi). W tych trumnach myśli Boże o was, które wami być miały samemi, a w któreście W. ś. nie podobałi. Krasin. Kościuszko wżywa ś. w epokę i potrzeby ojczyzny. Patrzył na ich biedę, wżywał ś. w ich tęsknoty. Ole. Wżywał ś., wmyślaj, wglębiaj w samego siebie. Brt.
- Wżymać, a, ał p. Wżąć.
- Wżzymanie, a, blm., czynność cz. Wżymać.
- Wżynac się, a ś., ał ś. p. Wżąć ś.
- Wżynanie się, a ś., blm., czynność cz. Wżynać ś.
- Wżywać się, a ś., ał ś. p. Wżyć ś.
- Wżywanie się, a ś., blm., czynność cz. Wżywać ś.

X

Ponieważ zasadą pisma głoskowego jest, ażeby każda pojedyncza litera odpowiadała pojedynczemu dźwiękowi, przeto niewłaściwie sobie postąpili nasi przodkowie, przyjmując do abecadła polskiego literę x dla wyrażenia dwu naraz spółgłosek: k i s lub (między samogłoskami) g i z. Zaczęto od pisania przez tę literę wyrazów greckich i łacińskich, jak: axamit, Xerxes, examen, ekukcja, exempja, exorbitancja, exmisja i t. d., przyczym, jak widzimy, nie uwzględniano pod wpływem obcej pisowni różnicy dźwięków ks i gz. Następnie literę x widzimy u dawnych naszych pisarzy i w niektórych wyrazach polskich, jak: xiążę wraz z wszystkimi od niego pochodniami, xieni, xiążka, xiężyc i t. d., przyczym pisarze ci (Jan Kochanowski, Januszowski, Górnicki i in.) wprost zachęcałi do używania tej litery w piśmie polskim. Deputacja Towarzystwa przyjaciół nauk radzi pisać wyrazy cudzoziemskie przez x, a polskie przez ks i gz. Krakowska Akademia umiejętności

również nie wyklucza użycia w pisowni polskiej itery x. Ponieważ tedy czytelnicy nasi znajdować będą w piśmiennictwie naszym wyrazy polskie pisane przez x, uważałem za potrzebne powiedzieć tych parę uwag, upoważniających słownikarza polskiego do wykazania wyrazów, o których mowa, zaczynających się dawniej od x, nie pod nieistniejącą prawie teraz w polskim abecadle literą X, ale pod literą K. W skrótcejach a) oznaczano literą x wyrazy: ksiądz, książe, książęcy, księga (= dział dzieła): X. pleban, X. biskup, XX. (księża) Bazylianie: w. x. litewski (= wielki książe litewski), W. X. Moś (= wasza książećca moś), tom II, x. 3, (= tom drugi, księga trzecia); b) X P S znaczyło Chrystus, tutaj bowiem X i P występują jako greckie litery = ch i r); c) w matematyce x oznacza ilość niewiadomą; d) jako liczba rzymska oznacza 10, przewrócone bokiem — tysiąc, prosto stojące z kreską u góry—10000.

Y

Y (ypsylon, ipsylon, igrek) jest literą przyjętą przez łacinników i Polaków od Greków. Używa ś. ona do wyrażania dźwięku jakby i po-

zbawionego jotacji, a ponieważ język polski każde i zaczynające wyraz lub zgłoskę jtuje, przeto nie może być wyrazów polskich (i zgło-

4.3. Mikrostruktura SW

Struktura artykułów hasłowych *Słownika warszawskiego* jest jego najbardziej złożonym komponentem. Chociaż przeprowadzona analiza wskazuje na istnienie licznych i szczegółowych dyrektyw praktycznych, którymi kierowali się autorzy, to nadal znaleźć można wiele odstępstw od tak zrekonstruowanego formalizmu. Nawet jeśli przymknąć oko na nieścisłości, które przy kompilacji tak szeroko zakrojonego i wymagającego dzieła były – zważywszy na możliwości techniczne i dostępne zasoby – nieuchronne, to zadanie uporządkowania i sklasyfikowania elementów, które w hasłach SW występują, nie sposób określić mianem trywialnego.

Rozszyfrowanie znaczenia informacji zawartych w słowniku tym nie stanowi jednak większego problemu dla człowieka – ten może się podeprzeć typografią hasła, ale także odnieść się od swojej językowej i pozajęzykowej wiedzy. Celem części praktycznej niniejszej tezy doktorskiej jest jednak automatyczna analiza treści SW i ekstrakcja zawartych w nim danych. O ile ludzki użytkownik słownika od początku dysponuje narzędziami niezbędnymi do przyswojenia zawartych w nim danych, to w komputer może te dane „zrozumieć” tylko poprzez analizę strukturalną.

W związku z powyższym przy rozpatrywaniu budowy artykułów hasłowych kluczowym jest, aby dobrać takie kryteria ich analizy, które będą miały przełożenie na zalgorytmizowanie procesu i pozwolą na odłączenie treści SW od formy, w jakiej dane zostały zaprezentowane w oryginale.

Mikrostrukturę artykułu hasłowego opisuje także Majdak (2018a: 118–225) w sposób szczegółowy i z szeroką egzemplifikacją, jednak czyni to na podstawie ekscerpcji całego materiału słownikowego, skupiając się przy tym na nieco innych kategoriach zagadnień, niż jest to czynione w niniejszej tezie. Choć ważną rolę przy wyodrębnianiu segmentów w obu pracach gra kryterium pozycyjne – „za definicję uznana będzie tu treść znajdująca się zwyczajowo w słowniku ogólnym w miejscu, w którym definicja jest spodziewana” (Majdak 2007: 224) – to konieczne jest sformułowanie dodatkowych kryteriów na potrzeby inżynierii językowej, która wymaga eksplicitnego wskazania precyzyjnych reguł.

4.3.1. Metodologia analizy artykułu hasłowego

Mając na uwadze specyfikę materiału badawczego, a za cel zaprogramowanie automatycznej segmentacji struktury hasłowej, wytypowano siedem kryteriów, na podstawie których zostanie przeprowadzona identyfikacja elementów hasła: **funkcjonalne, hierarchiczne, pozycyjne, referencyjne, symboliczne, sekwencyjne oraz opcjonalności.**

Kryterium funkcjonalne odnosi się do rodzaju treści zawartych w danym segmencie. Treść ta może zawierać informacje na ortografii, gramatyki, fonetyki, semantyki, słowotwórstwa czy etymologii.

Kryterium hierarchiczne opisuje poziom zagnieżdżenia elementów – innymi słowy relację nadrzędności, współrzędności lub podrzędności.

Kryterium pozycyjne służy do opisu absolutnego umiejscowienia elementu w hasle, oraz względem innych elementów na tego samego poziomu.

Kryterium referencyjne opisuje zakres, do jakiego odwołuje się dany segment – przede wszystkim czy dotyczy całości hasła, czy tylko jego części.

Kryterium symboliczne bierze pod uwagę występowanie wszystkich obiektów tekstowych i metatekstowych (symboli), które wskazują w sposób bezpośredni lub pośredni są pomocne przy interpretacji struktury.

Kryterium sekwencyjne służy weryfikacji, czy po danym elemencie może bezpośrednio występować element tego samego typu.

Kryterium opcjonalności służy do ewaluacji tego, czy dany element jest bezwzględnie obecny w każdym haśle.

Przy pomocy tychże można przystąpić do wyróżnienia elementów, które składają się na artykuły hasłowe. Koniecznym jest jednak tutaj nadmienić, że opis struktury jest oparty o poprawioną wersję tekstu ze względu na liczne anomalie strukturalne¹¹.

4.3.2. Segmentacja artykułu hasłowego

Na podstawie kryteriów opisanych w rozdziale poprzednim, zasadniczą budowę artykułu hasłowego można uznać za trójdzielną, to znaczy składają się z trzech segmentów najwyższego poziomu, które nazywane będą tutaj kolejno: **inicjalnym**, **centralnym** i **terminalnym**.

Segment inicjalny (pierwszy) funkcjonuje jako nagłówek hasła i, oprócz lemmy, zawiera informacje, które są wspólne dla wszystkich znaczeń hasła, przede wszystkim gramatyczne oraz te związane z rejestrem języka.

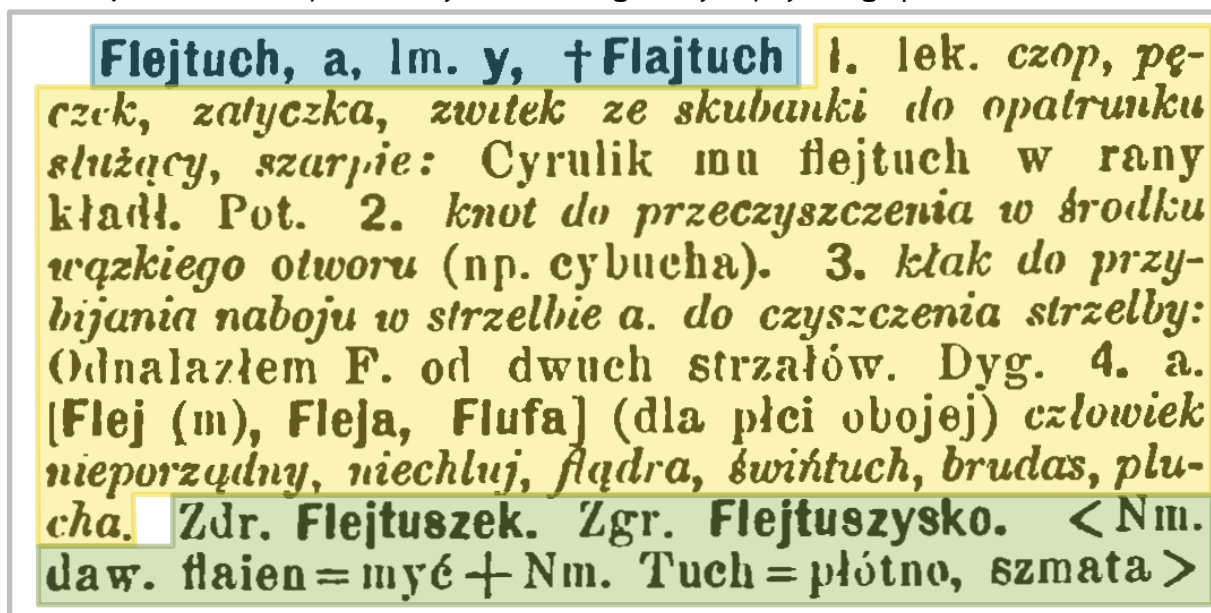
Segment centralny (drugi) poświęcony jest opisowi znaczeń lemmy, uwzględniając przy tym również wiadomości parasemantyczne związane z konkretnym użyciem wyrazu.

Segment terminalny (trzeci) wykazuje symetrię względem segmentu inicjalnego i dotyczy całego hasła, jednak zawiera informacje o słowotwórstwie, etymologii i powiązaniach z innymi hasłami.

Wizualna reprezentacja segmentów na przykładzie obrazu hasła **Flejtuch** widoczna jest na **ilustracji 4**, na której segment inicjalny podświetlono kolorem **niebieskim**, segment centralny kolorem **żółtym**, a terminalny – kolorem **zielonym**.

¹¹ Objasnienie działania podmian opisane jest w **podrozdziale 10.2.2.1**, a ich pełna lista w **Aneksie L**.

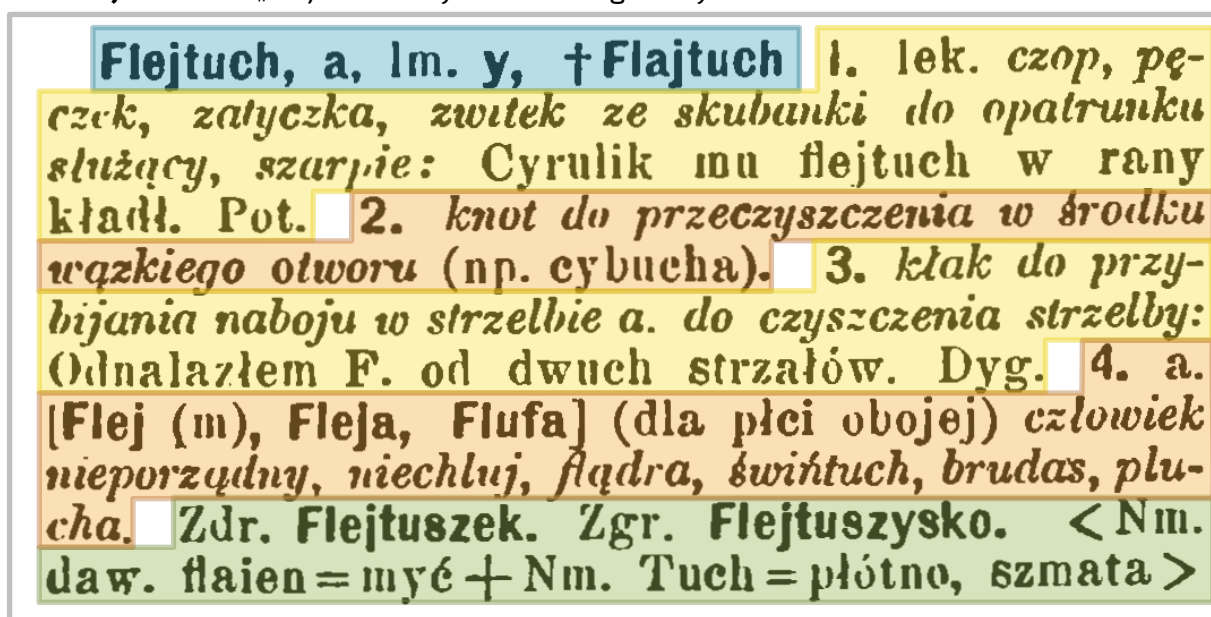
Ilustracja 4. Hasło „Flejtuch” – wyróżnione segmenty najwyższego poziomu.



Powyższą wizualizację oraz podane wcześniej informacje na temat segmentów najwyższego poziomu należy uzupełnić w tym miejscu o kilka uwag.

Po pierwsze segment centralny ma budowę sekwencyjną – oznacza to, że składa się z jednego lub więcej równorzędnych (a czasem i podrzędnych) opisów znaczeń. Tak dzieje się w przypadku hasła **Flejtuch**, które ma cztery znaczenia, co można traktować jako cztery równorzędne segmenty (zaznaczone na **ilustracji 5** kolorem **żółtym** i **pomarańczowym**).

Ilustracja 5. Hasło „Flejtuch” – wyróżnione segmenty wielokrotne.



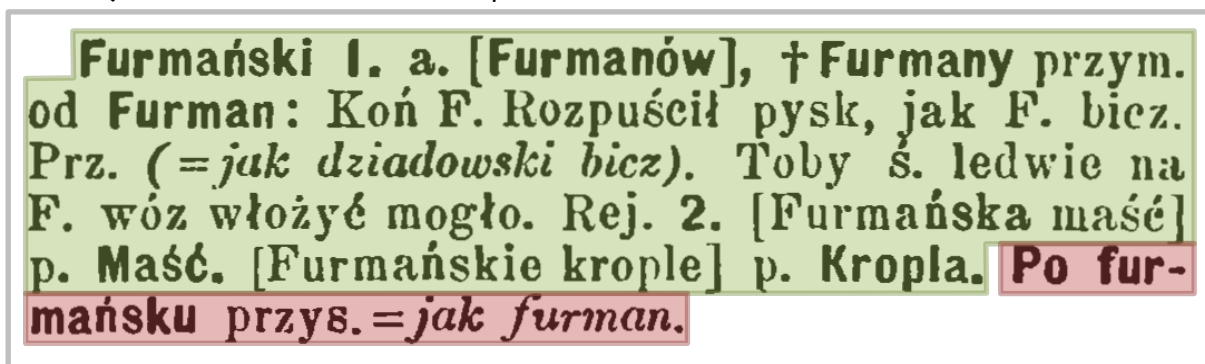
Po drugie niektóre rodzaje informacji – takie jak kwalifikatory, synonimy czy informacja gramatyczna – mogą pojawić się w segmencie inicjalnym lub centralnym (dla poszczególnych) znaczeń, zależnie od zakresu, do jakiego się aplikują; zakresy te będą zwane dalej **globalnym** (stosującym się do całości hasła) oraz **lokalnym** (stosującym się do jego fragmentu).

Mogą także wystąpić w obu segmentach jednocześnie – wówczas informacja lokalna (podana przy poszczególnych znaczeniach) uzupełnia lub zastępuje informację globalną, zależnie od zapisu. W przypadku hasła **Flejtuch** znaczenie **1.** ma kwalifikator lokalny *Lek.*, a znaczenie **4.** posiada dodatkowe synonimy o pochodzeniu gwarowym.

Po trzecie o ile segment pierwszy i drugi występują w każdym haśle, to segment trzeci (terminalny) jest w części haseł nieobecny.

Po czwarte budowę (segmentację) hasła komplikują **podhasła** (inaczej zwane tutaj również **hasłami zagnieżdżonymi**), które zgrupowane są wraz z **hasłami głównymi** (wewnątrz których się znajdują). Przykład takiej pary haseł widnieje na **ilustracji 6**, gdzie kolorem **zielonym** wyróżnione jest hasło główne **Furmański**, a kolorem **czerwonym** podhasło **Po furmańsku**.

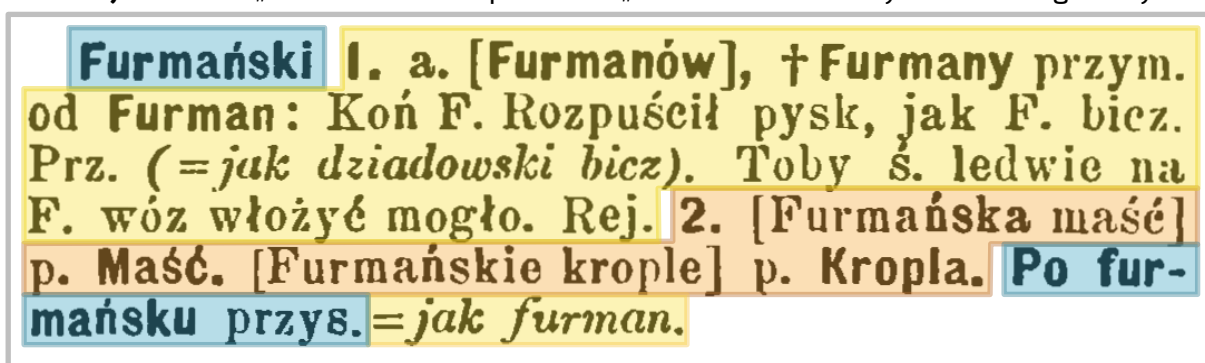
Ilustracja 6. Hasło „Furmański” oraz podhasło „Po furmańsku”.



Furmański I. a. [Furmanów], † Furmany przym. od Furman: Koń F. Rozpuścił pysk, jak F. bicz. Prz. (= *jak dziadowski bicz*). Toby ś. ledwie na F. wóz włożyć mogło. Rej. 2. [Furmańska maść] p. Maść. [Furmańskie krople] p. Kropla. **Po furmańsku** przys. = *jak furman*.

Jeśli do powyższego hasła zastosować wizualizację taką, jak na **ilustracji 5**, osiągnięty zostaje efekt widoczny na **ilustracji 7**.

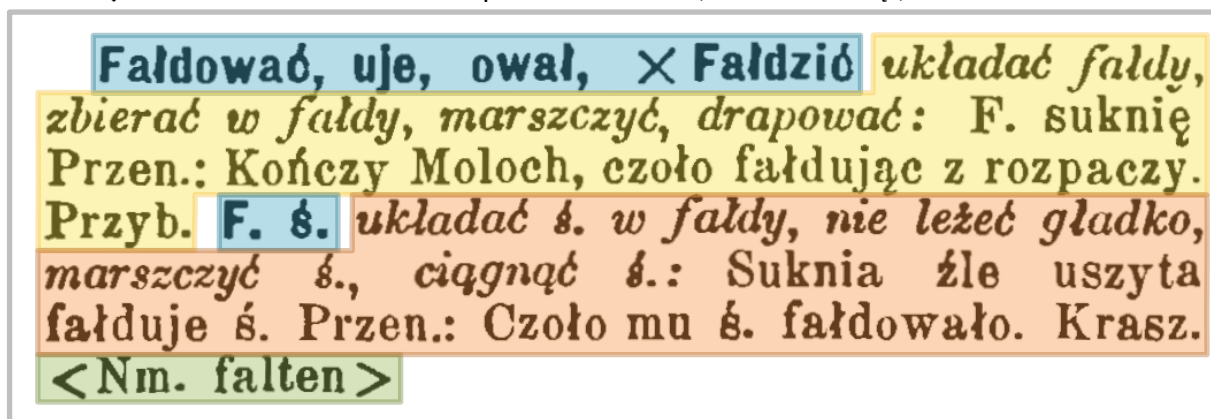
Ilustracja 7. Hasło „Furmański” oraz podhasło „Po furmańsku” – wyróżnione segmenty.



Furmański I. a. [Furmanów], † Furmany przym. od Furman: Koń F. Rozpuścił pysk, jak F. bicz. Prz. (= *jak dziadowski bicz*). Toby ś. ledwie na F. wóz włożyć mogło. Rej. 2. [Furmańska maść] p. Maść. [Furmańskie krople] p. Kropla. **Po furmańsku** przys. = *jak furman*.

Podhasła zagnieżdżone zlokalizowane są zawsze po ostatnim segmencie centralnym, ale przed segmentem terminalnym, co zostało przedstawione na **ilustracji 8**.

Ilustracja 8. Hasło „Fałdować” oraz podhasło „F. ś.” („Fałdować się”).



Na ilustracji tej widać również, że podhasła nie różnią się strukturalnie od haseł głównych, choć współdzielą z nimi segment terminalny; do ich opisu wystarczy zatem wprowadzony na początku tego podrozdziału trójpodział.

Szczegółowej analizie każdego z segmentów nadrzędnych oraz ich elementów składowych poświęcone są kolejne rozdziały. Dalsze przykłady będą na tekstowej reprezentacji haseł z zachowaniem oryginalnego formatowania.

4.3.3. Segment inicjalny

4.3.3.1. Lemma

Trzonem segmentu inicjalnego artykułu hasłowego jest lemma – jednostka leksykalna w formie słownikowej, składająca się przeważnie z jednego wyrazu, niekiedy tylko z połączenia wyrazów (abstrahując od czasowników i odsłowników z partykułą zwrotną „się”).

- **Fabryczny** przym. od **Fabryka**: Robotnik F. Robota fabryczna. (...)
- **Fajdanie się, a ś.**, blm., czynność cz. **Fajdać ś.**
- **Fiu bździu** 1. p. **Fiu**: To kobieta stateczna (...)

Choć w SW obecne są większe jednostki opisu językowego różnych rodzajów, to nie są traktowane one jako osobne hasła, lecz umieszczane są wewnątrz haseł w części przykładowej (także z opisem) – jest to zgodne z dzisiejszymi praktykami, które opisują m. in. Bergenholtz & Gouws (2010: 43):

Where idioms and fixed expressions are included as treatment units, a text production function compels the lexicographer to present items indicating the typical use of the idioms and fixed expressions.

Tam, gdzie idiomy i stałe połączenia wyrazowe są uwzględnione jako odrębne jednostki, funkcja tworzenia tekstu zmusza leksykografa do przedstawienia elementów wskazujących na typowe użycie idiomów i wyrażeń ustalonych.

Z lemmą współwystępuje często szereg elementów, z których część ją poprzedza, a część po niej następuje – podobna struktura zachowana jest w przypadku jednostek leksykalnych w innych częściach artykułu hasłowego.

4.3.3.2. Kwalifikator symboliczny

Po lewej stronie lemmy często znajduje się kwalifikator symboliczny: **krzyżyk** ×, **sztylet** †, **wykrzyknik** ! lub **asterisk** * – por.:

- ×**Fakultatywny dowolny, do wyboru zostawiony, nieobowiązujący**: (...)
- †**Falbierski** przym. od **Falbierz**: (...)
- !**Fędesjeklowy dotyczący końca XIX wieku, schyłkowy**: (...)

Poza tym w funkcji kwalifikatora stosuje się również **nawias kwadratowy** [], który to występuje wokół lemmy i przylegających do niej elementów:

- [**Fagólny**] **niepewny, na którym nie można polegać**: (...)
- [**Facjender jegier, facjendera jegra, lm. facjendery jegry**] (...)

Dana jednostka leksykalna może być opatrzona co najwyżej jednym kwalifikatorem symbolicznym, co jednak nie dotyczy asterysku, który pełni funkcję uwagi językoznawczej o tym, że podano domniemaną postać wyrazu, np.:

- [***Dworze, a, lm. a**] **w wyrażeniu**: Na dworzu (...)

4.3.3.3. Dezambiguator

W treści SW obecne są liczne homonimy, które stanowią przede wszystkim homografy, a które traktowane są jako odrębne artykuły hasłowe. Majdak (2008a: 111) udało się dzięki wnikliwej analizie zrekonstruować zasady redaktorskie, dzięki czemu jest w stanie wskazać, że homonimy wyróżniane są na podstawie czterech kryteriów:

1. etymologii wyrazu – np. **1. Tur** (zwierzę) i **2. Tur** (= „tura”);
2. podstawy słowotwórczej – np. **1. Gazowy** (od „gaz”) i **2. Gazowy** (od „gaza”)
3. kategorii leksykalnej – np. **1. [Tajj]** (rzeczownik) i **2. [Tajj]** (spójnik), oraz:
4. użytego kwalifikatora – np. **Troszka** i **†Troszka**.

Aby odróżnić od siebie wyrazy homonimiczne stosowane są dezambiguatory numeryczne, które umieszczane na pierwszej pozycji w segmencie (poprzedzając ewentualny kwalifikator symboliczny). Stosowane są one wyłącznie wówczas, kiedy para (lub grupa) homonimów nie różni się od siebie kwalifikatorami symbolicznymi (w tym ich brakiem), jak miało to miejsce w przykładach 1–3 powyżej – por.:

- **1. Futrówka, i, lm. i** 1. hut. **wyłożenie wnętrza szybu** (...)
- **2. Futrówka, i, lm. i** rub. 1. a. [**Futrunek**] (...)
- **1. [Futrunek, nku, lm. nki]** p. [**Futer**]. Por. **Furaż**.
- **2. [Futrunek, nku, lm. nki]** p. **2. Futrówka**: Dał mu F.
- **Falować, uje, ował, F. ś., ×Falić ś. poruszać ś. falą** (...)
- **×Falować, uje, ował** 1. i [**F. a. Felować**] **szwankować** (...)
- **Ikrzyca, y, lm. e** min. p. **Ikrowiec**.
- [**Ikrzyca, y, lm. e**] p. **Krzyca**. <Zam. Krzyca>
- **Flora, y, blm. ogół wszystkich a. pewnych gatunków roślin** (...)
- **1. †Flora, y, lm. y** **krepa, gaza**. <Nm. Flor>
- **2. †Flora, y, lm. y** p. **Flura**: **Włóczy ś. by flory. Jeżow.** (...)

- **1.** Fest, u, lm. y **1.** a. Festyn, Festyna, ×Festyn, ×Festa (...)
- **2.** Fest, u, lm. y *duża festona.*
- **1.** [Fest, Chwest] przym. *dziarski, gracki, zuchowaty, jędrny* (...)
- **2.** [Fest, Chwest] przys. **1.** *trwale, silnie, mocno, krzepko* (...)
- **1.** Frank, a, lm. i *jednostka monetarna we Francji = 50 gr.* (...)
- **2.** Frank, a, lm. **owie**, Firenk *nazwa Europejczyka z Europy* (...)
- **3.** Frank, a, lm. **owie** p. Frankista: Rękopis ks. Kitowicza (...)

Zastosowanie wykrzyknienia po lemmie **Fora** w charakterze oznaczenia wykrzyknienia nie jest jednak traktowane jako czynnik różnicujący hasła, i tutaj również zostały użyte dezambiguatory:

- **1.** **Fora!** **1.** w. żądający powtórzenia jakiegokolwiek ustępu na scenie a. estradzie = *bis, drugi raz, powtórnie, nanowo:* (...)
- **2.** **Fora**, y, lm. y p. **For:** *Dam ci trzydzieści for, bo* (...)

4.3.3.4. Informacja gramatyczna

Bezpośrednio po lemmie najczęściej występuje informacja gramatyczna. Ta może zawierać informację o:

1. Fleksji wyrazu:
 - **Fiksum** **nieod.** *fiksacja, bzik:* Ma F. (...)
 - †**Ferezyjka**, **i**, lm. **i** p. **Ferezja.**
 - **Fetorek**, **rka**, lm. **rki** p. **Fetor.**
 - ×**Frybiter**, **a**, lm. **owie a. rzy** p. **Frajbifer.**
 - **Fertanie się**, **a ś.**, blm., czynność cz. **Fertać ś.:** (...)
 - **Fajdać**, **a**, **ał**, F. **ś.**, †**Fejdać**, [**Fejtać**] *wypróżniać ś.* (...)
 - ×**Filtrum**, **w lp. nieod.**, lm. **a**, **ów** p. **2. Filtr.**
 - **Flader**, **dru**, lm. **dry** (...) **2. (tylko w 1-ym przypadku) kłamca, łgarz, krętacz, oszust, cygan, plotkarz:** (...)
2. Przynależności do części mowy:
 - †**Franca!** **w.** *do kata!, do djabła!, do licha!:* (...)
 - **1.** [Fest, Chwest] **przym.** *dziarski, gracki, zuchowaty* (...)
3. Rodzaju gramatycznym – wyłącznie w przypadku nieoczywistej klasyfikacji:
 - **Flaczysko**, **a**, lm. **a m. i ni.** p. **Flak:** F. obrzydliwe.
 - [**Fleja**, **i**, lm. **e**] **m. i ź.** p. **Flejtuch.**
4. Aspekcie gramatycznym – przy synonimach lub w segmencie centralnym:
 - **Fryknać**, **nie**, **nął**, **nied.** **Frykać** *wciągnąć smarki nosem.* (...)
 - **Ferować**, **uje**, **ował 1. (dok. i nied.)** *postanowić* (...)
5. Zbiorowym znaczeniu – częściej w segmencie centralnym:
 - **Furmaństwo**, **a**, blm. (...) **2. zb.** *furmani:* (...)
 - **Fajans**, **u**, [**Fajens**] (...) **2. lm. e zb.** *naczynia wyrabiane z tej gliny.* (...)

Nierzadko różne typy informacji gramatycznej współwystępują ze sobą:

- †**Francia**, **i**, blm. **zb.** = lm. od **Frant.**
- **Fraj** **nieod.** **przym.** **1.** a. ×**Frej**, ×**Frejowy**, [**Frajny**] (...)
- [**Fraj**] **nieod. rz.** *w wyrażeniach:* Chodzić na F. (...)

- **Flegmisko**, a, blm. ni. p. **Flegma**: (...)
- **Fugnąć**, nie, nął, nied. **Fugać** a. **Fugować** ucz. *uciec z lekcji*.
- **xFryczysko**, a, lm. a m. i ni. p. **Fryc**.

Zdarza się też, choć bardzo rzadko, że konkretne informacje gramatyczne opatrzone są kwalifikatorem – jest tak w przypadku zaledwie jednego hasła w materiale badawczym, ale za to aż pięć w materiale referencyjnym (litera I):

- **Fundament**, u, lm. a a. y, **†owie**, [Fudament, Felement] (...)
- **Imienie**, a, lm. a, **†imiona** 1. † *majętność rodowa, familijna*: (...)
- **Imię**, enia, 3 pp. **imieniu** a. **!imieniowi**, lm. ona a. [enia] (...)
- **Interesik**, u a. **xa**, lm. i p. **Interes**.
- **Iść**, idzie, **szedł**, **szła** a. **†szedła**, **szło**, nieos. **†szto** (...)
- **Iściec**, **śca**, **stca**, a. [**szca**] lm. **stce** a. [**szce**], **†Istca** 1. (...)

Stosunkowo niekonsekwentnie zostały potraktowane afiksy – por.:

- **Byle** (...) 2. **przyrostek** oznaczający wybór dowolny: *-kolwiek, -bądź, (jak) chcesz*: (...)
- **Izo-** chem. **przyrostek**, *którym odróżniają niek. związki izomeryczne*: (...)
- **Leuko-** chem. **przyrostek**, *dodawany często do nazw ciał bezbarwnych, a blisko spokrewnionych z barwnikami*: (...)
- **Meta-** 1. chem. **przyrostek**, *służący do oznaczenia*: (...)
- **Kolwiek**, **†Kolwie** **przyrostek**, *dodany do zaimka a. przysłówka, nadaje mu znaczenie rozciąglejsze bez wyboru, bez określenia, = -bądź*: (...)

Na powyższych przykładach można zaobserwować, że:

1. informacja o klasie nie jest wyrażana skróceniem, tylko stanowi albo część definicji (**Izo-**, **Leuko-**, **Meta-**), albo osobne wprowadzenie (**Byle**, **Kolwiek**);
2. czasem stosowany jest dywiz w miejscu przyłączania jednostki lewo- lub prawostronnie (**Izo-**, **Leuko-**, **Meta-**), a kiedy indziej nie (**Byle**, **Kolwiek**);
3. niejasny jest algorytm wyboru wyróżnianych morfemów.

Dodatkowo element **-bądź** wskazany jako komponent sufiksalny w haśle **Kolwie** w słowniku występuje jedynie jako spójnik.

Problem opis części mowy nie dotyczy jednak tylko afiksów, ale także wykrzyknień, które czasem wskazywane są skrótem i wykrzyknikiem, np.:

- **Fajt!** w., oznaczający szybką ucieczkę, bieg a. upadek = 1. *szust! myk! smyk! fiut!*: (...)
- **Fiu!** a. **Fiu fiu!** 1. w. *podziwu, pochwały = ho! ho!, fi! fi!, no! no!, oho!*: (...)
- **Fora!** 1. w. *żądanego powtórzenia jakiegokolwiek ustępu na scenie a. estradzie = bis, drugi raz, powtórnie, nanowo*: (...)
- [**Fort!**, **Furt!**, **Fort!**] w. 1. *precz!, za drzw!, won!, wont!, fora!* (...)
- **†Franca!** w. *do kata!, do djabła!, do licha!*: (...)
- **Frr!** w. p. **Fru!**: (...)

a kiedy indziej wyłącznie wykrzyknikiem – por.:

- **[Fuk!]** hop!, skok!: (...)
- **[Fert!, Fyrt!]** naśladowanie szybkiego ruchu nazewnątrz, myk!: (...)
- **[Fitu!, Fiutu!]** wyraz bez znaczenia, używany w zagadkach: (...)
- **Fyr!** p. Fru!
- **[Fyrt!]** p. Fert!

Wszystkie napotkane struktury i rodzaje informacji występujące w segmencie gramatycznym artykułu hasłowego zostały opisane w **Aneksie H**, z wyjątkiem informacji o walencji, która znajduje się w **Aneksie I**.

4.3.3.5. Walencja

W przypadku niewielkiej części artykułów hasłowych bezpośrednio po informacji gramatycznej podana jest także podstawowa walencja, która ogranicza się wyłącznie do związku rządu – wyrażonego zaimkami pytajno-względnyymi w funkcji zaimków nieokreślonych, zgodnie z dawniejszą tradycją polskiego językoznawstwa – i łączliwości z przyimkami, np.:

- †Fortunić, i, iż komu a. kogo, co = darzyć, szczęścić (...)
- Freskować, uje, ował co = malować freski na czym, pokrywać (...)
- 2. Futrować, uje, ował kogo 1. bić, okładać kijem. (...)
- Frasować (...) 2. F. ś. o kogo a. [za kogo]: Czy ś. za brata (...)

Walencja podawana jest tylko w przypadkach, gdy jest nietypowa, nieoczywista, lub gdy występuje zróżnicowanie rejestrowe. Ponadto powyższe cztery przykłady reprezentują wszystkie poświadczenia walencji w segmencie inicjalnym – częściej występuje ona w segmencie centralnym.

4.3.3.6. Ciąg synonimiczny

Wiele artykułów hasłowych ma podane szeregi wyrazów równoznacznych – na wyrazy grupowane w ten sposób składają się synonimy oraz, jak nazywał je sam Karłowicz (1876: 10f), „odmianki czyli waryjanty” (wyrazy alternatywne względem nagłówka artykułu hasłowego ortograficznie, fonetycznie lub słowotwórczo); uszeregowania tego rodzaju określane tu są mianem ciągów synonimicznych. Umieszczane są one bezpośrednio po informacji gramatycznej, a w przypadku jej braku – bezpośrednio po lemmie. Kolejne synonimy są przyłączona zazwyczaj po przecinku, a także pogrupowane według kwalifikatora symbolicznego, a dopiero w drugiej kolejności alfabetycznie:

- Fajerant, u, lm. y, **Fajerand, Fajrant** rzem. (...)
- **[Fafernuszki, szek, blp., Fafernuchy]** 1. drobne pierniki (...)
- Fajans, u, **[Fajens]** 1. blm. rodzaj gliny białej. (...)
- **[Fajn nieod., Fejn, Fań, Faj]** 1. a. **[Fajny]** przym. piękny (...)
- Fajerwerkerz, a, lm. e, **xFajerwernik, xFejerwernik, xFejerwerker** ten, co urządza fajerwerki (...)
- Fajdać, a, aź, **F. ś., †Fejdać, [Fejtać]** wypróżniać ś. (...)
- **xFajnzylber, bru, blm., xFejnzylber, xFajnzylbr, xFejnzylbr, †Fańzylbr, †Fańzylber, †Fańsilber, †Fanczliber** hut. (...)
- Falbana, y, lm. y 1. a. **†Falbaka, †Fafbala, †Falbula, †Fałbuła, [Falban, Falbela]** obszycie marszczone u spódnicy (...)

Synonim może być także przyłączony przy pomocy spójnika „i”, gdy jest to powtórzenie lemmy w nowym kontekście – np. z dodanym kwalifikatorem symbolicznym czy partykułą zwrotną „się”:

- **×Farbić, i, ił i [Farbić]** p. **Farbować: Nic-zawsze wiosna** (...)
- **×Forcować, uje, ował i ×F. ś.** p. **Forsować.**

Synonimy bywają niekiedy opatrzone dodatkową informacją – np. kwalifikatorem lub skróceniem:

- **[Forknąć, nie, nął, Furknąć, nied. Forkać, Furkać]** (...)
- **Frunąć, nie, nął, [Frujnąć, Fruknąć, Furnąć, Furgnąć, Furtnąć], nied. Fruwać, ×Furać, [Frujwać, Frugać, Frukać, Furwać, Furtać, Furgać]** 1. **wzlecieć na skrzydłach, wzbić ś. w powietrze** (...)
- **Furora, y, blm.** (zwykle tylko w 4 pp.), **lep. Furorę** *niestychane wrażenie i powodzenie, podziw, zapamiętanie, uniesienie:* (...)
- **Faktura** (...) b) a. **Faktur-księga a. lep. Księga fakturowa** (...)
- **Filharmoniczny, Filarmoniczny, niek. Filharmonicy** *muzyczny, miłujący, uprawiający muzykę:* (...)

Czy zawężeniem zakresu geograficznego:

- **Fosforon, u, lm. y, (w Galicji) Fosforyn** chem. *sól kwasu* (...)
- **Fluorowódor, oru, blm., Fluowódor, Kwas fluorowodorny, Kwas fluowodorny, (w Galicji) Fluorek wodu, a. Kwas fluorowodowy, †Kwas fluszpatowy związek fluoru z wodorem, HFL.** (...)

Jak widać na powyżej, w ciągach synonimicznych występują także leksemy wielowyrazowe, niekiedy w formie zagregowanej (częściowo skróconej):

- **Febrzysztret, u, lm. y, Drzewo angustury** bot. (...)
- **Filicyna, y, blm., Kwas filiksowy a. paprociowy** (...)
- **Figomorwa, y, lm. y, Figa morwowa a. egipska, Sykomor, Syk** (...)

Charakterystyka skróconej formy zapisu szeregów wielowyrazowych synonimów jest opisana w rozdziale następnym ze względu na jej znacznie częstsze występowanie w segmencie centralnym.

4.3.3.7. Wymowa

Przy niektórych hasłach widnieje także informacja o wymowie, umieszczona w nawiasie między informacją gramatyczną a ciągiem synonimicznym. Jest ona podawana tylko w nielicznych przypadkach, kiedy polska ortografia nie daje jednoznacznej informacji – chodzi przede wszystkim o rozbitcie przy pomocy dywizu tych połączeń literowych, które zwyczajowo są interpretowane jako dwuznaki, lub dla zaznaczenia rozłącznej wymowy samogłosek, np.:

- **Felchubel, bla, lm. ble (Felc-hubel)** (...)
- **Ferzegung, u, lm. i (Fer-zegung)** (...)
- **Fitobiologja, i, blm. (Fitobi-ologja)** (...)

W materiale badawczym poświadczono 10 haseł z podaną wymową.

4.3.3.8. Kwalifikator wyrazowy

W *Słowniku warszawskim* zaimplementowany jest bardzo bogaty system kwalifikatorów, nie tylko ze względu liczbę elementów, ale także mnogość klas i funkcji; klasyfikacja typów kwalifikatorów jest dokładnie opisana w **rozdziale 5.2**, a lista wszystkich jednostek jest dostępna w **Aneksie M** – w związku z tym ich charakterystyka zostanie tutaj pominięta. Natomiast ze względu na ich zasadniczo zbieżne funkcje zostaną one tutaj potraktowane zbiorczo.

Kwalifikatory są każdorazowo umieszczane na końcu segmentu inicjalnego, por.:

- **1. Facetować, uje, ował jub. szlifować drogie kamienie. (...)**
- **Fachowe, ego, blm. prow. opłata pocztowa od pólki listowej.**
- **Facjat, a, lm. ci zł. człowiek. <p. Facjata>**

W przypadku współwystąpienia dwu kwalifikatorów stosowane są dwie strategie zapisu stosownie do tego, czy owe kwalifikatory są tego samego typu. W przypadkach kwalifikatorów tego samego typu (np. dwóch dziedzinowych) stosowany jest spójnik „i”, por.:

- **xFarbnik, a, lm. i chem. i fizj. ciato, nadające barwę innym (...)**

Podczas gdy kwalifikatory różnych rodzajów (dla przykładu dziedzinowych i rejestrowych) oddzielone są tylko spacją, por.:

- **xFarmakoidy, ów, blp. chem. daw. nazwa gromadki pierwiastków (...)**
- **Figojadka, i, lm. i bot. bł. p. Przypołudnik. (...)**

Niemniej problemem pozostaje rozpoznanie typu kwalifikatorów, jako że autorzy nie wprowadzają żadnej ich klasyfikacji w sposób jawny. Klasyfikacja systemu kwalifikatorów w *Słowniku warszawskim* zostanie przedstawiona w **rozdziale 5.2**.

Same kwalifikatory mogą być również umieszczone w nawiasie – co komunikuje, że w przypadku danej lemmy kwalifikator jest opcjonalny, tzn. przynależność wyrazu do danej dziedziny, rejestru itp. jest niejednoznaczna, np.:

- **Fantastyczny, Fantazyjny (posp. i fil.) 1. a. †Fantastycki (...)**
- **Familijka, i, lm. i (zwykle ir.) p. Familja: (...)**

W drugim z przykładów przytoczonych powyżej obecny jest także wyraz modyfikujący zakres użycia kwalifikatora – jest to jedyne poświadczony przypadek tego rodzaju.


4.3.3.9. Informacja o literze

Na początku sekcji słownika poświęconych poszczególnym literom alfabetu często występuje informacja o samej literze – mogą to być uwagi o fonetyce, ortografii czy semantyce (o znaczeniu wyrazu składającego się wyłącznie z tejże litery). Sposób zapisu tej informacji jest jednak bardzo zróżnicowany, a niekiedy wręcz niekonsekwentny:

1. W przypadku sekcji poświęconych literom **F** oraz **V** nagłówkiem jest litera z kropką (dla wyróżnienia pojedynczej litery alfabetu), po której następuje typowy segment centralny z częścią definicyjną (por. **ilustracje 9 i 10**);
2. W przypadku sekcji poświęconej literze **I** nagłówkiem jest litera z kropką (jako symbolu) lub bez niej (w pozostałych użyciach) poprzedzona dezambiguatorem (por. **ilustracja 11**);


3. W przypadku litery **Ó** nie ma ani jednoliterowego hasła, ani uwagi ortograficznej (por. **ilustracja 12**);
4. W przypadku sekcji poświęconej literze **X** obecna jest wyłącznie uwaga o funkcji tejże litery w ortografii współczesnej autorom, która nie zaczyna się od tekstu „X” (por. **ilustracja 13**);
5. W przypadku sekcji poświęconej literze **Y** obecna jest uwaga o funkcji tejże litery w ortografii współczesnej autorom, która to zaczyna się od tejże litery zapisanej kursywą (por. **ilustracja 14**).

Ilustracja 9. Początek sekcji z literą „F”.



<p>F. l. spółgłoska wargowo-zębowa bezdźwięczna, czyli cicha (moena), powiewna, przeciągła. W piśmie występuje przeważnie w wyrazach przyswojonych, w polskich zaś rzadko i głównie w wyrazach dźwiękonaśladowczych, jak <i>fruwać, fiukać</i> i t. p.; w wymawianiu zaś ukazuje się często, chociaż się pisze <i>w</i>, mianowicie na końcu wyrazów i w środku po innych cichych, np. <i>kręw, łów, ... kwiat, twój, trwały, ...</i> które brzmią: <i>kręf, łuf, ... kwiat, twój, trwały...</i> 2. a. Fa muz. = a) czwarty stopień tonacji <i>c major</i>: F leży pomiędzy <i>e</i> a <i>g</i>. b) F dur a. major, F moll a. minor = nazwy tonacji a. gam majorowej (twardej) i minorowej (miękkiej): Gama f minor ma cztery bemole. Symfonia f major Beethovena. Piosnka (z) f major.</p>	<p>bryce czyli ułożeniu serea. Krup. Serea fabrykę opisywać zaczę. Krup. 8. a. Majątek fabryczny kość. <i>majątek ściśle kościelny w odróżnieniu od beneficjum.</i> Zdr. Fabryczka. <Łć. fabrica></p> <p>Fabrykacja, i, lm. e p. Fabrykowanie: F. cukrn, aktu. F. aniołków (= <i>srowadzanie śmierci na niemowlęta</i>). <Łć. fabricatio></p> <p>Fabrykacyjny przym. od Fabrykacja: Środki fabrykacyjne.</p> <p>Fabrykancki przym. od Fabrykant.</p> <p>Fabrykant, a, lm. ci i. ten, co fabrykuje, rękodzielnik, przemysłowiec, właściciel fabryki: F. wyrobów jedwabnych. Przen.: F. monety (= <i>fałszerz</i>). F. plotek (= <i>plotkarz</i>). Fabrykantów nowych słów i odmian okrywano obelgą. Śniad.</p>
---	---

Ilustracja 10. Sekcja z literą „V”.



<p>V. l. = w, zachowywane u nas niekiedy w piśmiennym wyrazach cudzoziemskich, jak: Vidal, Villiers i t. d. 2. † V = u, w. 3. rzymska cyfra 5; używa ś. u nas na oznaczenie klas, stuleci i in. 4. na monetach jest znakiem wybitcia ich w mieście Troyes. Sł. wil. 5. w skröceniach = a) albo, to jest, inaczej mówiąc, czyli (vel). b) wice: V. prezes, c) na nutach muz. = <i>skrzypce</i> (violino) a. <i>głos</i> (voce) oraz: <i>verte</i> (= <i>odwrót, przewrót karty</i>).</p>	<p>Veto, Weto, nieod. „<i>nie pozwalam</i>“, formuła zrywająca sejm w daw. Polsce: Siciński pierwszy dał przykład zerwania obrad użyciem bezprawnym <i>veto</i>. Mick. Przen.: Ja tu postawię swoje V.</p> <p>Viertelnik, a, lm. cy rodzaj urzędnika policyjnego, dzielnicowy: Straże miejskie, <i>viertelnicy</i> i dozorey. Krasz.</p>
--	--

Ilustracja 11. Początek sekcji z literą „I”.

I

I. I samogłoska podniebienna, nosi też nazwę samogłoski miękkiej z powodu, że ma własność miękczenia poprzedzającej ją w wyrazie spółgłoski, np. cisza, miły, sito, zima; dla tego w piśmie miękkość spółgłoski przed następnym *i* nie oznacza ś. dzisiaj osobnym znakiem miękczenia (kreską'), t. j. piszemy: cisza zam. ćisza, siwy zam. śiwy, zima zam. żima, ginąć zam. ginąć i t. p. W tych więc razach *i* jest jednocześnie samogłoską i znakiem miękczenia; w zgłosce *li* (list, byli...) wyobraża tylko samogłoskę. Pisze ś. więc po wszystkich spółgłoskach miękkich: b', ó, dź, ś... k', g', gdy tymczasem po spółgłoskach powstałych ze zmiękczenia: cz, dź, sz, rz... (dziś twardych) pisze ś. stale *y*. W wyrazach polskich *i* na początku ma znaczenie zgłoski *ji*, np. inny (=jinny), igła (-:jigła); to samo po samogłoskach w środku wyrazów i na końcu, np. stroić (=strojić), nadziei (=nadzieji), moi (moji) i t. p. Litera *i* bywa w środku wyrazów tylko znakiem miękczenia poprzedniej spółgłoski przed następną samogłoską, np. cieszyć (=ćeszyć), siano (=śano), ziarno (=żarno), i t. p.

Zamknij, i taki w izbie niewiele ciepła. Troc 12. [I]=a) *ani*: Zaliczył ś. pan starosta, ni miał i sielazka. Do mnie nie napisał i słowa. b) *też*: Cóż ja teraz i ucyniła! c) [I tak]=*mimo to, jednakże*: Tén mu gádá: nie bier piórka, bo bedzies miał biéde, ale un go i tak wzián. d) [I nie trzeba]=*i trzeba, aż tu, aż tu właśnie, jak na to*: I nie trzeba, że właśnie pod owe czasy szedł sobie chłopak. <I>

3. I w. I. *eh!*, *e!*: I! daj mi pokój. I! ja tego nie chcę. I! skądże ci wezmę. 2. *hal!*: I ja błazen, i ja głupi wierzyć! I to sprawa, rozum, rząd! †I to bieda. Kn. (=co za nieszczęście!). 3. [I] wyraża zdziwienie: I! I dzież! I eeh! I nu! <Okrzyk.>

4. I. fil. w logice: *symbol sądu szczegółowego twierdzącego* (jako druga samogłoska wyrazu *łać* affirmo, twierdząc).

†I, ja, je, z. p. II.

†I=go.

†Iba, y, lm. y p. Ibis.

[Iberszlag, u, lm. i] *przewyżka, nadwyżka*. <Nm. Ueberschlag>

Ilustracja 12. Początek sekcji z literą „Ó”.

Ó

[Ódka, i, lm. i] p. Wódka. <(W)ódka>
 Ólka, i, lm. i fl. p. Hólka. <p. Hólka>
 [Ólszka, i, lm. i] *mała olcha*: Spuściel tako ólske. <Od Olcha>
 [Ółtarz, a, lm. e] p. Ołtarz: Przystąpiła do ół-

wa arkusza złożonego z 16 kolumn, 8 kolumn. 8. fecht. *ruch palasza, opisującego końcem figurę liczby 8.* 9. muz. *ósma część nuty, nuta raz wiązana.* <Od Ósma, od Ósmy>

Ósemkowato przys. od Ósemkowaty.

Ilustracja 13. Sekcja z literą „X”.



Ponieważ zasadą pisma głoskowego jest, ażeby każda pojedyncza litera odpowiadała *pojedynczemu* dźwiękowi, przeto niewłaściwie sobie postąpili nasi przodkowie, przyjmując do abecadła polskiego literę *x* dla wyrażenia *dwu* naraz spółgłosek: *k* i *s* lub (między samogłoskami) *g* i *z*. Zaczęto od pisania przez tę literę wyrazów greckich i łacińskich, jak: axamit, Xerxes, examen, exekucja, exempcja, exorbitancja, emisja i t. d., przyczym, jak widzimy, nie uwzględniano pod wpływem obcej pisowni różnicy dźwięków *ks* i *gz*. Następnie literę *x* widzimy u dawnych naszych pisarzy i w niektórych wyrazach polskich, jak: *książę* wraz z wszystkimi od niego pochodniami, *xieni*, *książka*, *księżyc* i t. d., przyczym pisarze ci (Jan Kochanowski, Januszowski, Górnicki i in.) wprost zachęcali do używania tej litery w piśmie polskim. Deputacja Towarzystwa przyjaciół nauk radzi pisać wyrazy cudzoziemskie przez *x*, a polskie przez *ks* i *gz*. Krakowska Akademia umiejętności

również nie wyklucza użycia w pisowni polskiej litery *x*. Ponieważ tedy czytelnicy nasi znajdować będą w piśmiennictwie naszym wyrazy polskie pisane przez *x*, uważałem za potrzebne powiedzieć tych parę uwag, upoważniających słownikarza polskiego do wykazania wyrazów, o których mowa, zaczynających się dawniej od *x*, nie pod nieistniejącą prawie teraz w polskim abecadle literą *X*, ale pod literą *K*. W skróceniach a) oznaczało literą *x* wyrazy: *ksiądz*, *książę*, *książęcy*, *księga* (= dział dzieła): *X*. pleban, *X*. biskup, *XX*. (księga) Bazylianie: *w. x.* litewski (= wielki książę litewski), *W. X.* Mość (= wasza książęca mość), tom II, x. 3, (= tom drugi, księga trzecia); b) *X P S* znaczyło *Chrystus*, tutaj bowiem *X* i *P* występują jako greckie litery (= *ch* i *r*); c) w matematyce *x* oznacza ilość niewiadomą; d) jako liczbę rzymską oznacza *10*, przewrócone bokiem — *tysiąc*, prosto stojące z kreską u góry — *10000*.

Ilustracja 14. Początek sekcji z literą „Y”.



Y (ypsylon, ipsylon, igrek) jest literą przejętą przez łacinników i Polaków od Greków. Używa ś. ona do wyrażania dźwięku jakby *i* po-

zbawionego jotacji, a ponieważ język polski każde *i* zaczynające wyraz lub zgłoszkę jotuje, przeto nie może być wyrazów polskich (*i* zgło-

4.3.4. Segment centralny

Segment centralny stanowi najbardziej złożoną część artykułu hasłowego. Dzieje się tak ze względu na:

1. liczbę unikalnych elementów wszystkich poziomów, które mogą pojawić się w segmencie;
2. liczbę możliwych kombinacji tychże elementów;
3. występowanie wielopoziomowych struktur, na które elementy te się składają;
4. występowanie sekwencji struktur danego poziomu;
5. zagnieżdżania struktur wysokopoziomowych wewnątrz innych struktur wysokopoziomowych.

Dzieje się tak dlatego, iż podstawową jednostką w segmencie centralnym jest opis pojedynczego znaczenia lemmy; tych z kolei może być wiele, a każda z nich może mieć podznaczenia i przykłady użycia, w których może być zagnieżdżona samodzielna informacja semantyczna. W segmencie tym obecnych jest także część elementów z segmentu inicjalnego – ale z tą różnicą, że stosują się one do znaczenia, które poprzedzają, a nie lemmy głównej.

4.3.4.1. Definicja

Trzonem segmentu centralnego jest opis znaczeń lemmy – ten stanowią przede wszystkim definicje zapisywane kursywą, np.:

- [Fačka, i, blm.] *zielsko posiekane dla świń.* <?>
- xFaflować, uje, ował *mówić sposobem niezrozumiałym i nieprzyjemnym, paplać.* <p. Fafla>
- Fakcjonista, y, lm. *ści tworzący facje, stronnictwa; wichryciel, burzyciel, podżegacz, agitator, buntownik.*

Definicje mogą zawierać w sobie także takie elementy jak kwalifikatory symboliczne czy skrócenia, te jednak nie pełnią odrębnych funkcji strukturalnych – por.:

- [Fajczyć, y, yż] *palić fajkę a. wog. tytuń.*
- Farbka, i, lm. *i barwnik niebieski, używany przy praniu bielizny, farba berlińska, lazurek berliński, błękit berliński, [sinka], [łazurek].*
- [Furcicki, ckiego, lm. ccy] *człowiek [okrzętny].*

Niektóre definicje poprzedza także rodzaj komentarza zapisanego antykwą, który doprecyzowuje jej zakres lub znaczenie:

- Flejtuch (...) 4. a. [Flej m., Fleja, Flufa] *(dla płci obojej) człowiek nieporządny, niechluj, flądra, świńtuch, brudas, plucha.*
- Fluks, u a. a, lm. y (...) 2. a. †Flus lek. = a) *(według pojęć dawnych) napływ nieprawidłowy soków do pewnej części ustroju: (...)*

Częściej jednak zawiera on ekwiwalent terminologiczny w języku łacińskim, w tym taksonomię naukową:

- Fuks (...) 6. zool. *(vanessa polychloros, vanessa urticae) motyl dzienny.*
- Ficlaus, a, lm. y, Ficlauz, [Bekas kurka, Bekas mały] zool. *(limnocryptes gallinula) ptak podkasaty z rodziny bekasów. (...)*
- Figa, i, lm. i (...) 5. w lm. a. †Fiżki lek. *(condylomata in ano) sadzele bolesne w siedzeniu, szyszkowiny. Urzęd. (...)*

Tego rodzaju elementy są odrębne strukturalnie i powinny być traktowane jako takie – tak jednak nie jest, ponieważ z jednej strony wyodrębnienie tych dwóch kategorii nie przeprowadzając rozpoznania języka jest utrudnione, z drugiej zaś funkcjonalnie są zbyt różne, żeby można było je zgrupować. Inaczej jest w przypadkach opisanych w **podrozdziale 4.3.4.12**, które udało się wyodrębnić prostą regułą, a także w przypadku komentarzy pełniących zbliżoną funkcję, którym poświęcony jest **podrozdział 4.3.4.13**; niespójność ta jest dalej opisana w **podrozdziale 10.8.2**.

Definicja wyrazu hasłowego lub konkretnego znaczenia jest niekiedy pomijana w całości – informacja o semantyce może być zakodowana w przykładach użycia albo poprzez odsyłacze, które wskazują czytelnikowi, gdzie tej informacji szukać:

- **Farbierstwo**, a, blm. **1. p. Farbiarstwo.** (...)
- †Fairka, i, lm. **i p. Fajerka: F., stołowa do grzania potraw. Kn.** <p. Fajerka>

Definicje pojęć, zwrotów i wyrażen przykładowych wprowadzane są przy pomocy znaku równości – np.:

- **Federkantowy** tap. przym. od **Federkant**: Szpagat F. = *szpagat średniej grubości, używany do federkantu.*
- **Fuzyjka**, i, lm. i 1. p. **Fuzja**: Fuzyjką ś. bawić = *być amatorem polowania.* 2. [F.] *walcowate naczynie z wązkim otworem*: Portki na fuzyjkach = *spodnie wązkie, na nogach opięte.*

Jeśli w takich przypadkach obecny jest komentarz, to wówczas definicja umieszczana jest po nim – por.:

- **Felc** (...) 2. (...) *cieś. i stol. wręb, nacięcie brzegów drzewa a. kamienia dla ściślejszego spajania*: Dać co na F. (= *spoić brzegami*). F. w futrze od drzwi a. okna (*na pomieszczenie kitu*) = *przyłga.* (...)
- **Fizyk** (...) 4. **Ryzik** F. (*wyrażenie przysłowiowe*) = *na Los szczęścia, na chybił-trafił, co Bóg da.* (...)

Można jednak natrafić na odstępstwa od tej reguły:

- **Fijołkowaty, Fijołkowy** bot.: *Rośliny fijołkowe a. fijołkowe (violaceae) rodzina przyrodzona roślin dwuliścieniowych.*
- **Fukierowy** bot.: *Rośliny fukierowe (fouquieriae) plemię rodziny tamaryszkowatych.* (...)

Interesującą właściwością definicji jest fakt, że zawierają one mogą niemal dowolną treść, w tym taką, która bardziej pasowałaby do innego segmentu – np. końcowa część definicji hasła **Figuchna** zawiera informację zakrawającą o derywację lub etymologię, a przynajmniej odsyła do odrębnego hasła:

- **Figuchna**, y, lm. **y** *rzecz drobna; figa = mała osóbk.*

W niektórych przypadkach także informacja typograficznie odrębna bywa traktowana jako część definicji – dzieje się tak na przykład w przypadku poszerzonej informacji o części mowy (co zostało poświadczone tylko w przypadku wykrzyknień), ponieważ dotyczy ona znaczenia wyrazu:

- **Fe!** 1. a. **Pfe!**, **Fi!**, **Fuj!**, **Pfuj!**, **xFa!**, **†Pfa!**, **†Pfy!**, **†Pi!**, **†P'!** *w. obrzydzenia, wstrętu, nagany = tju! be! bwe! bla!:* (...)
- **1. Fora!** 1. *w. żądający powtórzenia jakiegokolwiek ustępu na scenie a. estradzie = bis, drugi raz, powtórnie, nanowo:* (...)

4.3.4.2. Numeracja

W przypadku, gdy dane słowo ma więcej niż jedno znaczenie, stosowana jest numeracja znaczeń. Ta w zależności od ich pogrupowania i hierarchizacji może zawierać cztery typy oznaczeń porządkujących – są to kolejno (od najwyższego poziomu do najniższego):

1. cyfry rzymskie (wraz z kropką);
2. cyfry arabskie (wraz z kropką);
3. litery łacińskie (wraz z zamknięciem nawiasu okrągłego);
4. litery greckie (wraz z zamknięciem nawiasu okrągłego).

Spośród wymienionych tylko cyfry arabskie są wyróżnione typograficznie – poprzez wytłuszczenie druku. Numeracja jest zawsze pierwszym elementem każdego podsegmentu opisującego pojedyncze znaczenie. Mimo teoretycznej możliwości zaistnienia czteropoziomowego opisu znaczeń sytuacja taka nie została poświadczona – w materiale badawczym maksimum wynosi trzy poziomy w jednym haśle.

Najczęstsza jest numeracja cyframi arabskimi, które służą do wyodrębnienia podstawowych znaczeń, np.:

- **[Fafernuszki, szek, blp., Fafernuchy]** **1.** drobne pierniki; orzeszki z piernika. **2.** nowolatki (kwiatki). (...)
- **Farfocel, cla, lm. cle** **1.** rub. strzęp, tachman wiszący: Podszewka ci ś. obwiesita i wiszą farfocle. **2.** [F.] część. w lm. szyszkowina, w lm. sadzele bolesne w siedzeniu. (...)
- **Fantownik, a, lm. cy** **1.** p. Fancierz. **2.** † zaciągający pożyczkę na fant. (...)
- **Figaro, a** **1.** lm. owie cyrulik, fryzjer. **2.** lm. a staniczek, wkładany na stanik, in. ~żuawka~. (...)
- **Faryniarka, i, lm. i** **1.** forma ż. od Faryniarz pod **1.** **2.** [F.] sprzedająca duszoną gęsinę a. flaki na gorąco.

W drugiej kolejności używane są litery łacińskie, które określają podznaczenia:

- **Facjendować, uje, ował** **1.** a) zajmować ś. facjendami, przeprowadzać interesa handlowe, spekulować. b) co za co = zamieniać. **2.** zż. dzielić ś. zdobyczą z kradzieży.
- **Folarz, a, lm. e** **1.** a. Foluszniak, [Foluśnik] blecharz sukna, blicharz, wątkarz, pilśniarz. **2.** × gór. = a) a. ×Fułarz, ×Forarzen, co zwozi węgle do pieca a. do fryszerki. b) p. Narażacz. (...)

Mogą być także stosowane jako pierwszy poziom podziału znaczenia, kiedy wprowadzają podkategorie (odmiany, gatunki, warianty) tego, co opisuje lemma główna:

- **Fluoryt, u, lm. y, Fluspat, †Szpat fluoryczny** min. fluorek wapnia, tworzący kryształy równoosiowe a. agregaty gruboziarniste i pręcikowate, najrozmaiciej zabarwione, w Anglii obracane na ozdoby i sprzęty, a w starożytności prawdopodobnie na tak zwane vasa murrhina. Odmiany: a) Chlorofan, fosforyzujący przez ogrzewanie; b) Flusztajn, zbity, używany jako topnik w robotach metalurgicznych, do otrzymywania fluorowodoru i rytowania na szkłe. (...)

Pierwszy segment oznaczony literami łacińskimi często jest wprowadzany przez znak równości, np.:

- **Fantazjować, uje, ował** **1.** a. ×Fantazować, †Fantastykować = a) rzędzić ś. fantazją, snuć z fantazji, tworzyć, roić, marzyć, zmyślać: F. w feljetonach. Koż. (...)
- **Furmanek, nka** **1.** × lm. nkowie = a) p. Furman. (...)

To samo tyczy się rzadkich przypadków znaczeń zagnieżdżonych w definicji głównej:

- **Filigranowy** przym. od **Filigran**; **xFilogranowy**, **!Filgranowy**: Robotą filigranową. Przen. = a) *delikatny, kruchy, łatwy do zniszczenia*: F. most, budynek. Filigranowe zdrowie. (...)

W przypadku zastosowania liter łacińskich do ponumerowania kolejnych elementów wprowadzonych w ramach typologii (opisanej w **podrozdziale 4.3.4.11**), kolejne segmenty oddziela kropkami lub średnikami, czasem niekonsekwentnie – por.:

- **Fasola** (...) bot. **1.** (phaseolus) *roś. z rodziny motylkowatych*. Gatunki: a) F. karłowata a. pieszka (*nizka*), pospolicie Piechotą zwana (ph. nanus); b) F. pospolita a. zwyczajna (*wysoka*), pospolicie F. tyczkowa, Bób turecki, Groch turecki a. włoski a. szablasty, Szabelbon, Szablak zwana (ph. vulgaris). c) ogr. F. szparagowa (*z miękkim strąkiem, bez pergaminu*). d) F. wielokwiatowa, także Szabelbon zwana (ph. multiflorus). (...)
- **Forma** (...) **15.** min. *bryła ograniczona pewną liczbą płaszczyzn*: a) F. całkowita; b) F. ćwiartkowa; c) F. krystaliczna; d) F. otwarta; e) F. pojedyncza (= *ograniczona jednakowymi płaszczyznami*); f) F. połowiczna; g) F. zamknięta; h) F. złożona (= *ograniczona niejednakowymi płaszczyznami*), in. kombinacja.

W materiale badawczym odnotowano tylko jedno hasło stosujące numerację przy pomocy cyfr rzymskich, przytoczone poniżej w całości:

- **Fluks**, u a. a, lm. y **I.** † = **1.** a. †**Flus** *wszystkie jednego koloru idące po sobie karty, sekwens*: Nuż jakie dziś gry nastały, z których płynie koszt niemały: dziwne fluksy, turmy, ruszę. Rej. **2.** a. †**Flus** lek. = a) (według pojęć dawnych) *napływ nieprawidłowy soków do pewnej części ustroju*: Wino z jabłek granatowych kwaśnych skuteczne jest przeciw płynieniu wewnętrznemu po członkach, to jest fluksom. Ślesz. F. w szczęce, gdy wszystkie zęby po jednej stronie bołą. Perz. (= *fluksja*). Przygody pijanych: podagra, F., scjatyka. Rej. Reuma, to jest niezyt, która jest matką, wszystkich niemocy, zwłaszcza fluksów. And. z Kobyl. b) *cieczenie ustawne*: Ryma i sapka jest fluksów ściekanie. Troc. F. ust = *ślinienie, ślinotok*. F. hemoroidalny = *krwawnice*. F. miesięczny = *miesiączka*. **3.** *cieczenie rzeki, odnoga*: Dunaj siedmiu fluksami w Czarne morze wpada. **4.** † *pęd (wody)*: Ocean o skopuły roztrąca okręty, na które kiedy już niewstrzymanym flukssem wpadają... Pot. **II.** [F.] = **1.** *ciecz, płyn*. **2.** *choroba bydłęca*: Mieć fluksa (o ludziach) = *być porywczym, nagłym, bez zastanowienia*. Por. **Flus**. <łć. fluxus>

Podobnie jest w przypadku uporządkowania znaczeń przy pomocy liter greckich:

- **Fijołek**, **łka**, lm. **łki**, **Fjołek** **1.** a. **xFijoł** bot. = a) a. **Fijałek**, †**Fijołka**, †**Fijałka**, [**Fijołka**, **Fijałka**, **Chwiałek**, **Chwilka**, **Chwiłek**, **Chwiołek**, **Chfiołek**, **Chwijałek**, **Chfijałek**] bot. (*viola*) *roś. z rodziny fijołkowatych, (i jej kwiat)*. Gatunki: F. dziki

(v. canina); F. pachnący a. wonny, zwany Fijołkami marcowymi a. brunatnymi (Viola odorata); ~F. trójkolorowy~ (v. tricolor), bardzo zmienny: α) odmiana ogrodowa, zwana Bratki, Braciszki, Brat z siostrą a. Brat i siostra; β) odmiana polna (arvensis), zwana u ludu Maślukami. b) F. alpejski p. Gduła. c) F. biały p. Gładyszek. d) Fijołki p. Podejźrzon. e) Fijołki nocne p. **Wieczornik**. f) Fijołki żółte p. **Lak**. (...)

W powyższym artykule hasłowym można także spostrzec elementy opisu przy, które – choć nie są zapisane kursywą – pełnią funkcję definicji.

Ze względu na rzadkość występowania trudno jest zrekonstruować reguły, którymi kierowali się autorzy stosując liczebniki rzymskie oraz litery alfabetu greckiego do numeracji. Poza tym odnotowano przypadki, w których odcień znaczeniowy jest wyróżniany bez numeracji, a jedynie przy pomocy kwalifikatora zapisywanego przeważnie wielką literą, po którym następuje dwukropek, np.:

- **Farba, y**, lm. **y**, [Farwa] 1. *kolor, barwa, maść*: Każdy promień, głos każdy z podobnym spojony harmonją, ogłasza przez farby i tony. Mick. Tylą farb różnych, żywych mak źrenicę mam. Mick. Farbą wstydu zarumieniła swe lice. Węg. Farbą wstydu zarumieniła swe lice. Węg. **× Przen.**: *cera, pteć*: Farbę ma na twarzy ceglaną. Rej. (...)

Numeracja bywa pomijana, jeśli kolejny odcień znaczenia jest wprowadzony przy pomocy kwalifikatora, po którym nie występuje dwukropek, np.:

- **Falować, uje, ował, F. ś.**, **×Falić ś.** *poruszać ś. falą, bałwanić ś., kotysać wodą*: Ocean faluje. **Przen.** *poruszać ś. naksztatt fali, chwiać ś., kotysać ś.*: Pierś jej falowała lekko. Zach. (...)
- **Folwarczny, ×Folwarczany, ×Folwarkowy** przym. od **Folwark**: (...) **Przen.** *grubijański, prostaczy, rubaszny*: Zachowanie ś. jego folwarczne. (...)

W przypadku haseł i podhaseł takich jak **Feruła** czy **Po farańsku** występuje inny rodzaj dewiacji od tej reguły: definicja wskazuje na dwa różne znaczenia – dosłowne i przenośne – które jednak strukturalnie są zgrupowane, a nie wyróżnione numeracją:

- **Farański** p. **Faraonowy**. **Po farańsku** = *jak faraon*; **przen.** *okrutnie*: (...)
- **Feruła, y**, lm. **y** *różga*; **przen.** *dyscyplina, regulamin, władza, powaga*: (...)

4.3.4.3. Przykład użycia

Przykłady użycia ilustrują znaczenie opisywanej lemmy w kontekście; niektóre są spreparowane przez autorów, inne zaczerpnięte w całości z literatury – wówczas występują wraz z podanym źródłem. Najczęściej znajdują bezpośrednio po definicji, od której oddzielone są dwukropkiem, i zapisywane są antykwą:

- [Fafecie, i, blp.] *suknie, szmaty*: **Zdyjmij, miła, twe fafeci, a weź miotłę, zamieć śmieci.**

- [Fafreği, ów, blp.] strzępy: Zdarła kosiule na F.

Egzemplifikacja może także wystąpić po innych elementach, takich jak odsyłacze czy kwalifikatory:

- †Fa! p. Fe!: Nie mogąc wstrzymać, zawołali: fa, fa! Sł. L. (...)
- ×Falcet, u, lm. y p. Falset: Baba wtóruje mu falcetem.
- Felcowy przym. od Felc: Spójnia felcowa.
- Ferotypowy fot.: Plata ferotypowa = czarna blacha, służąca za kliszę do fotografii.
- Fabrykant, a, lm. ci (...) 2. cukier. żart.: F. lodowaty = ten, co przyrządza lody w pracowni cukierniczej. (...)

Pojedynczy element może przyłączać wiele przykładów użycia:

- Fajka, i, lm. i 1. (...) przyrząd do palenia tytoniu, Lulka, cybuch z Lulką: F. porcelanowa, drewniana, piankowa. Palić, kurzyć, ćmić fajkę. Przestańże tam trąbić na fajce! Mick. Ma nos, jak węgierska F. Prz. (...)
- Fakt, u, lm. y a. a 1. rzecz dokonana, pewna; zdarzenie, wypadek; zjawisko; czyn; okoliczność: Faktu negować nie można. Fred. A. F. historyczny, naukowy. Przeciwno faktowi bezsilne są argumenty. Przypuszczenie stało się faktem. O odzyskaniu tego spadku mówił mi, jako o fakcie nadzwyczajnym. Krasz. F. spełniony, dokonany. Dopuszczyć się faktu niesłychanego. O prawdziwości tej pogłoski świadczy F., że pochodzi ona z wiarogodnego źródła. (...)
- Falanga, i, lm. i 1. pewien sztyk bojowy w Macedonji starożytnej: F. wielka miała szesnaście tysięcy żołnierzy. 2. przen. zastęp, gromada, tłum: F. przestępców, robotników. Zbita F. żywiołu żydowskiego. 3. stowarzyszenie socjalistyczne według systemu Fourriera: Czysty furjeryzm i zastosowanie falang do krajów nieznanych zdobycia. Słow. (...)

Niekiedy w przykładach użycia zawarte są warianty niektórych wyrazów (niekiedy nie samej lemmy głównej), co wskazywane jest przy pomocy nawiasu:

- Filant, a, blm. bzik, zajczek: Ma (nabył) filanta. Prz. (o figlarzu i awanturniku pijanym). (...)
- Fortepjanowy przym. od Fortepjan: (...) Układ (wyciąg) F. (Klavierauszug). Koncert F., a. na fortepjan.

Jeśli wyraz w przykładzie występuje w formie identycznej do formy hasłowej, wówczas jest skrącany do pierwszej litery zapisanej kapitalikiem; w przeciwnym razie jest zapisany w formie pełnej – por.:

- Feldmarszałek (...): Generał F. Wszedł napowrót w cesarską służbę jako F.-lejtant. Rzew. <Nm. Feldmarschall>
- Fletnia, i, lm. e 1. p. Flet: Gram na harfie, lub na fletni dmucham. Zabł. F. ze siedmiu piszczałek złożona. Bobr. Ucieszna fletni moja, niebios darze złoty, ty folgę sercu czynisz, ty koisz kłopoty. Nar. Sweśmy głowy sypali prochem, a dumając na swych fletniach o śmierci, gorzko narzekali. Gawiń. Na biesiadach

starzy przodków sławne dzieje przy graniu na **fletniach** śpiewali. Warg. (...)

- **Foremny** 1. odpowiedni normalnej formie, kształtny, regularny, prawidłowy, umiarowy: **F.** nos, budynek, czasownik, rzeczownik, trójkąt, wielokąt, wielościan. Czworokąt **F.** (= kwadrat). Figura **foremna**. Bryła **foremna**. (...)
- **Fru!** a. **Fru! fru!** w. (...) 4. †**F.** na **F.**, bru na bru. Prz. = *opór na opór, trafita kosa na kamień, klin klinem*. (...)
- **Forma** (...) 12. gram. postać wyrazu: **F.** wyrazu pełna, skrócona. **F.** gramatyczna. **F.** pierwotna, pochodna. **F.** osobowa, rzeczowa. **F.** prosta, złożona, opisowa. **F.** analogiczna. **F.** pewnego czasu a. konjugacyjna. **F.** deklinacyjna a. **F.** pewnego przypadku. **F.** zaimkowa przymiotnika. **F.** przymiotnika rzeczownikowa. (...)
- **Formowany** (...) 2. † p. **Formiany**: Ciasta **formowane**. Masło **formowane**.

W przypadku przykładów spreparowanych spotykane są zarówno formy słownikowe, jak i odmienione:

- **Fuzyjka**, i, lm. i 1. p. **Fuzja**: **Fuzyjką** ś. bawić = *być amatorem polowania*. (...)
- **Forsa** (...) 2. [**F.**] = a) *pieniądze, gotówka, fundusz, nakład*: Kto nie ma **forsy**, nie ma przyjaciół. Na to ś. rzucić, to nie żart: trzeba **forsy**. Jun. U nas we żniwa to zawdy je **F.** Wyprztykałem ś. z **forsy** = *wydałem wszystkie pieniądze*. (...)

Niektóre przykłady użycia bywają także poprzedzone własnymi kwalifikatorami, z których pierwszy zapisywany jest przeważnie wielką literą, a po których występuje dwukropek, np.:

- **Fontanna**, y, lm. y p. **Fontana**. **Przen.:** Krew tryska fontanny. (...)
- **Filozof**, a, lm. **owie** (...) 6. *człowiek poprzestający na małym, skromnych wymagań, przekładający niepodległość ducha nad dobra doczesne*: (...) **Ir.:** **F.**, bo mu palce z butów wyłażą. Prz. (...)
- **Fabryczny** przym. od **Fabryka**: Robotnik **F.** Robota fabryczna. Praw.: Prawa fabryczne = prawa, regulujące stosunki pomiędzy pracodawcami i robotnikami. **Kość.:** Majątek **F.** p. **Fabryka**.
- **Femiczny**: (...) **Żart.:** W usposobieniu umysłu femicznym, kiedy głos przeczuć bardzo głośno gada trzykrotnym głosem: Biada! biada! biada! Słow. (= podobnym do tego, jakie mieli skazani na śmierć przez sąd fetniczny). (...)
- **Fundamentalny** 1. a. **×Fundamentny** *zbudowany na mocnym fundamencie a. za fundament służący, gruntowny, trwały, silny, mocny, dobrze zbudowany*: Gmach **F.** Podstawa fundamentalna. **Przen. żart.:** Wolałbym zjeść coś fundamentalniejszego (= posilniejszego). (...)

Można jednak napotkać konteksty, kiedy kwalifikator taki zapisywany jest małą literą, zgodnie z interpunkcją:

- **Fałdowy** przym. od **Fałda**; **gieol.:** Góry, doliny fałdowe = *powstałe wskutek kurczenia ś. skorupy ziemskiej*.

Zilustrowanie użycia lemmy może być również poprzedzone kwalifikatorem symbolicznym:

- **Fiukać, a, aż** p. **Fiuknąć**: F. na konie. Ptaszyny fiukały. Sienk. [Niech se bez parę fiuka, niżli Franę dostanie].
- **Fałszowy** przym. od **Fałsz**: Sprawa fałszowa (= o sfalszowanie). †Głębokie jezioro w ryby bogate, a to fałszową miłością nazwane. Bron. (= fatszowy).

W ramach przykładów pojawiają się także zagnieżdżone terminy jedno- lub wielowyrzowe, które są opisane w **podrozdziale 4.3.4.10**. W niektórych przypadkach terminy te zastępują objaśnienie znaczenia, zwłaszcza gdy występowanie lemmy jest ograniczone do danego połączenia wyrazowego – por.:

- **Frażetowski**: **Wyroby frażetowskie** p. **Frażety**. (...)
- **Faluński**: **Brylanty faluńskie** = klejnoty szklane, wyrabiane w Szwecji. (...)

Warto zwrócić uwagę na fakt, że przykład użycia może niekiedy obejmować fragmenty dłuższe niż jedno zdanie czy równoważnik zdania – przypadki takie zostały podkreślone:

- **Fe! 1.** (...) **Fe! a to kto słyssał tak robić?** A fe do kata! Boh. (...)
- **Fiskał, a, lm.** (...) **3. † chciwiec, zdierca, wyzyskiwacz, drapieżnik: Go tych fiskałów u nas! co szlachcic, co dzierżawca, co arendarz, wszystek ich zamysł, aby worki natkali.** Birk. (...)
- **†Fortrop, a, lm. owie żołnierz nieznanego bliżej rodzaju broni: Z berdyszami od fortropów na janczarów wpadali. W tym czasie też fortropy z regimentów i kozacy z gór wystrzelali.** Djarjusz prawdziwej relacji z 1683 r.
- **Fuknąć, nie, nął, nied. Fukać 1.** (.....): **Książę fuknął ś.: A co mi u licha!** Pol. (...)

Niektóre spreparowane przez autorów przykłady zdają się być ponownie wykorzystane z niewielką modyfikacją, a przynajmniej nimi motywowane, np.:

- [Fiderle, i, blp.] p. **Kiderle: Panny fiderlami obwieszane.** (...)
- **Firleje** (...) **3. [F.] świecidetka, błyskotki, zabawki: Panny firlejami wystrojone.** (...)

Podobnie bywa w przypadku par niemal identycznych cytatów, co wyraźnie wskazuje na ingerencję autorów w co najmniej jeden z nich, np.:

- **Fechtować się** (...): **Pan Ranicki fechtował ś. gołą ręką z niewidzialnym przeciwnikiem.** Sienk. (...)
- **Fechmistrz** (...): **Pan Ranicki, wielki F., fechtował ś. gołą ręką z niewidzialnym przeciwnikiem.** Sienk. (...)
- **†Fremarczyć się** (...): **Łukasz fremarczył ś. z Taborem na sośnie.** Gac.
- **Frymarczyć** (...): **Łukasz frymarczył ś. na pszczoły ze Stanisławem.** Gac. **Andrzej frymarczył ś. z Janem barcią.** Gac.
- **xForman** (...): **W furmankach czterokonnych tylko jeden (koń) pod furmanem wierzchowy.** Kluk. (...)

- **xFórmanka** (...): **W fórmankach czterokonnych tylko jeden pod formanem wierzchowy. Kluk.**

W ostatniej parze pierwszy z przykładów zawiera także komentarz uzupełniający kontekst – komentarze wewnątrz przykładów mogą dotyczyć zarówno semantyki, jak i fleksji, np.:

- **1. Fala, i, lm. e** (...) **2. przen. nurt, woda:** **Ona śpi, falmi (lep. falami)** pokryta chłodnemi. Mick. (...)
- **Foremnie 1. przys. od Foremny:** **Uczony w piśmie kleryk F. (kartę) napisze. Sienk. (= pięknie).** (...)

Ze względu na brak strukturalnego ani typograficznego wyróżnienia tego rodzaju komentarzy, traktowane są one przez parser jako integralna część treści przykładów.

Może także wystąpić niespójność innego rodzaju – między aspektem lemmy a jej ilustracją w przykładzie – co jest o tyle zastanawiające, że aspekt czasownika jest wyraźnie różniony w siatce pojęciowej SW:

- **Folgować** (...) **†Rozbił Pan, ani ufolgował** ręki, wszystkie Joba ozdoby i sprzęty. Chr.
- **Filować** (...) **4. przen. oszukiwać, szachrować, kręcić:** **Zawsze mi koło Wielkanocy sfiluje** i chyba w resursie ś. spowiada. Wielis.

Podobny dysonans potrafi zaistnieć między zapisem lemmy w nagłówku hasła a zapisem wyrazu w przykładzie, np.:

- **Falena, y, lm. y** **motyl nocny:** **Anieli zdało ś., że ją widzi każda róża w ogniu i phalena.** Słow. (...)

Niekiedy przykład użycia stanowić może definicja, do której również podany jest przykład, tak jak jest to widoczne w cytowanych artykułach hasłowych:

- **1. Fala, i, lm. e 1. batwan, wał wody:** **F. morska. Spiętrzone fale. Przen.: Fale życia politycznego w Niemczech są silnie wzburzone. Fale przeciwnej fortuny. Pot. Przen.: Z falą płynąć (= postępować zgodnie z ogółem). Przen.: Etaláže sklepowe wypływają na trotuar falą porcelany. Zap. Przen.: tłum, chmara, masa posuwająca ś., kotyszca ś.: F. ludu. Krasień.** (...)
- **Furknąć** (...) **2. a. nied. myśl. Furnąć odlecieć raptem, frunąć z szelestem, zerwać ś. do lotu, pofrunąć:** **Zmieszane ptastwo tłumnie furknęło do góry. Mick. Strzepnąć skrzydłami i F. w świat szeroki. Krasz. Czapki nagle w powietrze furknęły. Przen.: pobiec, uciec: „Niemądryś waćpan!“ – zawołała Anusia i w tejże chwili furknęła, jak ptak spłoszony. Sienk.** (...)

Można też znaleźć pojedyncze przykłady, których forma wskazuje na to, że są one zaczerpnięte z literatury lub stanowią formę zagadki czy porzekadła ludowego, jednak nie jest to w ich przypadku wskazane, co z kolei sugeruje twórczość własną autorów SW:

- **[Fafecie, i, blp.] suknie, szmaty:** **Zdyjmij, miła, twe fafeci, a weź miotłę, zamieć śmieci.**

Korpus wszystkich przykładów użycia zaczerpniętych z materiału badawczego znajduje się w **Aneksie D**.

4.3.4.4. Źródło

Źródła, jako elementy artykułu hasłowego, są tutaj tożsame z oznaczeniami opisanymi w **podrozdziale 5.2.3.9**, gdzie też należy szukać ich dokładniejszego opisu. Obejmują one wszystkie informacje o tym, skąd dokonano ekscerpcji elementu, po którym są umieszczone, i zapisywane są antykwą.

Najczęściej źródłem opatrzony jest przykład użycia:

- **Feruła**, y, lm. y różga; *przen. dyscyplina, regulamin, władza, powaga*: **Pójdziemy na starość do szkół pod ferułę. Krasz. Wyemancypowany z pod feruły szkolnej. Orzesz.**
- **Fladrowaty** pełen fladrów, stojowaty: **Marmur F. Jawor, drzewo farby białej i fladrowatej. Haur.**
- **Foksalowy** przym. od **Foksał**: **Foksalowe zabawy. Niem.**

Niekiedy jest to sama lemma wraz z definicją:

- **Femińściarz**, a, lm. e żart. **zwolennik feminizmu. Prus.**
- **Folwarcznik**, a, lm. cy **właściciel folwarku. Orzesz.**
- †**Forer**, a, lm. owie **stajenny. H. Jana. (...)**

Lub lemma z odsyłaczem:

- †**Falbierz**, a, lm. e i [**Falbierz**] p. **Farbiarz: Glos. (...)**
- †**Fertrynkielt**, u, lm. y rzem. p. **Fertryngielt: Ust. cechu miechowników lubel. 1597.**

Typografia w przypadku przyłączania odsyłaczy nie jest jednolita – oprócz dwukropka do oddzielenia źródła stosuje się także zamiennie kropkę:

- †**Forlun**, u, lm. y p. **Furlon. Rej.**
- **Formularnie przys. od Formularny. Lel.**

Do danego elementu może być podanych więcej źródeł, jednak nie poświadczono ich nigdy więcej niż dwa naraz – por.:

- **Fersenkung**, u, lm. i teatr. p. **Ferzegung. Dmusz., Żółkow.**
- **Figownik (...)** 2. † lm. cy **ogrodnik, hodujący figi. Kn., Bobr.**
- ×**Fryjerstwo**, a, blm., †**Fryj**, †**Frej** **życie rozwiązte, rozpusta, nierząd; swawola: Dla fryjerstwa niechajby każdy swą miał żonę. Glicz., Glos.**

4.3.4.5. Kwalifikator wyrazowy

Użycie kwalifikatorów wyrazowych w segmencie centralnym jest zbliżone do tego, jak są stosowane w segmencie inicjalnym, z tą różnicą, że dotyczy wyłącznie konkretnych znaczeń, np.:

- **Fabrycznie (...)** 2. **przen. niedbale, licho, źle, Ladajako.**
- **Fach**, u, lm. (...) 4. **druc. kąt, tworzący ś. przez rozchyłanie ś. płaszczyzn nitek metalowych na maszynie tkackiej. (...)**
- **Fałsz (...)** 4. **kraw. część sukni z tańszego matarjału, nadstawiona w miejscu niewidocznym: (...)** 5. **muz. dźwięk mylny, pomyłkowy a. nieczysty, za wysokie a. za niskie brzmienie tonu. (...)**

- **1. Faza, y, lm. y 1. astr.** odmiana księżyca, Merkurego i Wenery.
- **2. elektrot.** różnica między momentami najwyższego napięcia i siły prądu, wyrażona w jednostkach obwodu koła.
- **3. fiz.** różny stopień rozwoju, różny stan zjawisk perjodycznie wracających: (...)
- **4. przen.** wejście w odmienne położenie, odmiana, zmiana, zwrot, kolej, stadjum, stan: (...)

W odróżnieniu od segmentu inicjalnego, opcjonalność kwalifikatorów jest tutaj zaznaczana nie tylko przy pomocy nawiasu, ale i spójnika „i”:

- **Fikcja, i, lm. e 1. fil.** = a) **(i posp.)** wymysł, urojenie, zmyślenie, pojęcie urojone, wytwór fantazji a. wyobraźni, któremu nic nie odpowiada w rzeczywistości: (...)
- **Fenol, u, lm. e chem. 1. a.** **Oksybenzol, Karbol, Alkohol fenilowy, xKwas fenilowy, xKwas karbolowy (i apt.)** produkt zastąpienia jednego atomu wodoru w benzolu przez hydroksyl, o silnym zapachu i palącym i ostrym smaku, używany jako środek dezynfekcyjny. (...)
- **Forma, y, lm. y 1. (i fil.)** ogół stosunków, w jakich coś istnieje a. coś poznajemy, sposób ułożenia szczegółów w jakiejś całości, kształt, postać, figura, wygląd, zarys, sposób: (...)

W przypadku szeregów kwalifikatorów zachowana jest ta sama metodologia, co w segmencie inicjalnym, tj. że kwalifikatory tego samego rodzaju są oddzielone spójnikiem „i”, np.:

- **Familija (...)** 5. x **bot. i zool.** rodzina, grupa. (...)
- **Faktor (...)** 4. ! lm. y **fil. i mat.** p. **Czynnik.** (...)
- **Farba (...)** 2. **(chem. i posp.)** materiał, nadający barwę innym ciałom, barwnik: (...)

Podczas gdy kwalifikatory różnych rodzajów oddzielone są spacją:

- **Fabrykant, a, lm. ci (...)** 2. **cukier. żart.:** F. lodowaty = ten, co przyrządza lody w pracowni cukierniczej. (...)

Występują jednak odstępstwa od tej reguły, jak w przypadku hasła **Faja**, w którym to dwa kwalifikatory rejestrowe oddzielone są wyłącznie spacją:

- **Faja (...)** 2. **rub. żart.** człowiek głupi, nieokrzesany; niedotęga, safanduta, fuszer, ciura, niezdara, cymbał, fujara: (...)

Kwalifikator nie musi poprzedzać definicji – może po nim wystąpić odsyłacz, np.:

- **Fłaszka, i, lm. i (...)** 2. **przen.** trunek, pijaństwo: Karty, F. gubią młodzież. Kras. Kto pisze fraszki, trzeba inu flaszki. Prz. 3. **bot. p. Flaszowiec.** (...)

Bezpośrednio po kwalifikatorze może także wystąpić przykład użycia – wówczas oddziela je dwukropek, np.:

- **Fałszywy (...)** 9. **muz.:** a) Fałszywe kwinty p. **Kwinta.** b) Fałszywy stosunek a. fałszywa relacja p. **Stosunek.** 10. **anat.:** xŻebra fałszywe p. **Żebro.** Nozdrza fałszywe (saceus v. tuba nasalis). 11. **bil.:** Sztoś F. p. **Fałszerz.** 12. **lek.:** Wody fałszywe p. **Woda.** †Bóle fałszywe p. **Ból.** Fałszywa chęć do jedzenia, in. przegłód, †przemór (bulimus) = **żartoczność chorobowa.** F. apetyt. (...)

4.3.4.6. Kwalifikator symboliczny

Kwalifikatory symboliczne, obok opisywania poszczególnych jednostek leksykalnych, mogą także zostać zastosowane do opisanego poszczególnych znaczeń lub użyć wyrazu – wówczas nie przylegają do następnego elementu, ale otoczone są spacjami, np.:

- **Fantom**, u, lm. y 1. × *mara, przywidzenie, widmo, upiór.* (...)
- **Figownik**, a (...) 2. † lm. cy *ogrodnik, hodujący figi.* Kn., Bobr.
- **Fraktura**, y, lm. y 1. ! *złamanie.* (...)

W przypadku kwalifikatora [] w nawiasie umieszczana jest pierwsza litera lemmy wraz z kropką, np.:

- **Folwark** (...) 4. [F.] *dom dla czeladzi.* (...)

Opcjonalność kwalifikatorów symbolicznych wyrażana jest spójnikiem „i” – por.:

- [Fałszywy] 1. i † p. **Fałszywy**: F. świadek. Leop. Bóg pomści niecnoty i fałszywej zdrady. Kochan. [Fałszywe wszystko na świecie]. (...)
- †**Francowaty** lek. 1. i [F.] *chory na przymiot.* (...)

Kiedy do danego znaczenia opatrzonego kwalifikatorem [] podane są synonimy oznaczone w ten sposób, zostają one zgrupowane ze spójnikiem „a[lbo]” lub bez niego:

- †**Fararz**, a, lm. e 1. i [F., **Falarz**] *proboszcz, pleban:* (...)
- ×**Falować**, uje, ował 1. i [F. a. **Felować**] *szwankować, niedomagać, mieć feler, podupadać, utykać, chwiać ś., staniać ś.:* (...)

Kolejne kwalifikatory symboliczne również przyłączone są spójnikiem „i”, a ich kolejność zdaje się być dowolna – por.:

- **Fasować**, uje, ował (...) 5. † i [F.] *bić, ćwiczyć, prac:* (...)
- **Fałeczny** 1. [F.] i † p. **Fałszywy**: Mamy światłość prawdziwą przeciwko fałecznej światłości mędrców tego świata. Białob. (...)

W przypadku obecności kwalifikatora wyrazowego, kwalifikator symboliczny umieszczany jest przed, a czasem po nim – por.:

- **Facjata** (...) 7. **żart.** × *tył, tyłek.* (...)
- **Fistuła**, y, lm. y (...) 5. ! **lek.** p. **Przetoka.** (...)

4.3.4.7. Ciąg synonimiczny

Ciągi synonimiczne są tu skonstruowane analogicznie do tego, jak wygląda to w segmencie inicjalnym – pierwszą różnicą w tak tym segmencie jest jednak fakt, iż zawsze przyłączane są przez skrócony spójnik „a.”:

- **Fantazja**, i, lm. e 1. blm. fil. = a) **a. Imaginacja, Wyobraźnia,** ×**Um** *władza duszy, tworząca i odtwarzająca wyobrażenia; zdolność kombinowania nowych wyobrażeń:* (...)
- **Frazes**, u, lm. y 1. **a. !Fraza,** ×**Frazesa** *zdanie, okres, wyrażenie, zwrot, ustęp, miejsce w mowie a. piśmie:* Nie pamiętam już ani frazesu. (...)

Powtórzenie lemmy w formie skróconej w celu przypisania kwalifikatora symbolicznego [] do danego znaczenia również kwalifikowane jest tu jako część ciągu synonimicznego:

- **×Falować, uje, ował 1. i [F. a. Felować]** szwankować, niedomagać, mieć feler, podupadać, utykać, chwiać ś., staniać ś.: (...)

Jeśli synonimy poprzedzają odsyłacz, to oddzielony jest on od nich średnikiem:

- **Filigranowy przym. od Filigran; ×Filigranowy, !Filigranowy:** Robota filigranowa. Przen. = a) *delikatny, kruchy, łatwy do zniszczenia*: F. most, budynek. Filigranowe zdrowie. (...)
- **†Faceletek, tka, lm. tki p. Facelet; †Faceletek.** Mur.
- **Faworyta, y, lm. y forma ż. od Faworyt; ×Faworka:** F. matki, królowej. Margrabina de Pompadour była faworytą Ludwika XV (= *kochanką*). (...)

Najczęściej synonimy są pogrupowane pod względem kwalifikatora symbolicznego, którym dane synonimy są opatrzone; dodatkowo w przypadku kwalifikatora [] wszystkie wyrazy umieszczone są w pojedynczym nawiasie:

- **×Fryjer, a, lm. y 1. a. ×Fryjerz, †Fryjarz, †Fryj, †Fryjerysta, [Frair, Frajerz, Frejer, Frejerz]** zalotnik, kochanek, gach. (...)
- **Fe! 1. a. Pfe!, Fi!, Fuj!, Pfuj!, ×Fa!, †Pfa!, †Pfy!, †Pi!, †P'!** w. obrzydzenia, wstrętu, nagany = *tju! be! bwe! bla!*: (...)
- **[Fałaszny] (...) 2. a. [Fałeczny, Fałesny, Feleśny]** chory, nadpsuty, chorowity: Ma fałeszna nogę (= *chory na nogę*).
- **Fernambuk (...) 2. blm. a. [Farnabuk, Faredborek, Drzewo fernambuckie]** wióry z tego drzewa do farbowania używane, in. brezylija, [drzewo czerwone]. (...)

W przypadku kwalifikatorów innych niż [] zasada ta nie zawsze jest jednak zachowana:

- **Fajerka, i, lm. i 1. a. ×Fejerka, †Fairka, ×Żarzelnia** piecyk przenośny, naczynie do rozpalania ognia: (...)
- **Faramuszka, i, lm. i 1. a. ×Farmuszka, ×Farmużka, ×Framużka, †Farmuska, [Framuszka, Warmuszka, Warmóz, Warmuz]** = a) zupa z chleba, rozgotowanego w wodzie a. w piwie, gramatka. (...)
- **Flegmatyk, a, lm. cy 1. † a. †Charchała, ×Charkała, ×Chrachała, †Chrapek, [Charkacz]** człowiek kaszlący i spluwający wiele. (...)

Co więcej, nie zawsze zachowana jest też kolejność alfabetyczna:

- **Figlarz (...) 4. a. Miliślepek, Figlarek** bot. (miraulus) roś. z rodziny trędownikowatych. (...)
- **Fujara, y, lm. y 1. a. [Fulara, Fujera, Fuira]** duża fujarka.

Kolejną różnicą jest fakt, że znacznie częściej niż w przypadku segmentu inicjalnego pojawiają się w ciągach synonimicznych terminy wielowyrazowe, np.:

- **Fenol, u, lm. e chem. 1. a. Oksybenzol, Karbol, Alkohol fenilowy, ×Kwas fenilowy, ×Kwas karbolowy** (i apt.) produkt zastąpienia jednego atomu wodoru w benzolu przez hydroksyl, o silnym zapachu i palącym i ostrym smaku, używany jako środek dezynfekcyjny. (...)
- **Fabryka (...) 8. a. Majątek fabryczny** kość. majątek ściśle kościelny w odróżnieniu od beneficjum. (...)

- **Funt**, a, lm. y (...) 5. a. **F. sterling, F. szterling** moneta angielska = przeszło 40 złp. (...)

W przypadku, w którym w skład synonimu wielowyrazowego wchodzi lemma w niezmienniczej formie, jest ona skraca do pierwszej litery:

- **Faktura**, y, lm. y 1. (...) b) a. **Faktur-księga a. lep. Księga fakturowa** księga kupiecka ze spisem należności za towary.
- **Foljo** nieod. 1. a. in. **F.** = format książki arkuszowy: (...)
- **Franca**, y, lm. e lek. 1. gm. i [**F.**] a. †**Pani F.**, †**F. zupełna**, †**F. ogólna** p. **Przymiot:** Niech cię **F.** stoczy! (...)
- **Fuknąć, nie, nął**, nied. **Fukać** 1. a. **F. ś.**, nied. **Fukać ś.** na kogo a. *kogo = *krzyknąć, ofuknąć ś., obwisnąć ś., obruszyć ś., wybuchnąć gniewem, oburzyć ś., rzucić ś. z gniewu, zgromić, złajać:* (...)

Jeśli do danego znaczenia podany jest kwalifikator symboliczny, to ciąg synonimiczny umieszczany jest po nim:

- [**Famuła**, y, lm. y] 1. i † a. †**Pamuła**, [**Pamuła**], ***Garnizorówka** potrawa z owoców rozgotowanych i chleba, gramatka (...)
- **Flamandzki** (...) 3. × a. ***Flamski** (o drzwiach) z krat złożony. (...)
- **Forsa**, y, lm. y 1. × a. ***Forca**, [**Forsz**] gwalt, przemoc, moc, siła, usiłowanie, wysiłek, natężenie sił, usilne popieranie, zabiegi: (...)

Jednocześnie ciągi synonimiczne są bywają umieszczane przed kwalifikatorami wyrazowymi:

- **Frysz** (...) 5. **Frysz, gór.** = a) miejsce wyznaczone na szyb. (...)
- 1. **Futrówka** (...) 5. a. **Futrówa ucz.** karteczka do odczytywania lekcji, czelbryk, in. ściągaczka.

Dwie powyższe reguły oznaczają, że synonimy zlokalizowane są między kwalifikatorem symbolicznym a kwalifikatorem wyrazowym, co w materiale badawczym i referencyjnym zostało poświadczony tylko raz:

- **Fracja** (...) 4. † a. †**Frakta** mat. *ułamek.* (...)

Niekiedy jednak kwalifikator znajduje się także przed ciągiem synonimów:

- **Falisty** (...) 2. **przen. a. Falowany, Falowaty** mający nierówności kształtu *fal:* (...)
- **Frajer**, a, lm. y 1. **rub. a. *Frejer, *Fryjer, *Fryjerz** człowiek nieznający gry w karty; (...)

W materiale badawczym nie poświadczono sytuacji, w której synonimy są poprzedzone kwalifikatorem symbolicznym i wyrazowym jednocześnie.

Jeśli zamiast ciągu synonimicznego powtórzona jest lemma wraz kwalifikatorem symbolicznym lub bez niego, i która nie jest poprzedzona spójnikiem „a”, to oznacza to, że synonimy globalne nie stosują się do danego znaczenia lemmy, np.:

- **Frysz**, u, lm. y, †**Fryst** (...) 5. **Frysz**, gór. = a) miejsce wyznaczone na szyb. (...)

- **Furta**, y, lm. y, **Fórta**, †**Forta** (...) 5. [**Furta**] *szcherba, szpara*: F. w zębach (= *szpara między rzadkimi zębami*). Krowa z furką. 6. **Furta** = a) hut. *drzwiczki do nabijania pieca płomiennego*. b) zdń. *wejście do pieca do wypalania kaflí*. (...)

Powyższe dwa hasła to jedyne poświadczenia takiego zabiegu w materiale badawczym; dzięki niemu nie jest konieczne podawanie tych wyrazów, których można użyć zamiennie we wszystkich pozostałych znaczeniach.

4.3.4.8. Informacja gramatyczna

Informacja gramatyczna segmencie centralnym jest kodowana w sposób zbliżony do tego, jak wygląda w segmencie inicjalnym:

- **Formalność**, i 1. **blm.** rz. od **Formalny**. 2. **lm.** i *sposób postępowania urzędownie przepisany, manipulacja*: Formalności kontraktowe, testamentowe, prawne. Formalności przepisane do badania świadków. (...)
- **Frej** 1. × **nieod. przym.** = a) p. **Fraj**. b) [**F.**] *wolny, swobodny*. 2. † **u**, **lm.** **e** rz. = a) p. **Fryjerstwo**. b) *kara pieniężna, grzywny*: Dawne było przysłowie u sądowych urzędników w Litwie: (...)

Pojawia się tutaj dodatkowo informacja różnicująca znaczenie lemmy w liczbie mnogiej, poświadczona w materiale badawczym 78 razy, np.:

- †**Figlik** (...) 2. **w lm.** *zbiór wierszy anegdotycznych, fraszki*.
- **Fanfara** (...) 3. × **w lm.** *odgłos trąb*: (...)

Także tutaj mogą wystąpić pełne formy fleksyjne, a nie tylko końcówki – w materiale badawczym został poświadczony tylko jeden taki przypadek – znajduje się on w segmencie centralnym:

- **Fum** (...) 4. **i w lm.** **Fumy, ów**, a. **Fuma** *pretensjonalne uroszczenie, próżność, zarozumiałość, nadętość, grymasy, duma, pycha*: (...)

Informacja gramatyczna zagnieżdżona wewnątrz definicji nie jest wyodrębniana:

- **Fakt** (...) 2. **dowód**, **w lm.** **dane**: *Przekonam cię faktami a. na faktach*. (...)

Wyjątkiem jest informacja o części mowy umieszczana na początku definicji – choć składniowo jest związana z resztą definicji, to jej wyodrębnienie z reguły nie stanowi problemu:

- **Fe!** 1. a. **Pfe!**, **Fi!**, **Fuj!**, **Pfuj!**, **×Fa!**, †**Pfa!**, †**Pfy!**, †**Pi!**, †**P'**! **w.** *obrzydzenia, wstrętu, nagany = tju! be! bwe! bla!*: **Fe!** *jaka to brzydota*. (...) 2. **przysłówkowo** = *wstyd!*, *hańba!*, *nietadnie!*, *brzydko!*, *do obrzydliwości!*, *aż strach!*: (...)

4.3.4.9. Walencja

Walencja, choć obecna także w segmencie inicjalnym, jest kilkakrotnie częstsza w segmencie centralnym – wynika to z charakterystyki polskiej syntagmatyki, ale jest zgodne ze współczesnymi teoriami leksykograficznymi, według których walencja stanowi część znaczenia.

Całość informacji o walencji zapisywana jest antykwą, a od definicji oddziela ją znak równości „=”, np.:

- **xFalować (...)** 2. **co** = *nabawiać szwanku, nadwerężyć, mieszać, mącić*: Pijaństwo zwykło ludzki rozum F. Haur. (...)
- **Fundować, uje, ował, [Fondować] (...)** 3. **x co na czym** = *opierać, zasadzać, gruntować, zakładać, pokładać*: Naukę fundował na mniemanych cudach. Troc.
- **Frymarczyć, y, ył 1. † a. †Frymarczać, †Farmarczyć co za co a. czym na co** = *prowadzić handel zamienny, mieniać, wymieniać, zamieniać, oddawać co za co*: (...)
- **Fasować, uje, ował 1. co w co a. co czym** = *ładować, napętniać, zapychać, pakować*: (...)

Niektóre formy mogą być także opatrzone kwalifikatorem symbolicznym – por.:

- **Fuknąć, nie, nął, nied. Fukać 1. a. F. ś., nied. Fukać ś. na kogo a. xkogo** = *krzyknąć, ofuknąć ś., obwisnąć ś., obruszyć ś., wybuchnąć gniewem, oburzyć ś., rzucić ś. z gniewu, zgromić, złażać*: (...)

Informacja o walencji zagnieżdżona w dalszej części hasła stanowi część definicji, a nie osobny element, na co wskazuje nie tylko nierozłączna strukturalna z sąsiadującym tekstem, ale także zapis kursywą:

- **xForsztelować (...)** 2. *awansować kogo, podnosić do wyższej rangi, posuwać na wyższy urząd*: (...)
- **Ignorować, uje, ował 1. nie zważać na kogo, za nic mieć, nie chcieć znać, lekceważyć. 2. przez szpary patrzeć na kogo a. na co.** (...)
- **Imputować, uje, ował zaliczać na rachunek; przypisywać komu co, posądzać go o co, wmawiać w niego co**: (...)

Zróznicowanie walencyjne pojedynczej lemmy jest bardzo dobrze zilustrowane w hasle **†Fołdrować**, w którym przy znaczeniach **1.** i **4.** podana jest prosta walencja, przy znaczeniu **2.** walencja opcjonalna, a przy znaczeniu **3.** walencja wyrażona alternatywą dwóch elementów:

- **†Fołdrować, uje, ował, †Foldrować, †Fordrować 1. na kogo** = *oskarżać kogo, instygować, donosić, oczerniać, nastawać, powstawać*: (...)
- 2. **(kogo)** *wszczynać proces, pozywać, dochodzić, prawować ś.*: (...)
- 3. **kogo a. co** = *stawać w interesie czego, bronić, popierać, gardłować, adwokatować*: (...)
- 4. **kogo** = *pomagać, wspierać, wspomagać, podtrzymywać, forytować*: (...)

4.3.4.10. Pojęcia zagnieżdżone

Na terminologiczne jednostki zagnieżdżone (które stanowią zarówno pojedyncze leksemy jak i całe związki wyrazowe) składają się dwa rodzaje elementów: wyróżnione terminy stanowiące część definicji oraz specjalny przypadek przykładów użycia. „Termin”, „pojęcie” oraz „wyrażenie” stosowane będą tutaj zamiennie.

Terminy umieszczone wewnątrz definicji, choć stanowią jej integralną część, można wyodrębnić na podstawie formatowania – w tekście wyróżnione są antykwą i rozstrzeleniem; dodatkowo mogą się także pojawić wyrazy wprowadzające.

Najczęściej wyrazem tym jest „inaczej” w formie skróconej, czyli in.:

- **Falset, u**, lm. **y**, **×Falcet** *głos męski, zakrawający na żeński, rejestr wyższy głosu śpiewaka, tak zwany „z głowy”, in. **fistuła**: Śpiewać, mówić falsetem. (...)*
- **Fizjologja, i**, blm. **1. a.** **×Rodnictwo, ×Rodotnictwo, ×Żywotnictwo, ×Żywotoznawstwo, ×Przyrodnictwo** *biol. nauka o czynnościach i objawach życiowych ustroju prawidłowego. Gram.: F. mowy ludzkiej = nauka o czysto zewnętrznej stronie wymawiania, o czynnościach fizjologicznych, których wynikiem są brzmienia czyli dźwięki języka ludzkiego, in. **antropofonika**. (...)*
- **1. Folać, uje, ował 1. a. Folać** = a) *wałkować, zbijać sukno, aby grubsze było, in. **spilśniać**.*

Innym wyrazem wprowadzającym jest **zwany / zwana / zwane**, który umieszczany jest przed szeregiem terminów lub po nim; on może występować także w połączeniu z kwalifikatorem modalnym **niek.**, czy nawet opisem zakresu:

- **Fankiel, kla a. klu**, lm. **kle, Fenkiel, Fankul** *bot. (foeniculum) roś. z rodziny baldaszkowych. Gatunek: F. pospolity, Koprem włoskim, Feniklem **zwany** (f. vulgare). (...)*
- **Fijołek (...)** **1. (...)** *bot. (viola) roś. z rodziny fijołkowatych, (i jej kwiat). Gatunki: F. dziki (v. canina); F. pachnący a. wonny, **zwany** Fijołkami marcowymi a. brunatnymi (Viola odorata); F. trójkolorowy (v. tricolor), bardzo zmienny: α) odmiana ogrodowa, zwana Bratki, Braciszki, Brat z siostrą a. Brat i siostra; β) odmiana polna (arvensis), **zwana** u ludu ~Maślukami~. (...)*
- **Firletka (...)** **d) F. poszarpana** (l. flos cuculi), **niek.** *Smolanką a. Kukułką, Kukułeczką **zwana**; (...)*

Pojęcie zagnieżdżone może też występować bez żadnego elementu wprowadzającego na początku definicji, np.:

- **Fasola (...)** **c) ogr. **F. szparagowa**** *(z miękkim strąkiem, bez pergaminu). (...)*
- **Fluoryt (...)** **a) **Chlorofan**, fosforyzujący przez ogrzewanie; b) **Flusztajn**, zbity, używany jako topnik w robotach metalurgicznych, do otrzymywania fluorowodoru i rytowania na szkło. (...)**

Zdarza się też, że element, który kwalifikuje się jako hasło zagnieżdżone, nie był wyróżnione – również nie były zmieniane, np. jednostka **ftaleina rezorcynowa** nie została wyróżniona w hasło **Fluoresceina**:

- **Fluoresceina, y**, blm., **Rezorcyftaleina** *chem. **ftaleina rezorcynowa**, ciało głośnie przez wspaniałą fluorescencję słabych jego roztworów w rozcieńczonym amonjaku, substancja krystaliczna barwy ceglastej, C₆H₄ (CO) (C(C₆H₃OH)₂O)O. <Z Fluorescencja>*

Termin zagnieżdżony może występować wraz z kwalifikatorem symbolicznym, np.:

- [Fąfrać, a, aź] *nadymacć ś., grymasić, in. [fonfry robić a. stroić].*
- Fernambuk (...) 2. blm. a. [Farnabuk, Faredborek, Drzewo fernambuckie] *wióry z tego drzewa do farbowania używane, in. brezylja, [drzewo czerwone].* (...)
- 1. Filtr, u, lm. y *warstwa materji porowatej, służącej do filtrowania, in. xścacek, xcedzidło, xprzesacznik.*

Poświadczono też pojedynczy przypadek, w którym użycie danej lemmy jest tylko w ramach struktury wielowyrazowej, umieszczonej w miejsce ciągu synonimicznego:

- Fiu! (...) 4. [Hecia fiu!] *odpędzanie owiec.* (...)

Inaczej wygląda to w przypadku terminów zagnieżdżonych w przykładach. Jako że te nie są w żaden sposób wyróżnione typograficznie, a strukturalnie pozostają nie do odróżnienia od zwykłych wyrażeń przykładowych – ich jednoznaczna detekcja nie jest przy pomocy stosowanej tu metody możliwa; z tego powodu próba ich ekstrakcji jest podejmowana tylko wtedy, kiedy mają wyróżniającą je strukturalnie informację, czyli np. definicję znaczenia:

- Fabryka, i, lm. i 1. a. xWyrobnia *znaczne przedsiębiorstwo przemysłowe, postugujące ś. motorami mechanicznymi i podziałem pracy na wielką skalę: F. sukna, cygar. F. stali = stalownia. F. cukru = cukrownia. F. piwa = browar. F. papieru = papiernia.* (...)
- Flet (...) 2. org.: F. szwajcarski = *jeden z głosów organu, pochodzący z ośmiostopowej piszczatki metalowej. F. major i F. minor = jeden z głosów fletowych organu, wydawanych przez piszczatki kryte.* (...)
- Forma (...) 14. mat. = a) (...), b) F. *algiebraiczna = funkcja catkowita wymierna jednorodna jednej a. więcej zmiennych:* (...)

Lub komentarz:

- †Fałesznie, [Fałesznie] przys. od Fałeszny: [Miłujesz F.] (= *falszywie, obłudnie*).

Czy nawet oba te elementy jednocześnie:

- Fukierowy bot.: *Rośliny fukierowe (fouquiereae) plemię rodziny tamaryszkowatych.* (...)

Wyrażenia zagnieżdżone w przykładach użycia oprócz definicji mogą także swoje własne przykłady użycia, np.:

- Femiczny: *Sąd F. = średniowieczny trybunał tajemny, karzący śmiercią winowajców bogatych i możnych: Sąd tajemny nazywał ś. jeszcze trybunałem femicznym. Mick.* (...)

W samych definicjach terminów zagnieżdżonych może się także znaleźć kolejny termin zagnieżdżony:

- xFundowy: *Wieża fundowa = więzienia kryminalne, in. xfunda: Skończy ś. na wieży fundowej, z której Wołodkowicz uciecze. Chodź.* (...)

Przy okazji warto tutaj napomknąć, że terminy zagnieżdżone można także niekiedy spotkać w segmencie terminalnym, np.:

- **xFundowy**: (...) <łć. fundus = dno; więzienie **in fundo** w dawnej Polsce dosł. = na dnie (wieży)>

Zarówno definicje, jak i przykłady mogą także zawierać ciągi pojęć zapisanych w formie skróconej – zazwyczaj poprzez niepowtórzenie elementu wspólnego – które nazywane są tutaj terminami zagregowanymi, np.:

- **Fasola** (...) 1. (...) a) **F. karłowata a. piesza (nizka)**, pospolicie Piechotą zwana (ph. nanus); b) **F. pospolita a. zwyczajna (wysoka)**, pospolicie F. tyczkowa, ~Bób turecki, **Groch turecki a. włoski a. szablasty**, Szabelbon, Szablak zwana (ph. vulgaris). (...)
- **Formacja** (...) 2. (...) c) F. aluwjalna a. Aluwjum. d) F. dewońska a. Devon. e) F. dyluwjalna a. Dyluwjum. f) F. eoceńska a. Eocen. g) F. hurońska a. Huron. h) F. jurajska a. Jura. i) F. kambryjska. j) F. kredowa. k) F. laurentyjska. l) F. miocenska a. Miocen. m) F. oligocenska a. Oligocen. n) F. permska a. Perm. o) F. pliocenska a. Pliocen. p) F. przechodowa. r) F. sylurska a. Sylur. s) F. trjasowa a. Trjas. t) F. węglowa. u) **Formacje azoiczne a. pierwotne = laurentyjska, hurońska i spodnia kambryjska.** w) **Formacje paleozoiczne a. pierwszorzędowe = sylurska, wierzchnia kambryjska, dewońska, węglowa i permska.** x) **Formacje mezozoiczne a. drugorzędowe = trjasowa, jurajska i kredowa.** y) **Formacje cenozoiczne a. trzeciorzędowe = eoceńska, oligocenska, miocenska i pliocenska.** (...)
- **Fałda** (...) 1. zmarszczka w tkaninie: **W fałdy a. fałdy układać = fałdować.** Suknia rzuca a. puszcza fałdy. (...)
- **Fałdowy** przym. od **Fałda**; gieol.: **Góry, doliny fałdowe = powstałe wskutek kurczenia ś. skorupy ziemskiej.**
- **Fenilowy** przym. od **Fenil**: **Alkohol, xkwas F. p. Fenol.**

Metoda, przy pomocy której parser rozwija warianty połączeń wyrazowych zapisanych w sposób zagregowany (skrótowy), opisana jest w rozdziale **10.3.10**.

Na koniec warto zaznaczyć, że można spotkać w materiale badawczym jednostki, które można by było zakwalifikować jako wyrażenia na poziomie analizy semantycznej, ale które nie zostały przez autorów *Słownika...* w żaden sposób wyróżnione – por. wybrane przykłady:

1. **mający dobrą firmę** jako element definicji:
 - **Firmowy** przym. od **Firma**: (...) Przen. **mający dobrą firmę = ustaloną stawę, znany, głośny**: (...)
2. **źarłoczność chorobowa** jako ekwiwalent poprzednich jednostek terminologicznych przyłączony w formie definicji:

- **Fałszywy** (...) 12. lek.: Wody fałszywe p. **Woda**. †Bóle fałszywe p. Ból. Fałszywa chęć do jadła, in. przegłod, †przemór (bulimus) = **żarłoczość chorobowa**. (...)
- 3. halucynacje, przywidzenia, **widziadła** w formie zagregowanej w formie definicji:
 - **Fantazyjny** (...) 1. *fantazji dotyczący*: (...) Objawy fantazyjne = **halucynacje i przywidzenia a. widziadła** (...)
- 4. **Figura płaska**, **Figura prostokreślna**, **Figura wielokątna**, **Figura wielościenne** w formie zagregowanej jako przykład użycia:
 - **Figura** (...) 16. mat. = a) *nazwa ogólna układu punktów, linii, powierzchni i t. p.*: **F. płaska, prostokreślna, wielokątna, wielościenne**. (...)

Należy też wspomnieć, że istnieje jeden poświadczony przypadek użycia tej metody do wyróżnienia lemmy, a nie terminu zagnieżdżonego – ma to miejsce w segmencie terminalnym:

- **[Formatyka, i, lm. i] wzory pisma kaligraficznego**. <Od wyrazu **Formuj** w pierwszym wierszu daw. wzorów kaligraficznych: „Najświętsza Panno, formuj moją rękę, Bym mógł opisać Syna Twego mękę“>

4.3.4.11. Podział typologiczny

W niektórych hasłach zamiast lub obok definicji wprowadzona jest typologia, w ramach której kolejne elementy – najczęściej opatrzone rozstrzeżeniem i numeracją kolejnego poziomu – składają się na opis znaczenia. W materiale badawczym zostały poświadczone trzy rodzaje określeń sygnalizujących taką strukturę hasła, które zapisywane są antykwą:

1. **gatunek / gatunki**:
 - **Fankiel** (...) bot. (foeniculum) *roś. z rodziny baldaszkowych*. **Gatunek**: F. pospolity, Koprem włoskim, Feniklem zwany (f. vulgare). (...)
 - **Feldspat** (...) min. *nazwa ogólna niektórych minerałów, przedstawiających bezwodne krzemiany glinu i potasu a. sodu, a. krzemiany glinu i wapnia*. **Gatunki**: 1. F. alkaliczny: a) F. potażowy p. **Ortoklas** i **Mikroklin**; b) F. sodowy~ p. **Albit**. 2. F. sodowowapienny p. **Labrador**. 3. F. szklisty p. **Ortoklas**. 4. F. wapiennosodowy p. **Oligoklas** i **Andezyn**. 5. F. wapienny = **Anortyt**. (...)
2. **odmiana / odmiany**:
 - **Ftanit** (...) *gieol. skała pojedyncza, łupkowa, złożona z kwarcu zbitego z domieszką gliny, węgla i tlenu żelaza, używana na bruk i do ostrzenia narzędzi stalowych*. **Odmiana**: **Lidy**t, używany na kamienie probiercze. (...)
 - **Fluoryt** (...) min. *fluorek wapnia, tworzący kryształy równoosiowe a. agregaty gruboziarniste i pręcikowate, najrozmaiciej zabarwione, w Anglii obracane na ozdoby i sprzęty, a w starożytności prawdopodobnie na tak zwane vasa murrhina*. **Odmiany**: a) **Chlorofan**, fosforyzujący przez ogrzewanie; b) **Flusztajn**, zbity, używany jako topnik w

robotach metalurgicznych, do otrzymywania fluorowodoru i rytowania na szkłe. (...)

3. w wyrażeniach:

- [Fraj] nieod. rz. w wyrażeniach: Chodzić na F. = *chodzić w zaloty*. Dać sobie F. = *dać sobie słowo, umówić ś.* (...)

4.3.4.12. Zakres

Przy niektórych elementach hasła – w tym poszczególnych definicjach lub całym segmencie semantycznym, a także przy przykładach użycia czy całych jednostkach leksykalnych, włączając w to synonimy i terminy zagnieżdżone, czy nawet w segmencie etymologicznym – może występować informacja o zakresie. Zakres pełni funkcję zbliżoną do kwalifikatorów wyrazowych – opisuje restrykcje geograficzne, rejestrowe, dziedzinowe itd. – jednak w przeciwieństwie do nich nie jest skodyfikowany, tylko stanowi kategorię otwartą. Od strony typografii zapisywany jest antykwą i umieszczany w nawiasie, a wyrazem, który tego rodzaju informację wprowadza, jest zawsze przyimek. W przypadku lemm i definicji zawsze je poprzedza – por.:

1. Przyimek „na”:

- 2. **Folga** (...) 8. druk. (na prasie ręcznej) papier, zakrywający tło między czcionkami formy drukarskiej, gdy ono smoli. (...)

2. Przyimek „o”:

- **Fremdować, uje, ował** (o czeladniku) przyjść z wędrownki: Czelnik, który fremdował, t. j. przyszedł z wędrownki, znalazł w gospodzie przytułek. Korz. (...)
- [Forknąć, nie, nął, Furknąć, nied. Forkać, Furkać] (o zwierzętach) parsknąć, czmychnąć: (...)
- **Fulminacja, i**, lm. e 1. (o złocie, srebrze a. prochu) 1. łoskot, szczęk, brzęk; trzask. (...)

3. Przyimek „u”:

- **Filistyn, a**, lm. owie 1. × (u studentów uniwersytetu) osoba, nie należąca do uniwersytetu, szczeg. kupiec. (...)
- **Forma** (...) 11. fil. (u Arystotelesa i scholastyków) pierwiastek metafizyczny, będący przeciwstawieniem materji, idea kształtująca, nadająca materji określoną postać bytu: Dzięki formie obojętna materia staje ś. bytem. (...)

4. Przyimek „z” (w wyrażeniach przysłówkowych):

- **Familja** (...) 4. [F. a. Famelija] (z ruska) nazwisko (rodowe). (...)

5. Przyimek „w”:

- **Fluorek** (...) 2. (w Galicji) F. wodu p. Fluorowodór.
- **Fosforon, u**, lm. y, (w Galicji) Fosforyn chem. sól kwasu fosforawego. (...)

Warto tutaj wspomnieć, że zakres wprowadzany przez ten ostatni przyimek może być także obecny w części etymologicznej segmentu terminalnego (również z pominięciem nawiasu):

- **xFundowy**: (...) <łć. fundus = dno; więzienie in fundo w dawnej Polsce dosł. = na dnie (wieży)>
- **Faro** (...) <Fr. (w Belgji) faro>

W niektórych sytuacjach zakres jest umieszczony po definicji i zapisany antykwą, np.:

- **Fantazja** (...) c) *zasada procesu wszechświata (w filozofji idealistycznej)*. (...)
- **Filipowiec**, wca, lm. wcy, **Filiponi** lm. *starowierca (w kościele wschodnim)*: Sekta filipowców dzieli s. na popowców i dżakowców. (...)
- **[Flażyc, y, yż]** 1. *padać (o deszczu ze śniegiem i wiatrem)*. (...)

W przypadku opisywania terminów wielowyrzowych i przykładów użycia informacja o zakresie następuje po tychże:

- **Facjata**, y, lm. y (...) 3. żart. *twarz, oblicze, gęba, pysk, morda*: Dać komu po facjacie. Kałmucką facjata cię nazwała. Fred. A. O ty błada facjato! Krasiń. *(o księżycu)*. (...)
- **Fajkarz**, a, lm. e (...) 2. p. **Fajczarz**: Arcydzieło, któreby u fajkarzy chińskich poklask wzięło. Mick. *(o puszczeniu dymu z fajki)*.
- **Fałszywo** p. **Fałszywie**: Półmisku F. pozłacany! Krasiń. *(o księżycu)*.
- **×Felbanka**, i, lm. i p. **Falbana**: W falbankach nowych, w ponsach tak ś. gnieździesz śmiało? L. *(o myszy)*.
- **Filant**, a, blm. *bzik, zajączek*: Ma (nabył) filanta. Prz. *(o figlarzu i awanturniku pijanym)*. (...)

W innych przypadkach ten sam rodzaj informacji zapisywany jest kursywą wraz z całą definicją:

- **[Fąfola, y, lm. y]** (...) 2. *łata, szmat (u sukni podartej)*.
- **Fes** nieod. muz. *nuta f z bemolem, czyli obniżona o pół tonu, równająca ś. e (na narzędziach utemperowanych)*: Fes rzadko w utworach muzycznych ś. spotyka. (...)

Problem z ekstrakcją zakresu na tej pozycji jest o tyle trudny, że trudno odróżnić go strukturalnie od nawiasów stanowiących część definicji, np.:

- **[Fąfa, y, lm. y]** *szczonek, prztyczek (w nos)*.

W niektórych przypadkach ten sam format jest wykorzystany do wskazania źródła, ale tylko w przypadku zagadek¹² – por.:

- **[Fimera, y, lm. y]** 1. *chimera*. 2. *świnia (w zagadce)*.
- **[Fyrka, i, lm. i]** *szczur (w zagadce)*.

Informacja o zakresie traktowana jest także jako część definicji, jeśli przez swoje położenie jest z nią strukturalnie nierozłączna:

- **Fawor** (...) 2. † i [F.] a. **[Fabor, Fafor]** *wstążka (szczeg. u szyi), tasiemka (u koszuli)*: Płacić mają od wstąg, od faworów. Vol. (...)

Pozostałe rodzaje informacji wyodrębnionej przy pomocy nawiasów, a niestanowiące części definicji – jak zostało to uwzględnione w rozdziale 4.3.4.1 – opisane są w dwóch kolejnych rozdziałach.

¹² Zagadki jako kategoria źródeł opisane są w rozdziale 5.2.3.9. **Oznaczenia źródłowe**.

Istnieje jeszcze jeden potencjalny sposób wprowadzenia informacji o zakresie, poświadczony w materiale badawczym tylko w jednym haśle, który zapisywany jest antykwą, bez nawiasu, a umieszczony bezpośrednio przez definicją i oddzielony od niej znakiem „=”:

- **xFalcydja, i, blm. praw. 1. w prawie ogólnym polskim = czwarta część dochodu. 2. w prawie miejskim = rozdział proporcjonalny sum dłużnika przy zbiegu wierzycieli. (...)**

Sam znak równości występuje znacznie częściej w definicji, a bez analizy treści i formatowania rozstrzygnięcie tego, jakiej segmentacji w haśle dokonać, jest trudne – por.:

- **Fasa (...)** 3. **pewna liczba przedmiotów = sto buntów:** F. noży. (...)
- **Figuchna, y, lm. y rzecz drobna; figa = mała osóbką.**
- **Fluidskrupuł, a, lm. y waga aptekarska = 1/3 fluidrachmy. (...)**
- **[Frąkać, a, ał, Frąchać, Fiąkać] nosem = głośno pociągać nosem. (...)**

4.3.4.13. Komentarz i glosa

Po niektórych pojęciach i wyrażeniach umieszczane są komentarze lub glosy, traktowane tutaj zbiorczo ze względu na wspólną typografię. Ich funkcją jest doprecyzowanie i ujednoznacznienie opisywanej jednostki leksykalnej, czego celem jest jej identyfikacja pośród innych podobnych jednostek, a nie zrozumienie jej znaczenia. Informacja ta jest zawsze umieszczona w nawiasie okrągłym i zapisywana antykwą.

Treść komentarza może stanowić zwykłe objaśnienie w języku polskim:

- **Falisty (...)** 2. (...) Anat.: **xFalot F. (tętnicowy) (...)**
- **Fizyk (...)** 4. Ryzyk F. **(wyrażenie przysłowiowe) = na los szczęścia, na chybił-trafił, co Bóg da. (...)**

Znacznie częściej element ten pełni funkcję glosy i zawiera treść w języku łacińskim:

- **Fijołek (...)** 1. a. **xFijoł bot. = a) (...)** bot. (viola) **roś. z rodziny fijołkowatych (i jej kwiat).** Gatunki: (...); F. pachnący a. wonny, zwany Fijołkami marcowymi a. brunatnymi (**Viola odorata**); F. trójkolorowy (**v. tricolor**), bardzo zmienny: (...)
- **Fijołkowaty, Fijołkowy bot.:** Rośliny fijołkowe a. fijołkowe (**violaceae**) **rodzina przyrodzona roślin dwuliścieniowych.**
- **Firletka (...)** bot. (lychnis) **roś. z rodziny lepicowatych.** Gatunki: a) F. gaształd a. gaształd (**l. chalcidonica**), niek. Płomieńczykiem, Gasztwą, Krzyżakami zwana; b) F. kąkol zwyczajny a. Kąkolnica (**l. githago**), niek. Czarnuchą a. Czarnuchą żytnią, Czarnuszką zwana; e) F. kwiat Jowisza (**l. flos Jovis**); (...)
- **Falisty (...)** 2. przen. a. Falowany, Falowaty mający nierówności kształtu fal: (...) Anat.: **xFalot F. (tętnicowy), in. wężykowy (plexus undulatus).** Lek.: Tętno faliste (**pulsus undosus**). (...)
- **Fałdzisty (...)** 2. **Żołądek F. (abomasum) = czwarty oddział żołądka zwierząt przeżuwających, śluz, trawieniec, żołądek właściwy. (...)**

- **Fasola** (...) bot. 1. (phaseolus) roś. z rodziny motylkowatych. Gatunki: (...) d) F. wielokwiatowa, także Szabelbon zwana (ph. multiflorus). (...)

Zarówno zakres, jak i komentarz dzielą wspólną typografię, różnią się za to rodzajem treści – jednak w niektórych przypadkach mogą być trudne do odróżnienia, np.:

- **Fasola** (...) bot. 1. (...) a) F. karłowata a. piesza (nizka), pospolicie Piechotą zwana (ph. nanus); (...) c) ogr. F. szparagowa (z miękkim strąkiem, bez pergaminu).
- **Felc** (...) 2. (...) cieś. i stol. wrąb, nacięcie brzegów drzewa a. kamienia dla ściślejszego spajania: Dać co na F. (= spoić brzegami). F. w futrze od drzwi a. okna (na pomieszczenie kitu) = przyłga. (...)
- **Fluks** (...) II. (...) 2. choroba bydłęca: Mieć fluksa (o ludziach) = być porywczym, nagłym, bez zastanowienia. (...)

4.3.4.14. Objaśnienie

Objaśnienia, podobnie jak zakres oraz komentarz i glosa, umieszczane są w nawiasie i mogą znaleźć się po definicjach, terminach zagnieżdżonych i przykładach użycia – na tym jednak podobieństwa się kończą. Pierwsza różnica jest funkcjonalna, ponieważ segment ten zawiera treść pomagającą w poprawnym zrozumieniu znaczenia elementu po którym następuje. Druga różnica jest typograficzna – objaśnienia są bowiem zapisywane kursywą i poprzedzone znakiem równości (=). Dla przykładu:

- **Fagocytowy** przym. od **Fagocyt**: Komórka fagocytowa (= fagocyt).
- **[Fagólny]** niepewny, na którym nie można polegać: To je chłop nie F. (= na którego można liczyć).
- **†Falić, i, ił i [Falić]** p. **Chwalić**: †F. ś. potrzebnoy (= jeżeli należy ś. chwalić). (...)
- **Farsiarz, a, lm. e, Farser, ×Farserz** (...) 2. ten, co prawi śmieszne niedorzeczności, puszcza puffy, kaczkę (= plotki): F. jesteś! był obraz, ale oblazł. Gam. (...)
- **Feldspatowy** przym. od **Feldspat**: Minerale F. (= z feldspatu złożony). Skała feldspatowa (= feldspat zawierająca).
- **Figowy** przym. od **Figa**: (...) Ser F. (= wyrabiany z fig, migdałów, orzechów, pistacji i korzeni). ×F. ogrodnik. Kn., Bobr. (= hodujący figi). (...)
- **Filozof** (...) 7. człowiek niejasno, zawile sądzący o rzeczach prostych, niepotrzebnie a. niefortunnie biorący na kiel w rzeczach prostych, oczywistych, gdzie wystarcza zdrowy rozsądek; człowiek niemądry, pozbawiony zdrowego rozsądku: To mi F. z ciebie! (= toś mi palnął głupstwo!). F., co lewarem kluski jada. (...)
- **Folwarczny, ×Folwarczany, ×Folwarkowy** przym. od **Folwark**: Grunt F. Konie folwarczne (= robocze). Ogród F. (= warzywny). Szlachta folwarczna (= posiadająca folwarki). (...)
- **Frytura, y**, blm. olej rzepakowy z krochmalem i łojem, używany na okrasę zamiast masła: Oszukują! smażą pieczeń na fryturze. F. płynna (= bez łaju). (...)

W ramach objaśnień poświadczono również termin zagnieżdżony:

- **Filtracyjny** przym. od **Filtracja**: Dno filtracyjne (= służące do oddzielania brzości od masy zaciernej, *in. dno cedne*). Każdą filtracyjną p. **Każd**.

Co więcej, w niektórych przypadkach można znaleźć objaśnienie dla konkretnego wyrazu wewnątrz definicji, co wykazuje pewną analogię względem podawania wariantów w przykładach użycia – wówczas parser informacji tej nie ekstrahuje:

- **Farba** (...) 7. myśl. *puszczać farbę (= juchę)*, *broczyć krwią, krwawić*: Kot farbuje. (...)

4.3.4.15. Odsyłacz

Zbiorczo potraktowane są tutaj dwa rodzaje odsyłaczy: **referencyjne** (wyłącznie wskazujące na relację między parą haseł) oraz **derywacyjne** (dodatkowo opisujące rodzaj relacji między danymi hasłami).

System odsyłaczy jest dość konsekwentny i tylko sporadycznie zdarzają się w nim błędy, takie jak odesłania donikąd lub brak odesłania w hasło głównym (Majdak 2018a: 116). Przegląd odsyłaczy obu typów w oparciu o całą treść *SW* można spotkać u Majdak (2008a: 106–109); piszący te słowa skupia się tutaj jednak na ich realizacjach znalezionych w materiale badawczym.

Odsyłacze referencyjne wprowadzane są najczęściej poprzez skrócenie p. (z rozwinięciem „patrz”), np.:

- **xFalić się, i ś., ił ś. p. Falować**: Szereg galopem fali ś. jak morze. Słow.
- **†Fijałka, i, lm. i i [Fijałka] p. Fijołek**. Glosy z w. XV. (...)
- **†Fortka, i, lm. i p. †Forta**. W. XV.

Niekiedy zamiast nich bywa użyty znak równości – por.:

- **xFejerwerkier, a, lm. rzy 1. p. Fajerwerkarz. 2. woj. = Fajerwerkier**.
- **Feldspat** (...) 5. F. wapienny = **Anortyt**. (...)

Odsyłacze derywacyjnych jest natomiast więcej rodzajów, opartych o kategorie gramatyczne:

- **xFacecjonowanie, a, blm., czynność cz. Facecjonować**.
- **Falsyfikatorski przym. od Falsyfikator**: Manipulacja falsyfikatorska.
- **Fałdzistość, i, blm. rz. od Fałdzisty**.
- **Familjantka, i, lm. i forma ż. od Familjant; [Famnelijantka]**.
- **Fantazyjnie przys. od Fantazyjny**.
- **Fantazyjność, i, blm. rz. od Fantazyjny**.
- **Fałdowany 1. im. od Fałdować**. (...)

Nie wszystkie odsyłacze tego rodzaju poświadczono w literze F, np.:

- **Ósmy licz. porządkowy od Osiem**: (...)

Na rozłączne funkcje obu rodzajów odsyłaczy wskazuje dodatkowo ich zdolność do współwystąpienia na tym samym poziomie wewnątrz pojedynczego hasła, np.:

- **!Filgranowy** przym. od **Filgran**, p. **Filigranowy**.
- **Filuterność**, i, blm. rz. od **Filuterny**, p. **Filuterja**.

Odesłanie do innych haseł zawsze uwzględnia kwalifikatory symboliczne:

- **†Fikowy** przym. od **†Fik** pod 1: Wziąwszy liście fikowe, uczyniłaśta sobie wieniki. B. Sz.
- [**Flach**, u, lm. y] p. [**Flek**].
- [**Forka**, i, lm. i] p. **†Fora**.

Podobnie jak dezambiguatory:

- [**Folg**, folgu, lm. folgi] p. **2. Folga**.
- **Fryza**, y, lm. y (...) 5. zdń. p. **1. Fryz**. (...)
- **1. ×Fraktowy** przym. od **1. Frakt**, p. **Frachtowy**.
- **2. ×Fraktowy** przym. od **2. Frakt**, *świecki, nie kościelny*: Śpiewania kościelnego, bądź na organach, bądź traktowego, bądź chóralnego, bardzo się dobrze u nich wyuczają. Birk.
- **Futrowanie**, a, lm. a **1.** blm., *czynność cz. 1. i 2.* **Futrować**.

Odsyłacze mogą się odnosić nie tylko do całego wyrazu, ale także do jego konkretnych znaczeń:

- **Fasolka**, i, lm. i **1.** p. **Fasola**. **2.** *rodzaj delikatnej fasoli, soja*.
- **Fladrowanie**, a **1.** blm., *czynność cz.* **Fladrować**. **2.** lm. a p. **Flader**.

Czy nawet terminów zagnieżdżonych:

- **†Francowaty** lek. (...) **2.** p. **Francowy**: *Niemoc francowata a. choroba francowata* p. **Przymiot**.

Odsyłacz może wskazywać także na konkretne znaczenie lemmy, do której odsyła, przy pomocy przyłączonego po lemmie przyimka *pod*:

- **†Farynny** przym. od **Faryna** pod 4: *Szczęście na farynnym stole tak trwałe, jak przepiórcze pole.* Koch.
- **Frantówka**, i, lm. i **1.** × *forma ż.* od **Frant** pod 1: *F. ta jakie ma serce chytre i zmyślane.* Lubom. (...)

W przypadku odsyłaczy referencyjnych wskazanie konkretnego znaczenia współwystępuje z użyciem znaku równości zamiast skrócenia p., np.:

- **Fizjoplastyka**, i, blm. = **Autoplastyka** pod 2.
- **Fizjotypja**, i, blm. = **Autoplastyka** pod 1.

Jeden odsyłacz może ponadto przyłączać dwie lemmy jednocześnie przy pomocy spójnika „i” – por.:

- **Faseczka**, i, lm. i **1.** p. **Fasa** i **Faska**. **2.** [F.] *wanna, wanienska do kąpieli*.
- **Feldspat** (...) **4.** F. *wapiennosodowy* p. **Oligoklas** i **Andezyn**. (...)

- [Febra, y, lm. y] p. **Febra i Zimnica**. (...)

Warianty odsyłaczy derywacyjnych, w przeciwieństwie do odsyłaczy referencyjnych, przyłączane są przy pomocy średnika, np.:

- **Federacyjny przym. od Federacja; Federalny**: Polacy przechowywali federacyjno-stanową formę uspołecznienia. T. K. (...)
- **Fortepjanistka, i, lm. i forma ż. od Fortepjanista; Pjanistka**.

W sposób identyczny przyłącza się także definicje po tego rodzaju odsyłaczach:

- **Facetka, i, lm. i forma ż. od Facet; dziewczyna Lekka**: To ci, panie, szyk F.! Gom. Co to za F.?
- **Futrzany przym. od Futro; podbity futrem**: Kołnierz F. Rękawice futrzane.

Tak samo jak kwalifikatory:

- **Fałdowy przym. od Fałda; gieol.**: Góry, doliny fałdowe = powstałe wskutek kurczenia ś. skorupy ziemskiej.

Można jednak napotkać sytuacje, w których zamiast średnika użyty jest przecinek, np.:

- **Flisowski, Flisarski, Flisi, Flisowy przym. od Flis, flisacki, orylski**: Jarzyna flisowska = pęczak, kasza jęczmienna, suto okraszona słoniną i skwarkami. (...)

Odsyłacze bywają też umiejscowione w innych miejscach hasła – zdarza się na przykład, że poprzedzają część definicyjną hasła, co oznacza, że odesłanie do innego artykułu hasłowego dotyczy wszystkich znaczeń, np.:

Fantazyjny (posp. i fil.) **przym. od Fantazja 1. fantazji dotyczący**: Proces F. Czynność fantazyjna. Objawy fantazyjne = halucynacje i przywidzenia a. widziadła. 2. p. **Fantastyczny**: Utwór, kapelusz F. Meble fantazyjne. Oficerowie zbytkowali nietylko w przepisanych mundurach, ale i w fantazyjnych półmundurach. Fred. A.

Można je także zaobserwować na pozycji bezpośrednio po definicji, w nawiasie lub po przecinku:

- **Fazonety, ów, blp. cukier, p. Gryzетки**. <Fr. fanchonnette>
- [Famuła, y, lm. y] 1. i † a. †Pamuła, [Pamuła], xGarnizorówka **potrawa z owoców rozgotowanych i chleba, gramatka (p. Chamuła)**: Ośrodku suchej z chleba warzyć, do tego jabłek a gruszek warzonych przydać, będzie potrawa, którą famułą nasi przodkowie zwali. Syr. (...)

Co interesujące, powyższa lemma wchodzi także w skład ciągu synonimicznego (zawierającego nieco inny komplet wyrazów) wprowadzanego przez znak =, ale nie wyodrębnionego z definicji w literze **G** tego samego tomu:

- **Gramatka, i, lm. i, Gramatyka polewka gęstawa z wody i tartego chleba (= famuła, faramuszka, garnizorówka, kapton, smelka, wodzianka), a. z chleba, piwa i żółtka**. (...)

4.3.5. Segment terminalny

Segment terminalny jest najprostszy w identyfikacji, a także pod względem strukturalnym. W jego skład mogą wchodzić łącznie cztery subsegmenty – trzy z nich rozpoczynają się skróceniem, a czwarty umieszczony jest w nawiasie trójkątnym. Chociaż występować może dowolna ich kombinacja, to ich kolejność jest zawsze następująca:

Zdrobnienie → Zgrubienie → Porównaj → Etymologia

W hasłach na literę **F** nie ma przykładów posiadających wszystkie subsegmenty, ale poświadczono jedno hasło w tomie I (**2. Głąb**), które je zawiera:

- **2. Głąb** (...) **Zdr. Głąbek, Głąbik, [Głębik]. Zgr. Głąbisko. Por. Głąbie, Kłąb. <? GŁAB>**

Pierwsze trzy subsegmenty są bardzo do siebie zbliżone strukturalnie i funkcjonalnie, dlatego też zostaną omówione wspólnie.

4.3.5.1. Segmenty derywacyjne

Zbioreczo potraktowane tutaj są trzy subsegmenty wprowadzające formy derywowane: kolejno **zdrobnienia** i **zgrubienia** (utworzone od lemmy danego artykułu hasłowego), a także **powiązane hasła** (ze względu na wspólne pochodzenie, podobne znaczenie lub sposób utworzenia). Wszystkie trzy wymienione subsegmenty składają się zasadniczo z dwóch głównych komponentów:

1. Skrócenia rozpoczynającego segment: **Zdr.**, **Zgr.** i **Por.**;
2. Lemmy lub grupy lemm wraz z elementami ją opisującymi.

Ta dwudzielna struktura jest ilustrowana na poniższych przykładach:

- [**Fujawica, y, lm. e**] (...) **Por. Fujawa.**
- [**Fajnowny, Fejnowny**] (...) **Por. [Fajn].**
- **Foryś** (...) **Por. 2. Forysic.** (...)
- [**Fujawa, y, lm. y**] (...) **Por. Fujawica, Kujawa.**
- [**Fitka, i, lm. i**] (...) **Por. Fiutka i Fuca.**
- **Fiuk** (...) **Por. Fiutka, Fucka i Fuca.**
- **Fum** (...) **Zdr. †Fumik.** (...)
- **Faeton** (...) **Zdr. Faetonik, ×Faetoncik, [Fajtonik].** (...)

Jak widać powyżej, oprócz samej lemmy – zapisywanej wielką literą – wystąpić może przed nią dezambiguator oraz kwalifikator symboliczny. W materiale badawczym nie zaobserwowano przypadku wystąpienia lemm w tej kategorii wraz z kwalifikatorami innego rodzaju¹³ ani informacją gramatyczną.

4.3.5.2. Segment etymologiczny

Subsegment ten jest w przeciwieństwie do poprzednich nie jest wprowadzany skróceniem, ale oznaczony nawiasem trójkątnym: < >¹⁴. W materiale badawczym 432 hasła z 3880 (czyli ich niecałe 9%) zawiera informację etymologiczną. Informacja ta jest silnie ustrukturyzowana,

¹³ Kwalifikatory i ich klasyfikacja są opisane w **rozdziale 5.2.**

¹⁴ Nawias ten jest obecny w *Spisie skrótów*, obok kwalifikatorów symbolicznych.

jednak poziom złożoności poszczególnych realizacji tego segmentu jest bardzo zróżnicowany – porównaj:

- [Fcieć, fce, fciał] (...) <CHOT>
- Flusztajn (...) <Nm. Flusstein>
- Fena (...) <Nm. Finne, Pinne>
- Fluowodorny (...) <Fluor + Wodór>
- Fąfel (...) <Zap. Fr. fanfan = dziecina>
- Fermel (...) <Zap. Nm., z łć. forma>
- Fińszczyzna (...) <Od nazwy plemienia Finnów>
- Filipcy (...) <Od nazwiska mineraloga Ang. Phillips †1874>
- †Far (...) <Od nazwy wyspy Gr. Fáros, na której urządzona była słynna latarnia morska>
- Firasy (...) <Może Gr. thýra (podług wymawiania Ngr. fíra) = drzwi, a. łć. foras = za drzwi>
- Fajans (...) <Fr. faïence, Nm. Fayence, od nazwy miasta Włos. Faenza, słynnego wyrobami z gliny>
- Folsz (...) <Śr. łć. fullaria, stąd Nm. Fuller a. Füller = wałkarz, z łć. fullo = folusznik>
- †Funehoreografja (...) <Now. z łć. funis = lina + Gr. choreía = taniec + grafía = -pis>

Mimo różnego poziomu złożoności informacji etymologicznej, można wyróżnić kilka elementów, na których jest ona zbudowana:

1. Wyraz rodzimy, od którego derywowano formę:
 - [Frebra, y, lm. y] (...) <Zam. Febra, pod wpływem Srebro>
2. Wyraz obcy, który jest podstawą zapożyczenia:
 - Fiolet (...) <Włos. violetto, Fr. violet, przez usta Nm.>
3. Słoworód, na podstawie którego powstał wyraz:
 - Fermentomierz (...) <Ferment + MIAR>
 - [Florek, rku, blm.] p. Chlorek. <Sld. do imienia Florek>
4. Nazwa własna, od której utworzono wyraz:
 - Fryc (...) <Nm. Fritz = zdr. im. Friedrich>
5. Objasnienie powstania i zmian semantycznych:
 - Fiksacja (...) <Z łć. fixus im., przez Fr. fixation = ustalenie, umocowanie; przejście znaczenia: obłęd na jednym punkcie, monomanja, w przeciwieństwie do Warjacja>
6. Cytat zamiast objaśnienia (tylko jedno poświadczenie w materiale badawczym):
 - xFingara (...) <„Fingara jest Febe, czyli księżyc wschodnich narodów.” Mick.>

Jednostki leksykalne będące źródłem zapożyczenia mogą być również opatrzone kwalifikatorem semantycznym.

- xFalmusy (...) <Nm. †Feimnuss, †Feimuiss>
- [Formedla, i, lm. e] (...) <Nm. [vormiedel]>

Elementy te mogą wchodzić w skład większych konstrukcji na dwóch różnych poziomach. Mowa tu nie tylko o rozbudowanie informacji o danym elemencie przez dołączenie dalszych informacji go opisujących, w tym m.in. informacji gramatycznej czy tłumaczeń:

- **xFlogochemja** (...) <Now. z Gr. flóx, 2 pp. flogós = płomień + Chemja>
- **Filistyn** (...) <łć. Philistinus, z Hb. Pliszthî, lm. Pliszthîm = Filistyn, por. Filister>
- **Fatygant** (...) <Fr. fatigant im. dosł. = utrudzający>
- **Film, u, lm. y** fot. błona żelatynowa. <Może = firm-, łć. firmare = umocowywać, utrwalać>

Ale także w ramach złożzeń wielowyrazowych:

- [**Fikołajki, ów, blp.**] (...) <Zmieszanie Fikać z Mikołajki>
- **Fotoheljograf** (...) <Now. z Gr. fô̄s, 2 pp. fôtós = światło + hélios = słońce + -gráfos = piszący>

Wszystko to powoduje, że możliwych struktur informacji etymologicznej jest bardzo dużo – w materiale badawczym zidentyfikowano ich 139¹⁵. Nawet w przypadku prostych konstrukcji występuje tutaj duże zróżnicowanie, co widać choćby na przykładzie wskazywania dwóch wariantów w zapożyczeniach, gdzie może się różnić zapis pierwszej litery lemm oraz sposób przyłączenia drugiej z nich:

- **Fuganek** (...) <Nm. Fügebank, Fügebaum>
- **Fota** (...) <Tur. fota, futa>
- [**Fus, a, lm. y**] (...) <Czes. fous a. vous>
- **Futrować** (...) <Nm. futtern i füttern>

W przypadku par wariantów, które różnicuje tylko obecność dodatkowej litery w jednym z nich, zastosowano zapis łączny – litery pojawiająca się tylko w jednym z wariantów, a więc opcjonalne w zapisie, umieszczone są w takich przypadkach w nawiasie:

- **Feldfebel** (...) <Nm. Feldwe(i)bel>
rozwińcie: Feldwe(i)bel → Feldwebel, Feldweibel
- [**Fodrować, uje, ował**] (...) <Nm. fo(r)dern = żądać>
rozwińcie: fo(r)dern → fodern, fordern
- †**Feret** (...) <Fr. ferret(te)>
rozwińcie: ferret(te) → ferret, ferrette
- **Fen** (...) <Nm. Fö(h)n>
rozwińcie: Fö(h)n → Fön, Föhn

Występują również sytuacje, w których wyraz ma taką samą formę wyrazu w więcej niż jednym języku i trudno rozstrzygnąć faktyczne źródło jego zapożyczenia, lub został on zapożyczony równoległe z kilku źródeł, np.:

- **Fakir** (...) <Tur., z Arab. fakir dosł. = ubogi>
- **Falbana** (...) <Fr., Włos., Hp., Pg. falbala>

I tutaj, podobnie jak w przypadku łączenia szeregu leksemów, sekwencje oznaczeń języków zapisywane są więcej niż jedną metodą, co widać porównując powyższe przykłady stosujące do tego celu przecinki z poniższym, wykorzystującym spójnik „i”:

- **xFohista** (...) <Chiń. i Japońs. Fo = nazwa Buddy>

¹⁵ Klasyfikacja ta została automatycznie wygenerowana na podstawie wyniku parsowania, opisanego w **rozdziale 10.4.7** Błąd! Nie można odnaleźć źródła odwołania.; ze względu jednak na ograniczenia czasowe oraz ich arbitralny charakter, piszący te słowa postanowił nie kontynuować tego wątku badań.

Ciekawym elementem stosowanym przez autorów *SW* są też tak zwane „słoworody”. Chociaż nie wyjaśniają oni, czym właściwie one są, to ich funkcja jest zbieżna z tematami słowotwórczymi. Lemma nie zawsze jednak zawiera je w dokładnie takiej formie, w jakiej zostały podane, ale mogą zachodzić oboczności względem słoworodu – zwłaszcza samogłoskowe, lub takie, które wynikają ze zjawisk związanych z obecnością samogłosek (przede wszystkim palatalizacji):

- BZD → Fiu bǫdziu
- CHOT → [Fcieć]
- CHRON → Falochron
- JAD → Figojadka
- ŁOM → Falołomny
- MIAR → Fermentomierz
- MOW → Fałszomówca
- MY → [Flaszkomyj]
- NOS → Filaronośny
- PIS → †Fałszopiszec, †Fraszkopis
- ROD → †Fałszorodny
- TROP → Farbotrop
- WŁOD → xFortunowładny
- ŻR → xFrancuzożerczy

Słoworody pod pewnymi względami przypominają rdzenie w kontekście, w jakim są one stosowane w gramatyce i leksykografii arabskiej – elementem wspólnym jest tutaj oparcie informacji przede wszystkim na spółgłoskach, podczas gdy samogłoski mogą się różnić lub być zupełnie pominięte (zwłaszcza w przypadku epentetycznego *e*).

4.3.6. Hasła zagnieżdżone

Przez hasła zagnieżdżone rozumie się podhasła, ale nazywane tutaj w ten sposób ze względu na terminologię programistyczną, niejako wbrew tradycji językoznawczej, gdzie element „zagnieżdżony” to taki, który został zdefiniowany wewnątrz innego elementu (znajduje się niżej w hierarchii). W przypadku reprezentacji podhasła w pliku XML są one zagnieżdżone względem hasła głównego.

Majdak (2008a: 109–110) wymienia pięć kategorii hasła, które z zasady nie stanowią odrębnych artykułów hasłowych:

1. wyrażenia quasi-przysłówkowe tworzone z członem „po”;
2. substantiva męskie i żeńskie;
3. czasowniki z „się”;
4. prefiksy, sufiksy i inne niesamodzielne;
5. związki frazeologiczne.

Typów hasła zagnieżdżonych jest jednak więcej; niezależnie od tego, można za zidentyfikować, ponieważ zachowują one te same reguły typograficzne, co hasła główne, nie rozpoczynają natomiast nowego akapitu, przez co są niejako „podczepione” pod główne artykuł hasłowy – por.:

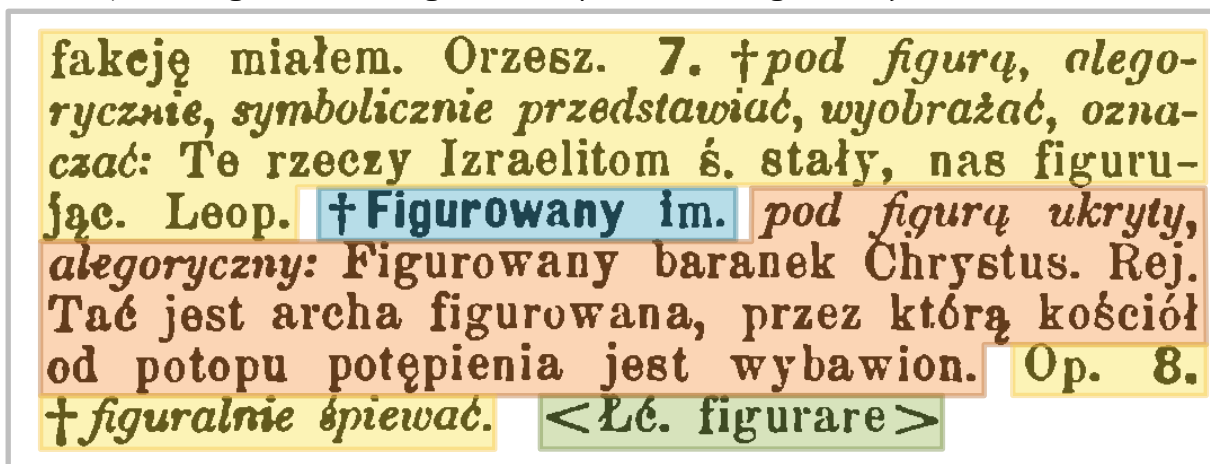
- **Fałdować, uje, ował, »Fałdzić układać fałdy, zbierać w fałdy, marszczyć, drapować:** F. suknię. Przen.: Kończy Moloch, czoło fałdując z rozpaczy. Przyb. **F. ś. układać ś. w fałdy, nie leżeć gładko, marszczyć ś., ciągnąć ś.:** Suknia źle uszyta fałduje ś. Przen.: Czoło mu ś. fałdowało. Krasz. <Nm. falten>
- **Fuks, a, lm. y 1. [F.] koń lisiej barwy (...) 6. zool. (vanessa polychloros, vanessa urticae) motyl dzienny. [Fuksem] przys. przypadkiem, trafem, przez zbieg okoliczności, mimowolnie, niechcący; potajemnie, podstępnie, bez pozwolenia:** Wykręcił ś. jakoś fuksem z ambarasu. Fuksem przejść z klasy do klasy (= przejść piecem, niezastuzenie). <Nm. Fuchs dosł. = lis>
- **Furmański 1. a. [Furmanów], †Furmany przym. od Furman:** Koń F. Rozpuścił pysk, jak F. bicz. Prz. (= jak dziadowski bicz). Toby ś. ledwie na F. wóz włożyć mogło. Rej. 2. [Furmańska maść] p. Maść. [Furmańskie krople] p. Kropla. **Po furmańsku** przys. = jak furman.

O ile podhasła umieszczane są przeważnie między segmentem centralnym a terminalnym hasła głównego, to w mogą się również znaleźć między jego poszczególnymi znaczeniami, np.:

- **Figurować, uje, ował (...)** 7. † *pod figurą, alegorycznie, symbolicznie przedstawiać, wyobrażać, oznaczać:* Te rzeczy Izraelitom ś. stały, nas figurując. Leop. †**Figurowany** im. *pod figurą ukryty, alegoryczny:* Figurowany baranek Chrystus. Rej. Tać jest archa figurowana, przez którą kościół od potopu potępienia jest wybawion. Op. 8. † *figuralnie śpiewać.* <łć. figurare>

Segmenty tego hasła można przedstawić tak, jak jest to widoczne na **ilustracji 15**:

Ilustracja 15. Fragment hasła „Figurować” z podhasłem „Figurowany”.



5. Skrócenia

Skrócenia użyte w *SW* są tematem niepozornym, któremu wystarczyłoby poświęcić akapit czy nawet krótki przypis – tak przynajmniej może się wydawać do momentu, w którym nie zajdzie potrzeba ich opracowania; dopiero wtedy skrócenia okazały się tematem zaskakująco wielowymiarowym i złożonym.

Potrzeba opracowania wspomnianych skrótów wynika z dwóch powodów. Pierwszym z nich jest fakt, iż skrócenia te stanowią część treści zasadniczej słownika, a więc są konieczne, aby jego użytkownik mógł poprawnie nawigować w tej treści – ich spis może pełnić formę legendy lub stanowić podstawę do konwersji skompresowanej informacji użytej w hasłach na jej pełną, rozwiniętą formę. Drugim natomiast wykorzystywanie skrótów przez parser podczas analizy składniowej treści *SW*, co jest opisane w **podrozdziale 10.3.4**.

Skróceniami w *SW* co prawda zajmowała się już Majdak (2018a: 146, 150–158), jednak ograniczając się do samych kwalifikatorów (pomijając w ogóle oznaczenia autorów i skrócenia w innych funkcjach); poza tym analiza w niniejszej pracy jest pogłębiona, ponieważ obejmuje nieścisłości wewnątrz samego systemu skrótów, a także błędy jego implementacji w samej treści.

A przeszkód, na które napotkano podczas opracowywania skrótów użytych w *SW*, jest wiele rodzajów: część z nich wynika ze specyfiki tegoż konkretnego dzieła i pomaga lepiej zrozumieć proces jego kompilacji (jak tworzenie własnego systemu skrótów i brak konsekwencji w poszczególnych tomach), pozostałe natomiast są reprezentatywne dla leksykografii tego okresu i można je uogólnić na inne słowniki (jak metoda ich tworzenia, w tym np. brak ujednoliconego zapisu spółgłosek miękkich).

Niniejszy rozdział poświęcony jest strategiom rozwiązywania problemów, które zostały napotkane podczas analizy skrótów, a także ich klasyfikacji.

5.1. Różnice pomiędzy tomami

Wszystkie użyte w słowniku skrócenia – przynajmniej w założeniu – powinny znaleźć się w *Spisie skrótów*. Taki spis jest obecny w każdym z ośmiu tomów *SW*, jednak spisy w poszczególnych tomach różnią się od siebie, i to pod wieloma względami, co należy uwzględnić przed omówieniem pozostałych aspektów skrótów w tym dziele zawartych.

Pierwszą różnicą jest już same umiejscowienie *Spisu skrótów* oraz jego erraty (nazywanej w omawianym słowniku *Dalszymi skróceniami*¹⁶) w każdym z tomów, co zostało wykazane w rozdziale **4.1** Megastruktura *SW*. Poza tym ich wewnętrzny układ w poszczególnych tomach nie jest wewnętrznie spójny. Ekstremalnym przykładem są *Dalsze skrócenia* w tomie V (widoczne na **ilustracji 16**), które składają się z dziewięciu odrębnych bloków tekstu, wyróżnionych na podstawie braku kontynuacji kolejności w układzie alfabetycznym. Ponadto skrócenie

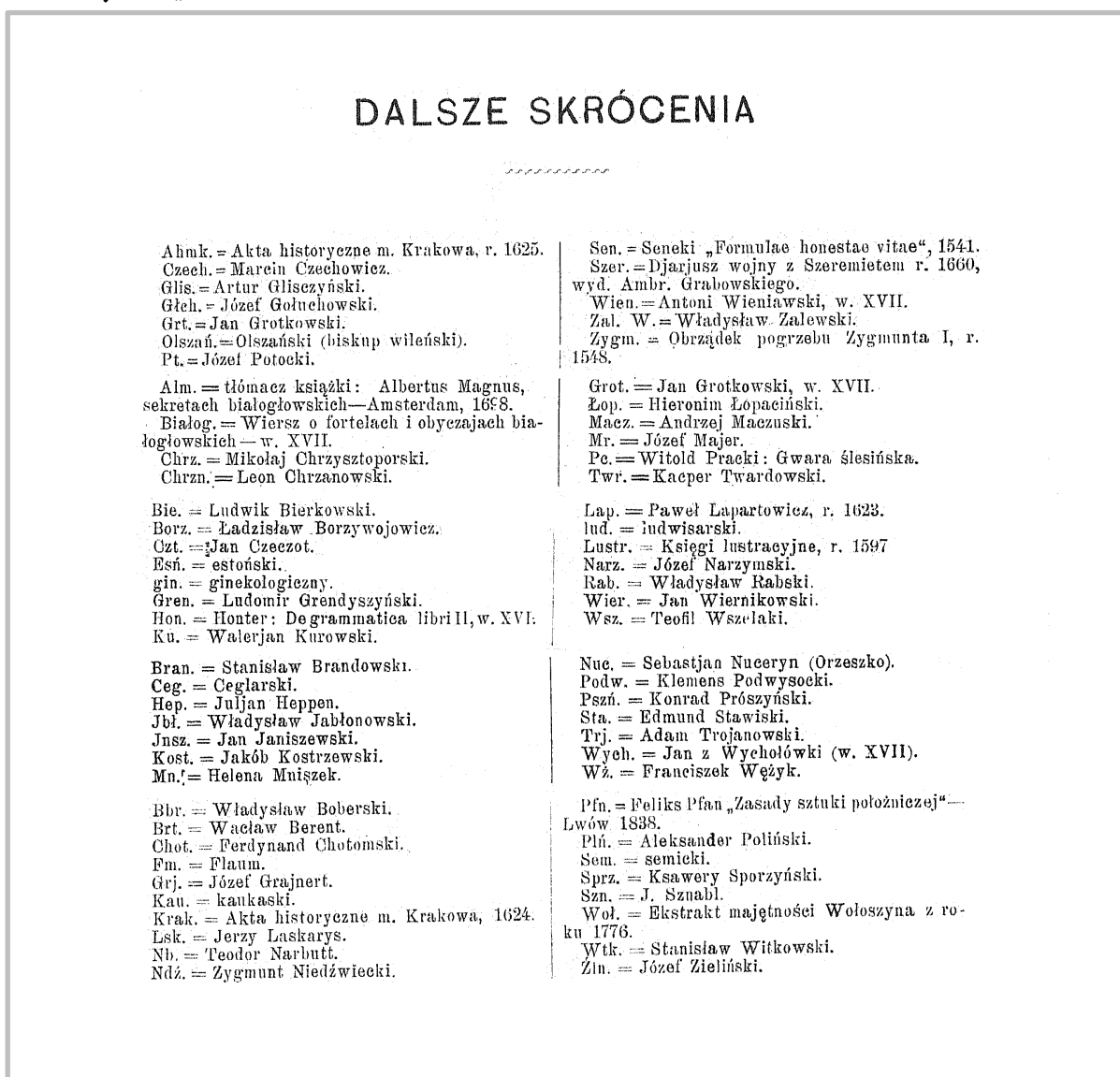
¹⁶ Należy tu zaznaczyć, że *Dalsze skrócenia* są obecne tylko w wersji *SW* z *Federacji Bibliotek Cyfrowych*, i to ona będzie tutaj traktowana za podstawę jako zawierająca więcej informacji. Temat wyboru źródła jest poszerzony w rozdziale **8.5.2**.

Grt. = Jan Grotkowski zostało omyłkowo umieszczone dwukrotnie: w pierwszym i szóstym bloku tekstu.

Jako że za makietę badawczą został przyjęty zbiór haseł zaczynających się na literę F, które to w całości zawierają się w tomie I słownika, to początkowo zaczerpnięto skrócenia ze spisu tylko tego tomu. Strategia ta okazała się błędem na wielu poziomach, który wynika z rozbieżności między:

1. Skróceniami użytymi w materiale badawczym a tymi zawartymi w *Spisie skróceń*¹⁷ w tomie I;
2. Skróceniami użytymi w materiale badawczym a tymi zawartymi w *Spisach skróceń* tomów dalszych, a przed wszystkim w tomie VIII (ostatnim);
3. Skróceniami zawartymi w *Spisach skróceń* poszczególnych tomów.

Ilustracja 16. „Dalsze skrócenia” w tomie V.



¹⁷ Piszący te słowa będzie traktował dalej *Dalsze skrócenia* jako część *Spisu skróceń*;

Należy tutaj wyjaśnić, że jako rozbieżności traktuje się tutaj poniższe przypadki:

1. W treści występuje skrót nieobecny w żadnym *Spisie skrótów*;
2. W treści występuje skrót o zapisie odmiennym od tego w *Spisie skrótów*;
3. Więcej niż jedno skrót odnosi się do tego samego autora lub źródła;
4. Zapis lub opis skrótu w *Spisie skrótów* jest zmienny pomiędzy tomami;
5. W *Spisie skrótów* występują nowe skróty, które nie są obecne w *Dalszych skrótach* danego tomu;
6. W *Dalszych skrótach* wprowadzone są nowe skróty, które nie są obecne w *Spisie skrótów* danego tomu;
7. W *Spisie skrótów* występuje skrót, który nie jest poświadczony w treści.

Wszystkie opisane wyżej rozbieżności zostały poświadczony w słowniku chociaż raz. Weryfikacja objęła spisy wszystkich ośmiu tomów. Należy tutaj jednak podkreślić, że asercja w punkcie 7 nie jest jednak w chwili obecnej w pełni weryfikowalna, ponieważ wykraczała poza opracowany materiał badawczy (litera F), a nawet poza zweryfikowane ręcznie dane tekstowe (litery F, I, Ó, V, X oraz Y, czyli łącznie niecałe 2% treści SW). Przeprowadzone przeszkanie pozostałych 98% pełnego tekstu wszystkich tomów SW w formie brudnego OCR zarówno metodą pełnego dopasowania (ang. *exact match*), jak i przybliżonego dopasowania (ang. *fuzzy match*), nie jest zatem w pełni miarodajne.

Poza tym w tomie VI pojawiło się następujące objaśnienie, zwiastujące wprowadzenie dodatkowych informacji przy nazwiskach autorów:

Podawany w poprzednich tomach i pomnożony w niniejszym wykaz „Autorów i źródeł” uzupełniamy obecnie wskazówkami chronologicznymi. Ułatwił, one poznanie dziejów wyrazu z materiału frazeologicznego.

Po imieniu i nazwisku autora dajemy rok śmierci, w razie zaś braku dokładniejszej wiadomości – przynajmniej stulecie albo inną wskazówkę.

W niektórych pozycjach wątpliwych musieliśmy poprzestać na znaku zapytania.

Nazwiska, przy których niema wskazówki chronologicznej, należą do autorów jeszcze żyjących.

Co jednak zastanawiające, rok śmierci pojawił się przy nazwiskach autorów dopiero w tomie VIII. Niestety, w żadnym z tomów nie uzupełniono informacji o źródłach eskerpcji.

Ze względu na dużą liczbę skrótów oraz przyjęte założenie próby automatyzacji możliwe największe liczby kroków w procesie, zamiast ręcznej kompilacji i analizy piszący te słowa opracował moduł do programu SWParser, który dokonuje tego sam, czego efektem jest **Aneks M: Skonsolidowany spis skrótów SW**. Proces generowania tego aneksu opisany jest w **podrozdziale 10.4.12**, a jego treść objaśniona jest w **podrozdziale 14.13**.

5.2. Klasyfikacja skrótów

Konieczność sklasyfikowania skrótów w SW może nie być oczywista, jeśli wziąć pod uwagę fakt, że użytkownikowi słownika nie jest ona konieczna do ich zrozumienia. Ich usystematyzowanie ma jednak pewną wartość badawczą – niemniej klasyfikacja ta z punktu widzenia

niniejszej tylko celem instrumentalnym; celem terminalnym jest takie opracowanie skrótów, które można będzie wykorzystać w analizie składniowej treści *SW*.

Aby jednak taką analizę skutecznie przeprowadzić, konieczne było rozróżnienie poszczególnych typów pragmatycznych oznaczeń występujących w tekście, których implementacja jest opisana w **podrozdziale 10.4.12**. Niniejszy rozdział natomiast jest próbą usystematyzowania skrótów zarówno pod względem pogłębienia wiedzy leksykograficznej, jak i umożliwienia ich implementacji na potrzeby programistyczne – jest to nie lada wyzwanie, ponieważ łączna liczba zidentyfikowanych oznaczeń wszystkich rodzajów wynosi 1797.

5.2.1. Terminologia

Przed sklasyfikowaniem skrótów w *SW* w pierwszej kolejności należy uporządkować terminologię i zdefiniować, czym one właściwie są. Sami autorzy *SW* wyróżniają w *Spisie skrótów* tylko dwie podkategorie:

- a) „Wyrazy”;
- b) „Autorowie i źródła”.

O ile druga z kategorii trafnie opisuje zawarte w niej jednostki, to pierwsza już nie, ponieważ obok wyrazów zawiera też połączenia wyrazowe i symbole. Problem z nazewnictwem dotyczy także kategorii nadrzędnej – *Spis skrótów* zawiera bowiem nie tylko skrócenia, ale czasem też pełne formy, zwłaszcza w przypadku krótkich nazwisk cytowanych autorów.

Kolejną kwestią jest funkcja elementów ze *Spisu* w tekście: niektóre używane są w ściśle określonych miejscach i opisują jedną z zamkniętej listy właściwości, inne zastępują pełną formę wyrazu lub grupy wyrazów. Nie można zatem kategorii tej opisać jako tożsamej z kwalifikatorami, chociaż je zawiera.

Z problemem tym borykają się także badacze innych słowników. Bilińska (2013: 122–125, 143–166) na przykład skrócenia na trzy podkategorie: (1) kwalifikatory, (2) skróty redakcyjne oraz (3) skróty nazw języków.

Najrozsądniejszym według piszącego te słowa jest zbiorcze opisanie elementów *Spisu skrótów* jako „oznaczenia”, co niweluje potrzebę odróżniania elementów pod względem liczby wyrazów składowych, ich formy (pełnej lub skróconej) oraz funkcji, a kiedy to konieczne – stosowanie szczegółowych kategorii, które opisane zostały w dalszej części tego rozdziału

5.2.2. Poziomy analizy skrótów

W kontekście niniejszej tezy doktorskiej analiza skrótów *SW* będzie przeprowadzona na dwóch płaszczyznach: **treściowej i formalno-składniowej**.

Przez **analizę treściową** rozumie się tu analizę skrótów jako integralnych elementów tekstu *SW*, a zarazem także i formy reprezentacji danych; z jednej strony stanowią one uzupełnienie treści w tym dziele zawartych, z drugiej zaś – są ich skompresowaną informacją, co pozwala zaoszczędzić miejsce (co jest szczególnie istotne w druku).

Pod pojęciem **analizy funkcjonalno-składniowej** natomiast rozumie się tu analizę nie warstwy językowej, ale warstwy tekstowej, a dokładniej: pozycji skrócenia w haśle i – co często z tej pozycji wynika – jego funkcji.

Oba poziomy analizy można zobrazować na podstawie następującego przykładu:

- **Figlarstwo** (...) Będziesz **ś.** wprawował przy mnie w gospodarstwo, odrzuciwszy te fochy i twoje **F. Zwr. M.**

W przykładzie tym obecne są trzy skrócenia trzech różnych rodzajów (spośród ich znacznie większego zbioru obecnych w *Słowniku*, których pełna klasyfikacja przedstawiona jest w **podrozdziale 5.2.3**). Na poziomie analizy treści należy zwrócić uwagę na to, że:

1. skrócenie **ś.** zastępuje konkretny wyraz pospolity: *się*;
2. skrócenie **F.** zastępuje wyraz hasłowy w formie słownikowej, zależnie od artykułu hasłowego, w którym się znajduje – tu: *Figlarstwo*;
3. skrócenie **Zwr. M.** wskazuje na pochodzenie przykładu użycia, odwołując się do nazwy źródła, które w *Spisie skrótów* widnieje w pełnej formie jako: *Zwroćenie Matjasza z Podola, zap. w. XVII.*

Analiza treści skrótów jest o tyle istotna, że określenie, gdzie w haśle może on wystąpić – to z kolei umożliwia pogłębioną analizę funkcjonalno-składniową. Zdolność do odpowiedniego zinterpretowania przez parser takich skrótów jest kluczowa do poprawnej segmentacji; parser musi bowiem rozpoznać, że:

1. **ś.** jest skróceniem, które może wystąpić w środku lub na końcu zdania;
2. **F.** jest skróceniem, które może wystąpić w dowolnym miejscu zdania;
3. **Zwr. M.** jest pojedynczym skróceniem umieszczanym po zdaniu, które opisuje.

Kryteria funkcjonalno-składniowe są tutaj o tyle istotne, iż błędne rozpoznanie tych cech może spowodować błędną segmentację tekstu przez parser, tj. „ucięcie” zdania w złym miejscu – w przeciwnym razie program mógłby zakwalifikować **Zwr.** jako źródło, a **M.** jako odrębne skrócenie na początku następnego zdania, czy nawet całego kolejnego zdania, składającego się wyłącznie z litery „M”.

Należy tutaj pamiętać, że parser dysponuje jedynie listą reguł dotyczących znaków i ich połączeń – żadna wiedza językowa i pozajęzykowa nie jest tutaj uwzględniana. Dlatego też, aby reguły takie mogły powstać, potrzebna jest statyczna, zamknięta lista skrótów wraz z przypisanymi kategoriami.

5.2.3. Kategoryzacja funkcjonalna

Zastosowany tutaj podział na kategorie, choć czerpie z typologii kwalifikatorów obecnej w literaturze, uwzględnia dodatkowo możliwość praktycznego wykorzystania nowych kategorii w parserze. Kategorie te będą zatem funkcjonalne, w głównej mierze oparte się o pozycję oznaczeń w mikrostrukturze i współwystępowanie z innymi elementami, a dopiero w drugiej kolejności o semantykę.

5.2.3.1. Oznaczenia językowe

Kategoria ta obejmuje te skrócenia w części a) *Spisu*, które zaczynają się wielką literą i wskazują na pochodzenie zapożyczenia. Oznaczenia w tej kategorii mają charakter kompozycyjny i można je podzielić na trzy dalsze podkategorie: nazwy, prefiksy oraz złożenia.

Pierwsza podkategoria zawiera nazwy trzech rodzajów:

1. języków, np.:
 - Czes. = czeski
 - Fr. = francuski
 - Grm. = giermański
 - Prs. = perski
 - Ros. = rosyjski
2. dialektów, np.:
 - Braz. = brazylijski
 - Gnm. = górnoniemiecki
3. grup języków, np.:
 - Ameryk. = amerykański
 - Ar. = aryjski, indoeuropejski
 - Sł. = słowiański
 - Sem. = semicki
 - Rus. = ruski wogóle (Ros., Brs., Ukr.)

W ramach tej podkategorii na uwagę zasługują tylko dwa skrócenia – ze względu na użycie przestarzałych form z punktu widzenia współczesnej polszczyzny, a mianowicie:

- Blg. = bułgarski (dziś: *bułgarski*), oraz
- Grm. = giermański (dziś: *germański*).

Formy te jednak są na tyle zbliżone do współczesnych, iż nie powinny sprawiać problemu współczesnym użytkownikom języka. Problemu interpretacyjnego może w praktyce za to parserowi sprawiać skrócenie *Cyg. = Cygański*, które odnosi się do języka (lub ich grupy), a którego forma jest zbieżna z oznaczeniem źródła *Cyg. = Mateusz Cygański (w. XVI)*.

Druga podkategoria to **oznaczenia prefiksalne**. Oznaczenia te nie występują samodzielnie, ale umieszczane są przed oznaczeniem nazwy, aby zawęzić zakres języka źródłowego do konkretnej epoki jego rozwoju. Oznaczeń tych jest tylko cztery:

1. N. = nowo- (poświadczone w materiale badawczym tylko raz):
 - <Ngr. Fanariótēs dosł. = mieszkaniec części Konstantynopola, zwanej po Ngr. Fanár dosł. = latarnia (morska)>
2. N = nowo- (poświadczone w materiale badawczym cztery razy):
 - <Ngr. Fanariótēs dosł. = mieszkaniec części Konstantynopola, zwanej po Ngr. Fanár dosł. = latarnia (morska)>
 - <Może Gr. thýra (podług wymawiania Ngr. fíra) = drzwi, a. tć. foras = za drzwi>
 - <Ros. fonart = latarnia, z Ngr. fandri>
3. S. = staro- (poświadczone w materiale badawczym tylko raz):
 - <Gr. foínix, 2 pp. foínikos, przez S. Słow. finikь>¹⁸
4. S = staro- (poświadczone w materiale badawczym dwa razy):
 - <Nm. Flöte (ze Sfr. flahute), Włos. flauto i Fr. flûte traversière>
 - <Rum. fluera, przez Słc. fujara>
5. Śr. = średniowieczny (poświadczone w materiale badawczym 46 razy):

¹⁸ Podane w przykładzie skrócenie S. Słow. jest błędne i przed parsowaniem jest poprawiane przez program, co jest wyjaśnione w **podrozdziale 10.2.2.1**.

- <Śr. łć. filialis>
- <Śr. łć. Franciscanus, od imienia Franciscus>
- <Śr. Grn. virlei = taniec z pieśniami, z Fr. daw. virelai>
- <Śr. Gr. phasol, łć. phaseolus>
- <Zap. z Ags. vîgle = wróżba, Śr. Dnm. wycheler = wróżbita; Miklosich wywodzi z Śr. łć. vigiliae = hałaśliwe obchody nocne>

Jak można zauważyć, w *SW* obecne są po dwa oznaczenia dla przedrostków *nowo-* oraz *staro-*, przy czym tylko warianty w formie osobnego wyrazu (z kropką i spacją) są obecne w *Spisie*, mimo że w materiale badawczym częściej używane są formy przyłączane bezpośrednio.

Trzecia podkategoria to złożenia. Mimo oddzielnego umieszczenia oznaczeń języków oraz form prefiksalnych w *Spisie skrótów*, występują w nim także dwie formy złożone: Śr. łć. oraz Śr. Grn.. Umieszczenie ich tam jest trudne do uzasadnienia z dwóch powodów: po pierwsze zważywszy na to, że mają generatywny charakter i ich utworzenie na podstawie poprzednich dwóch kategorii nie stanowi żadnego problemu, co czyni je redundantnymi, a po drugie ponieważ pozostałe formy złożone obecne w treści słownika nie zostały w *Spisie* uwzględnione.

Oprócz opisanego wyżej stosowania form prefiksalnych nieobecnych w *Spisie skrótów*, jedno ze złożzeń stosuje zapis Śr. Grn. = średniowieczno-górnoniemiecki., mimo że samodzielna nazwa języka ma zapis Gnm. = górnoniemiecki.

5.2.3.2. Oznaczenia stylistyczne

Kategoria ta służy do opisu stylu (lub rejestru) języka. Wszystkie oznaczenia stylistyczne znajdują się w pierwszej części *Spisu skrótów* i zapisywane są małymi literami; ich przypisanie do tej klasy wynika z analizy semantycznej podanych objaśnień. W ocenie piszącego te słowa dwa z nich wymagają jednak komentarza.

Pierwszym z nich jest nim *właśc.*. Jego zaklasyfikowanie stanowi pewną trudność, ponieważ pełni w tekście dwie różne funkcje. Z jednej strony skrócenie to opisuje styl, na co wskazuje samo objaśnienie skrócenia:

- *właśc.* = w znaczeniu właściwym; właściwie

Dodatkowo częste występowanie skrócenia przed definicją wskazuje na funkcję kwalifikatora; co więcej, na pozycji tej występuje wyłącznie obok kwalifikatora *przen.*, który bezspornie opisuje rejestr, jak widać na poniższych przykładach:

- **Mazać** (...) 5. *właśc.* i *przen.* *zamazywać, zmazywać, wykreślać, wymazywać, kasować, zacierać, ścierać, gładzić: (...)*
- **Mieścić** (...) 2. × *właśc.* i *przen.* *szukać miejsca, układać ś., gnieździć ś., sadowić ś.: (...)*
- **Mroczyć, y, ył** *właśc.* i *przen.* *zaciemniać, ćmić, zamraczać, pogrążyć w ciemnościach: (...)*
- **Parafja** (...) <łć. parochia z Gr. paroikia = *właśc.* to, co można objechać, obwód, objazd>
- **Wykąsać** (...) b) dok. *właśc.* i *przen.* *kąsając wystraszyć, wypędzić; wygryźć, wyjeść kogo skąd, dogryzając mu, wysadzić go: (...)*

Z drugiej strony oznaczenie to jest niekiedy stosowane jako zwykłe skrócenie wyrazu niemającego funkcji kwalifikatora:

- †Nad czczo (**właśc.** na tcz) p. Czcz. <Zam. Na czczo>
- Sąg (...) 2. gór. skała, **właśc.** sążeń skały do wykucia przeznaczony. (...)
- †Słēza, y, lm. y p. Śłōza. <**właśc.** Słza>
- w spisie: Jan Dymitr Solikowski †1603 (**właśc.** cytaty należą do Syrok. [Wł. Syrokomli], tłumacza na język polski).

Jest także poświadczony jeden przypadek, w którym oznaczenie to opisuje rejestr, ale nie znajduje się na pozycji właściwej kwalifikatorom:

- **Ostrowidz**, a 1. lm. **owie** a. e, †**Ostrowid** *człowiek bystrego wzroku, ostrooki, bystrooki, okal* (**właśc.** i przen.): (...)

Dodatkowym utrudnieniem jest fakt, że różnica między oznaczeniami **właśc.** a **dosł.** zdaje się nie być jasna. Autor tych słów uznał jednak, że zaliczenie oznaczenia do kwalifikatorów jest pragmatyczniejszym podejściem z punktu widzenia analizy składniowej, ponieważ te są wykorzystywane do segmentacji tekstu, ale tylko jeśli występują na określonej pozycji – nie wpływa to więc negatywnie na parsowanie haseł ze skróceniem **właśc.** w innej funkcji.

Drugim problematycznym oznaczeniem jest **bł.** = **błędnie**, które również zdaje się opisywać rejestr. Choć w całym słowniku kwalifikator ten jest użyty ponad dwieście razy, to w przyjętym materiale badawczym występuje tylko raz – jest to jednak najistotniejsze wystąpienie, ponieważ daje wystarczający kontekst na rozstrzygnięcie, jak był on traktowany przez samych autorów:

Figojadka, i, lm. i bot. **bł.** p. **Przypołudnik**. <Figa + JAD>

Jak widać na przykładzie tego hasła, przed kwalifikatorem **bł.** umieszczony jest kwalifikator **bot.** Bez oddzielającego je spójnika, który autorzy konsekwentnie stosowali między kwalifikatorami tego samego rodzaju:

- **Faktor** (...) 4. ! lm. y **fil. i mat.** p. **Czynnik**. (...)
- **Familja** 5. x **bot. i zool.** rodzina, grupa. (...)
- x**Farbnik**, a, lm. i **chem. i fizj.** ciało, nadające barwę innym ciałom, barwnik, pigment: (...)
- **Faszyha**, y, lm. y 1. a. †**Fascyna inż. i woj.** wiązka chróstu, używana do robót fortyfikacyjnych, do urządzania grobel i naprawy dróg, gać: (...)
- **Fazować**, uje, ował 1. **bud. i stol.** zbierać kanty na fazę. (...)
- **Fele** (...) 2. a. **Nut, Wpust, Paz, Paza, Pazwa, Żłobek, Rowek cieś. i stol.** wrąb, nacięcie brzegów drzewa a. kamienia dla ściślejszego spajania: (...)
- **Felcować** (...) 2. **cies. i stol.** spajać na felc: (...)
- **Fenacetyna**, y, blm., **Acetfenetydyna apt. i chem.** związek chemiczny w postaci szaro-białego krystalicznego proszku. (...)
- **Ferment**, u, lm. y 1. a. **Drożdzeń, Kisiel, xZaczyn chem. i piw.** materia organizowana a. nieorganizowana, powodująca fermentację. (...)

Nie jest to jednak jedyna pozycja dla tego oznaczenia – kwalifikator **bł.**, podobnie jak **lep.**, jest też często stosowany bezpośrednio przed jednostkami leksykalnymi, zwłaszcza w szeregach synonimicznych:

- **Czterykroć, czterechkroć, †Czterzykroć, †Cztyrkroć 1. bł. C. razy = cztery razy.**
- **Ciągliwy, Rozciągliwy, xCiąglawy, xCiągły, bł. Ciągły dający ś. wyciągać na cienutkie nici, druty, nie rwąc ś. przy tym: (...)**
- **Dziwło, a, lm. a, bł. Rozłóg bot. (...)**
- **Eleacki przym. od Eleata, bł. Elejski: (...)**

5.2.3.3. Oznaczenia dziedzinowe

Kategoria ta, jak nazwa wskazuje, zawiera kwalifikatory opisujące poszczególne dziedziny nauki, techniki, sztuki i sportu. Kryterium utworzenia tej grupy wynika z analizy semantycznej podanych objaśnień; podobnie jak w przypadku klasy poprzedniej, wszystkie skrócenia tego typu zapisane są małymi literami i znajdują się w pierwszej części *Spisu skrótów*. Wśród nich nie występują żadne oznaczenia, których klasyfikacja jest problematyczna, pojawiają się natomiast dwa, których zapis może być mylący.

Pierwsze oznaczenie to **ceg.** = **ceglarski**, które w tomach V i VI jest błędnie traktowane jako nazwisko i zapisane wielką literą, mimo że jest obecne w części a) *Spisu* i używane jest w treści *SW* na pozycji kwalifikatora. Błąd ten, który może wynikać z podobieństwa wyrazu pospolitego do nazwiska, został poprawiony w kolejnych tomach.

Drugie to skrócenia słowa „dentystyczny” – w spisie tomu I ma ono formę „dent.”, natomiast formę „den.” w spisach tomów kolejnych; pierwsza forma jest częstsza i ma poświadczenia w tomach I i IV, podczas gdy druga – w tomach VI i VII. Ze względu na brak wyraźnej przewagi frekwencyjnej jednej z form ostatecznie na potrzeby analizy składniowej zostały przyjęte obie.

5.2.3.4. Oznaczenia gramatyczne

Ta kategoria oznaczeń zawiera informacje o odmianie i kategorii leksykalnej, a skrócenia do niej należące występują w pierwszej części *Spisu* są pisane małymi literami.

Od tej ostatniej reguły istnieje jednak jedno odstępstwo, a mianowicie oznaczenie **Dźwn.** = **dźwiękonaśladowczy, onomatopeiczny**, które poświadczone jest wyłącznie w części zawierającej opis etymologiczny, występując raczej w charakterze opisu derywacji niż opisu części mowy.

Również dwa inne skrócenia powiązane z kategorią leksykalną wymagają tutaj komentarza. Pierwsze z nich to **w.** i w *Spisie skrótów* ma objaśnienie „wykrzyknik”, jednak zarówno w tekście słownika, jako i samym *Spisie* owo skrócenie zastępuje wyraz „wiek” (por. **podrozdział 5.2.3.7**). Oba użycia charakteryzują się jednak różnymi pozycjami w haśle, dzięki czemu ich rozróżnienie przez parser ani człowieka nie stanowi problemu. Analogicznie jest w przypadku skrócenia **im.**, które ma rozwinięcie „imiesłów” lub „imię, imienia”.

Drugim oznaczeniem jest **skróc.** = **skrócone, skrócenie**. Występuje ono wyłącznie w funkcji kwalifikatora i stanowi informację analogiczną do opisu części mowy – nie został

znaleziony ani jeden przypadek użycia go jako zwykłego skrócenia w tekście – jednak poświadczony jest tylko pięć razy:

- †Ktemu **skróć.** = ku temu (...)
- Pom. **skróć.** Pomocnik.
- Poń **skróć.** = *Po niego.* <Po + Ni(ego)>
- Ref. **skróć.** 1. referent 2. Ref. s. referendarz stanu.
- Reg. k. **skróć.:** *Registrator kolegiálny.*

Oprócz tego na uwagę zasługuje skrócenie **rod.** = rodzaj (gramatyczny), które ma tylko jedno poświadczenie – w tomie IV, w haśle *Pasówka (...)* <Od Pasowa, **rod.** ż. od *Pasowy*>. Należy jednak rozpatrywać cały fragment **rod.** ż. jako integralną całość opisującą proces słowotwórczy – kategoria ta jest opisana w rozdziale **5.2.3.12**.

Ciekawe jest również skrócenie **rozk.** = rozkazujący, które może występować samodzielnie, np.:

- **Brać, bierze**, [ja bierę a. bierzę, ty bieres, oni bierzą], **brał**, [**rozk.** bier, biercie, im. bierząc] p. **Wziąć**.
- **Dać, da, dał**, [Daci, 1 os. lp. dum, **rozk.** 2 os. lp. Da, (Daje = *dajże*, *Dóceż* = *dajcież*)], nied. **Dawać**, [Dajać, Daiwać], †Dajać (...)
- **Jeździć, i, ił**, [**rozk.** jeździj], [Jiźdzać] (...)
- **Obrać, obierze, obrał**, **rozk.** obierz a. [obier], †Obebrać, [Obebrać], nied. **Obierać** (...)

Ale równie często jest poprzedzone zbytecznym słowem „tryb”:

- **Cisnąć, śnie, snął** a. [cisł], **tryb rozk.** [ciś!] (...)
- **Cofnąć, nie, nął**, [**tryb rozk.** Cof, Cofcie], [Cafnąć, Cownąć, Cawnąć, Copnąć], nied. **Cofać**, [Cafać, Cować, Copać] (...)
- **Karać**, 1 os. lp. karzę a. [karam], karze, 3 os. 1m. karzą a. [karają], **karał**, **tryb rozk.** karz a. †karzy (...)
- **Oblizać, że, zał** (**tryb rozk.** obliż a. [oblizaj]), ×Obliznąć, nied. **Oblizywać**, ×Oblizować (...)

Brak koherencji autorów jest tutaj o tyle zaskakujący, że żadne znane wystąpienie skrócenia nie uzasadnia dodania wyrazu „tryb” choćby celem ujednoznacznienia – zdaje się być zupełnie nieznaczące.

Warto też zaznaczyć, że dwa oznaczenia tego rodzaju – **zgr.** = *zgrubiały.* oraz **zdr.** = *zdrobniały.*, pełniące funkcję gramatyczną, mają także swoje odpowiedniki zapisywane wielką literą. Są to nieobecne w *Spisie skrótów Zgr.* i *Zdr.*, które mają te same rozwinięcia, ale pełnią funkcję odsyłaczy: wskazują na augmentativa i deminutiva, które można utworzyć na podstawie lemy.

5.2.3.5. Oznaczenia frekwencyjne

Kategoria ta jest niewielka, a w jej poczet wchodzi tylko trzy oznaczenia: **część.**, **niek.**, oraz **zwykle.** Pierwsze dwa występują w pierwszej części *Spisu* i pisane są małymi literami, to ostatnie jednak skróceniem nie jest, ale pełni w hasłach analogiczną funkcję.

Decyzja o wyodrębnieniu tej kategorii wynika z łączliwości prawostronnej tych oznaczeń z kilkoma rodzajami elementów w haśle.

Po pierwsze oznaczenia te w charakterze kwalifikatorów występują przed jednostkami leksykalnymi:

- Benzol, u, lm. e, ×Fenilowódor chem. węglowódor, złożony z 6 atomów węgla i 6 atomów wodoru, niek. ×benzyna.
- Dwójnasób, część. W D., We D. (...)
- Dyfinitorjum, w lp. nieod., lm. ja, jów, część. Definitorjum (...)
- Dykterja, i, lm. e, część. zdr. Dykteryjka (...)
- Dzieciak, a, lm. i, Dzieciuch, Dzieciuk rub. 1. (...) 2. część. Dzieciuch (...)
- Filharmoniczny, Filarmoniczny, niek. Filharmonicy muzyczny, młujący, uprawiający muzykę: (...)
- Fraternalizować, uje, ował, część. F. ś. (...)
- †Fryj (...) 2. u, lm. e, część. w lm., p. Fryjerstwo: (...)

Po drugie mogą występować bezpośrednio przed opisem semantycznym (zwłaszcza definicją) artykułów hasłowych:

- Filicyna (...) niek. wyciąg eterowy z tegoż korzenia. (...)
- Fisk (...) niek. prokuratorja. (...)
- [Fryszki] (...) 2. niek. chytry, przebiegły. (...)

Po trzecie występują one przed innymi oznaczeniami, zwłaszcza jeśli pełnią one funkcję gramatyczną:

- [Baląg, część. lm. Balągi, ów] (...)
- [Bisiórka, i, część. i, lm. Wisiorki] (...)
- Blondyna, y, część. w lm. y, ×Blondynka (...)
- Dolec, dolegnie, doległ, ×Dolegnąć, †Doląc, część. nied. Dolegać, †Dolegać (...)
- Dowidzieć, i, ał 1. a. ×Dowidzać nied. (część. z przeczeniem) (...)
- Fatałaszek, szka, lm. szki (część. w lm.) (...)
- Federsztyft, u, lm. y 1. zwykle w lm. ślus. sztyfty kanciaste, na których utrzymuje ś. sprężyna zamkowa. (...)
- Fiok (...) 2. część. w lm. strój przesadny, oryginalny, zwracający uwagę na głowie kobiecej; lok, pukiel: (...)
- Flores, u, lm. y część. w lm. figiel rysunkowy, figlas, zawity deseń, ozdoba w kształcie gałązki z liśćmi i kwiatami, w lm. esy-floresy: (...)
- Furora, y, blm. (zwykle tylko w 4 pp.) (...)
- Gwiazda (...) 4. a. część. zdr. Gwiazdka plama biała na czole u konia: (...)
- †Górze, a, blm. (część. jako w.) nieszczęście, bieda, biada: (...)
- Przerznąć (...) Przen.: (część. nied.) P. co = przecinoć, ciągnąć ś., przechodzić przez co; dzielić, rozdzielać, przedzielać, przepasywać: (...)

Po czwarte skrócenia te mogą wystąpić w roli modyfikatora grupy wyrazów (dowolnego fragmentu tekstu):

- **Bam! Bom! Bum!** w., *naśladowący odgłos*: 1. dzwonu = *bambałam*, a **niek.** 2. dzwonka, zegara itp. = *dzyń-dzyń! dyń-dyń! deń-deń! dendeleń! den-dele!*
- **Budżet, u**, lm. **y** ekon. *wykaz dochodów i wydatków państwa a. instytucji (niek. osoby) w pewnym okresie czasu; preliminarz*: (...)
- **Egzystencja** (...) *posp. istnienie; część. możliwość a. sposób istnienia, byt, utrzymanie*: (...)

To ostatnie zastosowanie bywa niekiedy wyróżnione odpowiednim formatowaniem, jednak ich ekstrakcja ze struktury metodami heurystycznymi wymagałoby analizy treści na znacznie pogłębionym poziomie.

Wstępna analiza wskazuje także na to, że poszczególne oznaczenia nie są używane w pełni symetryczne – niektóre z nich zdają się współwystępować z elementami pewnych typów częściej niż z innymi, jednak nie można twierdzenia tego na chwilę obecną podeprzeć danymi liczbowymi ze względu na niewielki zakres materiału badawczego. Mimo to na uwagę zasługuje skrót **część.**, które niemal zawsze stosowane jest do opisu gramatycznego; w ramach tej kategorii najczęstszym połączeniem jest opis liczby gramatycznej lemmy, przede wszystkim w formie **część. w lm.**, np.:

- **Farfocel** (...) 2. [F.] **część. w lm.** *szyszkowina, w lm. sadzele bolesne w siedzeniu*. (...)
- **Fatałach, a**, lm. **y** (**część. w lm.**) *pog. 1. szmat gałgan, gałganek, rupieć, szpargał*. (...)
- **Flak, a**, lm. **i** 1. a. [Flaka] **część. w lm.** *kiszka, jelito, w lm. wnętrzości zwierzęce wog.; wnętrzości ludzkie, bebechy*: (...)

Poza materiałem badawczym zdarzają się także poświadczenia formy **część. lm.**, np.:

- **Cugiel, gła, część. lm.** (...)
- **×Czapraszka, i część. lm.** (...)

Ta druga forma, pomijająca przyimek *w*, jest ponad dwadzieścia razy rzadsza (ma tylko 6 znanych poświadczeń). Przyimek ten nie występuje w innych kontekstach przed skróceniem **lm.**, co sugeruje, że jest on dodany w funkcji *copuli* dla tej pary oznaczeń.

5.2.3.6. Oznaczenia modalne

Oznaczenia w tej kategorii opisują inne elementy w kategoriach modalności – są ich tylko dwa w pierwszej części *Spisu*, i zapisywane są małymi literami:

- **wzgl.** = względnie (respective).
- **zap.** = zapewne.

Pierwsze ze skrótów pojawia się w *Spisie* dopiero w tomie VI i dalszych i ma kilkadziesiąt poświadczeń w tekście – wszystkie w części etymologicznej haseł rozpoczynających się na literę S, i służy do wprowadzenia alternatywnego źródła zapożyczenia, np.:

- **Sczerwienić** (...) <S(z) + Czerwienić, **wzgl.** Czerwienieć>
- **Sennik** (...) <Od Sen, **wzgl.** Senny>
- **Spekulator** (...) <Fr. spéculateur, **wzgl.** łć. speculator>

Drugie skrócenie jest natomiast obecne w spisach i treści wszystkich tomów, ale również tylko w części hasła poświęconej etymologii – i to wyłącznie na jej początku. Jego funkcja to oznaczenie braku pewności co do podanej informacji o pochodzeniu wyrazu, np.:

- †**Ferenc** (...) <Zap. Węg. Ferenc = Franciszek>
- [**Ficówka**, i, lm. i] *szkopek*. <Zap. zam. *Świecówka = naczynie do robienia świec maczanych>
- †**Fotarłę** (...) <Zap. Węg. fattyú = bękart, dzieciuch>

W wyniku swojej inicjalnej pozycji w segmencie etymologicznym, skrócenie to zawsze rozpoczyna się wielką literą, przez co jego forma pokrywa się z oznaczeniem autora: Zap. = Gabryela Zapolska †1921., np.:

- **1. Fala** (...) Przen.: Etaláže sklepowe wypływają na trotuar falą porcelany. Zap. (...)
- **Figówka** (...) *rodzaj cykorji*: A toż kawa: sama cykorja i to F.! Zap. (...)
- **!Fota**, y, lm. y p. **Fot**: Ciotka popełniła wielką fotę, której nic nie naprawi. Zap. <p. Fot>

Tożsamość realizacji tych skrótów nie jest jednak problematyczna, ponieważ znajdują się one w innych segmentach nadrzędnych. W materiale badawczym jest jeden przypadek, w którym oznaczenie autora znajduje się w etymologii – jednakże nie znajduje się ono na pozycji inicjalnej w segmencie, przez co dezambiguacja nie jest problematyczna:

- **xFingara**, y, lm. y *księżyc*: Przy bladym świetle wschodzącej fingary widziano giaura. Mick. „,Fingara jest Febe, czyli księżyc wschodnich narodów.“ Mick.>

5.2.3.7. Oznaczenia przyliczbowe

Oznaczenia przyliczbowe zostały tak nazwane ze względu na swoje współwystępowanie z liczbami – przylegają do nich bezpośrednio lewo- lub prawostronnie. Oznaczeń tych jest niewiele i żadne z nich nie jest obecne w spisie (a przynajmniej nie w tej samej funkcji – patrz rozdział 5.2.3.4. Oznaczenia gramatyczne), przy czym są to wyłącznie skrócenia w powszechnym użyciu, a więc niewymagające objaśnienia:

- gr. = groszy
- r. = rok
- w. = wiek

Kategoria ta jednak została wyróżniona ze względu na przydatność przy segmentacji zdań w treści artykułów hasłowych. Jej praktyczna aplikacja jest dokładniej opisana w części inżynierskiej pracy, jednak regułę wykorzystującą tę kategorię można opisać w sposób następujący:

1. Jeśli skrócenie występuje po liczbie, a po nim jest spacja i wielka litera lub koniec segmentu, to kropka skrócenia jednocześnie stanowi końcową granicę obecnego zdania, np.:
 - †**Almakować** (...) <Almacks = bale modne londyńskie podczas sezonu a. u wód, od 1765 r.>
 - **Chrzest** (...) Mieczysław przyjął C. święty 965 r. Trzymać a. podawać dziecko do chrztu. (...)

- **Ekstraordynaryjny**, †**Ekstraordynarny**, ×**Ekstraordynałny nadzwyczajny**: Sejm E. warszawski 1768 r.
 - **1. Frank, a**, lm. i jednostka monetarna we Francji = 50 gr.
 - **Furman** (...) Furmani, którzy towary za granicę wywożą, płacą po gr. 15. Vol. Furmanem posłałem (= przez furmana). (...)
2. Jeśli skrócenie występuje po liczbie, a po nim następuje opcjonalny znak interpunkcyjny, a potem spacja i mała litera, to zdanie kończy się przy następnym terminalnym znaku interpunkcyjnym, np.:
- **Cwancygier, a**, lm. y *dawna moneta austrijacka - 40-42 gr. pol.*, in. [sorokowiec]. (...)
 - **Czartyzm, u**, blm. *ruch socjalno-demokratyczny w Anglii 1835-1857 r., oparty na tak zwanej karcie narodowej, zawierającej wymagania robotników.*
 - **Filipowiec** (...) <Nazwa sekty od imienia sekciarza Filipa Pustoswiata, który swych współwyznawców koło r. 1700 przywiódł na Litwę i do Polski>
3. Jeśli skrócenie występuje przed liczbą, ale nie jest poprzedzone inną liczbą, to zdanie kończy się przy następnym terminalnym znaku interpunkcyjnym, np.:
- [**Bałaguła, y**, lm. y, m., **Bałagura, Baraguła, Baragoła**] (...) 4. *junak, zawadjaka, Lekceważący przyjęte zwyczaje, koniarz; rodzaj postępowca około r. 1840 na Ukrainie*: (...)
 - **Burza** (...) Burze wstrząsnęły r. 1848 monarchją habsburską. (...)
 - [**Cesarzówka, i**, lm. i] *tak nazywano około r. 1850 w Wilnie książeczkę „O czci Cesarza,” używaną w szkołach tamecznych.*
 - **Cyrkularny** (...) Sąd C. (nazwa sądu miejskiego w r. 1807). (...)
 - ×**Dziennio, a**, lm. a p. **Dzienia**: (...) Przedał i na wieczność darował pszczoły w dwójniew w wyższym dziennie za gr. 40. Gac.
 - **Filipowiec** (...) <Nazwa sekty od imienia sekciarza Filipa Pustoswiata, który swych współwyznawców koło r. 1700 przywiódł na Litwę i do Polski>
 - **Filomat, a**, lm. ci *członek stowarzyszenia miłośników nauki, które istniało w uniwersytecie lateńskim przed r. 1823.* <Gr. filomathes>
 - **Furmanka** (...) Eksaktorowie, którzy furmankę wybierają, tylko po gr. 12 od jednego wozu brać powinni. Vol. (...)

Nie poświadczono natomiast sytuacji, w której oznaczenia tego rodzaju znajdują się w bezpośrednim sąsiedztwie liczb z obu stron, co mogłoby być problematyczne do rozstrzygnięcia przez analizator składniowy.

5.2.3.8. Oznaczenia ogólne

Oznaczenia, które znajdują się w tej kategorii, obejmują zarówno skrócenia obecne w pierwszej części *Spisu*, jak i skrócenia, które znalazły się w materiale badawczym. Nie pełnią one funkcji kwalifikatorów i nie stanowią części opisu gramatycznego, tylko są formami skróconymi

powszechnie stosowanymi w języku współnoodmianowym. W *SW* mają swoje realizacje przede wszystkim w treści artykułów hasłowych, czyli definicjach i przykładach; jedynym celem ich użycia jest zatem kompresja informacji.

Aby umożliwić parserowi prawidłową segmentację treści artykułów hasłowych, kategoria ta została podzielona dalej na trzy podkategorie funkcjonalne, wyróżnione ze względu na pozycję w tekście, na której mogą one występować:

1. **Terminalne** – występujące tylko na końcu segmentu tekstu;
2. **Nieterminalne** – występujące na początku i w środku, ale nie na końcu tekstu;
3. **Neutralne** – mogące wystąpić na dowolnej pozycji.

Oznaczenia terminalne są tylko dwa, a ich ostatnia kropka jednocześnie zawsze pełni funkcję kropki kończącej zdanie:

- **i t. p.** = i tym podobne.
- **i t. d.** = i tak dalej.

Oba skrócenia mają łącznie ponad 1500 poświadczeń w tekście wszystkich ośmiu tomów. Zapis ten – w którym słowa skrócone są indywidualne i oddzielone spacjami – jest dzisiaj uważany za przestrzały. Co ciekawe, w tekście występuje także zapis identyczny ze współczesnym, tj. bez spacji i tylko z jedną kropką – kolejno ok. 90 razy dla **itp.** i ok. 50 razy dla **itd.** – ale tylko w tomie I, wyłącznie w hasłach zaczynających się na litery A, B, C oraz D, np.:

- **Afinacja, i, blm., hut p. Wydzielanie.** <Z Fr. affiner = rafinować **itd.**>
- **Bunsen, a, lm. y** w skróceniu zamiast ogniwo Bunsena: Jeden bunsen, dwa bunseny **itd.** <Od nazwiska chemika Nm. Bunsen, ur. 1811>
- **Cydr, u, lm. y, Cyder napój wyskokowy z jabłek, in. jabłecznik.** <Gr. síkera, łć. sicera, Włos. cidro **itd.**>
- **Deponować, uje, ował składać przedmiot wartościowy w ręce osoby trzeciej do przechowania.** <łć. deponere = składać **itd.**>

Oznaczenia nieterminalne analogicznie nigdy nie występują na końcu zdania (nigdy nie wskazują jego granicy), ponieważ służą do przyłączenia informacji. Zaledwie 8 z 15 skrótów tego typu widnieje w *Spisie*.

I w tej kategorii występują alternatywne formy zapisu – tym razem są ich dwie pary:

1. W niektórych hasłach **i** jako komponent innych oznaczeń występuje skrócenie **ks.** (skrócenie ortograficzne, zgodne z normami dzisiejszymi), w innych zaś występuje **kś.** (skrócenie ortofonetyczne);
2. W niektórych hasłach występuje skrócenie **t. j.** (częstsze, zgodne z normami starszymi), jak i **tj.** (rzadsze, zgodne z normami dzisiejszymi).

Poza rozbieżnościami w pisowni problematyczna w tej grupie jest również analiza semantyczno-funkcjonalna niektórych zawartej w niej oznaczeń.

Pierwszym oznaczeniem tego rodzaju jest **dosł. = dosłownie**. Na podstawie przestudiowania samego znaczenia rozważano jego zakwalifikowanie do oznaczeń stylistycznych, w materiale badawczym jego użycie jest ograniczone do części etymologicznej, a konkretniej przed tłumaczeniem obcej jednostki leksykalnej, np.:

- **Fakir, a**, lm. **owie a. rzy asceta indyjski**. <Tur., z Arab. fakir **dosł.** = ubogi>
- [**Fefermenc, u**, blm.] p. **Wiatrówka**. <Nm. Pfefferminze **dosł.** = mięta pieprzowa>
- **Foljał (...)** <Z łć. in folio **dosł.** = w liściu, t. j. arkuszowy (format)>

Nie pojawia się jednak we wszystkich hasłach, a tylko tam, gdzie zaistniał dryf semantyczny – element ten pełni zatem jednocześnie funkcję strukturalną oraz informacyjną, np.:

- **Foljować, uje, ował** druk. *numerować stronicę w księdze* (szczeg. handlowej), *paginować*. <Z łć. folium = liść; arkusz>

Drugim oznaczeniem, które się wyróżnia, jest skrót **zag.** = zagadka. Ma ono tylko dwa znane poświadczenia:

- [**Choda, y**, lm. **y**] **1. świnia (w zagadce)**: Wisa wisi, C. chodzi; wisa spadła, C. zjadła (**zag.** o gruszcze i świni).
- **Nawrócić (...)** [Wyjechał, nigdzie nie zajechał, tylko nawracał] (**zag.** o oraczu). (...)

Skrót ten jest nietypowy nie tylko ze względu na niewielką liczbę wystąpień, czy zbliżoną formę do oznaczenia autora: **Zag.** = Włodzimierz Zagórski (†1902)., ale z zupełnie innego powodu. Otóż wykorzystywane jest ono do zapisu rodzaju źródeł, którymi są zagadki; źródła tego typu są opisane w rozdziale **5.2.3.9**.

Warto też wspomnieć o skrótach **ur.** = urodzony. – dodanym przez piszącego te słowa do *Spisu* (które użyte nie tylko w hasłach, ale i w samym *Spisie*) – ze względu na to, że tylko wielkością pierwszej litery różni się od skrótu źródłowego **Ur.** (a. Psł.) = Urywek protokołów sądu radzieckiego sławkowskiego z lat 1713–1723.

Oznaczenia neutralne natomiast to takie, które mogą występować w każdym danego segmentu tekstu, ponieważ zastępują one wyrazy pospolite. Zidentyfikowano pięć elementów tego typu:

- **poch.** = pochodny, pochodzi.
- **roś.** = roślina.
- **ś.** = się.
- **wł.** = imię własne.
- **wog** = wogóle.

Pierwsze ze skrótów ma tylko jedno poświadczone wystąpienie, w którym rozszerza zakres odsyłacza o derywaty podanego leksemu:

- [**Hańśba, y**, lm. **y**] *opłata za skradzenie to nocy z pola zboża*. <? por. Chąśba i **poch.**>

Skrót **wł.** natomiast ma kilkanaście poświadczeń i opisuje nazwy własne w części hasła poświęconej etymologii, np.:

- **Haniflta, y**, lm. **ci**, **Hanifita** *sekcjarz muzulmański*. <Od Arab. **wł.** Abu-Kanifa>
- **Hamilton, a**, lm. **y** kraw. *rodzaj sukni męskiej*. <Od Ang. **wł.** Hamilton>

Skrócenie pojawia się trzy razy przed nazwą języka, zamiast po niej (w hasłach rozpoczynających się od litery A):

- **Arjostyczny** (...) <Od **wł.** Włos. Ariosto>
- **Aresowy** (...) <Od **wł.** Gr. Ares = Mars >
- **Arystarch** (...) <Od **wł.** Gr. Arístarchos, krytyka pieśni Homerycznych>

W jednym hasle skrócenie to zostało też redundantnie użyte w połączeniu „imię wł.”:

- **Bękart** (...) <Nm. Bank(h)art, Bankert, niby imię **wł.** (na wzór Gebhart, Reinhart) od Bank = ławka, więc Bankhart = niby ławczak, na ławce, na łtrycie (nie na łóżu) spłodzony>

Dominującym podejściem jest jednak używać wyrazów „imię” lub „nazwa” w pełnej formie, i tylko tego rodzaju przypadki mają poświadczenia w materiale badawczym, np.:

- **Faustyzm** (...) <Od imienia Faust, bohatera tragedji Götheego>
- **Fińszczyzna** (...) <Od nazwy plemienia Finnów>
- **Florans** (...) <Fr. florence, od nazwy miasta Włos. Florencji>

Ostatnie skrócenie, **wog.** = **wogóle**, zapisywane w SW jeszcze podług starej, łącznej pisowni i ma złożoną funkcję – używane do odwołania się do ogółu przedmiotów, zjawisk lub cech. Może ono występować jako część treści definicji na wszystkich pozycjach, np.:

- **[Fajczyć, y, ył]** *palić fajkę a. **wog.** tytuń.*
- **Filozofja** (...) 2. *nauka, badająca zasady bytu, poznania i działania, badająca istotę wszechrzeczy, dążąca do wytworzenia jednolitego poglądu na świat, badająca krytycznie zasadnicze pojęcia nauk specjalnych i syntezująca najogólniejsze wyniki tych nauk, teoria nauki **wog.:** (...)*
- **Fryszować, uje, ował** chem. i hut. *przygotowywać żelazo kowalskie z surowca, in. świeżyć; **wog.** oczyszczać kruszce przez topienie w ogniskach a. piecach. <Nm. frischen>*

W tym także przed pojęciami zagnieżdżonymi w definicji:

- **Figuralny** (...) 2. *muz. używający figur muzycznych: Śpiew, kontrapunkt F., **wog.** muzyka figuralna = urozmaicenie melodji użyciem nut zamiennych a. przechodnich: (...)*

Oraz na pozycji inicjalnej:

- **Fałd** (...) 2. ***wog.** zmarszczenie, np. krawędzi blachy przy pokrywaniu dachu. (...)*
- **Fińszczyzna** (...) 2. ***wog.** wszystko co fińskie, świat fiński.*
- **Furdyga** (...) 2. ***wog.** zamknięcie, karcer: Bywało ojciec na pokutę osadzi go w furdydze. Rzew. (...)*

Ze względu na występowanie w różnych częściach definicji, niekonsekwentne stosowanie kursywy oznaczającej treść definicji, a także niewielką przydatność dla parsera – nie jest ono traktowane jako kwalifikator.

5.2.3.9. Oznaczenia źródłowe

Kategoria ta obejmuje wszelkie oznaczenia dotyczące źródeł informacji w haśle, zwłaszcza zaczerpniętych przykładów, ale niekiedy i innych elementów. Chociaż z punktu widzenia analizatora składniowego funkcja wszystkich oznaczeń źródłowych jest jednolita, to ich format już nie. Poza tym rozróżnienie niektórych kategorii może być przydatne z punktu widzenia użytkownika ludzkiego. Mając to na uwadze, wyróżniono pięć podkategorii, a są nimi:

1. **Autorzy;**
2. **Dziela;**
3. **Gatunki literackie;**
4. **Daty;**
5. **Zagadki.**

Tylko pierwsze trzy kategorie widnieją w *Spisie skrótów* (w części drugiej) i wszystkie rozpoczynają się wielką literą. Pozostałe dwie natomiast nie zostały weń uwzględnione, ponieważ są kategoriami otwartymi, użytymi w treści zasadniczej słownika; podawane w pełnej, a nie skrótowej, formie, i w treści nie są przed nimi stawiane kropki końca zdania, którego dotyczą.

Oznaczenia autorów informują, że przykład został zaczerpnięty z ogółu dzieł danego twórcy, nie wskazując przy tym żadnego konkretnego dzieła. Jest to najliczniejsza grupa, ponieważ ich liczba wynosi aż 1215, czyli stanowią prawie 68% wszystkich oznaczeń. Jak na tak liczną grupę pojawia się tutaj stosunkowo niewiele oznaczeń problematycznych, jednak poniżej opisano ich dwa godne uwagi rodzaje.

Pierwszym z nich jest niekonsekwencja między oznaczeniami w *Spisie skrótów* oraz tymi użytymi w treści. Tak jest w przypadku oznaczenia **Les. p. łs.**, które nie ma poświadczeń w jakimkolwiek tomie i jest to błędna wersją skrócenia imienia Józef Franciszek Łęski (bez znaków diakrytycznych), ale widnieje w *Spisie*. Odwrotnie jest w przypadku skrócenia **Mikl. = Franciszek Miklosicz**, które przestało pojawiać się w spisie od tomu IV, mimo że są poświadczenia jego użycia w części z informacją etymologiczną hasła w tomach I i II.

Drugim rodzajem są niejasne informacje bibliograficzne. Oznaczenie **Ponęt. = Jan Ponętowski (Sejm lubel. 1569)** jest przykładem takiej niejasności, ponieważ tytuł w rozwinięciu jest niezgodny z oryginałem – jego oryginalna wersja brzmi: „Jana Ponętowskiego wierszowana relacja o unii lubelskiej i sejmie 1569 roku”.

W tekście pojawia się też czasem skrót **Koch.**, chociaż jako taki nie występuje w spisie – najprawdopodobniej chodzi o Jędrzeja Kochanowskiego (oznaczanego **Koch. J.**) lub Wespazjana Kochowskiego (oznaczanego **Koch. W.**), jednak jest to nierozstrzygalne na podstawie treści samego *SW*.

Pojawiają się również sytuacje, w których granica między podkategoriami źródła nie są jasne. W przypadku oznaczenia **Głb. = Hieronim Gołębiowski (autor „Obrazków rybackich”, r. 1910)** może chodzić tylko o wskazane dzieło, albo wszystkie dzieła tegoż autora, jeśli takowe istnieją – to jednak nie wynika z rozwinięcia w *Spisie*. Z kolei mimo że pod skróceniem **Pot. = Wacław Potocki (†1696)** rozumie się całokształt prac autora, to osobno wymienione jest jedno z jego dzieł: **Merk. = Wacława Potockiego (p. Pot.) „Mercurjusz Nowy”**.

Oznaczenia dzieł, analogicznie do podkategorii poświęconej autorom, wskazują, że informacja została zaczerpnięta z konkretnej publikacji, niekiedy wskazując przy tym autora w samym skróceniu. W kategorii tej pojawia się również kilka skrótów, w przypadku których nie udało się jednoznacznie ustalić, czy podany jest tytuł, czy opis dzieła – są to:

- KRyb. = Komedja rybałtowska z r. 1615. oraz Rybał. = Komedja rybałtowska, 1615. (jest to gatunek literatury sowizdrzalskiej); jednocześnie skrócenie ma tylko 1 poświadczenie w formie KRyb. (w tomie III) i 4 poświadczenia w formie Rybał. (w tomie IV).
- Sap. J. = Dziennik Jana Piotra Sapiehy, 1608–1611 (†1611).;
- Strzel. = Ustawa towarzystwa strzeleckiego krakowskiego (1564).;
- Suk. = Książka cechowa cechu tkacko-sukienniczego (Sławków) (?).;
- Szw. = Statut będziński cechu szewskiego (r. 1595).;
- Woł. = Ekstrakt majątności Wołożyna z r. 1776..

Poza tym dzieło zatytułowane *Zdanie wolnego szlachcica na sejmiki* (r. 1608). dwie różne formy skrócenia w poszczególnych tomach – Zd. Szl. i Zd. Sz. – z których żadne nie zostało poświadczone.

Oznaczenia gatunków obejmują tylko dwa skrócenia:

- Psł. = przysłowie.
- Pśń. = pieśń.

Podkategoria ta, obok niskiej liczebności, jest ciekawa jeszcze z dwóch innych powodów. Po pierwsze – skrócenia zapisane są wielką literą jak inne źródła, mimo że są obecne w części A *Spisu*. Po drugie – nie odnoszą się do nazw własnych, ale do przykładów użycia zaczerpniętych z tego, co można by było określić mianem anonimowych i integralnych elementów języka i kultury: nie można bowiem wskazać konkretnych autorów ani w przypadku przysłów (w przeciwieństwie do cytatów), ani pieśni ludowych (w przeciwieństwie do pieśni jako dzieł literackich).

Oznaczenia dat składają się zawsze ze skrócenia (r. dla roku i w. dla wieku) w połączeniu z wartością liczbową i w tekście przeważnie umieszczane są w nawiasie. W materiale badawczym obecne są trzy takie oznaczenia:

- †Fancz (...) Ani na F., ani na przyjaźń nic ś. nie oglądając... (z r. 1604). (...)
- Futryna (...) Nikt nie śmie w deszcz szufłować błota, gdyż ś. futryny zacierają (r. 1650). (...)
- Felpa (...) F. farbista, neapolitańska (w. XVII). (...)

Oznaczenia zagadek również odwołują się do anonimowych treści istniejących w świadomości użytkowników języka. Odwołania do nich są zawsze umieszczane w nawiasie okrągłym i przeważnie poprzedzone wyrażeniem „zagadka o”, np.:

- [Hneda] p. Wnet: Malutkie to, ćarniutkie to, H. cały świat obleci (zagadka o oku). (...)
- [Kokosiczek, czka, lm. czki] kogutek, kurek: Kokosiczka jak doniczka, K. tyćki (zagadka o kapuście).
- [Macu] wyraz bez znaczenia?: M. po gorącym placu (zagadka o piecu). <Może od Macać>

- [Podkrzesanek, nka, lm. nki] p. Podkasanek: Chudy pánek P. z lasa owce żenie (zagadka o grzebieniu). (...)
- [W nagu] p. Nago: W. po świecie chodzi (zagadka o serze).

W rozdziale 5.2.3.8 wymienione zostały jedyne poświadczenia, w przypadku których zamiast pełnego wyrazu „zagadka” zostało użyte jego skrócenie – w hasłach [Choda] i Nawrócić. Poza tym przed wyrazem tym może pojawić się przydawka, np.:

- [Blenk nieod.] p. ×Blank: Bene (= zewnqtrz) B., a westrzód sztek mięsa męk (= nadto, do tego) (kaszubska zagadka o pierścionku).

Znany jest też jeden przypadek, w którym informacja o zagadce jest formie zbliżonej do komentarza – na początku nawiasu pojawia się znak równości, jednak w odróżnieniu od komentarza jego treść nie jest zapisana kursywą:

- [Oddyrdać, a, ał] (...) Popiérdám, popiérdám, zaraz wam tu oddyrdam (= zagadka o maślnicy). (...)

Te dwie ostatnie kategorie – oznaczenia dat i zagadek – są kategoriami półotwartymi, dlatego też są wychwytywane przez parser dynamicznie.

5.2.3.10. Oznaczenia symboliczne

W tej kategorii jest tylko pięć elementów – są to wyłącznie symbole umieszczone na końcu części A *Spisu skrótów* i pełnią funkcję kwalifikatorów:

- † = staropolski.
- ! = wyraz, którego unikać należy.
- × = mało używany.
- * = domniemana postać wyrazu.
- [] = gwarowy.

Symbole te umieszczane są bezpośrednio przed modyfikowaną jednostką leksykalną (bez spacji), np.:

- †Falbuła, i, lm. e p. Falbana.
- !Falacja, i, lm. e oszukaństwo, fortel: (...)
- ×Falowanie, a, blm., czynność cz. ×Falować.
- Falisty (...) ×Splot F. (tętnicowy), in. wężykowy (plexus undulatus). (...)
- [Ficówka, i, lm. i] szkopek. <Zap. zam. *Świecówka = naczynie do robienia świec maczanych>

Podczas przetwarzania obrazu na tekst – co jest dokładniej opisane w rozdziale 9 – postanowiono jednak, aby oznaczenia te odseparować od większych segmentów, aby ułatwić programowi analizę składniową:

- Fajfura (...) 2. × zb. wyroby gliniane.
- Fakin, a, lm. owie 1. ! tragarz: (...)
- Falc, u, lm. e 1. p. Felc. 2. † p. Fałda. (...)

Należy tutaj zauważyć, że ostatnie z oznaczeń jest nietypowe, ponieważ nie poprzedza lemmy bądź segmentu, który modyfikuje, ale je okala:

- [Fejn] p. [Fajn].

- [Faborek, rka, lm. rki] p. Fabor.
- [Fajn nieod., Fejn, Fań, Faj] (...)
- Fach (...) 3. [F.] przeszło w budynku. (...)
- Faeton (...) Zdr. Faetonik, xFaetoncik, [Fajtonik]. (...)
- [Fijorówka, i, lm. i, Fijórówka] sukmana z wyszyciem z tasiemek kolorowych, [fijorów].
- [Fajer, u, blm.] (...) b) Fajeru dać a. zadać = *oporzędzić, dać pieprzu, [dychtu], [fałagi]*. (...)¹⁹

Wraz z powyższymi symbolami jest umieszczony także nawias trójkątny < >, ale nie jest to faktyczny kwalifikator, tylko element strukturalny, którego funkcja jest omówiona w **podrozdziale 5.2.3.13**.

5.2.3.11. Oznaczenia referencyjne

W tej kategorii znajdują się skrócenia, które pełnią funkcję strukturalną, tj. wskazują odniesienie do innej lemmy, przy czym część zawiera także dodatkową informację o relacji między danym artykule hasłowym a wskazywaną lemmą:

- an. = analogja, analogicznie, analogiczny.
- p. = patrz.
- przec. = pojęcie przeciwne.
- por. = porównaj.
- zam. = zamiast.
- Zdr. = zdrobniały.
- Zgr. = zgrubiały.

Pierwsze pięć skrótów jest obecnym w pierwszej części *Spisu skrótów*, pozostałe dwa natomiast pełnią funkcję strukturalną (wskazują segmenty poświęcone kolejno zdrobnieniom i skróceniom). Kategoria ta jednak nie jest wykorzystywana przez analizator składniowy – poszczególne rodzaje odsyłaczy są bowiem parsowane w różnych kontekstach funkcjonalnych i strukturalnych. Poza tym granica odsyłaczy bywa nieostra, np.:

- **Bunsen, a, lm. y w skróceniu zamiast ogniwo Bunsena: (...)**

W powyższym przykładzie trudno jest określić, czy „zamiast” pełni taką funkcję – zwłaszcza że formatowanie połączenia wyrazowego „ogniwo Bunsena” jest sformatowane jak treść definicji (kursywą), a nie jak pojęcie zagnieżdżone.

5.2.3.12. Oznaczenia derywacyjne

Nazwa tej kategorii jest pod pewnymi względami myląca, ponieważ obejmuje nie tylko derywację w znaczeniu słowotwórstwa, ale i jednostki przyłączające zapożyczenia. Kategoria ta jest zresztą nietypowa z innych powodów. Jednym z nich jest fakt, że użycie tego rodzaju oznaczeń ogranicza się do części hasła poświęconej etymologii. Innym jest fakt, że w części A *Spisu* obecne są tylko dwa skrócenia do niej należące.

Pierwsze z nich to an. = analogja, analogicznie, analogiczny. i ma pojedyncze poświadczenie, znajdujące się w tomie III:

¹⁹ Warto zwrócić tutaj uwagę na fakt, że forma „fałagi” nie występuje jako hasło w *SW* – w obu kopiach cyfrowych zdaje się występować wyraźnie litera Ł, a nie L. Istnieje jednak hasło „fałagi”.

- [Obąść się, obędzie ś., obył ś.] p. Obyć ś. <Zam. Obyć ś., an. do Obędzie według typu Przędzie, Prząść>

Z jednej strony można by było potraktować an. jako zwykłą formę skróconą wyrazu ze względu na pojedyncze wystąpienie, jednak wchodzi ono w skład elementu an. do wprowadzającego derywaty, który należy do szerszej grupy połączeń wyrazowych używanych przez analizator składniowy.

Drugie ze skrótów natomiast ma aż trzy warianty:

- sld., sł. lud. = słoworód ludowy (nowotwór, analogizm).
- sł. lud. = p. sld.
- sld. = słoworód.

Pierwsza z form, sld., jest najczęstsza i ma prawie 200 poświadczeń, sł. lud. ma ich około 50, natomiast ostatnia z form nie została poświadczona w żadnym tomie, co może wskazywać na omyłkowy zapis skrócenia.

Wariantywność tego skrócenia dotyczy jednak nie tylko zapisu, ale i kontekstu użycia, a dokładniej zróżnicowania elementów derywacyjnych, w skład których ono wchodzi. Bazując głównie na tomie I są to np.:

1. warianty sld. do oraz sł. lud. do:
 - Arkiposada (...) <Fr. (eau d')arquebusade dosł. = (woda) arkabuzowa, tj. dla arkabuzerów, sł. lud. do Arcy + Posada>
 - [Bzder] (...) <Por. Czes. na vzdor = na przekorę; zap. sld. do BZD>
 - Cybulizować się (...) <Żartobliwy sld. do Cybula>
 - [Dornica, y, lm. e, Dorna, Dwornica] izba. <Zap. Sł. *gornica, przez Nm. [durniz, dünnitz, dornitz i in.]; [dwornica] jest sld. do dwór>
 - xDyszkiepcja (...) <Przestawka wyrazu despekacja = despekt, sld. do Kiepc>
 - [Letki] (...) <Zam. Lekki, sł. lud. do LOT>
2. wariant sld. do imienia:
 - [Gregoraszek, szka, lm. szki, Gregoraczek] (...) <Dźwn. i sld. do imienia Grzegorz>
3. warianty ze sld. do oraz ze sł. lud. do:
 - Biedrzeniec (...) <Śr. łć. pi(m)pinella, Fr. pimprenelle, stąd Nm. Bibernel; przetłumaczono Nm. Biber = bóbr i zrobiono biebrzeniec, a potem, ze sld. do biodro, biedrzeniec>
 - Capagi (...) <Ros. sapogi, ze sld. do cap.>
 - Cięgiel (...) <Nm. Züngel = dosł. języczek, ze sld. do CIĘG>
 - Lichwa (...) <Z Got. leihvan, S. Gnm. lihan (Nm. leihen) = pożyczać, ze sł. lud. do Lichy>
4. warianty sld., zamiast oraz sld., zam.:
 - [Bładoróżga, i, lm. i, Latoróżga] latorośl, różga. <BLAD + Różga; może sld., zamiast latoróżga?>
 - Branica (...) <Może sld., zam. Dranica>
 - xChwaligóra (...) <Zap. sld., zamiast Chwalibur(a)>

5. wariant z ... sld.:
 - **Barwinek** (...) <Z łć. pervinea, przez Czes. barvínek, a to z Nm. sld. Bärwinkel>
6. wariant przez ... sld.:
 - **Bieldrzan** (...) <Śr. łć. Taleriana, przez. Nm. sld. Baldrian>
7. wariant umyślny sł. lud. do oraz umyślny sld. do:
 - ×**Aresztagrata** (...) <Umyślny sł. lud. do arystokrata>
 - [**Chłaścic**, a, lm. e] *szlachcic*. <Umyślny sld. do CHLAST>

Skrócenie **sld.** występuje niekiedy na pozycji finalnej, także w ramach wielowyrazowych, acz nieco trudniejszych do wyekstrahowania struktur:

- †**Archandurja** (...) <Tur. akyndży = żołnierz do rabunku, przez różne sł. lud.>
- [**Czy Żyw?!**] naśladowanie głosu wróbla. <Dźwn. ze sld.>
- **Gierlasz** (...) <Z Ros. jerałas = mieszanina; gra w karty, stanowiąca mieszaninę wista z preferansem (z Tatars.); Gierylas jest sld.>
- **Gnea** (...) <Niby łć. gnetum; gniot jest umyślny sld.>
- **Kłokotka** (...) <? KŁOK, sld.>

Co ciekawe, żadne z tych skrótów nie pojawia się w materiale badawczym, czyli hasłach rozpoczynających się od litery *F*. Jednak wymienione wyżej struktury wpisują się w dużo szerszą klasę **złożonych oznaczeń derywacyjnych**, cechującą się bardzo wysokim zróżnicowaniem. W przypadku materiału badawczego jest ich 39:

- | | |
|----------------------|--------------------------|
| ▪ +; | ▪ skrócenie z; |
| ▪ czyli; | ▪ skrócone z; |
| ▪ czynność cz.; | ▪ sld. do; |
| ▪ forma ż. od; | ▪ sld. do imienia; |
| ▪ im. od; | ▪ to zaś z; |
| ▪ lm. od; | ▪ toż, co; |
| ▪ od; | ▪ w rodzaju; |
| ▪ od wykrzyknika; | ▪ w związku z; |
| ▪ pod wpływem; | ▪ wywodzi z; |
| ▪ podług wymawiania; | ▪ z; |
| ▪ przeinaczone z; | ▪ z podprowadzeniem pod; |
| ▪ przekręcenie; | ▪ zam.; |
| ▪ przerobione z; | ▪ zamiast; |
| ▪ przerobiono z; | ▪ zdr. od; |
| ▪ przez; | ▪ ze; |
| ▪ przez usta; | ▪ ze zdr. imienia; |
| ▪ przez wymawianie; | ▪ zmieszanie; |
| ▪ przym. od; | ▪ zwanej po; |
| ▪ przys. od; | ▪ złącińszczenie. |
| ▪ rz. od; | |

Lista ta nie jest jednak wyczerpująca w skali całego słownika, ponieważ zbiór oznaczeń tego rodzaju ma charakter półotwarty, tzn. kolejne elementy były dodawane przez autorów *ad*

hoc. W samych zacytowanych wyżej przykładach pojawiają się takie elementy jak Przewaga wyrazu, przetłumaczono czy zrobiono.

Oznaczenia tego rodzaju są bardzo niejednolite i nie stanowią kategorii w z punktu widzenia języka naturalnego, jednak ich wyróżnienie jest przede wszystkim istotne dla analizatora składniowego, który wykorzystuje je do interpretacji struktury informacji etymologicznej.

5.2.3.13. Oznaczenia niesklasyfikowane

Ostatnia grupa oznaczeń zawiera te z nich, których nie udało się w sposób satysfakcjonujący przyporządkować do innych kategorii ze względu na ich mocno nietypową charakterystykę. Z tego powodu nie są traktowane przez parser jako kategoria, ale indywidualnie.

Pierwszym z oznaczeń < > = uwaga o pochodzeniu wyrazu. Choć umieszczone w spisie (obok kwalifikatorów symbolicznych), ma ono funkcję strukturalną, ograniczając segment poświęcony etymologii znakiem < z jego lewej strony, a znakiem > ze strony prawej.

Drugim oznaczeniem tego rodzaju jest Now. = w nowszych czasach dowolnie utworzony. Skrót ten służy uzupełnieniu opisu etymologicznego, przeważnie w połączeniu z językiem, np.:

- **Fantasmagorja** (...) <Now. z Gr. *fántasma* = widmo + *agoreúō* = mówię>
- **Fortepjan** (...) <Now. z Włos. *forte* = głośno + *piano* = cicho>
- **Fosforoskop** (...) <Now. z Fosfor + Gr. *skopós* = -widz>²⁰

Trzecim oznaczeniem natomiast jest wym. = wymawia się. Skrót ten poprzedza informację o wymowie wyrazu i wraz z nią jest umieszczane w nawiasie, np.:

- **Chicha** (...) <Hp. *chicha* (wym. *czycza*) = napój odurzający z rośliny tejże nazwy>
- **Donkiszot** (...) <Z imienia Hp. bohatera powieści Cervantesa *Don Quixote* (wym. *Kichot*)>
- **Dyska** (...) <Z Fr. *dix* (wym. *dis*) = dziesięć>

Oznaczenie to ma zaledwie kilka poświadczeń – wszędzie tam, gdzie funkcja nawiasu może nie być strukturalnie jednoznaczna. Znacznie częściej pozycja, użycie nawiasu oraz jego zawartość jednoznacznie wskazuje na informację o wymowie, zwykle w połączeniu z dywizem w funkcji separatora liter, które oznaczają głoski, które mają być czytane osobno:

- **Felchubel, bla**, lm. **ble** (**Felc-hubel**) *hebel do robienia felców*. <Nm. *Falzhobel*>
- **xFirzac, a a. u**, lm. **e** (**Fir-zac**) *blacha, stawiana przed otworem pieca hutnego*. <Nm. *Vorsatz*, daw. *Fürsatz*>
- **Fitobiologja, i**, blm. (**Fitobi-ologja**) *bot. biologja roślin, nauka o życiu roślin*. <Now. z Gr. *fyton* = roślina + *Biologja*>

W jednym przypadku skrót ten jest jednak redundantny:

- **Lwiogon, a**, lm. **y** (wym. **lwi-ogon**) *bot. 1. p. Lwian. 2. p. Serdecznik*.

²⁰ W oryginalnym tekście jest brakujące oznaczenia językowego Gr. – przykład jest tutaj przytaczany w formie poprawionej przez parser (opis ingerencji w oryginalną treść SW jest w **rozdziale 10.2.2.1**).

5.2.3.14. Skrócony zapis haseł

Oprócz oznaczeń o stałych rozwinięciach w *SW* stosuje się również skrócony zapis hasła w postaci pierwszej (wielkiej) litery wyrazu wraz z kropką (czasami także wraz ze skróconą partykułą zwrotną ś.). Skrócenie takie pojawia się jednak tylko w sytuacji, gdy jest zastępuje ono formę słownikową wyrazu z nagłówka artykułu hasłowego wewnątrz tego artykułu – dzieje się tak, ponieważ – w przeciwieństwie do pozostałych skrótów – nie ma ono stałego denotatu, jako że ten jest uwarunkowany kontekstowo.

Zapis ten wykorzystywany w kilku sytuacjach – przede wszystkim:

1. w przykładach użycia:
 - **Faktorować, uje, ował, Faktorzyć** (...) 2. za kim = *przemawiać, usprawiedliwiać, wstawiać ś., adwokatować: F.* za współuczniem nieukiem.
 - **Fernepiksum** nieod. ni. (...) 2. lek, lekarstwo: *Gotowa posiać jej jakiego F., i ona mnie przeżyje. Gliń.*
 - †**Folunk, u, lm. i p. Filung**: *Ma być (w stole pisarskim) F. we środku głęboki. Akta grodzkie. (...)*
2. w połączeniach wyrazowych:
 - **Fabrycznie 1.** przys. od **Fabryczny**: *Wyrabiać F. (...)*
 - **Filuterny, xFiluterski frantowski, chytry, przebiegły a na pozór ujmujący**: *Uśmiech F. (...)*
3. jako element składowy terminów wielowyrazowych:
 - **Fałd, u, lm. y 1.** (...) b) (...) **xF.** *postawowy = środek, założony w postawie sukna, sztaba, bruch. (...)*
 - **Flasza, y, lm. e** (...) 3. † (u złotników) **F.** *lewna = prasa do ściskania formy. Ślesz. (...)*
 - **Fugasowy** przym. od **Fugasa**: *Pocisk F. (= rozrywający).*
4. dla powtórzenia hasła z innym kwalifikatorem:
 - **Flankierować, uje, ował** (...) 3. [**F.**] = a) *watęsać ś., chodzić z kąta w kąt. (...)*
5. dla wskazania synonimicznej formy zwrotnej:
 - **xForcować, uje, ował i xF. ś. p. Forsować.**
6. dla zagnieżdżenia formy pochodnej (zwłaszcza zwrotnej) hasła:
 - **Formować, uje, ował** (...) **F. ś. 1.** *x kształcić ś., ćwiczyć ś., wyrabiać ś., układać ś., nabierać ogłady, poloru, ocierać ś., polerować ś.: (...)*

Taka forma skracania haseł do inicjału jest problematyczna, ponieważ wchodzi w interferencję ze skrótami źródłowymi składającymi się z pojedynczej litery, co zostało dokładnie opisane w rozdziale 5.5.2 Dezambiguacja skrótów.

5.3. Uzupelnienie źródeł

Podczas opracowywania materiału badawczego natrafiono w hasłach na informacje o 13 nowych źródłach, które nie widnieją w *Spisie skrótów* – są one przeważnie zapisywane w pełnej

formie bądź mieszając pełne i skrócone formy wyrazów. Poniżej wymienione są one w kolejności wystąpień w tekście²¹:

1. *Przywilej lubel. 1570*:
 - **Farba** (...) 3. † blichtr, pokrywka, pozór, płaszczyk: Cnocie farby obce szkodzą. Warg. Aby żadnych myt pod jakąkolwiek zasłoną a. farbą nie wybierali. **Przywilej lubel. 1570**. (...)
2. *Księgi kaliskie 1410*:
 - †**Fatać** (...) †**F. ś. chwytać ś.**: Jakom ś. nie fatał Wirzbięty za jego gardło, ani go siepał w jego domu gwałtem. **Księgi kaliskie 1410**. (...)
3. *Akta grodzkie (XV w.)*:
 - †**Fertryngielt** (...): Ojciec (gospodny) więcej towarzyszowi żadnemu borgować nie powinien, jonu szynk a F. **Akta grodzkie (XV w.)**. (...)
4. *Ust. cechu miechowników lubel. 1597*:
 - †**Fertryngielt, u, lm. y** rzem. p. **Fertryngielt**: **Ust. cechu miechowników lubel. 1597**.
5. *Akta grodzkie*:
 - †**Filetka** (...) Rura (rusznicy) w śliwowe drzewo osadzona, z filetkami i macią perłową. **Akta grodzkie**. (...)
6. *Gorzka wolność młodzieńska 1680*:
 - [**Flura, y, lm. y**] (...) 2. i † rozpustnica, nierządnicą, ładacznica: Jak F. chodzi. **Gorzka wolność młodzieńska 1680**. (...)
7. *Ustawa cechu miechowników lubel. 1613 r.*:
 - †**Forborg** (...): Trzy miechy nasadzono w jednej sztuce, ukrojone pod forborgiem. **Ustawa cechu miechowników lubel. 1613 r.** (...)
8. *Roty przyś. Poznań. r. 1388*:
 - †**Fordrować** (...): Gdy wójt, fordruje sirotną; rzecz, może li z nią czynić jako iściec. Ort. Jako Jan nie fordrował Krzywosanda, kiedy wziął pod Wiencencem konie. **Roty przyś. Poznań. r. 1388**. (...)
9. *Djarjusz prawdziwej relacji z 1683 r.*:
 - †**Fortrop** (...) Z berdyszami od fortropów na janczarów wpadali. W tym czasie też fortropy z regimentów i kozacy z gór wystrzelali. **Djarjusz prawdziwej relacji z 1683 r.** (...)
10. *Akta grodzkie r. 1399*:
 - **Forwark** (...) 2. † p. **Folwark**: Jako Wociech miał wrócić F. Kuszowi, kiedyby u niego Wir wykupił. **Akta grodzkie r. 1399**. (...)
11. *Ps. 1535*:
 - **Fundament** (...) †Fundamentowie a. założenia. **Ps. 1535**. (...)
12. *Taksa 1573*:
 - †**Furmany** p. **Furmański**: Do wozu skarbowego lub furmanego. **Taksa 1573**.

²¹ Źródła te są wypisane również w kolejności alfabetycznej w **Aneksie M**.

13. *Ust. cechu haftarskiego lubel. 1669:*

- †**Frysztuk** (...): Brat młodszy świeżo cech przymujący ma mieć F. a. czas pewny niech będzie wolny od takich posług. **Ust. cechu haftarskiego lubel. 1669.** (...)

Co więcej, podczas tworzenia reguł dla oznaczeń przyliczbowych (**podrozdział 5.2.3.7.**) oraz dat w oznaczeniach źródłowych (**podrozdział 5.2.3.9.**) przeprowadzono eksperyment, którego celem była weryfikacja, czy zaprogramowane w analizatorze składniowym reguły opracowane na podstawie materiału badawczego wystarczą do odpowiedniego sparsowania podobnych przypadków w hasłach spoza badanego zakresu.

Eksperyment ten przeprowadzony był na całej treści tomu I (w formie brudnego OCR²²). Do wyszukania tego rodzaju skrótów wykorzystano wyrażenia regularne zaczerpnięte z parsera lub zbliżone do nich. Wynik eksperymentu był pomyślny, ponieważ pomyślnie rozpoznano przykłady takie, jak poniżej:

- †**Głōza**, y, lm. y = **Głōsa**. 1541 r.
- †**Gniży**, [**Gniży**] p. **Zgniży**: †Mięso gniłe = *zgorzele, gangrena*. 1541 r.
- **Galanterja** (...) G. stołowa z przystawkami trzema złocistemi (r. 1667). (...)
- **Goży** (...) Rozjemy goże, a nie zapisane (r. 1531). (...)

Jednak jego interesującym skutkiem ubocznym było zidentyfikowanie aż 18 nowych źródeł, których forma w tekście zawiera datę, a które nie figurują w *Spisie*:

1. *Rota krak. 1400:*
 - **Dzierzenie** (...) Był w dzierzeniu piątej części karczmy 15 lat. **Rota krak. 1400** (= w posiadaniu).
2. *Dypl. Lub. r. 1647:*
 - †**Garbuzowy** przym. od 1. **Garbuz**: G. rów idący od garbuzu. **Dypl. Lub. r. 1647.**
3. *Rękopis z 1734 r.:*
 - ×**Gardebur**, a, lm. **owie** p. **Gardekur**: Jeden z gardeburów, zakochawszy ś. w Henrykcie, myślał ś. z nią żenić. **Rękopis z 1734 r.**
4. *Taksa dla kuśnierzy r. 1573.:*
 - †**Garła**, i, lm. **e** *jakieś ubranie*: Szlamowego z garli, aby żaden nie śmiał czynić błanu, jedno z dwudziestu sześciu błanu grzbietowego. **Taksa dla kuśnierzy r. 1573.** <GAR(DŁ)>
5. *Ks. Jabczyński. Rys hist. m. Dolska 184 akt 1664 r.:*
 - †**Gilnik**, a, lm. **i** p. **Gielnik**: Prasołowie, gdy miewają sól twardą, powinni dawać na każdy targ a. jarmark po 4 gilniki soli a. po mierze. **Ks. Jabczyński. Rys hist. m. Dolska 184 akt 1664 r.** <p. Glon>
6. *Ust. cechu Lub. 1618 r.:*
 - **Glejtować**, uje, **ował** kogo (...) 2. † a. z kim = *odprowadzać kogo ceremonjalnie, dla honoru*: Ma zaraz z nim G. do mistrza. **Ust. cechu Lub. 1618 r.**

²² Objasnienie problematyki brudnego OCR znajduje się w **rozdziale 9**.

7. *Ust. gosp. stol. Lub. 1671 r. oraz*
8. *Ust. mieczn. Lubel. 1618:*
 - **xGlejtowny** (...) Magister powinien wieczerzą, sprawić glejtowną. *Ust. gosp. stol. Lub. 1671 r.* Aby na glejtowną, wieczerza, nie zamieszkiwali. *Ust. mieczn. Lubel. 1618* (= na wieczerzę, którą wyprawiają dla wędrownych towarzyszków). (...)
9. *Ustawa 1560 r.:*
 - **Głownik**, a. lm. **cy 1.** † p. **Głownik**: O mężobojcach a. głownikach. *Ustawa 1560 r.*
10. *Lustracje z r. 1602:*
 - **†Gnojowe** (...) Mieszczanie brali w jarmark od kramarskich wozów miastu nieprzyległych gnojowego per gr. 1. *Lustracje z r. 1602.*
11. *Stat. Wawrz. z Prażmowa 1531 r.:*
 - **Gody** (...) Ona posłowica od gód do gód ma być rozumiana od nowego lata. *Stat. Wawrz. z Prażmowa 1531 r.* (...)
12. *Testament Tob. Morsztyna z 1664 r.:*
 - **†Gont**, u, lm. **y miara długości**: Frandzli jedwabnej żółtej z czerwonyra dwa gonty. *Testament Tob. Morsztyna z 1664 r.*
 - **Grzebieniarz** (...) G. złotogłowy bogaty z herbami. *Testament Tob. Morsztyna z 1664 r.*
13. *Relacja Aleksandra Śliźnia z r. 1620:*
 - **†Gończyk**, a, lm. **i p. Goniec**: W Możajsku drugie półtory niedziele, dokąd póki ś. G. ich z Warszawy nie wrócił. *Relacja Aleksandra Śliźnia z r. 1620.*
14. *Osądzenie (= kalendarz z r. 1528):*
 - **†Gradny** p. **Gradowy**: G. czas. *Osądzenie (= kalendarz z r. 1528).*
15. *Akta wikar. Włocławskich 1688 r.:*
 - **Gratyny**, yn, blp. **1.** × część promu, bal służący za pomost do promu: Zrobić prom i G. do niego. *Akta wikar. Włocławskich 1688 r.* (...)
16. *Księgi peregrynackie, rękopis 1584 r.:*
 - **xGrobgrynowy** przym. od **Grobgryn**, p. **Grobrynowy**: Grobgrynowe szaty. *Księgi peregrynackie, rękopis 1584 r.*
17. *List Garwaskiego do Jana Zamojskiego r. 1602:*
 - **Gromnica** (...) Przybył do mnie w środę przed Gromnicą. *List Garwaskiego do Jana Zamojskiego r. 1602.* (...)
18. *Cennik dla rzemieślników miasta Książa 1538 r.:*
 - **†Gryfnik**, a, lm. **cy wędliniarz**, in. † **sadelnik**. *Cennik dla rzemieślników miasta Książa 1538 r.*
19. *List Jana Sobieskiego z r. 1671.:*
 - **Gienerał-kwatermistrz**, a, lm. **owie kwatermistrz gieneralny**, **główny**. *List Jana Sobieskiego z r. 1671.* (...)

To ostatnie źródło jest bardzo podobne do innego, które jest obecne w *Spisie skrótów*: Drsz. = List Doroszenki do Jana Sobieskiego, r. 1671, jednak według piszącego te

słowa są to raczej dwa odrębne teksty. Inaczej jest w przypadku dzieła, które występuje w tekście w pełnej formie: *Catalogus Schnebergera z r. 1557.*, ale figuruje także w *Spisie z nieznacznie innym zapisem: Cat. = „Catalogus” Sznebergera z r. 1557:*

- †*Dziembrenożki, ów, blp. roś. bodziszek czerwony, in. dziebrenosek. Catalogus Schnebergera z r. 1557.* <?>²³

Wszystkie te oznaczenia źródeł zdają się być przeoczone w procesie kompilacji, na co wskazuje nie tylko ich pełna lub półpełna forma w tekście, ale także fakt, że ograniczają się niemal zupełnie wyłącznie do haseł zaczynających się od litery **G**, oraz kilka dalszych związanych z nimi nieścisłości – nie jest np. jasne, czy:

1. czy *Testament Tob. Morsztyna z 1664 r.* wskazuje na tego samego autora co obecne w *Spisie* skrócenie *Mor. Tob. = Tobjasz Morsztyn (r. 1664)*;
2. czy *Taksa 1573* i *Taksa dla kuśnierzy r. 1573* wskazują na to samo źródło;
3. czy źródło *Akta grodzkie* wskazuje na to samo, co *Akta grodzkie (XV w.)* lub *Akta grodzkie 1399*, czy są to trzy niezależne źródła;
4. jak rozwinąć skrócenia takie jak *Dypl. Lub. r. 1647* czy *Ust. gosp. stol. Lub. 1671 r.*.

Autor tych słów pragnie w tym miejscu podkreślić, że wyszukiwanie było przeprowadzone na niesprawdzonym rezultacie rozpoznawania tekstu, co mogło efektywnie umożliwić znalezienie wszystkich źródeł zawierających tekst *r.*. Poza tym źródła takie najprawdopodobniej stanowią tylko podzbiór wszystkich dzieł, które zostały podane w tekście w pełnej formie – nie udało się jednak opracować metody ich skutecznego wyszukiwania, nawet na sprawdzonych i poprawionych danych tekstowych.

5.4. Reguły tworzenia skrótów

O ile część skrótów *SW* jest w powszechnym użyciu lub ich formacja nie budzi zastrzeżeń, to stosowana metodologia ich tworzenia nie zawsze jest konsekwentna. Niekonsekwencja ta przejawia się na dwóch poziomach: systemowym – poprzez wykorzystanie konkurencyjnych, a nawet sprzecznych ze sobą strategii formowania skrótów, oraz indywidualnym – poprzez arbitralne decyzje w przypadku poszczególnych skrótów.

5.4.1. Skrócenia ortograficzne i ortofonetyczne

Najliczniejszą grupą, która przejawia brak spójności metodologicznej w zapisie, są wyrazy, których skrócenia kończą się na spółgłoskę zmiękczoneą przy użyciu litery „i”. Autorzy stosowali tutaj dwie strategie skracania.

Pierwsza strategia – ortograficzna – pozostawia oryginalny zapis spółgłoski, pomijając przy tym zmiękczone „i”:

- *Brodz.* = *Brodziński.* [dzi → dz]
- *Choc.* = *Józef Chociszewski.* [ci → c]
- *Czern.* = *Czernicki.* [ni → n]

²³ Pełna forma źródła ma też drugie poświadczenie – w tomie II, w haśle †*Kaza*.

- **Nies.** = Kacper Niesiecki. [si → s]
- **Pien.** = Krzysztof Pieniążek. [ni → n]
- **Radz.** = Michał Radziwiłł. [zi → z]
- **Ron.** = Bohdan Ronikier. [ni → n]
- **Waśn.** = Józef Waśniewski. [ni → n]
- **karc.** = karciarski. [ci → c]
- **kości.** = kościelny, teologiczny. [ci → c]
- **późn.** = późniejszy. [ni → n]
- **właśc.** = w znaczeniu właściwym; właściwie. [ci → c]

Nieco częściej stosowana jest jednak druga strategia – fonetyczna – która zmienia zapis ostatniej samogłoski w skrócie na jej formę diakrytyczną, zachowując przy tym jej wymowę:

- **Bliż.** = Józef Bliziński. [zi → ź]
- **Chęć.** = Jan Chęciński. [ci → ć]
- **Gąś.** = Wacław Gąsiorowski. [si → ś]
- **Kości.** = Tadeusz Kościuszko. [ci → ć]
- **Rogoż.** = Stefan Szolc Rogoziński. [zi → ź]
- **koń.** = koniarski. [ni → ń]
- **łoż.** = Walery Łoziński. [zi → ź]
- **łźń.** = Władysław Łoziński. [zi → ź]
- **ś.** = się. [si → ś]
- **Żłń.** = Józef Zieliński. [zi → ź]

Na uwagę zasługuje tutaj skrócenie nazwy autora (jego imienia, nazwiska lub pseudonimu), które posiada dwa kontrastujące ze sobą warianty, utworzone przy pomocy różnych strategii zapisu miękkości:

- **Jan Kaz. a. J. Każ.** = Jan Kazimierz, król. [zi → z/ź]

Obecny jest również pojedynczy przypadek skrócenia, które zachowało finalną samogłoskę zmiękczonej przy użyciu litery *i*:

- **Prusi.** = Jan Prusinowski

Zapis fonetyczny dotyczy również kilku innych skrótów, w przypadku których ortografia samogłosek nie odpowiada ich fonetyce:

- **Balc.** = Oswald Balzer. [z → c]
- **Klej.** = Juljusz Kleiner. [i → j]
- **Tretj.** = Józef Tretiak †1923. [i → j]

5.4.2. Dyferencjacja nazw autorów

Biorąc pod uwagę liczbę autorów, których twórczość stanowiła podstawę dla opracowania *SW*, nie powinno dziwić, iż niektórzy z nich dzielili to samo nazwisko. Kompilatorzy słownika zastosowali kilka strategii do odróżnienia od siebie poszczególnych autorów.

Pierwszą strategią jest użycie skrócenia samego nazwiska dla pierwszego z autorów i dodanie inicjałów dla kolejnych, np.:

- Bud. = Szymon Budny.
- Bud. B. = Bieniasz Budny.
- Chodk. = Karol Chodkiewicz.
- Chodk. A. = Aleksander Chodkiewicz.
- Chodk. K. = Krzysztof Chodkiewicz.
- Grab. = Ambroży Grabowski.
- Grab. B. = Bronisław Grabowski.
- Grab. P. = Piotr Grabowski.
- Hube = Michał Jan Hube.
- Hube R. = Romuald Hube.
- Koż. = Kajetan Koźmian.
- Koż. A. = Andrzej Edward Koźmian.
- Niem. = Juljan Niemcewicz.
- Niem. A. = Andrzej Niemojowski.
- Ryb. = Maciej Rybiński
- Ryb. J. = Jan Rybiński
- Sur. = Wawrzyniec Surowiecki.
- Sur. K. = Karol Surowiecki.
- Wys. = Szymon Wysocki
- Wys. W. = Włodzimierz Wysocki
- Zbyl. = Piotr Zbylitowski.
- Zbyl. A. = Andrzej Zbylitowski.

Również niektóre podobne do siebie – ale nie identyczne – nazwiska zostały skrócone do tej samej formy, co – jak w przypadku metody pierwszej – wymagało dodania inicjałów lub formy skróconej imienia:

- Biel. = Tomasz Bielawski
- Biel. J. = Joachim Bielski.
- Biel. M. = Marcin Bielski.
- Biel. P. = Piotr Bieliński.
- Szym. = Józef Szymanowski.
- Szym. A. = Adam Szynański.
- Szym. W. = Wacław Szymanowski.
- Boh. = Franciszek Bohomolec.
- Boh. J. = Jan Bohomolec.
- Boh. K. = Ksawery Bohusz.
- Boh. M. = Michał Bohusz.
- Mor. = Hieronim Morsztyn.
- Mor. A. = Andrzej Morsztyn.
- Mor. K. = Kazimierz Morawski.
- Mor. T. = Teodor Morawski.
- Mor. Tob. = Tobjasz Morsztyn.

Skracanie różnych nazwisk do tej samej formy może dziwić zważywszy na to, że autorzy niekiedy stosowali kolejną metodę, polegającą na skracaniu tego samego nazwiska do odmiennych form:

- DyB. = Benedykt Dybowski.
- Dyb. = Władysław Dybowski.
- Garcz. = Stefan Garczyński (pisarz XIX wieku).
- Garczyń. = Stefan Garczyński (pisarz XVIII wieku).
- Gill. = Agaton Giller.
- Gil. = Paweł Gilowski.
- LeW. = Włodzimierz Lewicki..
- Lew. = Fryderyk Henryk Lewestam †1879.
- Kotb. = Edward Kotłubaj.
- Kotł. = Henryk Kotłubaj.
- Rdz. = W. Eljasz-Radzikowski.
- Rc. = Radziwiłł Czarny.

Napotkać można również przypadki mieszane, w której połączone są dwie poprzednie metody, czyli jednocześnie to samo nazwisko ma różne formy skrócenia dla różnych autorów, i różne nazwiska dzielą to samo skrócenie, odróżnione przy pomocy inicjałów:

- Czer. = Rafał Józef Czerwiakowski †1846.
- Czerm. = Wiktor Czerniak †1913.
- Czern. = Gustaw Czernicki †1911.
- Czer. S. = Stanisław Czerski w. XIX.
- Kras. = Ignacy Krasicki.
- Kras. A. = Adam Stanisław Krasiński.
- Krasiń. = Zygmunt Krasiński.
- Ostror. = Jan Ostroróg (Myśliwstwo z ogary).
- Ostr. S. = Stanisław Ostrowski.
- Wys. = Szymon Wysocki.
- Wysp. = Stanisław Wyspiański.
- Wys. W. = Włodzimierz Wysocki.

5.4.3. Pozostałe osobliwości

Oprócz opisanych wcześniej problemów systemowych przy tworzeniu skrótów jest też kilka nieścisłości na mniejszą skalę, które zostały opisane poniżej.

5.4.3.1. Obce znaki diakrytyczne

Jedną z najbardziej podstawowych zasad skracania, według której wybór pozostawionych liter jest w dużym stopniu dowolny, to ich kolejność musi być zachowana. W *Spisie* zawarte są jednak dwa skrócenia autorów, których nazwiska zawierają obce znaki diakrytyczne unaoczniające złamanie tej zasady:

- Mát. = Karol Matyás.
- Mey. = Leopold Méyet.

5.4.3.2. Zapis inicjałów

W przypadku kilku skrótów inicjały imion i nazwisk autorów pisane są bez spacji i z jedną kropką na końcu:

- AB. = August Bielowski †1876.
- JK. = Jan z Koszyczek w. XVI.
- KK. = Karol Kaczkowski †1867.
- KP. = K. Przyborowski †1879
- LJ. = Józef Lompa †1863.
- SS. = Stanisław ze Szczodrkowic w. XVI.

Jednak spotkać też można inicjały oddzielone spacją, z kropkami umieszczonymi po każdej literze:

- W. A. = Antoni Waga †1890.
- T. K. = Tadeusz Korzon †1918.
- Z. A. = Zygmunt August †1572.

Ten sam problem dotyczy również niektórych nazw źródeł:

- HJ. = Historja o św. Józefie, r. 1530.
- HR. = Hube: Roty przysięg z w. XIV i XV
- Sł. = Statut łaskiego, r. 1541.
- K. G. = Kancjonał Gdański, r. 1628.
- S. K. = Słownik łacińsko-polski wyrazów lekarskich (...)

5.4.3.3. Arbitralność typograficzna

Na koniec warto zwrócić uwagę na przypadkowość zapisu niektórych oznaczeń. O ile większość z nich wpisuje się w format, zgodnie z którym każdy z jego skróconych członów rozpoczyna się wielką lub małą literą (zależnie od zapisu oryginalnego wyrazu), po której mogą być tylko małe litery oraz kropka, to są dwa rodzaje odstępstw od tej reguły.

Pierwszym jest stosowanie wielkich liter do oznaczenia fragmentu kolejnych wyrazów wchodzących w skład skrócenia:

- BdC. = Jan Baudouin de Courtenay.
- DyB. = Benedykt Dybowski.
- IGost. = Inwentarz starostwa-gostyńskiego, r. 1726.
- JGr. = Podkanie Jonasa z Gregorjasem, r. 1598.
- LeW. = Włodzimierz Lewicki.
- PmK. = Prawa i przywileje miasta Krakowa, r. 1560.

Drugim odstępstwem natomiast jest stosowanie dywizu zamiast spacji:

- Or-Ot = Artur Oppman.

Nie należy tej formy zapisu mylić z poniższą, która składa się z inicjałów dwuczłonowego nazwiska autora:

- P.-N. = Stanisław Piołun-Noyszewski.

5.5. Standaryzacja zapisu

Standaryzacja *Spisu skrótów*, której się podjęto, nie ma na celu naprawienia problemów metodologicznych i kompilatorskich autorów *SW*, ale jego uzgodnienie względem treści słownika oraz ujednoczenia typografii – zarówno samych skrótów, jak i ich rozwinięć – tak, aby analiza

cyfrowej wersji tekstu nie sprawiała problemu ani użytkownikowi ludzkiemu, ani rozwiązaniom programistycznym (w tym opisywanemu dalej analizatorowi składniowemu).

W sumie standaryzacja objęła cztery kategorie zmian:

1. Przywrócenie zgodności między *Spisem skrótów* a artykułami hasłowymi;
2. Ujednoznacznienie formy skrótów w przypadkach ambiwalentnych;
3. Wprowadzenie jednego standardu oznaczania wariantów skrótów;
4. Ujednolicenie zapisu i typografii w rozwinięciach.

Każdej z tych kategorii poświęcono osobny rozdział.

5.5.1. Uzgodnienie treści *Spisu* i haseł

Niewielka część skrótów i rozwinięć w *SW* ma niejednorodną, a niekiedy wewnętrznie sprzeczną, treść, co zachodzi:

1. między *Spisami* poszczególnych tomów;
2. między *Spisami* a treścią artykułów hasłowych;
3. między informacjami podanymi w *Spisach* a źródłami zewnętrznymi.

Poniżej przedstawiono kilka przykładów najlepiej ilustrujących tego typu problemy:

1. W spisach tomów V i VI występuje skrót *Alm.* = *tłómacz książki: Albertus Magnus, o sekretach białogłowskich – Amsterdam, 1698.*, która ma co najmniej jedno poświadczenie w tomie V, natomiast w spisach tomów kolejnych występuje forma *Am.*, która nie ma żadnych znanych poświadczeń.
2. W spisach tomów I–VI figuruje wpis *Gór.* = *Adam Górczyński.*, którego rozwinięcie w tomach jest skorygowane na *Gór.* = *Adam Gorczyński (błędnie: Górczyński).* Skrót został jednak zmieniony na *Gorczyń.* = *Adam Gorczyński (†1876).* ze względu na identyczny skrót *gór.*, które występuje w tekście częściej o ponad rząd wielkości oraz fakt, że uzgodnienie skróta z rozwinięciem do formy *Gor.* powodowałoby z kolei konflikt z *Gor.* = *Antoni Gorecki (†1861).*
3. W tomach I–III słownika figuruje forma *Mac. z Róż.* = *Maciej z Różana (w. XV).*, natomiast w tomach dalszych *Mac. z Roż.* = *Maciej z Rożana w. XV.* Na podstawie źródeł zewnętrznych ustalono, że formą poprawną jest jednak ta pierwsza, i taką przyjętą w ostatecznej formie *Spisu skrótów.*
4. Skrót *Jbn.* ma rozwinięcie *Anna Jabłonowska (†1802).* we wszystkich tomach, w których występuje, ale w *Dalszych skrótach* w tomie VIII ma formę *Anna Jabłowowska*; poprawną formę nazwiska autorki ustalono na podstawie źródeł zewnętrznych.
5. W tomach III–VII skrót *Krm.* ma rozwinięcie *Olbracht Karmanowski*, a w tomie VIII *Olbrycht Karmanowski.* Na podstawie źródeł zewnętrznych ustalono, że to ta ostatnia forma jest poprawna.
6. Dodane w tomie VIII skrót *Śwż.* ma rozwinięcie *Ernest Świeżawski †1919* w *Spisie skrótów*, a *Ernest Świeżowski* w *Dalszych skrótach*;

na podstawie źródeł zewnętrznych ustalono, że to Świeżawski jest poprawną formą nazwiska.

Nie uzgadniano jednak wariantów skrócenia do jednej formy – zarówno formy Feliń. = Zygmunt Feliński. jak i Fel. S. = p. Feliń. – pierwsze z nich ma 7 poświadczeń w tekście słownika, podczas gdy to drugie ma ich 5.

5.5.2. Dezambiguacja skrótów

W *Spisie* obecne są dwa skrócenia, których format to pojedyncza wielka litera oraz kropka:

- L. = Słownik Lindego, 1807–1814.
- W. = Jakób Waga †1872.

Format ten jest identyczny z zapisem skróconej wersji lemmy w przykładach, np.:

- **Lagrowanie**, a, blm., czynność cz. **Lagrować**: L. kamieni na pół czysto. L. proste, ukośne. L. piwa.
- **Wywar**, u, lm. y (...) 3. a **Wywarzyny** to, co pozostało po wygotowaniu: Wywary w gorzelni, resztki zboża, ziemniaków, pozostałe po oddestylowaniu spirytusu. Wywary z gorzelni ryną sprowadzać trzeba do wołowni. Świt. W. gorzalczany. Odesłano woły na W. Torz. (= aby ś. pasty wywarami, na karmienie, na wypas, na opas). Żart.: [Jeżeli dziewczyna brzydka, to zaglądnij jej w oczy, mówią, że brzydka i stara, i wysadzają za drzwi na W.]. 4. [W.] dół zrobiony przez wodę: Woda narobiła w polu wywary.

Zbieżność zapisu w tym przypadku jest wyjątkowo problematyczna dla parsera, ponieważ niekiedy nie da się określić funkcji takiego skrócenia bez analizy treści. Mimo iż problem ten nie dotyczy litery F, to postawiono rozwiązać ten problem przyszłościowo i zmodyfikować zapis obu skrótów zarówno w tekście²⁴, jak i *Spisie skrótów* w taki sposób, aby nie pozostawiał on żadnych wątpliwości:

- Sł. L. = Słownik Lindego, 1807–1814.
- W. J. = Jakób Waga †1872.

5.5.3. Warianty skrótów

W treści *SW* do niektórych źródeł odwołuje się więcej niż jedno oznaczenie – w takich przypadkach jeden z wpisów traktowany jest jako główny i ma także podany wariant, natomiast ten drugi ma w miejscu rozwinięcia odsyłacz do wpisu głównego, np.:

- Ext. a. Godl. = Exterus (t. j. Ludwika Godlewska) †1901
- Godl. = p. Ext.

Warianty te nie zawsze jednak są podawane w sposób jednolity – w *Spisie skrótów* zastosowano aż cztery różne metody ich podawania:

1. Najczęstszym sposobem jest użycie skrótu „a.” zastępującego „albo”, np.:
 - Ant. z Przem. a. Przem. = Antonin z Przemyśla.

²⁴ Proces opisany jest w **rozdziale 10.2.2.1.**

- Blw. a. AB. = August Bielowski †1876.
 - Cyw. a. Glc. = Ustawy cywilne dla Galicji Zachodniej, r. 1797
 - Ext. a. Godl. = Exterus (t. j. Ludwika Godlewska) †1901.
2. W dwóch przypadkach wariant wraz ze skrótem go przyłączającym został umieszczony w nawiasie:
- AB. (a. Blw.) = August Bielowski †1876.
 - Ur. (a. Psł.) = Urywek protokołów sądu radzieckiego sławkowskiego z lat 1713–1723
3. Jest też pojedynczy przypadek, w którym użyto spójnika „albo” w pełnej formie:
- Gad. albo Gadom. = Jan Gadomski.
4. Dwa skrócenia miały także umieszczony wariant po przecinku:
- Grot., Grt. = Jan Grotkowski, w. XVII.²⁵
 - sld., sł. lud. = słoworód ludowy (nowotwór, analogizm).

W cyfrowej wersji *Spisu skrótów* ujednociono i uproszczono zapis w taki sposób, że wszystkie warianty oddzielone ukośnikami, jak widać poniżej:

- Ant. z Przem. / Przem.
- AB. / Blw.
- Gad. / Gadom.
- sld. / sł. lud.

5.5.4. Zapis treści rozwinięć

Ostatnią formą ingerencji w *Spis skrótów* jest ustandaryzowanie typografii i oznaczeń użytych w rozwinięciach. Zmiany te mają w głównej mierze charakter kosmetyczny, a tylko niekiedy ujednoznacziający, i obejmują:

1. Przeniesienie autorów przed tytuł dzieła i oddzielenie go myślnikiem (w oznaczeniach źródeł);
2. Umieszczenie tytułów dzieł w cudzysłowie;
3. Umieszczenie wszystkich dodatkowych informacji w nawiasach – w tym dat, miejsca wydania i komentarzy;
4. Usunięcie kropek na końcu objaśnień;
5. Zastąpienie wszystkich odsyłaczy p. strzałką: →;
6. Umieszczenie w nawiasie i po znaku równości prawdziwe nazwiska autorów obok pseudonimów (jeśli oznaczenie oparte jest o pseudonim);
7. Umieszczenie w nawiasie i po słowie „pseudonim” i dwukropku pseudonimów autorów (jeśli oznaczenie jest oparte o faktyczne nazwisko autora).

Powyższe zmiany ilustrują poniższe pary wpisów w *Spisie skrótów*, z których pierwszy reprezentuje formę oryginalną, a drugi formę po modyfikacji:

1. Alm. = tłumacz książki: Albertus Magnus: „0 sekretach białogłowskich” – Amsterdam, 1698.
Alm. = tłumacz książki: Albertus Magnus – „0 sekretach białogłowskich” (Amsterdam, 1698)

²⁵ W tej formie skrócenie figuruje w tomach od V do VII; w tomie VIII zmieniono zapis na Grot. a. Grt..

2. Alb. = Albertus z wojny” r. 1596.
Alb. = „Albertus z wojny” (r. 1596)
3. Chdk. = Zorjan Dołęga Chodakowski (czyli Adam Czarnocki) †1825.
Chdk. = Zorjan Dołęga Chodakowski (= Adam Czarnocki) (†1825)
4. Chrzasz. = Pamiętniki Chrzaszczewskiego, w. XVIII.
Chrzasz. = Pamiętniki Chrzaszczewskiego (w. XVIII)
5. Czn. = Compendium ferculorum, Stanisława Czernieckiego, Kraków, r. 1682.
Czn. = Stanisław Czerniecki - „Compendium ferculorum” (Kraków, r. 1682)
6. Drz. = Józef Drzewiecki †1857.
Drz. = Józef Drzewiecki (†1857) - „Pamiętniki”
7. Erl. = Zabytek jęz. pols. z początku w. XVI z rękop. bibliot. uniw. w Erlangen (wyd. L. Malinowski).
8. Erl. = „Zabytek języka polskiego z początku wieku XVI. Z rękopisu Biblioteki Uniwersytetu w Erlangen” (wyd. L. Malinowski)
9. Eww. = Ewweli-Bej (pseudonim) w. XIX.
Eww. = Ewweli-Bej (pseudonim) (w. XIX)
10. Fors. = Choroby drzew Forsytha (przekład) Warszawa r. 1791.
Fors. = „Choroby drzew Forsytha” (przekład - Warszawa, r. 1791)
11. Hon. = Honter: De grammatica libri II, w. XVI.
Hon. = Honter - „De grammatica libri II” (w. XVI)
12. H. Tr. = Historia Trojańska (z r. 1563).
H. Tr. = „Historja Trojańska” (r. 1563)
13. Karcz. = Wacław Karczewski (pseudonim: Marjan Jasieńczyk) †1911.
14. Karcz. = Wacław Karczewski (pseudonim: Marjan Jasieńczyk) (†1911)
15. Łęt. = Julian Łętowski (pseudonim: Władysław Książek) †1897.
Łęt. = Julian Łętowski (pseudonim: Władysław Książek) (†1897)
16. Mik. = Kacper Mikosza ? w. XVIII (u Lindego).
Mik. = Kacper Mikosza (? w. XVIII) (u Lindego)
17. Mir. a. Psm. = Mirjam (t. j. Zenon Przesmycki).
Mir. / Psm. = Mirjam (= Zenon Przesmycki)
18. Ms. = Słownik anatomiczno-fizjologiczny J. Majera i F. Skobla (Kraków, 1838).
Ms. = J. Majer i F. Skobel - „Słownik anatomiczno-fizjologiczny” (Kraków, 1838)
19. P. of. = „Pamiętniki oficera polskiego” Henryka Brandta †1868, przekład polski.
P. of. = Henryk Brandt (†1868) - „Pamiętniki oficera polskiego” (przekład polski)
20. Ponęt. = Jan Ponętowski (Sejm lubel. 1569).
Ponęt. = Jan Ponętowski - „Sejm lubel. 1569”
21. Prus = Bolesław Prus (t. j. Aleksander Głowacki) †1912.
Prus = Bolesław Prus (= Aleksander Głowacki) (†1912)
22. Roż. = Rożańskiego „Sztuka babienia” - Warszawa, r. 1786.
Roż. = Rożański - „Sztuka babienia” (Warszawa, r. 1786)

23. Sap. J. = Dziennik Jana Piotra Sapiehy, 1608–1611, †1611.
Sap. J. = Dziennik Jana Piotra Sapiehy, 1608–1611 (†1611)
24. Spos. = Sposób, jako zabezpieczyć, aby pieniądze wywożone nie były. ?
Spos. = „Sposób, jako zabezpieczyć, aby pieniądze wywożone nie były” (?)
25. St. a. Srg. = Andrzej Strug (właśc.: Tadeusz Gałęcki).
St. / Srg. = Andrzej Strug (= Tadeusz Gałęcki)
26. Sul. = Sulima = Walery Przyborowski, p. Przybor.
Sul. = Sulima (= Walery Przyborowski) → Przybor.
27. Urs. = Jan Ursynus (a. Jan Ursinus Leopolda) †1613.
Urs. = Jan Ursynus (a. Jan Ursinus Leopolda) (†1613)
28. Wejs. = p. Weys.
Wejs. = → Weys.
29. Zsz. = Ludwik Zejszner (Zeuschner) †1871.
Zsz. = Ludwik Zejszner (= Zeuschner) (†1871)

Ponadto wyjątkowo usunięto kropkę po liczbie w oznaczeniu Ps. 1540. = Psałterz z r. 1540., ponieważ jej obecność jest nieuzasadniona.

5.5.5. Inne uwagi

Istotną inspiracją dla autorów *SW* zdaje się być *Słownik języka polskiego* Bogumiła Lindego, co widać nie tylko w zaczerpniętych zeń hasłach, ale także jest jawnie wskazane przez nich samych przy niektórych pozycjach w *Spisie skrótów* ostatniego tomu:

- Cer. = „Ceremonje kościoła rzymskiego” (u Lindego) zap. w. XVIII.
- Fund. = Fundamenta wiary (u Lindego).
- In. I. = Inwentarz ilski (u Lindego).
- Mik. = Kacper Mikosza ? w. XVIII (u Lindego).
- Przędz. = Przędziwa, lnu, konopi przyprawa, (u Lindego).
- Salin. = Wojciech Salinarjusz ? (u Lindego).
- Tyt. pr. magd. = Tytuły prawa magdeburskiego ? (u Lindego).

Zarówno jednak źródła zaczerpnięte od Lindego, jak i te dodane już przez samych kompilatorów *SW* często nie są opisane dostatecznie dokładnie, aby można je było – przynajmniej w sposób relatywnie łatwy – zidentyfikować i znaleźć. Piszący te słowa nie zdołał określić, czy „Końskie lekarstwo” i „Sprawa a lekarstwa końskie” to różne tytuły tego samego dzieła – a jest to zaledwie jeden przykład pośród wielu, do których nie udało się dotrzeć. W kilku potwierdzonych przypadkach tytuły podane są w formie skróconej albo wręcz zmienionej względem oryginału.

Nawet jeśli dzieło można zidentyfikować i do niego dotrzeć, autorzy nie podają wydania ani stron, co wymaga ręcznego przeszukania całego tekstu przez osobę pragnącą dotrzeć do oryginalnego kontekstu zaczerpniętych haseł i przykładów. Jeszcze trudniejsza jest sytuacja, gdy źródło podaje tylko autora – wówczas konieczne jest przeszukanie całokształtu jego twórczości.

6. Metodologia retrodigitalizacyjna

Sformalizowanie metodologii retrodigitalizacyjnej przed rozpoczęciem prac nad przedsięwzięciem jest kluczowe, ponieważ pozwala uniknąć rewizji założeń oraz przerabiania kodu programu – reorganizacja pracy może być bowiem w późniejszych etapach bardzo kosztowna.

6.1. Przegląd podejść

Na samym początku istotne jest określenie docelowego poziomu digitalizacji. Żmigrodzki (2008: 102–103) na przykład wyróżnia ich trzy:

1. **digitalizację pierwszego stopnia** – do postaci graficznej;
2. **digitalizację drugiego stopnia** – do postaci graficznej z indeksowaniem i możliwością przeszukiwania;
3. **digitalizację trzeciego stopnia** – do postaci tekstowej lub bazodanowej z zaawansowanymi możliwościami przeszukiwania.

W ramach digitalizacji pierwszego stopnia – „dygitalizacji graficznej” – Majewska (2020) wyróżnia następujące kroki:

1. skanowanie słownika;
2. weryfikację kompletności i numeracji skanów;
3. prezentację danych (np. w postaci strony internetowej).

Hawkins & Dalmau & Bauman (2018)²⁶ natomiast proponują nieco dokładniejsze rozróżnienie na pięć etapów:

1. **poziom pierwszy:** tekst został wygenerowany za pomocą programu do OCR-u i został podporządkowany plikowi graficznemu i nie przewiduje się jego samodzielnego udostępniania,
2. **poziom drugi:** tekst został wygenerowany za pomocą programu do OCR-u i został podporządkowany plikowi graficznemu, jednak jest wykorzystywany do tworzenia elementów nawigacyjnych (nagłówki, sekcje w tekście).
3. **poziom trzeci:** tekst został wygenerowany albo za pomocą programu do OCR-u, albo przepisany i zostały oznaczone pewne jego elementy strukturalne; może współistnieć z plikami graficznymi albo stanowić samodzielną formę udostępniania tekstu,
4. **poziom czwarty:** tekst został wygenerowany albo za pomocą programu do OCR-u (i został poddany korekcie), albo przepisany i może być

²⁶ Tłumaczenie za Bilińską (2013: 28).

samodzielnie wykorzystywany do celów edukacyjnych i ogólnie czytania bez dodatkowych plików graficznych,

5. **poziom piąty**: tekst został wygenerowany albo za pomocą programu do OCR-u (i został poddany korekcie), albo przepisany i może być wykorzystywany samodzielnie jak tekst z poziomu czwartego, ale dodatkowo został otagowany/oznaczony strukturalnie przez specjalistów w danej dziedzinie.

Poziomy te są w pewnym sensie hierarchiczne: każdy kolejny poziom opiera się o poprzedzający go. Bień & Bilińska & Sarnecki (2014a: 3; 2014b: 9) wykorzystują ten fakt w swojej propozycji metody inkrementalnej digitalizacji, w ramach której dodaje się kolejne warstwy informacji w osobnych iteracjach – tych wyróżniają cztery:

1. anotację graficzną;
2. reprezentację w formie czystego tekstu;
3. reprezentację w formie tekstu otagowanego;
4. indeksowanie.

Powyższe opisują jedynie stan końcowy, a nie proces uzyskania danych. Opisu kroków, które muszą zostać wykonane w ramach warsztatu digitalizacyjnego, podejmuje się Kopicová (2011: 18–28) – wymienia ich sześć:

1. przygotowanie dokumentu do digitalizacji;
2. skanowanie i rozpoznanie dokumentu;
3. weryfikacja i walidacja danych;
4. indeksowanie danych;
5. zdeponowanie danych w repozytorium;
6. archiwizacja i utylizacja (oryginalnych) dokumentów.

Dwa ostatnie – zarówno w języku oryginału, czyli czeski, jak i w języku polskim – mogą być mylące od strony terminologicznej: archiwizacja jest w wielu kontekstach równoznaczna z deponowaniem i dotyczy też dokumentów cyfrowych; Kopicová jednak stara się tutaj odseparować to, co dzieje się z kopią cyfrową od tego, co dzieje się z oryginalnym nośnikiem.

Niestety, opisując metodologię *stricte* retrodigitalizacyjną trudno jest się oprzeć o literaturę, ponieważ o ile można znaleźć publikacje o digitalizacji tekstu w ogóle, to publikacji traktujących o problematyce retrodigitalizacji – a tym bardziej retrodigitalizacji leksykograficznej – prawie nie ma. Najbliższa celom tej pracy jest metodologia skompilowana przez Federację Bibliotek Cyfrowych w artykule zatytułowanym *Przygotowywanie cyfrowych wersji obiektów dziedzictwa kulturowego*, którą się tu – obok przytoczonych źródeł – posiłkowano.

Zgoła innym problemem jest fakt, że szczegółowa egzekucja digitalizacji może przyjąć różne formy zależnie od źródła, dostępnych narzędzi i pożądanego docelowego poziomu digitalizacji. W przypadku niniejszej pracy podstawą źródłową jest *Słownik warszawski* (a właściwie jego fragment), celem natomiast jest uzyskanie digitalizacji najwyższego poziomu; jest to jednak cel instrumentalny, a nie terminalny, przez co trudno

określić parametry na wyjściu – a to struktura i treść samego dzieła dostarcza parametrów na wejściu. Narzędzia przyjmuje się tu za zmienną: zostaną dobrane lub opracowane tak, aby zrealizować zamierzony cel.

Na koniec warto podkreślić, że skuteczna digitalizacja to nie tylko taka, która składa się z wielu warstw opisu danych, ale także taka, która cechuje się niewielką liczbą błędów. Te można podzielić na cztery kategorie w zależności od tego, na jakim etapie wystąpiły:

1. **błędy autorów** – czyli błędy merytoryczne w treści samego słownika;
2. **błędy druku** – czyli błędy w reprezentacji oryginalnych danych poprzez literówki, niewłaściwą segmentację czy niekonsekwentne zastosowanie typografii;
3. **błędy OCR** – czyli niepoprawnie rozpoznane znaki lub ciągi znaków, a także wyniki fałszywie dodatnie, czyli rozpoznanie elementów graficznych i artefaktów z kopii cyfrowej jako znaków;
4. **błędy parsowania** – czyli niewłaściwe zaklasyfikowanie elementów artykułów hasłowych słownika na poziomie analizy składniowej.

Warto tutaj nadmienić, że za element prac retrodigitalizacyjnych można także uznać weryfikację opisu leksykograficznego (Majewska 2020: 321). Może to obejmować włączenie jednostek do siatki hasłowej, spójne traktowanie rubryk czy analizę treści artykułów hasłowych.

Szczegółowego porównania metodologii prac nad słownikami *born digital* oraz digitalizowanymi dokonuje Majewska (2020: 268–273). Żadna ze wskazanych we wstępie prac referencyjnych nie poświęca zagadnieniom technicznym w procesie digitalizacji od zbyt wiele uwagi.

6.2. Przyjęta metodologia

W ramach niniejszej tezy *Słownikowi warszawskiemu* nadano postać cyfrową w wyniku procesu digitalizacji, a dokładniej digitalizacji właściwej, czyli zmianie informacji zapisanej analogowo na jej reprezentację cyfrową (Ober 2005). Opracowane zostały wszystkie warstwy, jednak nie w ramach oddzielnych iteracji, lecz jako kolejne poziomy w obrębie jednego procesu digitalizacji. Jako że zakłada się wierność treści, a nie formie, błędy merytoryczne autorów nie zostały poprawione, natomiast dokonano korekty pozostałych ich typów na poszczególnych etapach.

Poniżej przedstawiony został algorytm, który został wykorzystany do pełnej retrodigitalizacji *Słownika warszawskiego*, ale który został sformalizowany tak, aby stosował się do pozostałych dzieł leksykograficznych. Nie wyczerpuje on tym samym przestrzeni wszystkich kroków, które są niezbędne, aby przeprowadzić retrodigitalizację danego zasobu, ale wyodrębnia te z nich, które są zasadnicze i można je uogólnić na cały gatunek, jakim są słowniki. Choć algorytm ten jest w założeniu linearny, to nie sposób

uniknąć sytuacji, w których dopiero przy krokach dalszych ujawni się konieczność zmodyfikowania kroków je poprzedzających.

Czynności, które weszły w skład metodologii wykorzystanej w niniejszej tezie, są następujące:

1. Wybór źródła;
2. Analiza źródła;
3. Wybór narzędzi;
4. Pozyskanie kopii cyfrowej;
5. Przygotowanie do digitalizacji
6. Optyczne rozpoznanie tekstu;
7. Korekta rezultatów OCR;
8. Formatowanie tekstu cyfrowego;
9. Analiza składniowa (parsowanie);
10. Kontrola danych wyjściowych;
11. Indeksowanie;
12. Generowanie danych pochodnych;
13. Prezentacja i interpretacja danych.
14. Archiwizacja i utylizacja;

Koniecznym jest tutaj zaznaczyć, że przedstawiona kolejność nie jest *sensu stricte* chronologiczna, ale logiczna (przyczynowo-skutkowa) i pokazuje przede wszystkim zależność między danymi krokami – w przypadku opisywanych tu badań często problemy napotkane w dalszych krokach wymagały rewizji kroków poprzednich, i to niejednokrotnie w kilku iteracjach.

Poniżej zostanie omówiony każdy z wymienionych etapów.

6.2.1. Wybór źródła

Poprzez wybór źródła można rozumieć co najmniej trzy różne rzeczy: (1) wybór dzieła, które będzie przedmiotem retrodigitalizacji, (2) wybór konkretnej kopii (konkretnego nośnika lub konkretnej reprezentacji cyfrowej), oraz (3) wybór zakresu opracowania dzieła.

Wybór dzieła powinien być oparty o jasne kryteria, jednak w przypadku niniejszej tezy narzucony z góry i oczywisty – jestem nim *Słownik warszawski*.

Wybór kopii może dotyczyć oryginalnych nośników lub kopii cyfrowych i jest kluczowy on nie tylko ze względu na zakres treści (różnice między wydaniem), ale także ze względu na jakość warstwy graficznej, która wykorzystana będzie do rozpoznania tekstu – zwłaszcza, że czytelność danego tekstu dla człowieka nie jest tym samym co jego czytelność dla oprogramowania.

W przypadku *SW* autor tych słów miał dostęp do tzw. wydań faksymilowych, które Hartmann & James (2002: 56) definiują w sposób następujący:

facsimile edition

The identical reproduction of a DOCUMENT, with or without ANNOTATIONS. Facsimile editions of early dictionaries in manuscript or printed form allow their comparative-historical study.

wydanie faksymilowe

Identyczna reprodukcja DOKUMENTU, z ADNOTACJAMI lub bez nich. Edycje faksymilowe wczesnych słowników w formie rękopisów lub druków umożliwiają ich badanie porównawczo-historyczne.

Reprodukcje tego rodzaju nie tylko zachowują wszelkie artefakty drukarskie oryginału, ale cechują się również zwiększeniem poziomu degradacji danych. Oprócz tego dostępne są dwie kopie cyfrowe *Słownika* oparte o różne nośniki fizyczne: wersja z Federacji Bibliotek Cyfrowych oraz wersja z biblioteki Uniwersytetu Torontońskiego. Te z jednej strony niwelują potrzebę skanowania dzieła, z drugiej strony nakładają ograniczenia na parametry obrazów, z którymi będzie przebiegała praca.

Więcej informacji o wyborze źródła na potrzeby rozpoznania tekstu dostępnych jest w **rozdziałach 8 i 9**.

Wybór zakresu, w jakim opracowane będzie dzieło, również stanowi pewne wyzwanie. Oczywiście można postawić sobie za cel retrodigitalizację całego dzieła, ale w przypadku ograniczonej mocy przerobowej może być to praktyczne awykonalne. Jeśli natomiast ograniczyć się do fragmentu, to trzeba najpierw wiedzieć jaki fragment to będzie i na jakiej podstawie został wyselekcjonowany.

Tego rodzaju zawężenie zakresu było konieczne w przypadku niniejszej pracy której celem jest stworzenie makiety cyfrowej reprezentacji *SW*. Wybór ten jest wynikiem następującego procesu decyzyjnego:

1. postanowiono opracować próbkę nie więcej niż 100 stron *Słownika* ze względu na ograniczenia czasowe;
2. zgodnie z praktyką leksykograficzną wybrano pojedynczą literę, ponieważ jest to koherentna jednostka makrostruktury części hasłowej *Słownika*, a także ponieważ:
3. dyrektywy redaktorskie, którymi kierowano się w opracowywaniu haseł rozpoczynających się na poszczególne litery, zdają się różnić od siebie (co zostało opisane w **rozdziale 4**), a to utrudniłoby opracowanie makiety;
4. ponadto odrzucono możliwość opracowania litery podzielonej między tomy, jako że metody redaktorskie zmieniały się także między tomami;
5. zdecydowano opracować pierwszy fragment spełniający kryteria.

Podane kryteria zostały spełnione przez sekcję zawierającą hasła rozpoczynające się na literę **F**, która obejmuje zakres od początku strony 708 do połowy strony 788 w tomie I *Słownika*, i to właśnie ona posłużyła za materiał badawczy. Opracowanie haseł na tę literę mają też dodatkową zaletę względem choćby haseł zaczynających się na literę **A**, ponieważ te bywają nadreprezentowane przez wyrazy o pochodzeniu obcym.

Analogiczne kryteria zastosowano do wyboru materiału referencyjnego, za który miała posłużyć sekcja z literą **I** (od początku strony 71 do połowy strony 119 w tomie II), jednak materiał ten nie został w pełni wykorzystany. Opracowano za to literę **Ó** (znajdującą się w t. III), a także **V**, **X** oraz **Y** (wszystkie trzy znajdujące się w tomie VIII).

Taki dobór materiału badawczego niesie ze sobą pewne problemy. Kaszewski (2014: 12–13) w kontekście korpusów pisze, że analiza leksyki na jedną literę wiąże się z niereprezentatywnością. Uwaga ta stosuje się również do analizy siatki hasłowej: z synchronicznego punktu widzenia kryterium literowe jest wysoce arbitralne. I tak w obrębie litery **F** nie ma żadnych przyimków ani przedrostków o pochodzeniu słowiańskim, co może mieć wtórny wpływ na strukturę haseł. Przyjęto jednak, że na potrzeby analizy strukturalnej jednostek hasłowych nie zrównoważony zasób leksykalny ma jedynie marginalne znaczenie.

6.2.2. Analiza źródła

Po doborze źródła konieczna jest jego wnikliwa i wszechstronna analiza, ponieważ silnie zdeterminuje ona dalsze kroki. Mowa tutaj nie tylko o samej treści i charakterystyce dzieła, ale także o poszczególnych poziomach struktury i elementach wchodzących w jej skład, reprezentacji typograficznej i graficznej, czy wreszcie o wszelkiego rodzaju anomaliamiach.

W przypadku *SW* konieczne było zwłaszcza bardzo uważne przyjrzenie się jego strukturze, ponieważ ta wykazuje złożoność nieporównywalną z większością słowników, a także stosunkowo wysoką nieregularność, co determinuje strategię przyjęte w analizie składniowej.

Aby to uczynić, najpierw wynotowano uwagi na podstawie wstępnego zapoznania się z tekstem – konieczna jednak była weryfikacja założeń podczas pogłębionej analizy, która oparta była o treść w formie cyfrowej, uzyskanej w jednym z dalszych kroków. Konieczne okazało się pogrupowanie próbki haseł pod względem podobnej struktury i posortowanie pod względem dwóch kryteriów: częstotliwości (od najwyższej) i poziomu skomplikowania (od najniższego). Na tej podstawie rozpisana została wstępna struktura danych wyjściowych, która uzupełniona została o problemy, które wyodrębniono dopiero podczas opracowywania programu parsującego, opisanego w **rozdziale 10**.

Konieczne też było zaprojektowanie innych narzędzi, związanych przede wszystkim z pracą na warstwie graficznej kopii cyfrowych *Słownika*, które wynikały z wyzwań zaobserwowanych w trakcie analizy całościowej materiału badawczego, co zostało opisane w rozdziale następnym.

6.2.3. Wybór narzędzi

Wybór narzędzi obejmuje zarówno urządzenia fizyczne, jak i oprogramowanie. W przypadku wyboru oprogramowania chodzi nie tylko o dobór spośród dostępnych rozwiązań, ale niekiedy także ich adaptację czy wręcz samodzielne opracowanie. Narzędzia te mogą

albo adresować problem przy digitalizacji źródła, albo pozwalać wygenerować dane źródłowe w ramach przyjętych założeń.

W przypadku SW konieczne było zarówno wykorzystanie dostępnych narzędzi, jak i stworzenie własnych, dopasowanych do charakterystyki i materiału badawczego. Jeśli chodzi zaś o ich zastosowanie, to można je pogrupować w sześć głównych kategorii:

1. skanowanie;
2. obróbki graficznej;
3. konwersji obrazów na tekst;
4. pracy z tekstem niesformatowanym;
5. analizy składniowej (parsowania);
6. przechowywania danych;
7. prezentacji danych.

Wszystkie narzędzia, które zostały wykorzystane w tejże pracy doktorskiej, zostały opisane w **rozdziale 7**, a opracowaniu dwóch z nich poświęcono osobne rozdziały: **rozdział 8**, poświęcony ekstrakcji kart hasłowych, oraz **rozdział 12**, poświęcony analizie składniowej i prezentacji danych.

6.2.4. Pozyskanie kopii cyfrowej

Pozyskanie kopii cyfrowej może odbyć się trzema ścieżkami: kopię tę można (1) stworzyć poprzez skanowanie lub zrobienie zdjęć, można (2) pobrać gotową wersję przygotowaną przez kogoś innego, lub w skrajnych przypadkach (3) dokonać ręcznego przepisanie treści.

W przypadku SW eksperymentowano ze stworzeniem własnej kopii przy pomocy różnych urządzeń skanujących, jednak okazało się, że nie można uzyskać jakości wyższej od już istniejących kopii, ponieważ ich rozdzielczość jest wystarczająco wysoka, a artefakty w głównej mierze pochodzą się z kopii faksymilowych, które są powielone między dostępnymi egzemplarzami – dlatego też wykorzystano jedną z takich wersji cyfrowych.

6.2.5. Przygotowanie do digitalizacji

Etap ten obejmuje wszystkie czynności, które pozwolą wykorzystać obrazy w celu uzyskania wysokopoziomowej digitalizacji, czyli zanotowanej i zindeksowanej wersji tekstu. Zasadniczo można wyróżnić dwa podetapy: obróbkę graficzną i segmentację danych.

Obróbka graficzna ogranicza się tutaj do podjęcia kroków, które ułatwiają zadanie programowi do rozpoznawania znaków i tym samym zwiększają jakość rozpoznanego tekstu: chodzi tu o usunięcie artefaktów, kadrowanie, korektę przekosów, rotacji, niwelację zniekształceń geometrycznych, wyprostowania linii tekstu i redukcję kolorów. Wiele z tych rzeczy można osiągnąć w programach do OCR, jednak funkcje te bywają bardziej rozbudowane w programach do tego dedykowanych, poza tym odrębne rozwiązania pozwalają na automatyczne, ale kontrolowane przygotowanie obrazów.

Odrębną kwestią jest segmentacja (ang. *chunking*), która nie musi dotyczyć wyłącznie danych graficznych, ale w przypadku niniejszej tezy została dokonana także na tym poziomie. Strategia ta ułatwia anotację dokumentów poprzez podzielenie ich na mniejsze i przez to łatwiejsze do opracowania fragmenty (Bouda & Cysouw 2012: 17). W przypadku pracy nad *SW* oznaczało to dwie rzeczy: (1) podzielenia skanów stron *Słownika* na fragmenty odpowiadające poszczególnym artykułom hasłowym, a także (2) parsowanie artykułów pojedynczo, w zestawach odpowiadających sekcji z daną literą w oryginalnej treści.

Oba rodzaje segmentacji sprawiają, że można operować na mniejszych jednostkach danych, co upraszcza pracę z nimi, a jednocześnie zwiększa dokładność ich przetwarzania – w przypadku analizatora składniowego oznacza to, że ten ma jasno określony początek i koniec każdego artykułu hasłowego, co nie jest trywialne do rozpoznania przy pomocy opisanych dalej metod.

6.2.6. Optyczne rozpoznanie tekstu (OCR)

Mając do dyspozycji przygotowane obrazy, można przystąpić do procesu ich konwersji na tekst. Można to uczynić przy pomocy różnych programów, które w poszczególnych przypadkach uzyskują lepsze lub gorsze rezultaty; dodatkowo na trafność rozpoznawania może wpłynąć odpowiednie skonfigurowanie i sparametryzowanie programu.

Niektóre rozwiązania oprócz interfejsu graficznego posiadają także API programistyczne, co pozwala na ich podłączenie do rozwiązań programistycznych, co pozwala na seryjnie przetwarzanie wielu elementów i zautomatyzowanie pewnych procesów, zupełnie wyeliminowując udział człowieka. W przypadku opracowywania pojedynczych dzieł jest to jednak zbyt cenne, a poza tym eliminacja ludzkiej weryfikacji nie zawsze jest pożądana.

Należy przy tym pamiętać, że tekst taki prawie nigdy nie jest odczytany w pełni skutecznie, a w przypadku tekstów starszych problem ten jest nasilony. Nie inaczej jest w przypadku *SW*, którego OCR okazał się wyjątkowo dużym wyzwaniem zarówno z powodu jego treści, jak i formy.

Zagadnieniom z problematyki technologii OCR, obejmującym krok niniejszy oraz następny, poświęcono **rozdział 9**Błąd! Nie można odnaleźć źródła odwołania..

6.2.7. Korekta rezultatów OCR

Korekta rezultatów OCR może objąć dwa procesy: weryfikację i walidację; różnica między nimi polega na tym, że weryfikacja jest ręczna i obejmuje ręczne sprawdzenie rozpoznania poszczególnych znaków, podczas gdy walidacja jest automatyczna i obejmuje wykonanie zadanych reguł. Jeśli celem jest uzyskania tekstu cyfrowego o maksymalnej wierności, najistotniejszy jest ten pierwszy, drugi może natomiast pełnić funkcję pomocniczą.

Większość komercyjnych programów posiada interfejs graficzny, który pozwala na weryfikację tekstu w jego oryginalnym kontekście, zestawiając rozpoznany fragment tekstu z fragmentem obrazu, który jest źródłem owego tekstu. Etap ten w doświadczeniu piszącego te słowa jest kluczowy, ponieważ to tutaj użytkownik ma dostęp do największej liczby danych pozwalających na weryfikację rezultatu OCR. Oczywiście błędy można znaleźć też i na innych etapach, np. po wyeksportowaniu tekstu do pliku tekstowego.

O ile niektóre błędy, jak większość literówek, nie będzie miało znaczenia (oczywiście poza niską wiernością względem oryginalnego tekstu), to niektóre mogą wprowadzić parser w błąd lub też całkowicie uniemożliwić przetworzenie danego fragmentu. Z tego powodu walidacja danych może być bardzo pomocna: o ile tworzenie reguł podmian na podstawie zaobserwowanych prawidłowości nierzadko miewa nieoczekiwane rezultaty, to odfiltrowanie słów na podstawie słownika lub „podejrzanych” połączeń znaków pozwala wyeliminować wiele błędów.

W przypadku *SW* duży problem stanowiły zwłaszcza znaki nietypowe, jak nietypowe dla polskiego litery diakrytyczne, czy znaki niealfanumeryczne – nie tylko OCR miał z nimi problemy, ale były też trudne do wprowadzenia na klawiaturze podczas weryfikacji, co doprowadziło do opracowania własnego układu klawiatury (opisanego w **rozdziale 7.9**), co problem ten niemal całkowicie rozwiązało.

6.2.8. Format tekstu cyfrowego

W rozwiązaniach programistycznych wykorzystuje się zazwyczaj tzw. czysty tekst, czyli tekst niesformatowany, lub jego warianty; sam format pliku zależy jednak od jego zastosowania i zaimplementowanych rozwiązań. Jednak nawet po zapisaniu rezultatów OCR w odpowiednim formacie nie zawsze jest on automatycznie gotowy o przetwarzania – wynika to z dwóch powodów.

Po pierwsze należy się zastanowić, co będzie istotne dla analizatora składniowego – ten może wykorzystywać różne separatory przetwarzanych jednostek, np. znaki nowej linii, ale także i wiele innych elementów, które pozwalają na prawidłową interpretację struktury. Po drugie tekst ten może wymagać wzbogacenia na potrzeby jego końcowej prezentacji, jak np. odzyskania pewnych informacji o jego formatowaniu.

Rozwiązanie zastosowane w przypadku *SW* zostało szczegółowo opisane w **podrozdziale 10.2.1**.

6.2.9. Analiza składniowa (parsowanie)

Analiza składniowa pozwala nadać zwykłemu tekstowi nową jakość poprzez rozpoznanie jego struktury, a tym samym umożliwić komputerowi dostęp do danych wcześniej dla niego niedostępnych, co z kolei pozwala na manipulowanie nimi w sposób automatyczny, odpowiednio do potrzeb badacza.

W przypadku *SW* jest to etap bezsprzecznie najbardziej złożony i czasochłonny, ale którego rezultat pozwala na niedostępne do tej pory zastosowania treści *Słownika*.... Cały proces parsowania został opisany szczegółowo w **podrozdziale 10.3**.

6.2.10. Kontrola danych wyjściowych

Po wygenerowaniu danych konieczna jest ich skontrolowanie – o ile jednak w niewielkim zakresie można dokonać ręcznej weryfikacji, dużo rozsądniejszą strategią w przypadku danych mało czytelnych dla człowieka jest walidacja. Ta może przebiegać jeszcze w trakcie samego parsowania, jak i po nim, i przybrać różne formy.

W przypadku *SW* walidacja jest dokonywana w trakcie samej analizy składniowej, a wtórnie także podczas interpretacji jej wyników; rozważano także walidację składni wyjściowego pliku XML, ale w ostatecznej funkcjonalności ta nie została zaimplementowana z kilku powodów, które opisane są w **podrozdziale 10.8.8**.

6.2.11. Indeksowanie

Indeksowanie polega na dodaniu do anotacji informacji o źródle (Bouda & Cysouw 2012: 17), czyli jest formą metadanych. Te ostatnie są istotne w humanistyce cyfrowej – jak zaznaczają Bouda & Cysouw (2012: 16):

[I]t is of central importance that everybody working with the digitized data should always be able to trace back the information to the original source.

[Z]asadniczą kwestią jest, aby każda osoba pracująca z danymi cyfrowymi zawsze była w stanie prześledzić informacje do oryginalnego źródła.

W przypadku retrodigitalizacji możliwość wskazania dokładnego miejsca, któremu odpowiadają dane tekstowe, nie tylko ułatwia weryfikację wierności danych, ale także zwiększa wiarygodność ich cyfrowej kopii i pozwala badaczowi na wyciągnięcie własnych wniosków w stosunku do ewentualnych różnic między oryginalną prezentacją tekstu źródłowego, a jego cyfrową reprezentacją.

W przypadku *SW* informacja o lokalizacji każdego artykułu hasłowego została umieszczona w znacznikach XML, co zostało wykorzystane do wygenerowania **Aneksu F**, który służy jako zasób fotodokumentacyjny parując artykuły hasłowe z fragmentami ich oryginalnego kontekstu.

6.2.12. Generowanie danych pochodnych

W momencie, w którym ma się do dyspozycji ustrukturyzowane dane, można przystąpić do ich przetwarzania i interpretacji. Oznacza to nie tylko uzyskanie danych ilościowych, ale także – przy odpowiedniej algorytmizacji – uzyskanie nowej jakości, jak filtrowanie danych na podstawie zadanych reguł, ich automatyczna generalizacja czy klasyfikacja.

W przypadku danych *SW* dane pochodne zostały otrzymane przez dekompresję informacji zawartej w *Słowniku* i odpowiednie reguły ich przetworzenia, w innych zaś

wykorzystano dane pozasłownikowe, co pozwoliło na analizę treści w zupełnie nowych wymiarach, co zostało opisane w **rozdziale 10.4**.

6.2.13. Prezentacja i interpretacja danych

Opracowane dane, jeśli mają podlegać analizie, powinny zostać zaprezentowane formie, który będzie łatwy do odczytu przez człowieka. Oznacza to nie tylko graficzny układ danych i ich typografię, ale także opatrzenie odpowiednimi opisami w języku naturalnym. Efekt taki można uzyskać manipulując danymi w sposób manualny lub automatyczny.

W przypadku niniejszej pracy doktorskiej zaprojektowano program przetwarzający dane *SW* w taki sposób, aby samodzielnie generował fragmenty opisu w języku naturalnym na podstawie zakodowanych reguł, a także żeby prezentował dane w przyjaznej i czytelnej formie na podstawie szablonów generujących kod \LaTeX . Rezultatem tego rozwiązania są w pełni automatycznie generowane **aneksy** od **A** do **P**, co ilustruje zalety i możliwości przetwarzania danych ustrukturyzowanych.

Etap ten może objąć także ewaluację opisu leksykograficznego, co też w niniejszej pracy miało miejsce, jednak proces ten był rozproszony między innymi etapami jeszcze na etapie opracowywania narzędzi programistycznych – analiza składniowa bez interpretacji struktury słownika była niemożliwa.

6.2.14. Archiwizacja i utylizacja

Końcowym etapem jest archiwizacja i utylizacja, czyli odpowiednie zdeponowanie istotnych danych i nośników, a także usunięcie tych, które są już zbędne; dotyczy to zarówno kopii fizycznych, jak i cyfrowych.

W przypadku retrodigitalizacji zachowanie oryginalnego nośnika fizycznego jest bardzo istotne ze względu na jego wartość historyczną. Inaczej wygląda to w przypadku danych: jednym z celów rozpoznawania tekstu jest kompresja informacji – graficzna reprezentacja tekstu ma większość objętość na dysku niż tekst cyfrowy.

Jednocześnie w dokumencie zatytułowanym *Some Best Practices for Retrodigitization* (2006: 2) autorzy wskazują, że „the un-retouched original scans should be archived. Future technological advances may be easy to apply to the original scanned images; rescanning will be impossible in many instances” (‘„nieretuszowane oryginalne skany należy zarchiwizować. Przyszłe postępy technologiczne mogą być łatwe do zastosowania na oryginalnych zeskanowanych obrazach; ponowne skanowanie będzie w wielu przypadkach niemożliwe.’). Zachowanie danych przejściowych, nie tylko skanów, może być z wielu powodów pożądane – nie tylko dokumentacyjnych, ale i pragmatycznych. Konieczne jest więc zastanowienie się, które dane (i metadane) będą istotne w przyszłości i do jakich celów.

W przypadku tejże pracy problem archiwizacji nośnika fizycznego nie zaistniał, ponieważ od początku operowano na kopii cyfrowej *SW*. Na płaszczyźnie cyfrowej napisany program zapisuje rezultaty każdego etapu przetwarzania danych (logując przy tym

zaistniałe błędy). Jako że opisywany program jest stanowi rozwiązanie deterministyczne, to zachowane być muszą jedynie jego kod oraz dane wejściowe, aby móc odzyskać poszczególne etapy aż po dane wyjściowe; dane te są zapisane w standardowych, co pozwala, na ich odczytanie bez specjalnego oprogramowania.

7. Narzędzia

W rozdziale tym zostały przedstawione narzędzia oraz technologie – zarówno w postaci oprogramowania autorskiego, jak i pochodzącego od stron trzecich – które zostały wykorzystane do opracowania danych badawczych będących przedmiotem niniejszej tezy doktorskiej.

7.1. Środowisko programistyczne

7.1.1. Języki programowania Python i Ruby

Języki programowania Python i Ruby łączy wiele cech wspólnych – oba z nich:

- są językami skryptowymi,
- są językami dynamicznymi;
- są językami interpretowanymi;
- są językami wysokopoziomowymi;
- są językami ogólnego przeznaczenia;
- należą do paradygmatu programowania obiektowego;
- należą do paradygmatu programowania imperatywnego;
- należą do paradygmatu programowania funkcjonalnego;
- są silnie typowane;
- posiadają system bibliotek rozszerzających ich funkcjonalność;
- posiadają bogatą dokumentację.

Języki te jednak dzieli kilka istotnych różnic, które spowodowały, że na potrzeby niniejszej tezy zostały wykorzystane oba, acz w różnym zakresie.

7.1.1.1. Język programowania Python

Język Python²⁷ cechuje się przejrzystością kodu, ponieważ kładzie nacisk na eksplicytny zapis, co ułatwia jego naukę oraz proces debugowania. Poza tym posiada znacznie bogatszy system bibliotek (ze względu na dużą popularność tego języka), co też było powodem wykorzystania tego języka do stworzenia programu IndexCardSplitter, opisanego w **rozdziale 8**.

Do rozwiązań zastosowanych w niniejszej pracy doktorskiej wykorzystano Pythona w wersjach od 3.6 do 3.9.1.

²⁷ Strona domowa języka Python: <https://www.python.org/>.

7.1.1.2. Język programowania Ruby

Język Ruby²⁸, w przeciwieństwie do Pythona, jest językiem w pełni obiektowym, a także pozwalającym na znacznie większą elastyczność i jednoczesną zwięzłość zapisu, co spowodowało, że to właśnie on został wykorzystany do opracowania programu SWParser, który jest opisany **rozdziale 10**.

7.2. Standard Unikod (Unicode)

Wszelkie przetwarzanie języka z użyciem komputera wymaga pracy z tekstem cyfrowym – te stoją w opozycji do tekstów drukowanych. Bień (1999: 4) stosuje pojęcia „tekstów fizycznych” i „tekstów elektronicznych”:

[W]yróżnimy natomiast dwa podstawowe typy tekstów: „teksty fizyczne” i „teksty elektroniczne”. Za charakterystyczną cechą tekstów elektronicznych uznajemy fakt, że nie zmieniają one swojej tożsamości przy zmianie ich nośnika fizycznego (...).

W przypadku retrodigitalizacji mowa o tekście, którego cyfrowy zapis ma charakter wtórny:

Niektóre teksty elektroniczne są po prostu mniej lub bardziej wiernym zapisem pewnych tekstów fizycznych – możemy mówić wtedy o tekście pierwotnym (fizycznym) i wtórnym (elektronicznym). (Bień 1999: 5)

Bień (2016a) pisze więcej o problematyce stworzenia systemu cyfrowego, który by odpowiadał systemowi analogowemu w odrębnym artykule, jednak w pisać te słowa skupia się w tym podrozdziale na jego przechowaniu i prezentacji. Zasadniczo teksty elektroniczne dzieli się na dwa typy: tekst czysty i tekst sformatowany; ten pierwszy zawiera wyłącznie informację o znakach, które na ten tekst się składają, natomiast ten drugi zawiera też informację na temat prezentacji tekstu. Jak opisuje to Bień (1999: 6):

W terminologii anglosaskiej przyjęło się ten fakt odzwierciedlać rozróżniając *plain text* (dosłownie zwykły tekst) z jednej strony i *fancy text* (dosłownie tekst wymyślny)²⁹ lub *rich text* (dosłownie tekst bogaty); dla *plain text* dość szeroko stosowane – również przeze mnie – jest tłumaczenie czysty tekst, termin *rich text* proponuje tutaj tłumaczyć jako tekst wzbogacony.

Standard UNICODE (...) stwierdza (...), że czysty tekst reprezentuje podstawowa treść tekstu w wymiennej – to znaczy nie związanej z konkretnym oprogramowaniem – postaci. W konsekwencji czysty tekst nie ma określonego swojego wyglądu – kwestie kroju i wielkości znaków, podziału na wiersze i strony, koloru, odsyłaczy hipertekstowych, a nawet języka, w którym tekst lub jego fragmenty są zapisane, należą do domeny tekstu wzbogaconego.

²⁸ Strona domowa języka Ruby: <https://www.ruby-lang.org/>.

²⁹ Autor niniejszej tezy nigdy się z określeniem tym nie spotkał.

Niezależnie od formy, to wspomniany powyżej standard Unicode (Unikod) stanowi współczesny standard zapisu tekstu elektronicznego. Pozwala on w sposób spójny wyświetlić szeroką gamę znaków językowych (należących do różnych systemów) i symboli pozajęzykowych. Jak podsumowuje to Strzelczyk (2013: 41):

[J]est standardem kodowania znaków zaprojektowanym do wymiany, przetwarzania i wyświetlania dokumentów pisanych w każdym języku i dotyczących każdej dziedziny współczesnego świata, w tym również tekstów w językach wymarłych i historycznych.

Kawałek dalej pisze o założeniach standardu (Strzelczyk 2013: 41):

Założenia standardu Unicode obejmują: ogólnoświatowy (szeroki) zestaw znaków, ich logiczny porządek, efektywność przetwarzania dokumentów, ich jednolitość, stabilność standardu, zachowanie semantyki znaków i niezależność od ich wyglądu. Standard obejmuje także algorytmy łączenia znaków (np. litery akcentowane), ich dekompozycje, a także wersje znaków w zależności od kontekstu (co dotyczy głównie języków afroazjatyckich).

Standard ten jest obecnie na tyle rozpowszechniony i kompleksowy, że z perspektywy współczesnego użytkownika jest w zasadzie niewidoczny:

Przeciętnemu użytkownikowi komputera standard Unicode zapewnia wygodną obsługę tekstów w dowolnym języku, ich przenośność pomiędzy programami i systemami operacyjnymi, a także wymianę dokumentów z osobami z całego świata. Upowszechnienie tego standardu pozwala zapomnieć o koszmarze „standardów kodowań” (dla samego języka polskiego używaliśmy ponad dziesięć stron kodowych, w tym: ISO-8859-2, Windows 1250, Latin 852 i Mazovia). (Strzelczyk 2013: 42)

Więcej o genezie i działaniu Unikodu od strony technicznej pisze Bień (1993: 6ff; 2019: 7), tutaj jednak kwestie te zostaną pominięte.

Strzelczyk (2013: 43) zauważa, że standard ten potrafi przysporzyć wiele trudności osobom pracującym z szeroko pojętym składem tekstu i jego kodowaniem. Wynika to przede wszystkim z jego struktury.

Podstawową jednostką systemu Unikod jest tzw. znak kodowy, któremu przyporządkowana jest współrzędna kodowa (zapisywana numerycznie w systemie szesnastkowym). Znaki kodowe tworzą zestawy (ang. *character sets*), czyli zbiory znaków i przyporządkowanych im kombinacji bitowych, a Unikod jest jednym z nich. Graficzna realizacja znaku natomiast nazywana jest glifem (Bień 2016b: 67–68). Innymi słowy, znaki w danym zestawie nie mają kształtu *per se* – są tylko abstraktami, a kształtu nadają im ich realizacje, czyli glify (Bień 2011b: 20). Można więc mówić o relacji między znakiem abstrakcyjnym (ang. *abstract character*) a znakiem kodowym (ang. *encoded character*).

Znaki takie dzielić można na **proste** i **złożone**; te ostatnie składają się ze znaku podstawowego i „znaku dostawnych”³⁰ (ang. *combining character*). Od strony binarnej wiąże się to z podaniem więcej niż jednej współrzędnej. Dla niektórych połączeń (np. liter ze znakami diakrytycznymi) istnieją odrębne współrzędne kodowe, co czyni je znakami prostymi. Oznacza to, że niektóre znaki abstrakcyjne mogą być zakodowane znakami kodowymi na więcej niż jeden sposób.

Standard Unikod mówi jednak, że formy „ekwiwalentne kanonicznie”, czyli mające takie samo znaczenie, są na poziomie binarnym traktowane tak samo niezależnie od tego, czy są to pojedyncze znaki czy sekwencje znaków (*Unicode Standard Annex #15*). Przykładowo, e z makronem może być zapisane w postaci jednego znaku ē (współrzędna kodowa U+0113), lub poprzez połączenie litery e (współrzędna kodowa U+0065) ze znakiem diakrytycznym dostawnym makron, którego samodzielna realizacja ma postać ◌̄. W przypadku znaków z dwoma diakrytykami można do znaku połączonego dodać znak dostawny, co też przyjęto za standard zapisu takich przypadków w niniejszej tezie.

Nakreślona wyżej relacja dotyczy jednak tylko abstraktu i cyfrowej wersji informacji o tym abstrakcie; te pierwsze są zawsze powiązane ze znakami pisarskimi (znaki użyte w piśmie i druku również są realizacją tych abstraktów), a przypadku retrodigitalizacji – konkretne znaki kodowe odpowiadają znakom pisarskim (a przynajmniej starają się je odwzorować). Jak pisze Bień (1999: 8):

Wyróżnienie podstawowych elementów pisma (czyli znaków pisarskich lub czcionek — angielski termin *character* oznacza jedno i drugie) jest w dużym stopniu kwestia konwencji, wywodzącej się z tradycji lub uwarunkowań technicznych (...)

Bień (2016a: 1) ukonkretnia, że „Najbardziej konkretnym punktem wyjścia do badania repertuaru znaków piśmiennych są czcionki drukarskie”.

Unikod boryka się obecnie z wieloma problemami: **technicznych**, które opisuje Strzelczyk (2013: 43), **konceptyjnych** – kolejne wersje dodają bowiem coraz więcej emotikon, i można by było polemizować, czy ich dodawanie do standardu poświęconego dotychczas językom naturalnym i symbolom technicznym jest dobrą decyzją³¹ – a także, co najistotniejsze w kontekście tekstu wtórnie cyfrowego, pojawiają się w nim trudności związane z **kodowaniem starszych form ortografii** polskiej oraz reprezentacji znaków w poszczególnych fontach (Bień 2010a; 2011a).

Bień (2019: 7) opisuje dwie alternatywne strategie, które adresują ten problem: (1) zapisywanie znaku graficznego z tekstów historycznych przy pomocy najbardziej zbliżonych do nich znaków cyfrowych o innej semantyce, oraz (2) stworzenie czcionki, w której reprezentacja graficzna danych znaków jest zmodyfikowana, ale ich semantyka

³⁰ Termin „znaki dostawne” został użyty przez Bienia (2011b: 21), z zastrzeżeniem, że autorem tego zgrabnego sformułowania jest kto inny.

³¹ I piszący te słowa podziela te wątpliwości, uważając skupianie się na emotikonach za rodzaj infantylicyzacji prac nad standardem.

pozostaje bez zmian. Bień pisze też próbie uwzględnienia znaków historycznie obecnych w alfabecie polskim (Bień 2019: 17ff).

Według części opracowań cytowanych przez Bienia (2011b: 18–20) tekst składa się z abstrakcyjnych grafemów analogicznie do tego, jak wymowa składa się z fonemów – można więc mówić o obocznościach i alografach. Segmentem tekstu dla Bienia (2011b: 21) jest zatem „zbitka grafemiczna”, czyli ciąg grafemów pojedynczych grafemów.

Odpowiednie kodowanie znaków jest istotne nie tylko ze względu na ich graficzne odwzorowanie, ale także ze względu na ich rolę w parsowaniu. Saloni (2007) opisuje na przykład problemy stosowania „znaków konwencjonalnych” w tekstach (także w kontekście leksykografii, a konkretnie: *SGJP*), w tym: apostrofu (’), ukośnika (/), łącznika (-) oraz kropki (.). Znaków, które będą pełniły funkcje inne niż zapisywanie dźwięków języka czy interpunkcji jest zresztą więcej.

W przypadku retrodigitalizacji *SW* na szczęście nie było większych problemów z kodowaniem znaków, ponieważ stosuje on ich formy tożsame ze współczesnym standardem, natomiast specjalnego potraktowania wymagały niektóre znaki złożone.

7.3. Język znaczników XML

Cyfrowa reprezentacja ustrukturyzowanych danych tekstowych może być dokonana na jeden z trzech sposobów: poprzez (1) bazy danych, (2) język znaczników, lub (3) opracowanie własnych rozwiązań (Stokes & Pierazzo 2009: 207). Biorąc pod uwagę, jak wiele zostało zainwestowane w dwa pierwsze rozwiązania, opcja trzecia jest mało realna.

Hartmann & James (2002: 34) definiują bazy danych następująco: „A facility for the electronic storing and manipulation of linguistic and other DATA” (‘Urządzenie do elektronicznego przechowywania i manipulacji DANYCH językowych i innych’). Bazy danych dzielą się na: relacyjne, nierelacyjne i rozwiązania mieszane; bazy pierwszego rodzaju bardzo dobrze się sprawdzają w przypadku silnie ustrukturyzowanych danych, jednak w przypadku zmienności strukturalnej lepszym rozwiązaniem jest język znaczników XML, który stanowi dla nich skuteczną alternatywę (Stokes & Pierazzo 2009: 208).

XML (ang. *Extensible Markup Language*) to uniwersalny język znaczników przeznaczony do reprezentowania różnych danych tekstowych w ustrukturyzowany sposób. Język ten jest oparty o SGML (ang. *Standard Generalized Markup Language*), czyli standardowy uogólniony język znaczników służący do definiowania i ujednocnienia struktury i formatu różnego typu informacji (Bień 1993: 10ff). Zarówno SGML jak i XML należą do standardu ISO, a ten drugi jest dodatkowo rekomendowany przez organizację W3C.

Szafran (2007: 20–21) opisuje język XML jako służący do reprezentacji struktury tekstu, i wymienia jego trzy główne cechy: (1) opisowy charakter znaczników (co pozwala na ich oderwanie struktury od reprezentacji), (2) definicja typu dokumentu (określająca reguły stosowania znaczników), (3) niezależność od środowiska sprzętowego i systemowego (co pozwala na przenoszenie dokumentu bez straty informacji).

Dane przechowywane w formacie XML (najczęściej w plikach o tym samym rozszerzeniu: `.xml`) są niezależne od platformy i czytelne zarówno dla ludzi, jak i dla komputerów. Jego prostota, ogólność, elastyczność, czytelność, wielość zastosowań i wieloplatformowość uczyniły go jednym z dominujących standardów wymiany informacji w Internecie; z tych samych powodów to właśnie w tym języku przechowywane są dane opracowane na potrzeby niniejszej pracy doktorskiej.

Oprócz XML istnieją również pod pewnymi względami zbliżone rozwiązania, takie jak JSON czy YAML, które służą do wymiany danych, nie są jednak językami znacznikowymi. XML został wybrany, ponieważ jest lepszym rozwiązaniem (1) anotacji, (2) stosowania atrybutów, (3) rozwiązań serwerowych³², (4) reprezentacji struktury w formie drzewa³³.

Dzięki znacznikowaniu XML można anotować wszelkie komponenty strukturalne hasła (główne i peryferyjne) poprzez znacznikowanie XML (Martinez 2013: 226f), co pozwala nie tylko na odzwierciedlenie struktury hasła, ale także na jej wzbogacenie (np. o lematyzację), jednocześnie dodając możliwość łatwego jej przeszukiwania (Gasiglia 2013a: 11). Co więcej, dobre zakodowanie danych w strukturze XML pozwala na różne sposoby odczytu i prezentacji danych (Stokes & Pierazzo 2009: 219), a tym tę samą strukturę reprezentować przez różne interfejsy, które mogą być dopasowane do indywidualnego użytkownika (Ecker 2013: 105f). Ta ostatnia zaleta jest istotna w kontekście rozszerzenia funkcji słownika, a tym samym i zakresu odbiorców (Ecker 2013: 108).

Na XML oparte są też różnego rodzaju formalizmy, takie jak RDF(ang. *Resource Description Framework*), OWL (ang. *Web Ontology Language*) (Heid 2014: 25), czy dużo bardziej znany HTML (ang. *HyperText Markup Language*); najistotniejszy z nich w kontekście niniejszej pracy jest jednak system TEI, który może być przydatny mimo swoich „ograniczeń konceptualnych” (Lemnitzer et al. 2013: 1207–1208), które jednak zdecydowały o jego niezastosowaniu w niniejszej pracy.

Kolejną zaletą XML jest liczba dostępnych narzędzi do pracy z nim (Heid 2014: 25), poczynając od edycji, poprzez generowanie i odpytywanie, a kończąc na walidacji; tej ostatniej można dokonać m.in. przy pomocy rozwiązań takich jak Schematron, W3C XML Schema, DTD (ang. *Document Type Definition*), czy XSD (ang. *XML Schema Definition*). Przykłady kodowania różnego typu informacji słownikowych przy pomocy schematów DTD/XML podaje m.in. Trap-Jensen (2006).

Ostatnią zaletą XML, dowodzącą jego elastyczności, jest fakt, że jest on dobrym formatem przejściowym, a dane w nim zawarte w razie potrzeby można zakodować w ramach bazy danych. Przykładem podejścia bazodanowego, alternatywnego względem XML, ale i zasadniejszego ze względu na swoją strukturę, jest digitalizacja *Słownika języka Jana Chryzostoma Paska*, którą opisuje Majewska (2010).

³² Tutaj należy zaznaczyć, że JSON i YAML są lepiej dostosowane do interfejsów programistycznych i pełnienia funkcji plików konfiguracyjnych, także ze względu na kompaktowość zapisu.

³³ Por.: <https://www.educba.com/json-vs-xml/>.

7.4. Technologia DjVu

Technologia DjVu została opracowana już niemal ćwierć wieku temu, a jej głównym zadaniem jest archiwizacja i publikacja zeskanowanych dokumentów w Internecie. Technologię tę Le Cun et al. (2001: 2) definiują w następujący sposób:

[A]n image compression technique, a document format, and a software platform for delivering documents images over the Internet.

[T]echnika kompresji obrazu, format dokumentu i platforma oprogramowania do dostarczania obrazów dokumentów przez Internet.

Szczegóły funkcjonowania formatu DjVu i jego aspekty techniczne można znaleźć w przytoczonym wyżej artykule. Istotnych jest tu jednak tylko kilka kwestii.

Od strony strukturalnej dokument DjVu posiada warstwę graficzną – składającą się z trzech warstw składowych: tła (ang. *background*), pierwszego planu (ang. *foreground*) oraz szablonu (ang. *stencil*) – oraz ukrytej warstwy tekstowej (ang. *hidden text layer*) (Bień 2009a: 164). Pozwala to na przeszukiwanie i kopiowanie tekstu z jednoczesnym zachowaniem oryginalnego wyglądu dokumentu.

Dokument DjVu może też przyjąć jedną z dwóch postaci:

1. pojedynczych plików `.djvu` odpowiadających poszczególnym stronom dokumentu, wraz z opcjonalnym plikiem indeksującym w tym samym formacie, zawierającym informacje o kolejności stron;
2. skonsolidowaną (ang. *bundled document*), która zawiera wszystkie strony w pojedynczym archiwum `.djvu`.

Forma „rozbita” ma przede wszystkim zastosowanie w przypadku przeglądania dokumentów w Internecie, ponieważ nie wymaga ładowania wszystkich stron, a jedynie dynamicznie pobiera te strony, które użytkownik chce przejrzeć. Z drugiej strony forma skonsolidowana jest prostsza w przesyłaniu i archiwizowaniu. Bień (2009a: 164) pisząc o „słownikach jako tekstach” zwraca jednak uwagę na inne zagadnienie z tym związane:

Every DjVu viewer allows for searching the hidden text layer, but for unbundled documents it is unefficient as it defeats the purpose of splitting the document into separate pages: to access the hidden text, all the pages have to be loaded, and if the search is repeated, they are reloaded multiple times.

Każda przeglądarka DjVu umożliwia przeszukiwanie warstwy tekstu ukrytego, ale w przypadku dokumentów niepowiązanych jest to nieefektywne, ponieważ niweczy cel dzielenia dokumentu na pojedyncze strony: aby uzyskać dostęp do ukrytego tekstu, wszystkie strony muszą zostać załadowane, a jeśli wyszukiwanie jest powtarzane, są ładowane wielokrotnie.

Dzięki swojej dwoistej naturze, tj. posiadaniu warstwy graficznej i tekstowej, dokumenty w formacie DjVu mogą zarówno być produktem procesu OCR, jak i stanowić dane wejściowe dzięki zawieraniu danych graficznych. Od 2001 roku również i format

PDF posiada taką strukturę. Nie zmienia to faktu, że w FBC w 2009 r. 72% wszystkich archiwów było w formie DjVu (Bień 2009a: 164) i format ten nadal funkcjonuje obok plików PDF oraz archiwów z obrazami rastrowymi. Pliki DjVu i PDF są w dużym stopniu zamienne, a jeśli pominąć warstwę tekstową, to nie różnią się praktycznie od archiwum z obrazami.

To jednak właśnie w technologii DjVu udostępniona została zeskanowana wersja *Słownika warszawskiego*, i to ona w tejże pracy doktorskiej posłużyła za źródło, z którego zaczerpnięty będzie materiał badawczy.

7.5. Oprogramowanie i język LaTeX

LaTeX, a właściwie L^AT_EX (wym. /'la:tɛx/, pol. [latech])³⁴, to oprogramowanie oraz system makr wykorzystywany do zautomatyzowanego składu tekstu. L^AT_EX wykorzystujemy system TeX, w którym sam też jest napisany. Oprócz makr standardowych posiada także bogate repozytorium, z których automatycznie pobiera te wskazane przez użytkownika.

Zasadniczą strukturę języka przedstawia Bień (1990: 132–133):

W systemach *adnotacyjnych* (ang. *markup systems*) właściwy tekst opatruje się – w trakcie wprowadzania go do komputera za pomocą osobnego programu nazywanego *edytorem* – specjalnymi adnotacjami, stanowiącymi mniej lub bardziej szczegółowe instrukcje opisujące wygląd finalnej postaci tekstu.

(...)

W systemie TeX sposób formowania tekstu opisujemy, wplatając we właściwy tekst odpowiednie *komendy*, co – z punktu widzenia użytkownika – czyni go bardzo zbliżonym do systemów adnotacyjnych.

Język ten nie nadaje się do innych zastosowań niż skład tekstu, ponieważ nawet funkcjonalność programowania warunkowego nie stanowi jego podstawowego wyposażenia – jest jednak silnie ugruntowany w swojej niszy, ponieważ praktycznie nie istnieją dla niego żadne alternatywy.

Do przeglądania i testowania kodu pomocniczo wykorzystany został darmowy program TeXstudio³⁵ w wersji 3.1.2.

7.6. Graficzna obróbka haseł

Piszący te z kilku powodów stwierdził, iż każde hasło w formie tekstowej warto byłoby wzbogacić powiązując je z odpowiednim fragmentem obrazu, który ilustrującego jego wygląd w oryginale. Nie udało się jednak znaleźć narzędzi stosownych do ekscerpcji

³⁴ Strona domowa systemu L^AT_EX: <https://www.latex-project.org/>.

³⁵ Strona domowa oprogramowania TeXstudio: <https://texstudio.org/>.

fragmentów SW odpowiadającym poszczególnym kartom. W związku z tym zostało opracowane własne oprogramowanie do automatycznej ekscerpcji, które zostało szczegółowo opisane w **rozdziale 8**.

W przypadku niektórych obrazów konieczna była też ich ręczna obróbka zarówno na wejściu, jak i na wyjściu rozwiązania automatycznego. Jako że komercyjne edytory graficzne wiążą się często z dużymi kosztami, a nie oferują niezbędnych funkcji względem oprogramowania darmowego, podjęto decyzję o użyciu w tym celu programu GIMP (GNU Image Manipulation Program)³⁶ w wersji 2.8. Zakres, w jakim program ten został zastosowany, oraz wykorzystane funkcje również opisane są we wspomnianym powyżej **rozdziale 8**.

7.7. Parsowanie struktury hasłowej

Jako że nie udało się przystosować żadnego oprogramowania do takiego przetworzenia warstwy tekstowej SW, aby dało pożądaną rezultat, zostało napisane własne dedykowane narzędzie, które jest przystosowane do treści i struktury badanego dzieła, i które w kolejnych krokach wykorzystuje sparsowaną treść do wygenerowania danych pochodnych. Dokładny opis kluczowych elementów działania programu znajduje się w **rozdziale 10**.

7.8. Optyczne rozpoznawanie znaków

Aby uniknąć żmudnego ręcznego przepisywania tekstu, wykorzystana została technologia optycznego rozpoznawania znaków (ang. *optical character recognition*, OCR), zwana także rozpoznawaniem tekstu, która jest jedną z poddziedzin widzenia komputerowego. Technologia ta jest zagadnieniem samym w sobie – więcej o metodologii pisze Bień (2012a).

Cały proces rozpoznawania znaków jest dosyć złożony i nie będzie tutaj szczegółowo opisywany, opiera się on jednak o trzy zasadnicze współgrające ze sobą mechanizmy:

1. rozpoznawanie próbek obrazów poszczególnych znaków i porównywanie ich z bazą przykładowych rastrowych realizacji znaku;
2. rozpoznawanie kształtów poszczególnych znaków i porównywanie ich z wyidealizowaną, wektorową budową znaku;
3. porównywanie rozpoznanych wyrazów ze słownikiem.

Na rynku dostępne są zarówno rozwiązania komercyjne, jak i darmowe (tu warto wspomnieć o silniku Tesseract³⁷). Piszący te słowa posłużył się komercyjnym programem ABBY FineReader (wersje 12, 14 i 15) do rozpoznania tekstu *Słownika warszawskiego*, a wybór ten motywuje trojako:

³⁶ GNU Image Manipulation Program: <https://www.gimp.org/>.

³⁷ Strona projektu w repozytorium GitHub: <https://github.com/tesseract-ocr/tesseract>.

1. relatywnie wysoką jakością danych wyjściowych w przypadku materiału badawczego;
2. preferencją co do interfejsu użytkownika i sposobu prezentowania danych;
3. wieloletnim doświadczeniu z oprogramowaniem tego producenta.

Program ten posiada także możliwość trenowania rozpoznawania tekstu pod nadzorem, co potrafi pozytywnie wpłynąć na jakość rozpoznawania. Jednak materiał badawczy analizowany w tej pracy jest bardzo mały, przez co wytrenowanie mechanizmu na kilku–kilkunastu stronach doprowadziło do pogorszenia się rezultatu względem wbudowanych schematów.

Niestety, programy do OCR – w tym ABBY FineReader – posiadają słowniki wspierające rozpoznawanie tekstu tylko dla współczesnych form języka, nieuwzględniających historycznych zmian w ortografii, przez co wiele wyrazów zostało błędnie „poprawianych” na formy współczesne lub do nich zbliżone. Słowniki takie można jednak tworzyć. O ile opracowanie takiego zasobu na podstawie *SW* do lepszego rozpoznania tekstu w tymże dziele jest przykładem myślenia cyrkularnego, to pełną cyfrową treść tegoż dzieła można by wykorzystać do retrodigitalizacji opublikowanych tekstów, które są względem *Słownika...* współczesne lub od niego starsze. Efekt taki można osiągnąć, z pewnymi dodatkowymi krokami, poprzez połączenie ze sobą danych zawartych w **Aneksie G** z danymi zawartymi w **Aneksach A** i **C**, co pozwoliłoby na uzyskanie pełnych tablic paradygmatycznych.

7.9. Metoda wprowadzania danych

Jednym z problemów podczas tworzenia cyfrowej reprezentacji treści *Słownika warszawskiego* jest konieczność wprowadzania poprawek do rezultatu rozpoznawania tekstu – o ile większość znaków składających się na tekst jest łatwych do wprowadzenia na standardowej klawiaturze polskiej, to część z nich może przysparzać pewnych problemów³⁸.

Przez znaki niestandardowe rozumie się tu znaki, które należą do standardu Unicode, ale które nie występują w standardowym polskim układzie klawiatury. Mowa tu przede wszystkim o literach, które nie występują w polskim alfabecie – w tym o literach łacińskich ze znakami diakrytycznymi nieobecnych w polszczyźnie – oraz znakach specjalnych, które nie są alfanumeryczne ani interpunkcyjne.

Znaki takie pojawiają się niekiedy w treści haseł, ale stosowane są także do oznaczenia haseł w funkcji kwalifikatorów, np. × czy †. W „Indeksie *a tergo* haseł *Słownika warszawskiego*” Wołosz obszedł ten problem stosując własny zapis takich symboli³⁹:

³⁸ Lista wszystkich znaków występujących w materiale badawczym znajduje się w **Aneksie E**.

³⁹ „Indeks *a tergo* haseł *Słownika warszawskiego*” został opublikowany przez Magdalенę Derwojedową 15.01.2017 r. na stronie *Automatyczna analiza fleksyjna polszczyzny XIX wieku* (dostępnej pod adresem: <http://www.f19.uw.edu.pl/2017/01/lista-hasel-slownika-warszawskiego/>); piszący te słowa opiera się również na nieopublikowanym opisie znaczników i instrukcji do jej tworzenia, udostępnionym przez Roberta Wołosza, za co pragnie mu podziękować.

- † jako /+;
- ! jako /!;
- × jako /x;
- [] jako /g;
- * jako /*.

W podobny sposób badacz ten zapisał także dezambiguatory numeryczne: **1.** jako /1, **2.** jako /2 itd., np.:

- /1 ekstra;
- ekstra-;
- /1 /x ekstra;
- /2 /x ekstra.

Autor tych słów postanowił jednak od początku wprowadzać wszystkie znaki w docelowej formie, aby uniknąć problemów z ich konwersją na dalszych etapach. Do tego celu zaprojektował własny układ klawiatury przy pomocy programu Microsoft Keyboard Layout Creator (MSKLC) w wersji 1.4⁴⁰. Przez układ klawiatury rozumie się tutaj funkcjonalne mapowanie fizycznych przycisków klawiatury na znaki, które przy ich wciśnięciu zostają wprowadzone na komputerze.

Zaprojektowany układ został nazwany Polyglot i pozwala na wprowadzenie wszystkich znaków występujących w SW więcej niż raz. Oparta jest ona o najczęściej stosowane rozwiązanie dla języka polskiego o nazwie *Klawiatura programisty (QWERTY)*. Główne zmiany względem oryginału obejmują dodanie przydanych znaków niealfanumerycznych⁴¹ oraz umożliwienie wprowadzania dziewięciu rodzajów tzw. składających znaków diakrytycznych (widocznych w górnym rzędzie na **ilustracji 19**, zaznaczonych jaśniejszym odcieniem koloru szarego), które automatycznie łączą się z bezpośrednio wprowadzoną po nich literą w niemal dowolnej kombinacji (np. makron $\bar{\text{a}}$ oraz akcent ostry $\acute{\text{e}}$ w połączeniu z literą e pozwalają wyświetlić znak $\acute{\text{e}}$). Pełen schemat możliwych do wprowadzenia znaków przedstawiony jest na **ilustracjach 17–20**.

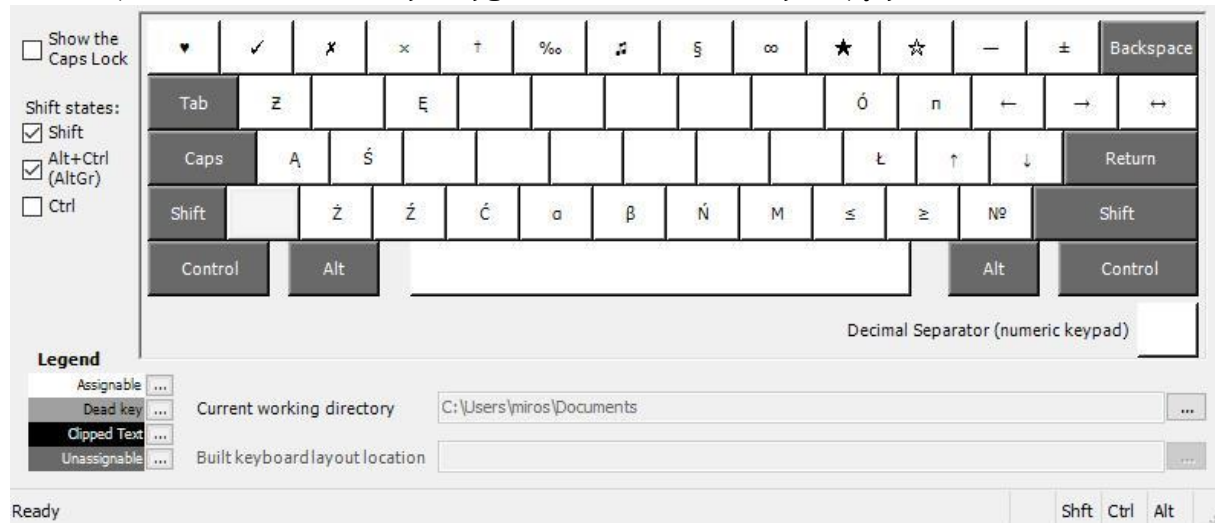
Po opracowaniu pliku źródłowego i skompilowaniu go w programie MSKLC uzyskiwany jest instalator, który pozwala zainstalować układ klawiatury na każdej maszynie z systemem operacyjnym firmy Microsoft od Windows XP wzwyż.

Dzięki klawiaturze Polyglot praca nad korektą rezultatu rozpoznawania tekstu została usprawniona, a jego dalsza przetwarzanie nie było konieczne.

⁴⁰ Program jest darmowy i dostępny do pobrania na stronie firmy Microsoft: <https://www.microsoft.com/en-us/download/details.aspx?id=102134>.

⁴¹ Należy tutaj zaznaczyć, że układ klawiatury Polyglot była stosowana także do innych celów, dlatego zawiera również znaki nieprzydatne z punktu widzenia digitalizacji *Słownika warszawskiego*.

Ilustracja 20. Układ klawiatury Polyglot z klawiszami modyfikującymi SHIFT oraz ALT.



7.10. Edycja tekstu

Do edycji danych opartych o zwykły tekst (ang. *plain text*), czyli niezawierający formatowania takiego jak kroje czcionek ani obiektów nietekstowych (tabeli, obrazów itd.), wykorzystany został program Notepad++⁴². Program ten jest przeznaczony do edycji kodu źródłowego, jednak jego zastosowania są dużo szersze dzięki funkcjonalności takiej jak pionowe zaznaczanie tekstu, wbudowany system makr, wielojęzyczne sprawdzanie pisowni czy obsługa wyrażeń regularnych (którym poświęcony jest **podrozdział 7.12**).

W ramach niniejszej tezy doktorskiej został wykorzystany m.in. do drugiej fazy czyszczenia tekstu powstałego w wyniku OCR, a także edycji plików wejściowych (.txt), wyjściowych (.xml) oraz plików pomocniczych w różnych formatach (.tab, .csv oraz .swp) generowanych przez parser⁴³ autorstwa piszącego te słowa.

Do edycji samego kodu źródłowego został wykorzystany jednak komercyjny program Sublime Text 3⁴⁴, który oferuje nowocześniejszy interfejs i liczne rozwiązania ułatwiające pisanie kodu programu.

7.11. Porównywanie tekstu

Specjalnym przypadkiem edycji tekstu jest jego porównywanie. Jest to szczególnie istotne w przypadku modyfikacji kodu programu, którego pliki wyjściowe są w formatach tekstowych, i kiedy modyfikacje te zmieniają otrzymywane dane w trudny do śledzenia sposób ze względu na nieliniową propagację.

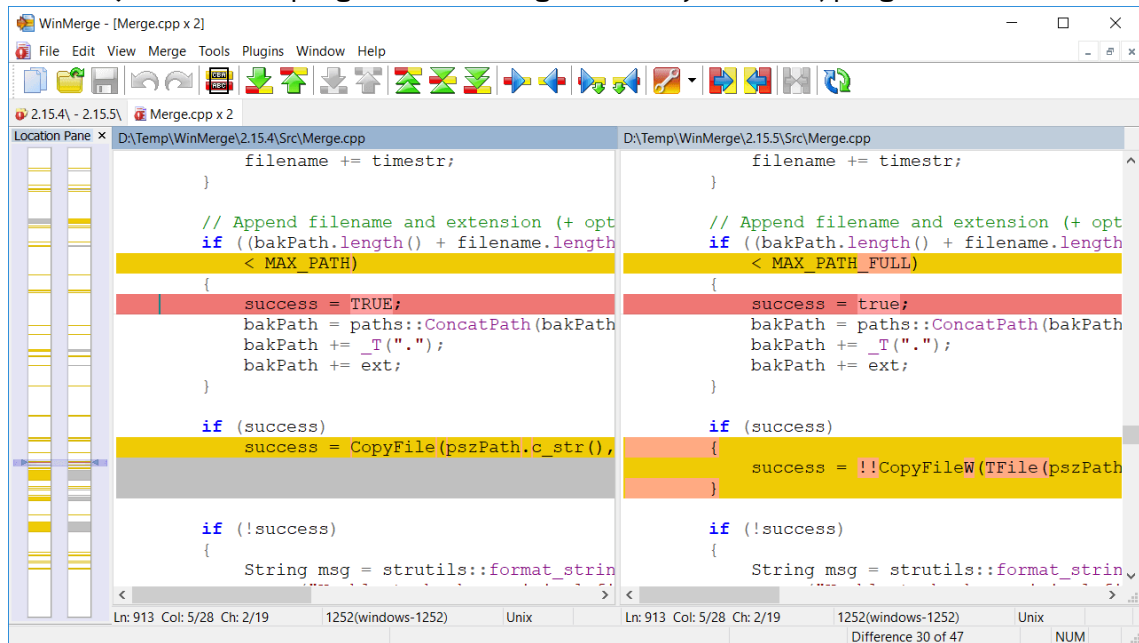
⁴² Program jest darmowy i do pobrania na stronie: <https://notepad-plus-plus.org/downloads/>.

⁴³ Parser ten jest opisany w **rozdziale 10**.

⁴⁴ Strona programu: <https://www.sublimetext.com/>; z programu można bezpłatnie korzystać przez okres próbny.

Do śledzenia takich zmian między kolejnymi iteracjami wykorzystany został otwartoźródłowy program WinMerge⁴⁵, który umożliwia porównywanie zawartości plików, prezentując graficznie różnice między wskazanymi dokumentami.

Ilustracja 21. Zrzut z programu WinMerge ze strony domowej programu.



Porównywanie tekstu zostało także zaprogramowane jako część programu SWParser, co jest opisane w **rozdziale 10.4.11**, a co ma za zadanie posłużyć prezentacji rezultatów.

7.12. Wyrażenia regularne

Wyrażenia regularne (ang. *regular expressions*, RegEx)⁴⁶ to rodzaj gramatyki formalnej, dzięki której można przeprowadzić złożone wyszukiwanie i modyfikację tekstu spełniającego określone kryteria. Oznacza to, że przy ich pomocy jest możliwe nie tylko znalezienie konkretnych ciągów znaków, ale także elementów o zmiennej treści⁴⁷. Mimo pewnych problemów⁴⁸, przez mnogość zastosowań – od walidacji poprawności adresów e-mail, przez automatyczną edycję tekstu aż po gry⁴⁹ – wyrażenia regularne są praktycznie wszechobecne.

⁴⁵ Strona programu: <https://winmerge.org/>.

⁴⁶ Źródłem poświęconym wyrażeniom regularnym, do którego piszący te słowa wraca najczęściej ze względu na jego kompleksowość, jest strona: <https://www.regular-expressions.info/>.

⁴⁷ Kanał Computerphile na Youtube poświęcił temu zagadnieniu aż trzy odcinki z prof. Brailsfordem, które stanowią solidne wprowadzenie do tego zagadnienia.

⁴⁸ Popularną krytyką wyrażen regularnych jest stwierdzenie: „Regular expressions are anything but regular” (Wszystko można powiedzieć o wyrażeniach regularnych, tylko nie to, że są regularne).

⁴⁹ Bardzo ciekawym sposobem na oswojenie się z wyrażeniami regularnymi jest strona *Regex Crossword*: <https://regexcrossword.com/>, która uczy ich poprzez grę zbliżoną zasadami do sudoku.

W ramach tej pracy zostały one wykorzystane dwójako: po pierwsze do półautomatycznego wyszukiwania i poprawiania błędów OCR za pomocą programu Notepad++, a po drugie do parsowania struktury hasłowej dzięki integracji wyrażeń regularnych z językiem programowania Ruby, co zostało opisane w **rozdziale 10.3**.

Wyrażenia regularne mają pewne istotne ograniczenia – przede wszystkim tworzenie zapytań negatywnych czy restrykcje w patrzeniu w przód oraz w tył (ang. *lookahead* i *lookbehind*). Słyną one także z niskiej wydajności, jednak operacje na tekście nie są dla współczesnych maszyn istotnym wyzwaniem obliczeniowym w większości zastosowań. W związku z tym sugeruje się takie ich tworzenie, w którym priorytetem jest ich czytelność, zrozumiałość i możliwość utrzymania. Nie oznacza to, że nie można wydajności tej zwiększyć, zwłaszcza poprzez eliminację tzw. *runaway regular expressions* czy zastosowanie tzw. grup atomowych.

Wyrażenia regularne powinny zatem wejść do repertuaru narzędzi każdego specjalisty pracującego z komputerowym przetwarzaniem tekstu.

8. Ekstrakcja kart hasłowych

8.1. Wprowadzenie

Jak już wcześniej wspomniano, praktyczna część niniejszej tezy doktorskiej ma na celu przede wszystkim ekstrakcję danych w taki sposób, aby zdjąć ograniczenia nałożone przez tradycyjną formę słownika i uzyskać je tym samym w nowym, hipertekstowym formacie. Wymaga to jednak wyekstrahowania z warstwy graficznej słownika (czyli obrazów będących skanami poszczególnych stron) warstwy tekstowej przy pomocy technologii OCR.

Nie oznacza to jednak, że oryginalna informacja – a raczej: jej cyfrowa reprezentacja – staje się zupełnie bezwartościowa. Obrazy stron oryginalnego słownika nie oferują nic poza pokazaniem autentycznego kontekstu (Dummer & Michaelis & Schlaefer 1998: 196). Sprawa wygląda nieco inaczej w przypadku graficznego wyodrębnienia poszczególnych haseł – może tu uprościć proces przetwarzania treści słownika oraz indeksację.

O wykorzystaniu elektronicznych fiszek słownikowych oraz ich indeksowaniu – także w powiązaniu z treściami o innej formie i pochodzeniu – pisze m.in. Bień (2016c). Zauważa on, że już w trakcie przetwarzania obrazów można poczynić pierwszy krok w stronę hipertekstowości i tym samym odejść od linearnego charakteru układu siatki hasłowej. Choć wykonane skany są odbiciem kart słownika i zachowują oryginalny układ haseł, to, jako że stanowią dane cyfrowe, pozwalają na manipulację; w tym przypadku oznaczałaby ona podzielenie obrazu z zeskanowanymi stronami na mniejsze, odpowiadające poszczególnym hasłom.

W większości kontekstów można by ten krok zwyczajnie pominąć – można przecież rozpoznać tekst na obrazach, obrazy odrzucić, i dopiero warstwę tekstową podzielić na poszczególne artykuły hasłowe. Są jednak co najmniej dwa powody, dla których w przypadku SW warto wygenerować obrazy odpowiadające poszczególnym hasłom.

Pierwszy powód to korzyści dla osoby przeprowadzającej retrodigitalizację SW. Główną z nich fakt, że dostępne wersje cyfrowe tego dzieła są skanami przedruku wcześniejszego wydania, którego jakość była ograniczona ówczesnie dostępną techniką drukarską. Oznacza to, że obrazy zawierają zakumulowaną ilość szumu i artefaktów pochodzenia analogowego. To z kolei sprawia, że oprogramowanie do OCR niepoprawnie rozpoznaje granice części akapitów (odpowiadających artykułom hasłowym), i piszącemu te słowa nie udało się w sposób dostatecznie spolegliwy przywrócić ich odpowiednią segmentację w sposób automatyczny. Problem ten jest szczególnie istotny, ponieważ miejsce początku i końca artykułu hasłowego są brane pod uwagę w trakcie parsowania jego struktury. Mając jednak do dyspozycji obrazy z poszczególnymi artykułami hasłowymi problem segmentacji jest rozwiązany.

Dodatkową korzyścią takiego rozwiązania jest natomiast fakt, że znacznie łatwiej jest weryfikować poprawność rozpoznanego tekstu względem obrazu źródłowego, jeśli porównuje się mniejsze, powiązane ze sobą fragmenty.

Drugi powód zorientowany jest na odbiorcę wygenerowanych danych. Wynika z założenia, iż zestawienie ustrukturyzowanych danych tekstowych z informacją w formie fragmentu zeskanowanego obrazu z oryginalnym zapisem treści artykułu hasłowego może być wartościowe dla użytkownika. Rozwiązanie takie ma więc zarazem wartość fotodokumentacyjną, jak i indeksacyjną, jako że pozwala na powiązanie fragmentu informacji graficznej z odpowiadającą informacją tekstową.

Samą ekstrakcję pożądaných fragmentów można dodatkowo wzbogacić o ich obróbkę graficzną. Obróbka taka miałaby trzy główne cele:

1. Zwiększenie czytelności obrazu dla człowieka.
2. Zwiększenie czytelności obrazu dla maszyny.
3. Skompresowanie ilości danych.

Aby tego dokonać, konieczne jest dobranie odpowiednich metod przetwarzania obrazów, które opisane są w rozdziale następnym.

8.2. Cyfrowe przetwarzanie obrazów

W niniejszym rozdziale piszący te słowa będzie poruszał się po dziedzinie, której nazwa brzmi cyfrowe przetwarzanie sygnałów (ang. *digital signal processing*, DSP); dziedzina ta obejmuje wszelką manipulację sygnałami informacyjnymi – czyli obrazami, dźwiękiem, video, a także pomiarami naukowymi czy transmisjami sterującymi. W ramach DSP najistotniejsze w tym rozdziale będzie oczywiście przetwarzanie obrazów, któremu poświęcona jest poddziedzina o nazwie cyfrowe przetwarzanie obrazów (ang. *digital image processing*, DIM). Szczególnym przypadkiem przetwarzania obrazów będzie widzenie komputerowe (ang. *computer vision*), czyli zbiór metod służących do rozpoznania elementów na obrazie.

Zasada działania cyfrowego przetwarzania obrazów jest ściśle powiązana z tym, w jaki sposób komputer przechowuje informacje. Otóż robi to przy pomocy tzw. kontenera (ang. *container*), który jest zorganizowaną strukturą danych; do różnych typów danych stosuje się różne kontenery. W przypadku obrazów są to przede wszystkim macierze⁵⁰. Macierz to, najprościej mówiąc, dwu- lub wielowymiarowy układ liczb, który w tym przypadku opisuje właściwości poszczególnych elementów obrazu – czyli pikseli (Rafajłowicz & Rafajłowicz 2008: 45ff).

Wszystkie operacje graficzne będą zatem niczym innym jak operacjami matematycznymi na macierzach. W ramach obróbki graficznej stosuje się przede wszystkim operacje:

⁵⁰ Co prawda można również zastosować tablice tablic lub listy tablic, jednak logiczna struktura zasadniczo nie różni się od macierzy.

- filtracji;
- binaryzacji;
- segmentacji;
- transformacji geometrycznej;
- transformacji pomiędzy przestrzeniami barw;
- morfologiczne;
- kompresji.

Tego typu czynności przeprowadzić z reguły można przekształceniem punktowym lub morfologicznym.

Przekształcenie punktowe (ang. *point transformation, coordinate transformation*) jest rodzajem operacji linearnej, a więc możliwą do zapisania przy pomocy funkcji matematycznej. Operacje tego rodzaju „przekształcają bezpośrednio każdy punkt obrazu źródłowego w odpowiadający mu punkt obrazu docelowego” (Rafajłowicz & Rafajłowicz 2008: 56). Przekształcenia takie zatem dla określonych danych wejściowych zawsze wygenerują takie same dane wyjściowe.

Podstawowymi przykładami operacji punktowych w kontekście przekształcania obrazów są:

- obrócenie;
- odbicie lustrzane;
- zmiana jasności;
- zmiana kontrastu;
- przetwarzanie histogramowe;
- wyrównanie kolorów.

Przekształcenie morfologiczne (ang. *morphological transformation*) polega natomiast na nieliniowym operowaniu na cechach obrazu. Oznacza to, że zamiast uwzględniać statyczne wartości konkretnych elementów, porównywane są one względem wartości elementów sąsiednich (Efford 2000: 271ff; Rafajłowicz & Rafajłowicz 2008: 193ff; Szczurek 2003: 70ff; Świtoński & Stapor 2004: 5ff; Kupidura et al. 2008: 323ff). Innymi słowy, algorytm, przetwarzając konkretny region grafiki, może wygenerować inne dane wyjściowe w zależności od tego, jakie inne regiony obrazu weźmie pod uwagę (np. mierząc średni kolor fragmentu obrazu).

Podstawowymi operacjami morfologicznymi są:

- erozja (ang. *erosion*), zwana także zwężaniem;
- dylatacja lub dylacja (ang. *dilation*);
- otwieranie (sekwencyjne wykonanie erozji i dylatacji);
- domykanie (sekwencyjne wykonanie dylatacji i erozji);
- rekonstrukcja;
- trafi-nie trafi (ang. *Hit and Miss*).

Mając do dyspozycji powyższe metody można przystąpić do pierwszego etapu przetwarzania obrazu, a mianowicie zwiększenia jego czytelności; mowa tutaj o łatwiejszej interpretacji jego zawartości zarówno dla człowieka, jak i algorytmów

rozpoznawania obrazu. Krok ten często skutkuje zmniejszeniem się rozmiaru danych, co wynika usunięcia szumów (ang. *noise*) i artefaktów (ang. *artifacts*), uproszczenia kształtów i ograniczenia kolorów, przez co komputer musi przechowywać mniej informacji. Jest to szczególnie zauważalne w przypadku analogowego pochodzenia obrazu.

Pierwszą czynnością jest redukcja kolorów zeskanowanego obrazu. Jest to istotne, ponieważ rozpoznawanie kształtów ma przede wszystkim charakter binarny: po zredukowaniu obrazu do czerni i bieli granica kształtów (takich jak na przykład litery) sprowadza się do znalezienia miejsc, gdzie piksele białe graniczą z pikselami czarnymi; wartości tych pikseli to kolejno zera i jedynki. Efekt taki można uzyskać poprzez progowanie (inaczej binaryzację), które to zamienia każdy piksel na biały, jeśli numeryczna reprezentacja jego koloru jest poniżej pewnej wartości, albo na czarny, jeśli powyżej.

Następnie należy wyrównać orientację tekstu poprzez obrócenie całego obrazu, co umożliwi wyznaczenie najmniejszych możliwych prostokątnych obszary, które będą obejmować treść poszczególnych artykułów hasłowych. Rezultatem takiego procesu są *de facto* cyfrowe fiszki z hasłami.

Po podzieleniu obrazów stron na mniejsze wskazane jest odfiltrować artefakty, czyli elementy nieznaczące – w tym przypadku zabrudzenia strony, skutki uboczne redukcji kolorów, a także fragmenty haseł sąsiadujących w przypadku, gdy oryginalna geometria tekstu nie była należycie skorygowana.

Sformalizowany w ten sposób zestaw instrukcji zarówno ułatwia badaczowi zadanie poprzez zdefiniowanie operacji do przeprowadzenia oraz ich ustaloną kolejność kolejność, jak i otwiera możliwość przetworzenia obrazu przy pomocy dwóch metod: manualnej i automatycznej. W pierwszej kolejności dokonano symulacji metody manualnej ze względu na to, że materiał badawczy stanowi procentowo niewielką część *SW*.

8.3. Symulacja metody manualnej

Do testu metody manualnej została wybrana pierwsza strona z hasłami na literę **I**, czyli strona 72 tomu II. Została ona przyjęta za reprezentatywną po pierwsze ze względu na jednolity charakter graficzny wszystkich stron z hasłami na literę **F**, a po drugie ze względu na to, że zawiera dokładnie 37 bloków tekstu⁵¹, co odpowiada średniej liczbie haseł na stronę⁵².

Literę **I** zaś wybrano zamiast **F** ze względu na początkowy plan, w którym przetwarzanie warstwy graficznej byłoby opracowane na podstawie haseł zaczynających się od tejże litery i przetestowane na hasłach na literę **F**; w przypadku parsowania warstwy tekstowej natomiast to litera **F** została przyjęta za podstawę, a litera **I** miała posłużyć do

⁵¹ Zazwyczaj jeden blok tekstu odpowiada jednemu artykułowi hasłowemu, jednak hasła podzielone między sąsiednie kolumny składają się z dwóch segmentów, co też miało miejsce na tej stronie – co odpowiada 36 definicjom (w tym jedna podzielona między dwie kolumny tekstu)

⁵² Jeśli przyjąć deklarowaną liczbę 280 000 artykułów hasłowych za prawdziwą i podzielić ją przez łączną liczbę 7 541 stron z hasłami we wszystkich ośmiu tomach.

weryfikacji rozwiązania. Ostatecznie jednak pomysł ten zarzucono ze względu na ograniczenia czasowe.

W ramach manualnej ekstrakcji fragmentów z artykułami hasłowymi wykonano następujące czynności przy pomocy programu GIMP:

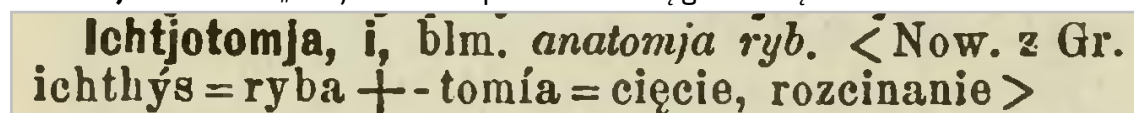
1. Zredukowano kolory do czerni i bieli przy pomocy progowania;
2. Skorygowano orientację i geometrię strony (kiedy było to konieczne);
3. Wyodrębniono regiony grafiki zawierające poszczególne artykuły hasłowe i umieszczono je w osobnych plikach;
4. Ręcznie usunięto artefakty.

Ukończenie wszystkich wymienionych czynności na stronie testowej zajęło łącznie około 30 minut, czyli średnio ok. 1 minuty na hasło. Najpierw jednak piszący te słowa kilkukrotnie przećwiczył cały proces na różnych stronach tak, aby wynik nie uwzględniał sobie czasu potrzebnego na naukę.

Prędkość obróbki takiej strony może jednak być zmienna przede wszystkim z dwóch powodów. Po pierwsze po przyzwyczajeniu się do zadania osoba nim obciążona będzie wykonywała je coraz szybciej, choć tylko do pewnego pułapu. Z drugiej strony powtarzalny charakter pracy może łatwo to doprowadzić do znużenia, co prędkość tę obniży. Po drugie ustalić różne optima pomiędzy czasem spędzonym nad daną partią obrazów poprzez kontrolę nad jakością danych wyjściowych.

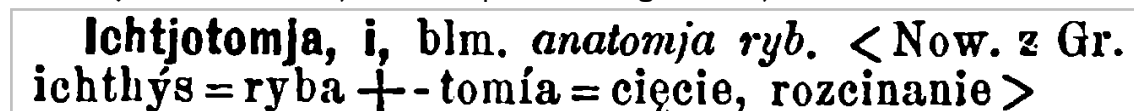
Jeśli jednak optymalizować cały proces pod kątem maksymalnej jakości, metoda ta pozwala nie tylko na precyzyjne usunięcie wszystkich artefaktów, jednocześnie zachowując lub zwiększając czytelność danego fragmentu tekstu. Optymalny rezultat ręcznej obróbki obrazu (**ilustracja 23**) dla artykułu hasłowego **Ichtjotomja** ze strony 72. Tomu II SW (**ilustracja 22**) przedstawiony jest poniżej:

Ilustracja 22. Hasło „Ichtjotomia” – przed obróbką graficzną.



Ichtjotomja, i, blm. anatomja ryb. < Now. z Gr. ichthýs = ryba † - tomía = cięcie, rozcinanie >

Ilustracja 23. Hasło „Ichtjotomia” – po obróbce graficznej.



Ichtjotomja, i, blm. anatomja ryb. < Now. z Gr. ichthýs = ryba † - tomía = cięcie, rozcinanie >

8.4. Wybór metody

Przed obraniem sposobu obróbki obrazów należy rozważyć wszystkie słabe i mocne strony metody manualnej i automatycznej w taki sposób, aby można była dobrać najwłaściwsze w kontekście rozważanego zadania podejście. Spośród czynników relewantnych warto wymienić rodzaj danych wejściowych, ich jakość oraz ilość, ograniczenia czasowe, ograniczenia sprzętowe, dostępne zasoby ludzkie oraz pożądany rezultat wyjściowy. Przewidywane rezultaty natomiast są tutaj oceniane z pięciu kategoriach: koszt wdrożenia, nakład pracy na przetworzenie pojedynczej strony, jakość wykrywania błędów, skalowalność rozwiązania i jakość obrazu wyjściowego.

8.4.1. Koszt wdrożenia

Ocena nakładu pracy koniecznego do wdrożenia rozwiązania jest uzależniona od kilku zmiennych. W przypadku metody manualnej najistotniejszą z nich jest początkowy poziom umiejętności osoby, której zadaniem jest obróbka graficzna obrazów. Jednak nakreślony wcześniej proces wymaga bardzo podstawowych umiejętności w zakresie pracy z grafiką. Jako że mowa o podstawowej funkcjonalności, obecnej we wszystkich wiodących programach do manipulacji obrazami, takich jak komercyjny Adobe Photoshop, czy darmowy i otwarty jak GIMP, oprogramowania ma znikomy wpływ na rezultat i tempo obróbki obrazów dla tak sparametryzowanej obróbki obrazów.

W przypadku metody automatycznej jest to nieco bardziej złożony problem: zależy od języka programowania, dostępnych bibliotek, a także umiejętności przewidzenia nieintuicyjnych dla człowieka problemów, które mogą wystąpić podczas przetwarzania materiału wejściowego poprzez algorytmy komputerowe. Rozstrzał między potencjalnym początkowym nakładem pracy dla obu metod jest dosyć istotny: nawet w przypadku doświadczonego programisty można śmiało zakładać, że stworzenie zautomatyzowanego rozwiązania będzie wymagało więcej czasu, choćby ze względu na jego testowanie.

8.4.2. Nakład na przetworzenie pojedynczej strony

Mając już wcześniej oszacowany nakład pracy potrzebny na manualne przetworzenie jednej strony, pozostaje oszacowanie nakładu pracy na pojedynczą stronę przy pomocy rozwiązania automatycznego. Można by było przyjąć, że w przypadku metody automatycznej czas ten będzie odpowiadał okresowi, który komputer potrzebuje na wykonanie algorytmu na danym obrazie na sprzęcie o konkretnych parametrach. W chwili pisania tych słów na przeciętnym komputerze osobistym przetworzenie jednej strony powinno, w zależności o parametrów programu, zająć od kilku sekund do minuty.

Powyższe podejście zakłada jednak pełną automatyzację, co może nie być do osiągnięcia, jeśli przyjąć pewną dolną granicę jakości danych wyjściowych. Oznacza to, że konieczny byłby udział człowieka w dane wejściowe lub dane wyjściowe. Nakład pracy w takim hybrydowym systemie mógłby być różny, jednak nie większy niż przy obróbce manualnej. Nie posiadając uprzedniego doświadczenia w dziedzinie komputerowego widzenia trudno przewidzieć, czy przy tak niewielkim zakresie danych

(kilkadziesiąt obrazów) nakład pracy na opracowanie metody automatycznej nie przekroczy sumarycznego czasu zaoszczędzonego na przetwarzaniu poszczególnych stron.

8.4.3. Podatność na błędy

Obie metody wykazują pewną podatność na błędy, jednak ich charakter jest zgoła różny. W przypadku obróbki przez człowieka błędy będą ograniczać się niemal tylko i wyłącznie do takich, które wynikają z nieuwagi lub są związane z ograniczeniami ludzkiego zmysłu wzroku.

W przypadku metody automatycznej natomiast błędy wynikać będą z tego, że ich autor podczas opracowywania nie uwzględnił wszystkich parametrów, które występują w danych. Problem ten dotyczy zarówno rozwiązań heurystycznych, jak i uczenia maszynowego. W bardziej ekstremalnych przypadkach program może zwrócić wyjątek, częściej jednak dane wyjściowe zwyczajnie będą inne od oczekiwanych. Można jednak temu w dużym stopniu zaradzić poprzez staranne dobranie próby badawczej. Metoda automatyczna pozwala przy tym na typową komputerom wysoką precyzję obrazów końcowych, jednocześnie eliminując szansę na błędy typowo ludzkie.

8.4.4. Skalowalność

W przypadku możliwości skalowania rozwiązanie automatyczne przeważa w sposób oczywisty. Naturalnie, skalowalność będzie zależała od czynników takich jak to, ile potrzeba ingerencji człowieka w dane wejściowe oraz wyjściowe, a także, jak wspomniano wyżej, na ile materiał testowy jest reprezentacyjny. Niezależnie od tego bezwzględny nakład pracy człowieka nie rośnie wcale albo rośnie w niewielkim stopniu względem ilości materiału do przetworzenia. Natomiast w przypadku metody alternatywnej jedyny sposób zwiększenia wydajności to skalowanie liniowe, tj. zwiększenie liczby osób przetwarzających obrazy – wówczas koszt przetworzonej jednostki pozostaje jednak ten sam niezależnie od wolumenu.

8.4.5. Jakość obrazu wyjściowego

W przypadku obu metod ostateczna jakość obrazu wyjściowego jest silnie uzależniona od parametrów, które zostaną przyjęte. Zważywszy na to, że jakość obrazu oceniana jest w głównej mierze przez korę wzrokową – która jest jednocześnie, obok oprogramowania, głównym narzędziem do obrabiania grafiki – w niemal każdym przypadku łatwo jest zmaksymalizować jakość danych wyjściowych (tj. ich czytelność oraz estetykę) przy odpowiednim nakładzie pracy.

W przypadku metody automatycznej silnie uzależnione jest to od problemów w materiale źródłowym oraz rozwiązań, które będą zastosowane. Nawet, jeśli przyjęte metody będą w pewnym stopniu adaptacyjne, nadal istnieje ryzyko napotkania zniekształceń, w przypadku których implementacja w pełni automatycznej metody może być bardzo trudna; mowa tutaj przede wszystkim o zniekształceniach geometrycznych, z którymi nie zawsze radzą sobie nawet wiodące programy do OCR.

8.4.6. Wnioski

Problem, przed którym staję tutaj piszący te słowa, jest dobrze znany informatyce i można sformułować go następująco: czy czas spędzony na automatyzacji (lub generalizacji czy optymalizacji) zadania nie przekroczy czasu, który dzięki temu byłby zaoszczędzony względem wykonania zadania bez tejże automatyzacji⁵³.

Jeśli zrobić rachunek zysków i strat ograniczając się tylko do materiału badawczego, to ocena nakładu czasu jest trudna: metoda automatyczna może go skrócić, wydłużyć, lub dać zbliżony rezultat. W przypadku niniejszej pracy zakładany materiał obejmuje 81 stron dla litery **F** (a wcześniej także 49 stron dla litery **I**, co dałoby łącznie 130 stron). Potencjalny zysk czasu wynikający z zastosowania metody automatycznej byłby w tym przypadku znikomy – i to przy założeniu, że wypracowanie i przetestowanie rozwiązania zajmie nie więcej niż 50–80 godzin, oraz że jego skuteczność (mierzona jakością danych wyjściowych) będzie na tyle wysoka, że ewentualna konieczność naniesienia poprawek na obrazy nie wydłuży znacząco łącznego czasu pracy.

Biorąc pod uwagę ryzyko włożenia istotnego nakładu wysiłku w opracowanie programistycznego rozwiązania, które może okazać się fiaskiem, co pociągnęłoby za sobą konieczność manualnej obróbki, być może rozsądniejszym byłoby pominąć próbę automatyzacji. Powyższe podejście ignoruje jednak dwa kluczowe założenia niniejszej pracy.

Po pierwsze ma ona stanowić nie tylko makietę, ale także przyczynek do digitalizacji *SW* w pełnym wymiarze. Jeśli opracować ręcznie cały *SW*, to liczba wymaganych roboczogodzin wyniosłaby prawie 4 000; przy takiej skali korzyści z automatyzacji są jest dużo bardziej prawdopodobne.

Po drugie, dla autora niniejszej tezy istotny jest walor edukacyjny, który niewątpliwie rozwiązanie automatyczne posiada. Perspektywa spędzenia (szacowanych) kilkunastu godzin na próbie znalezienia gotowych lub wypracowania własnych rozwiązań nie wydaje się przesadnym kosztem w skali ogółu nakładu pracy włożonej w niniejszą pracę doktorską.

8.5. Implementacja metody automatycznej

8.5.1. Strategia programistyczna

Jeszcze przed przystąpieniem do samego programowania piszący te słowa musiał podjąć szereg decyzji. Pierwszą z nich był wybór nadrzędnej strategii – czy podjąć próbę stworzenia rozwiązania algorytmicznego, heurystycznego, czy przy pomocy uczenia maszynowego.

⁵³ Problem jest na tyle rozpowszechniony, iż jeden z ulubionych komiksów piszącego te słowa, **xkcd**, poświęcił mu aż trzy różne numery – por. [xkcd 974: The General Problem](#), [xkcd 1205: Is It Worth the Time?](#) oraz [xkcd 1319: Automation](#), a także ich objaśnienia: [explain xkcd 974: The General Problem](#), [explain xkcd 1205: Is It Worth the Time?](#), oraz [explain xkcd 1319: Automation](#).

Ostatnie z wymienionych, czyli uczenie maszynowe (ang. *machine learning*), polega na zastosowaniu sieci neuronowych, które same uczą się na danych wejściowych próbując osiągnąć wskazany rezultat końcowy. Proces ten nazywa się trenowaniem. Jest wiele rodzajów trenowania sieci neuronowych. Wspólną bolączką niemal wszystkich z nich jest jednak fakt, że potrzebują stosunkowo dużych zbiorów danych. Nawet w chwili pisania tych słów w rozpoznawaniu obrazów za najmniejszą możliwą kolekcję uważa się liczbę tysiąca obrazów. Wdrożenie tego rozwiązania wymagałoby zatem znacznego poszerzenia zakresu materiału badawczego, a i tak nie gwarantowałyby pożądanego rezultatu. Nie mniej istotnym jest fakt, że w momencie początków pracy nad programem dostępność gotowych rozwiązań była niska, a próg uzyskania odpowiednich umiejętności w tej dziedzinie nie należał do najniższych. W związku z tym pomysł ten odrzucono.

Pozostały zatem metody algorytmiczne i heurystyczne. Metoda algorytmiczna to taka, w której określane są obliczenia, które musi podjąć komputer, aby uzyskać pożądanego efekt. Metoda heurystyczna natomiast to taka, która generuje efekt zbliżony do pożądanego w sytuacji, w której algorytmiczne uzyskanie najlepszego możliwego efektu jest trudne lub czasochłonne. Ostateczne rozwiązanie zostało uzyskane z połączenia tych metod.

8.5.2. Wybór źródła

Kolejnym problemem był wybór źródła, z którego zostaną zaczerpnięte obrazy. Do dyspozycji były cztery opcje:

1. wykorzystanie kopii cyfrowej dostępnej w Federacji Bibliotek Cyfrowych (FBC) na podstawie egzemplarza ze zbiorów Uniwersytetu Warszawskiego;
2. wykorzystanie kopii cyfrowej dostępnej na stronie instytucji Internet Archive wykonana na podstawie egzemplarza ze zbiorów Uniwersytetu w Toronto (UT);
3. wykonanie własnej kopii cyfrowej na podstawie egzemplarza ze zbiorów Biblioteki Uniwersyteckiej przy Uniwersytecie im. Adama Mickiewicza (BU UAM);
4. pozyskanie własnego egzemplarza i wykonanie własnej kopii cyfrowej na jego podstawie.

Najpierw podjęto próbę wykorzystania już istniejących kopii cyfrowych. Inicjalnie wybór padł na wersję FBC, ponieważ była ona już po wstępnym przetworzeniu obrazu. Niestety okazało się, że część informacji, które zostały utracone podczas przetwarzania, będzie potrzebne w procesie automatycznej obróbki. Dlatego też w drugiej kolejności wybrano wersję cyfrową UT.

Wykonując jednak w dalszym etapie testy OCR na obu wersjach (FBC i UT) jakość rozpoznanego tekstu była jednak na tyle niska, że piszący te słowa podjął próbę wykonania własnej kopii, aby jakość OCR wzrosła. Motywowane to było doświadczeniem z digitalizacji kilkuset różnych źródeł książkowych, w tym kilkudziesięciu słowników wydanych na przestrzeni od kilku dekad przed wydaniem SW do wybuchu II wojny

światowej. Niestety, skanery dostępne w BU UAM miały zbyt niską rozdzielczość, żeby wykonać kopię dającą lepsze rezultaty. Ponadto okazało się, iż problemy z dostępnymi wersjami wynikają przede wszystkim z jakości źródłowego przedruku, a nie samego procesu digitalizacji. W rezultacie zrezygnowano z pozyskania własnego egzemplarza i wykorzystano kopię słownika z UT.

Ostatecznie program został zoptymalizowany pod kątem litery **F**, ale opracowany w pierwszej kolejności na podstawie litery **I**.

8.5.3. Środowisko programistyczne

Inicjalnie piszący te słowa chciał napisać program w swoim preferowanym języku programowania, czyli języku Ruby⁵⁴. Niestety, język ten nie posiadał wówczas tak wystarczająco rozbudowanego wsparcia dla rozpoznawania obrazów. W wyniku tego wybór padł na podobny do Ruby'ego język programowania, Python⁵⁵, który takie wsparcie oferował.

Podobnie jak Ruby jest to język typowo skryptowy i dynamicznie interpretowany. Języki tego rodzaju, choć wydajnościowo ustępują językom kompilowanym, pozwalają na łatwiejsze korygowanie działania programów, zwłaszcza tych mniej wymagających obliczeniowo. Poza tym część bibliotek do Pythona – czyli gotowych rozwiązań dla pewnych klas problemów opracowanych przeważnie przez społeczność – napisanych jest w wydajniejszych od niego językach kompilowanych (takich jak C czy C++), co pozwala na połączenie najlepszych cech obu podejść.

Spośród wykorzystanych bibliotek należy wymienić dwie, które są absolutnie kluczowe: NumPy oraz OpenCV. Obie biblioteki zostały napisane w językach kompilowanych, przez co pozwalają programom napisanym w Pythonie korzystać z funkcji i klas przeznaczonych do wydajnego wykonywania operacji na liczbach.

NumPy (Numerical Python)⁵⁶ to biblioteka zawierająca zestaw narzędzi do przeprowadzania obliczeń na potrzeby naukowe. Najważniejszą, z punktu widzenia zaplanowanych w niniejszej pracy zadań, zaletą tej biblioteki jest możliwość przeprowadzania skomplikowanych obliczeń na tablicach wielowymiarowych; jak wcześniej wspomniano, to właśnie w takiej formie przeważnie przechowuje się dane graficzne.

OpenCV (Open Source Computer Vision Library)⁵⁷ to zestaw narzędzi przeznaczonych do manipulacji obrazów uczenia maszynowego opublikowanych na licencji BSD. Dzięki niej większość podstawowych i część zaawansowanych operacji na obrazach jest bardzo prosta.

Chociaż Ruby posiada swoje odpowiedniki funkcji z NumPy, to nie posiada wsparcia dla trzeciej generacji biblioteki OpenCV, która wprowadza wiele istotnych

⁵⁴ Więcej informacji o języku Ruby na stronie: <https://www.ruby-lang.org/pl/>.

⁵⁵ Więcej informacji o języku Python na stronie: <https://www.python.org/>.

⁵⁶ Więcej informacji o bibliotece NumPy na stronie: <http://www.numpy.org/>.

⁵⁷ Więcej informacji o bibliotece OpenCV na stronie: <http://opencv.org/about.html>.

udogodnień względem generacji poprzedniej, a z których część była kluczowa z punktu widzenia pisanego programu.

8.5.4. Dobre praktyki

W środowisku programistów bardzo rozpowszechnione jest pojęcie tzw. dobrych praktyk (ang. *good practices*). Oznacza to zbiór dyrektyw, dzięki którym unika się częstych w dziedzinie problemów. Dobre praktyki nie są silnie zdefiniowane – różnią się między języka, są zależne od zastosowania, a poza tym często brak konsensusu. Piszący te słowa wyznaje zasadę sformułowaną przez Kevlina Henney, która brzmi: „Dobry kod nie bierze się z dobrych praktyk. Dobry kod bierze się ze złego kodu.”. Jej przesłaniem jest to, i decydującym czynnikiem jest podczas pisania spójnego kodu jest doświadczenie, a nie reguły. Nie oznacza to jednak, że nie są one pomocne – szczególnie dla początkujących. Dlatego też autor tych słów pozwolił sobie wprowadzić kilka najbardziej podstawowych, którymi się kierował.

Modułowość. W programowaniu funkcjonalnym, jak sama nazwa wskazuje, ważną rolę odgrywają funkcje. Funkcje to zbiory instrukcji, do których można się odnieść w (niemal) dowolnym momencie programu. Można oczywiście podać wszystkie instrukcje sekwencyjne, ale wykorzystanie funkcji daje dwie istotne korzyści: pozwala uczynić logikę kodu prostszą, przez co jego rozwój i utrzymanie są łatwiejsze, a także wykorzystać ten sam kod w różnych kontekstach.

Przejrzystość. Nierzadko programista stoi przed następującym dylematem: czy napisać kod w sposób jak najbardziej przejrzysty, czy w sposób jak najbardziej wydajny. Piszący te słowa kierował się jednak zasadą najprzystępniej opisaną przez Harolda Abelsona, Geralda Sussmana i Julie Sussman, która brzmi: „(...) programs must be written for people to read, and only incidentally for machines to execute.” [(...) programy są pisane dla ludzi, a tylko niekiedy są wykonywane przez maszyny.] (1996: 15 p.)⁵⁸. Zgodnie z tą zasadą piszący te słowa starał się przede wszystkim zachować czytelność i łatwość modyfikacji kodu, a nie osiągnąć jego maksymalną wydajność.

Parametry. Wiele wartości, na podstawie których wykonywany jest program, jest zmienna zarówno w czasie produkcji kodu, jak i po niej. Modyfikowanie wartości wewnątrz kodu bywa problematyczne i często prowadzi do błędów. Dlatego warto jest odzielić logikę programu od parametrów; w najprostszej postaci oznacza to deklarowanie zmiennych dla najważniejszych parametrów, do których wykorzystuje się potem odwołania w różnych częściach programu.

Nazewnictwo. Większość dominujących języków programowania powstało w oparciu o język angielski, co narzuca pewne formalne ograniczenia w nazewnictwie właśnie pod kątem tego języka. Co więcej, międzynarodowa społeczność programistyczna jest prężna pod względem wzajemnej pomocy, przez co zarówno kod, jak i komentarze do niego warto pisać w tym własnym języku.

⁵⁸ Cytat pochodzi z jednego z ważniejszych dzieł w kanonie literatury programistycznej, co odzwierciedla zwyczajowa nazwa książki – *The Wizard Book*, czyli dosłownie ‘Księga czarodziejów’.

8.6. Działanie programu IndexCardSplitter

Poniżej opisane jest działanie programu o nazwie IndexCardSplitter (zwanego dalej ICS), który służy do ekstrakcji obrazów z poszczególnymi artykułami hasłowymi *SW*, będąc zarazem praktyczną implementacją opisanej wcześniej metodologii⁵⁹.

Program ten zaczyna swoje działanie od wczytania listy wszystkich zeskanowanych stron *SW*, które zostały wyeksportowane z dokumentu w formacie DjVu⁶⁰ do osobnych plików w formacie `.png`. Następnie otwiera pętlę, w ramach której dla każdego elementu listy wczytuje odpowiadający mu plik, a następnie wykonuje na nim zaprogramowane operacje (które są podzielone na kilka etapów), po czym zapisuje rezultat przetwarzania obrazu do plików wyjściowych i przechodzi do następnego pliku, i tak aż do ostatniego z elementów w pętli. Za przykład wczytywanej strony niech posłuży strona 72 tomu II *SW*, przedstawiona na **ilustracji 24**.

8.6.1. Etap pierwszy

Pierwszą operacją, jaką program wykonuje na obrazie, jest zredukowanie jego kolorów do odcieni szarości. Następnie zachodzi dalsza redukcja kolorów obrazu przy pomocy progowania, które wykonuje funkcji OpenCV o nazwie `cv2.threshold()`. Funkcja ta binaryzuje obraz, czyli sprowadza wartość każdego piksela do 0 lub 1, zależnie od sprecyzowanych parametrów. W tym przypadku posłużono się tzw. metodą Otsu⁶¹ – jest to rodzaj algorytmu binaryzacyjnego, który dla każdego piksela obrazu wylicza wartość pikseli sąsiadujących; na tej podstawie określa, które piksele są tłem, a które są częścią pierwszego planu, a następnie przypisuje im kolory (kolejno biel i czerń, czyli wartości 0 i 1). Rezultat binaryzacji strony 72. metodą Otsu można zobaczyć na **ilustracji 25**.

Spośród wypróbowanych sposobów progowania algorytm Otsu sprawdził się najlepiej – dzieje się tak, ponieważ metoda ta dynamicznie wylicza wartość progową. Jest to szczególnie istotne, gdy zeskanowane obrazy różnią się jasnością tekstu i ustawienie wspólnej wartości granicznej dla wszystkich z nich jest niepraktyczne: dla jednych wartość ta będzie za niska, przez co na obrazie pozostaną niechciane artefakty – podczas gdy dla innych będzie za wysoka, co skutkuje degradacją tekstu. Na **ilustracji 26** widać oryginalny skan strony 110., a na **ilustracji 27** rezultat progowania o parametrach, takich jakie były zastosowane na stronie 72.

8.6.2. Etap drugi

W ten sposób przetworzona strona pozwala na łatwe wyszukanie konturów, które posłużą do wyszukania tekstu na obrazie. Aby uzyskać taki efekt, najpierw konieczne jest odwrócenie kolorów obrazu przy pomocy funkcji `cv2.bitwise_not()`, a następnie przekazanie

⁵⁹ Powstawanie jego kodu było równoległe z nauką podstaw języka Python oraz specyfiki funkcjonowania bibliotek NumPy i OpenCV przez piszącego te słowa, co wiązało się z przetestowaniem szeregu metod, z których większa część została ostatecznie odrzucona. Opisany tu został efekt końcowy.

⁶⁰ O współczesnym użytkowaniu DjVu i obejściu problemów technicznych pisze Bień (2020).

⁶¹ Nazwanej tak od nazwiska jej twórcy – Nobuyuki Otsu.

rezultatu takiego przekształcenia funkcji `cv2.getContours()`. Ta ostatnia wyszukuje regiony, które wyznaczają granice czarnych i białych pikseli na obrazie, następnie je obrysowuje, po czym zwraca tablicę ze współrzędnymi znalezionych na obrazie obiektów.

Obraz został jednak przedtem poddany erozji. Erozja zamienia czarne piksele na białe, jeśli w ich sąsiedztwie dominują te drugie. Innymi słowy, usuwane są czarne piksele na brzegach kształtów, a także samotne wyspy na ekranie, usuwając tym samym mniejsze artefakty. Zbyt silna erozja może jednak doprowadzić do degradacji tekstu – podobnie jak w przypadku progowania – który jest wykorzystywany przy kolejnej operacji, skutkiem czego koniecznością było ustawienie erozji na takim poziomie, który pozostawiał część artefaktów.

Uzyskane kontury są w dalszej kolejności użyte do obrysowania tekstu przy pomocy funkcji `cv2.minAreaRect()`. Funkcja ta generuje najmniejszy możliwy prostokąt obejmujący wszystkie podane mu koordynaty. Jej istotnym atrybutem jest to, że zwraca kąt rotacji tego prostokąta: jeśli jego krótsze boki są idealnie poziome, a dłuższe idealnie pionowe, wówczas kąt rotacji wynosi 0; jeśli prostokąt jest obrócony w prawą lub lewą stronę, wartość ta przyjmuje kolejno wartości dodatnie lub ujemne.

Niekoniecznie dzieje się tak jednak w sytuacji, gdy prostokąt obrysowuje artefakty poza obszarem tekstu, których nie udało się usunąć przy pomocy erozji. Aby zmniejszyć ryzyko ich obrysowania, program ignoruje wszystkie obrysowane regiony, które znajdują się zbyt blisko krawędzi obrazu.

Rozwiązanie to nie zawsze jednak wystarczające, ponieważ artefakty mogą znaleźć się daleko od krawędzi, ale jednocześnie w takiej odległości od tekstu, co doprowadza do obrysowania prostokąta o złych wymiarach i wykryciu błędnej rotacji. Przykład takiej sytuacji przedstawia **ilustracja 29**, na której pojawił się taki artefakt (wskazany strzałką). W przypadku niektórych obrazów wejściowych w analizowanym zakresie najskuteczniejszym rozwiązaniem było ręczne usunięcie większych artefaktów, dzięki czemu wielkość i orientacja prostokątów została wykryta poprawnie, jak jest to widoczne na **ilustracji 30**.

Po wykryciu orientacji program na jej podstawie obraca cały obraz w taki sposób, aby boki prostokąta były równoległe z krawędziami obrazu. W większości przypadków jest to wystarczające, żeby skorygować orientację tekstu, co widać na **ilustracji 31**.

8.6.3. Etap trzeci

Po wyrównaniu orientacji obrazu, program przystępuje do przycięcia jego marginesów. Aby to osiągnąć, program mierzy średnią jasność pikseli najpierw dla każdego wiersza od góry do dołu, aż napotka na wiersz, którego jasność spadnie poniżej pewnej wartości granicznej – jest dobrana tak, aby była wyższa niż średnia jasność w linii z nagłówkiem czy numerem strony, ale niższa niż wartość linii, w którym zaczyna się tekst artykułów hasłowych. Następnie ta sama czynność wykonywana jest w pionie, tzn. dla każdej kolumny pikseli.

ICHNOGNOMIKA

× **Ichnognomika**, i, blm. sztuka dochodzenia śladów, tropienia. St. wil. <Now. z Gr. ichnos = ślad + gnōmē = poznawanie >

× **Ichnograficzny** przym. od **Ichnografja**.

× **Ichnografja**, i, blm. sztuka kreślenia planów. <Gr. ichnografja >

Ichtjofag, a, lm. owie człowiek, żywiący ś. tylko rybami. <Gr. ichtjofagos >

× **Ichtjograficzny** przym. od **Ichtjografja**.

× **Ichtjografja**, i, blm. opisanie ryb. St. wil. <Now. z Gr. ichtjys = ryba + grafia = opis >

Ichtjol, u, blm., **Sulfoichtjolan** amonu apt. związek amonjalkny kwasu sulfoichtjolowego, otrzymywanego przez suchą destylację kopalnych pozostałości ryb i zwierząt morskich. <Now. z Gr. ichtjys = ryba + lē. oleum = olej >

× **Ichtjolit**, u, lm. y skamieniały szkielet, żab, skamieniała ość ryby przedpotopowej. St. wil. <Now. z Gr. ichtjys = ryba + lithos = kamień >

× **Ichtjolitograficzny** przym. od **Ichtjolitografja**.

× **Ichtjolitografja**, i, blm., × **Ichtjolitologja** opisanie skamieniałości rybich. St. wil. <Now. z Gr. ichtjolit + grafia = opis, opisanie >

× **Ichtjolitologja**, i, blm. p. **Ichtjolitografja**. St. wil. <Now. z Gr. ichtjys = ryba + lithos = kamień + logia = znanstwo >

Ichtjolog, a, lm. owie badacz, zajmujący ś. nauką o rybach. <Now. z Gr. ichtjys = ryba + lōgos = uczyony >

Ichtjologiczny przym. od **Ichtjologja**.

Ichtjologja, i, blm. zool. nauka o rybach. <Now. z Gr. ichtjys = ryba + logia = nauka >

× **Ichtjomancja**, i, blm. wróżenie z ryb, z wnętrzości rybich. St. wil. <Now. z Gr. ichtjys = ryba + manteia = wróżba >

Ichtjomorficzny mający kształt rybi. <Now. z Gr. ichtjys = ryba + morfē = kształt >

Ichtjomorfit, u, lm. y, **Ichtjotopolit** odcisk rybi na kamieniu. St. wil. <p. Ichtjomorficzny >

Ichtjotomiczny przym. od **Ichtjotomja**.

Ichtjotomja, i, blm. anatomja ryb. <Now. z Gr. ichtjys = ryba + tomia = cięcie, rozcinanie >

Ichtjopolit, u, lm. y p. **Ichtjomorfit**. St. wil. <Now. z Gr. ichtjys = ryba + typos = ślad, odcisk + lithos = kamień >

[**Ioki**, ów, blp.] **pejsy**, loczki żydowskie. <Od imienia Żyd. Icek >

Iok, ioka, lm. ickowie p. Icek.

[**Iczme**, enia, lm. enie] p. **Ięczmień**.

Iozoglan, a, lm. i, **Iozoglan** pokojowiec turecki: Przy boku baszy grono iczoglanów czyli pokojowców tureckich. Szaj. Wezyr sadowił ś. na kosztownym dywanie, otoczony drabantami, iczoglanami, strażą honorową. Ról. <Tur. iczoglan >

Iozoglan, a, lm. l p. **Iczoglan**. Tw.

† **Ić**, idzie, szedł p. **Iść**: Abych była dostojna s tobą ić s duchwatością do królestwa. Ż. E. A ja też stąd już ić mogę. Rej. <ID >

Idący i. im. od **Iść**. 2. † a. † **Ruszysty** ruchomy: Dobra ruszyste a. idące. Imię idące p. **Imienie**.

Idea, i, lm. e, × **Ideja** i. † obraz, wizerunek, postać czegoś. 2. stan świadomości wog.: Idee-siły = stany świadomości, rozważane jako czynniki dynamiczne. 3. wyobrażenie: I. konkretna, ogólna. I. stała a. uporczywa (idée fixe) p. **Urojenie**. Kojarzanie idei (lep. wyobrażeń). 4. pojęcie. 5. w idealizmie platonowskim: pojęcie ogólne, rodzajowe; istota rzeczy; prototyp a. pierwowzór rzeczy zmysłowych. 6. w filozofji Kanta: pojęcie rozumo-

IDEALIZOWAC

we, któremu nic w doświadczeniu nie odpowiada; pojęcie wrodzone: Idee teoretyczne (pojęcia Boga, wolności, nieśmiertelności). Idee praktyczne (normy moralne). 7. myśl, pomysł, myśl przewodnia a. kierownicza: Powziąć ideę. Kierować ś. ideą. Poświęcić ś. dla idei. Działać w imię idei. I. narodowa, polityczna, naukowa, religijna. Żart.: Niema idei tak robić (= rozumnego celu). <Gr. idéa >

Ideacja, i, lm. e odtwarzanie obrazów i wyobrażeń w rozmaitych związkach i następstwie; gra wyobrażeń; działanie wyobraźni: I bezwiedna a. podświadoma. Por. **Cerebracja**. <p. Idea >

Idealk, u, lm. i p. **Ideal**: Życzę panu, aby mu ś. jego I. przyśnił u nas. Bał.

Idealista, y, lm. ści fil. 1. zwolennik, wyznawca, przedstawiciel idealizmu w filozofji a. estetyce. Przec. **Realista**. 2. ten, co w myślach i postępowaniu kieruje ś. ideami lub idealami; ten, co dla idei i idealów lekceważy sobie rzeczywistość, nie uwzględnia interesów praktycznych; marzyciel. Przec. **Materiałista**. <Fr. idéaliste, Nm. Idealist >

Idealistka, i, lm i fil. forma ż. od **Idealista**: Więc ty, o idealistko, swoją miłość rozciągasz aż na córy Helwecji? Orzesz.

Idealistycznie przys. od **Idealistyczny**: Stała ś. nagle rozmarzoną, I. nastrojoną. Kow.

Idealistyczny fil. przym. od **Idealizm**: Pogląd, kierunek I. Filozofja idealistyczna (bt. filozofja idealna). Czy to siła przekonania? czy idealistyczne zaślepienie? Gers. Przec. **Realistyczny**.

Idealizacja, i, blm. fil. p. **Ideallzowanie**: Ten realizm starego Willa jest na posługach u najpotężniejszej, jaka może istniała, idealizacji. Śpas. <p. Ideal >

Idealizm, u, blm. fil. 1. filozofja a. pogląd filozoficzny, który uznaje tylko pierwiastek idealny za był prawdziwie rzeczywisty i zaprzecza realności świata zmysłowego, in. I. metafizyczny. Przec. **Realizm**. I. podmiotowy a. subiektywny = uznający świat zmysłowy tylko za objaw, za wytwór ducha a. jaźni (Fichte). I. przedmiotowy a. obiektywny = według którego idea tkwi w samej rzeczywistości zmysłowej, będącej jej rozwinięciem (Hegel). I. realny = próby kompromisu, poglądy usiłujące pogodzić a. zszyteżować realizm z idealizmem. 2. w teorii poznania: pogląd, według którego wiedza nasza nie jest poznaniem rzeczywistości w istocie swojej niepoznawalnej, lecz jest systemem idei a. pojęć systematyzujących i symbolizujących doświadczenie nasze. 3. bt. filozofja uznająca był rzeczywisty tylko substancji duchowej. 4. w estetyce: kierunek a. doktryna estetyczna, podnosząca idealizowanie rzeczywistości jako konieczny warunek piękna, wymagająca od sztuki, by w dziełach swoich przeobrażała rzeczywistość w duchu idealów, podnosiła rzeczywistość do ideału, wcielała idee w obrazy rzeczywistości. Przec. **Realizm**. 5. kierunek myśli a. usposobienie lubujące ś. w ideach i idealach, skłonność do kierowania ś. ideami i idealami, z pominięciem warunków i interesów rzeczywistych; marzycielstwo: Chorować na I. Przec. **Materiałizm**. <Now. niby Łć. idealismus >

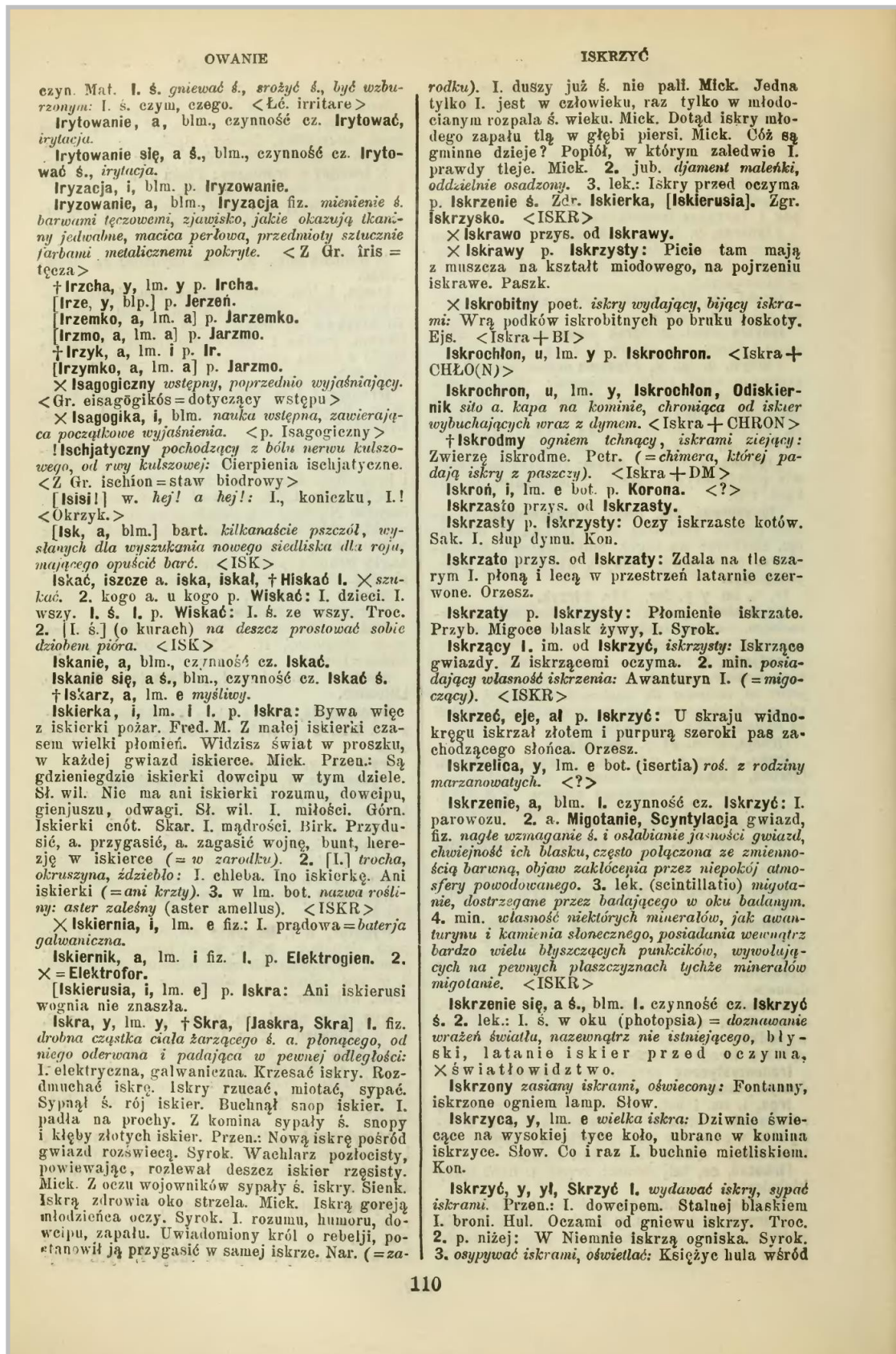
Idealizować, uje, ował fil. 1. wyobrażać sobie, postaciować rzeczywistość zgodnie z ideą; przenosić coś w dziedzinę idei a. ideału; przeobrażać coś, wzorując ś. na ideale. 2. podnosić coś do godności a. wartości ideału, ułuskonalac, usuwać a. pomijać braki, wady rzeczy a. kogoś. 3. przeceniać coś; nadawać przesadną wartość, rzeczywistości nieodpo-

ICHNOGNOMIKA

× **Ichnognomika**, i, blm. sztuka dochodzenia śladów, tropienia. Sł. wil. <Now. z Gr. ichnos = ślad + gnōmē = poznawanie >
 × **Ichnograficzny** przym. od **Ichnografja**.
 × **Ichnografja**, i, blm. sztuka kreślenia planów. <Gr. ichnografía >
Ichtjofag, a, lm. owia człowiek, żywiący ś. tylko rybami. <Gr. ichtjofagos >
 × **Ichtjograficzny** przym. od **Ichtjografja**.
 × **Ichtjografja**, i, blm. opisanie ryb. Sł. wil. <Now. z Gr. ichtjys = ryba + -grafia = opis >
Ichtjol, u, blm., Sulfoichtjolan amonu apt. związek amonjalkalny kwasu sulfoichtjolowego, otrzymany przez suchą destylację kopalnych pozostałości ryb i zwierząt morskich. <Now. z Gr. ichtjys = ryba + Lc. oleum = olej >
 × **Ichtjolit**, u, lm. y skamieniały szkielet, żąb, skamieniała ość ryby przedpotopowej. Sł. wil. <Now. z Gr. ichtjys = ryba + lithos = kamień >
 × **Ichtjolitograficzny** przym. od **Ichtjolitografja**.
 × **Ichtjolitografja**, i, blm., × **Ichtjolitologia** opisanie skamieniałości rybich. Sł. wil. <Now. z Gr. Ichtjolit + -grafia = opis, opisanie >
 × **Ichtjolitologia**, i, blm. p. **Ichtjolitografja**. Sł. wil. <Now. z Gr. ichtjys = ryba + lithos = kamień + -logia = znawstwo >
Ichtjolog, a, lm. owie badacz, zajmujący ś. nauką o rybach. <Now. z Gr. ichtjys = ryba + -lōgos = uczoney >
Ichtjologiczny przym. od **Ichtjologia**.
Ichtjologia, i, blm. zool. nauka o rybach. <Now. z Gr. ichtjys = ryba + -logia = nauka >
 × **Ichtjomancja**, i, blm. wróżenie z ryb, z wnętrzości rybich. Sł. wil. <Now. z Gr. ichtjys = ryba + manteia = wróżba >
Ichtjomorficzny mający kształt rybi. <Now. z Gr. ichtjys = ryba + morfē = kształt >
Ichtjomorfizm, u, lm. y, **Ichtjotypolit** odcisk rybi na kamieniu. Sł. wil. <p. Ichtjomorficzny >
Ichtjotomiczny przym. od **Ichtjotomia**.
Ichtjotomia, i, blm. anatomja ryb. <Now. z Gr. ichtjys = ryba + -tomia = cięcie, rozeźnianie >
Ichtjotypolit, u, lm. y p. **Ichtjomorfizm**. Sł. wil. <Now. z Gr. ichtjys = ryba + týpos = ślad, odcisk + lithos = kamień >
Icki, ów, blp. pejczy, loczki żydowskie. <Od imienia Żyd. Icek >
Ick, ioka, lm. ickowie p. Icek.
Iczme, enia, lm. enie] p. Iczmień.
Iczoglan, a, lm. i, Iczoglan pokojowiec turecki: Przy boku baszy grono Iczoglanów czyli pokojowców tureckich. Szaj. Wezyr sadowił ś. na kosztownym dywanie, otoczony drabantami, Iczoglanami, strażą honorową. Rol. <Tur. iczoglan >
Iczoglan, a, lm. i p. Iczoglan. Tw.
 † **Ić**, idzie, szedł p. Iść: Abych była dostojna s tobą ić s duchwałością do królestwa. Ż. E. A ja też stąd już ić mogę. Rej. <ID >
Idący i. im. od Iść. 2. † a. † **Ruszyste** ruchomy: Dobra ruszyste a. idące. Imienie idące p. Imienie.
Idea, i, lm. e, × **Ideja** i. † obraz, wizerunek, postać czegoś. 2. stan świadomości wog.: Idee - siły = stany świadomości, rozważane jako czynniki dynamiczne. 3. wyobrażenie: I. konkretna, ogólna. I. stała a. uporczywa (idée fixe) p. **Urojenie**. Kojarzenie idei (lep. wyobrażeń). 4. pojęcie. 5. w idealizmie platonowskim: pojęcie ogólne, rodzajowe; istota rzeczy; prototyp a. pierwowzór rzeczy zmysłowych. 6. w filozofji Kanta: pojęcie rozumno-

IDEALIZOWAC

we, któremu nic w doświadczeniu nie odpowiada; pojęcie wrodzone: Idee teoretyczne (pojęcia Boga, wolności, nieśmiertelności). Idee praktyczne (normy moralne). 7. myśl, pomysł, myśl przewodnia a. kierownicza: Powiąż ideę. Kierować ś. ideą. Poświęcić ś. dla idei. Działać w imię idei. I. narodowa, polityczna, naukowa, religijna. Żart.: Niema idei tak robić (=rozumnego celu). <Gr. idéa >
Ideacja, i, lm. e odtwarzanie obrazów i wyobrażeń w rozmaitych związkach i następstwie; gra wyobrażeń; działanie wyobraźni: I bezwiedna a. podświadoma. Por. **Cerebracja**. <p. Idea >
Idealk, u, lm. i p. **Ideal**: Życzę panu, aby mu ś. jego I. przysnił u nas. Bał.
Idealista, y, lm. ści fil. i. zwolennik, wyznawca, przedstawiciel idealizmu w filozofji a. estetyce. Przec. **Realista**. 2. ten, co w myślach i postępowaniu kieruje ś. ideami lub idealami; ten, co dla idei i idealów lekceważy sobie rzeczywistość, nie uwzględnia interesów praktycznych; marzyciel. Przec. **Materialista**. <Fr. idéaliste, Nm. Idealist >
Idealistka, i, lm i fil. forma ż. od **Idealista**: Więc ty, o idealistko, swoją miłość rozciągasz aż na córy Helwecji? Orzesz.
Idealistycznie przys. od **Idealistyczny**: Stała ś. nagle rozmarzoną, I. nastrojoną. Kow.
Idealistyczny fil. przym. od **Idealizm**: Pogląd, kierunek I. Filozofja idealistyczna (bl. filozofja idealna). Czy to siła przekonania? czy idealistyczne zaślepienie? Gers. Przec. **Realistyczny**.
Idealizacja, i, blm. fil. p. **Idealizowanie**: Ten realizm starego Willa jest na posługach u najpotężniejszej, jaka może istniała, idealizacji. Spas. <p. Ideal >
Idealizm, u, blm. fil. i. filozofja a. pogląd filozoficzny, który uznaje tylko pierwiastek idealny za byt prawdziwie rzeczywisty i zaprzecza realności świata zmysłowego, in. I. metafizyczny. Przec. **Realizm**. I. podmiotowy a. subiektywny = uznający świat zmysłowy tylko za objaw, za wytwór ducha a. jaźni (Fichte). I. przedmiotowy a. obiektywny = według którego idea tkwi w samej rzeczywistości zmysłowej, będącej jej rozwinięciem (Hegel). I. realny = próby kompromisu, poglądy usiłujące pogodzić a. zsyntezować realizm z idealizmem. 2. w teorii poznania: pogląd, według którego wiedza nasza nie jest poznaniem rzeczywistości w istocie swojej niepoznawalnej, lecz jest systemem idei a. pojęć systematyzujących i symbolizujących doświadczenia nasze. 3. bl. filozofja uznająca byt rzeczywisty tylko substancji duchowej. 4. w estetyce: kierunek a. doktryna estetyczna, podnosząca idealizowanie rzeczywistości jako konieczny warunek piękna, wymagająca od sztuki, by w dziełach swoich przeobrażała rzeczywistość w duchu idealów, podnosiła rzeczywistość do ideału, wcielała idee w obrazy rzeczywistości. Przec. **Realizm**. 5. kierunek myśli a. usposobienie lubujące ś. w ideach i idealach, skłonność do kierowania ś. ideami i idealami, z pominięciem warunków i interesów rzeczywistych; marzycielstwo: Chorować na I. Przec. **Materializm**. <Now. niby Lc. idealismus >
Idealizować, uje, owal fil. i. wyobrazać sobie, postaciować rzeczywistość zgodnie z ideą; przenosić coś w dziedzinę idei a. ideału; przeobrażać coś, wzorując ś. na ideale. 2. podnosić coś do godności a. wartości ideału, ułoskonaląć, usuwać a. pomijać braki, wady czegoś a. kogoś. 3. przeceniać coś; nadawać przesadną wartość, rzeczywistości nieodp-



OWANIE

czyn. Mat. I. s. gniewać s., srożyć s., być wzburzonym: I. s. czym, czego. <Łć. irritare>

Irytowanie, a, blm., czynność cz. Irytować, irytacja.

Irytowanie się, a s., blm., czynność cz. Irytować s., irytacja.

Iryzacja, i, blm. p. Iryzowanie.

Iryzowanie, a, blm., Iryzacja fiz. mienienie s. barwanami tęczowymi, zjawisko, jakie okazują tkaniny jedwabne, macica perłowa, przedmioty sztucznie farbami metalicznymi pokryte. <Z Gr. iris = tęcza>

† Irzcha, y, lm. y p. Ircha.

[Irze, y, blp.] p. Jerzeń.

[Irzemko, a, lm. a] p. Jarzemko.

[Irzmo, a, lm. a] p. Jarzmo.

† Irzyk, a, lm. i p. Ir.

[Irzymko, a, lm. a] p. Jarzmo.

× Isagogiczny wstępny, poprzednio wyjaśniający.

<Gr. eisagogikós = dotyczący wstępu>

× Isagogika, i, blm. nauka wstępna, zawierająca początkowe wyjaśnienia. <p. Isagogiczny>

! Ischjatyeczny pochodzący z bólu nerwu kulszowego, od ruy kulszowej: Cierpienia ischjatyeczne.

<Z Gr. ischion = staw biodrowy>

[Isisil] w. hej! a hej!: I, konieczku, I!

<Okrzyk.>

[Isk, a, blm.] bart. kilkanaście pszczół, wystawnych dla wyszukania nowego siedliska dla roju, mającego opuścić barć. <ISK>

Iskać, iszcze a. iska, iskał, † Hiskać I. × szukać. 2. kogo a. u kogo p. Wiskać: I. dzieci. I. wszy. I. s. I. p. Wiskać: I. s. ze wszy. Troc. 2. [I. s.] (o kurach) na deszcz prostować sobie dziobem pióra. <ISK>

Iskanie, a, blm., czynność cz. Iskać.

Iskanie się, a s., blm., czynność cz. Iskać s.

† Iskarz, a, lm. e myśliwy.

Iskierka, i, lm. i I. p. Iskra: Bywa więc z iskierki pożar. Fred. M. Z malej iskierki czasem wielki płomień. Widzisz świat w prosku, w każdej gwiazd iskierce. Mick. Przen.: Są gdzieniegdzie iskierki doweipu w tym dziele. Sł. wil. Nie ma ani iskierki rozumu, doweipu, gienjuszu, odwagi. Sł. wil. I. miłości. Gór. Iskierki enót. Skar. I. mądrości. Birk. Przydusić, a. przygasić, a. zagasić wojnę, bunt, herezję w iskierce (= w zarodku). 2. [I.] trocha, okruszyna, ździebło: I. chleba. Ino iskierkę. Ani iskierki (= ani krzty). 3. w lm. bot. nazwa rośliny: aster zaleśny (aster amellus). <ISKR>

× Iskierka, i, lm. e fiz.: I. prądowa = bateria galwaniczna.

Iskiernik, a, lm. i fiz. I. p. Elektrogien. 2. × = Elektrofor.

[Iskierusia, i, lm. e] p. Iskra: Ani iskierusi wognia nie znaszła.

Iskra, y, lm. y, † Skra, [Jaskra, Skra] I. fiz. drobna cząstka ciała żarzącego s. a. płonącego, od niego oderwana i padająca w pewnej odległości: I. elektryczna, galwaniczna. Krześć iskry. Rozdmuchać iskrę. Iskry rzucać, miotać, sypać. Sypnął s. rój iskier. Buchnął snop iskier. I. padła na prochy. Z komina sypały s. snopy i kłęby złotych iskier. Przen.: Nową iskrę pośród gwiazd rozświeca. Syrok. Waehlarz pozłocisty, powiewając, rozlewał deszcz iskier rześisty. Mick. Z oczu wojowników sypały s. iskry. Sienk. Iskrą zdrowia oko strzela. Mick. Iskrą goreją młodzieńca oczy. Syrok. I. rozumu, humoru, doweipu, zapału. Uwiadomiony król o rebelji, postanowił ją przygasić w samej iskrze. Nar. (= za-

ISKRZYĆ

rodku). I. duszy już s. nie pał. Mick. Jedna tylko I. jest w człowieku, raz tylko w młodocianym rozpala s. wieku. Mick. Dotąd iskry młodego zapału tlą w głębi piersi. Mick. Cóż są gminne dzieje? Popiół, w którym zaledwie I. prawdy tleje. Mick. 2. jub. diament maleńki, oddzielnie osadzony. 3. lek.: Iskry przed oczyma p. Iskrczenie s. Zdr. Iskierka, [Iskierusia]. Zgr. Iskrzysko. <ISKR>

× Iskrawo przys. od Iskrawy.

× Iskrawy p. Iskrzysty: Picie tam mają z muszeza na kształt miodowego, na pojrzaniu iskrawe. Paszk.

× Iskrobitny poet. iskry wydający, bijący iskrami: Wraż podków iskrobitnych po bruku łaskoty. Ejs. <Iskra + BI>

Iskrochlón, u, lm. y p. Iskrochron. <Iskra + CHŁO(N)>

Iskrochron, u, lm. y, Iskrochlón, Odiskiernik sito a. kapa na kominie, chroniąca od iskier wybuchających wraz z dymem. <Iskra + CHRON>

† Iskrodmy ogniem tchnący, iskrami ziejący: Zwierzę iskrodme. Petr. (= chimera, której padają iskry z paszczy). <Iskra + DM>

Iskron, i, lm. e bot. p. Korona. <?>

Iskrzasto przys. od Iskrzasty.

Iskrzasty p. Iskrzysty: Oczy iskrzaste kotów. Sak. I. słup dymu. Kon.

Iskrzato przys. od Iskrzaty: Zdala na tle szarym I. płoną i lecą w przestrzeń latarnie czerwone. Orzesz.

Iskrzaty p. Iskrzysty: Płomień iskrzate. Przyb. Migocze blask żywy, I. Syrok.

Iskrzący I. im. od Iskrzyć, iskrzysty: Iskrzące gwiazdy. Z iskrzącemi oczyma. 2. min. posiadający własność iskrczenia: Awanturyn I. (= migoczący). <ISKR>

Iskrzeć, eje, al p. Iskrzyć: U skraju widnokregu iskrzał złotem i purpurą szeroki pas zachodzącego słońca. Orzesz.

Iskrzelica, y, lm. e bot. (isertia) roś. z rodziny marzanowatych. <?>

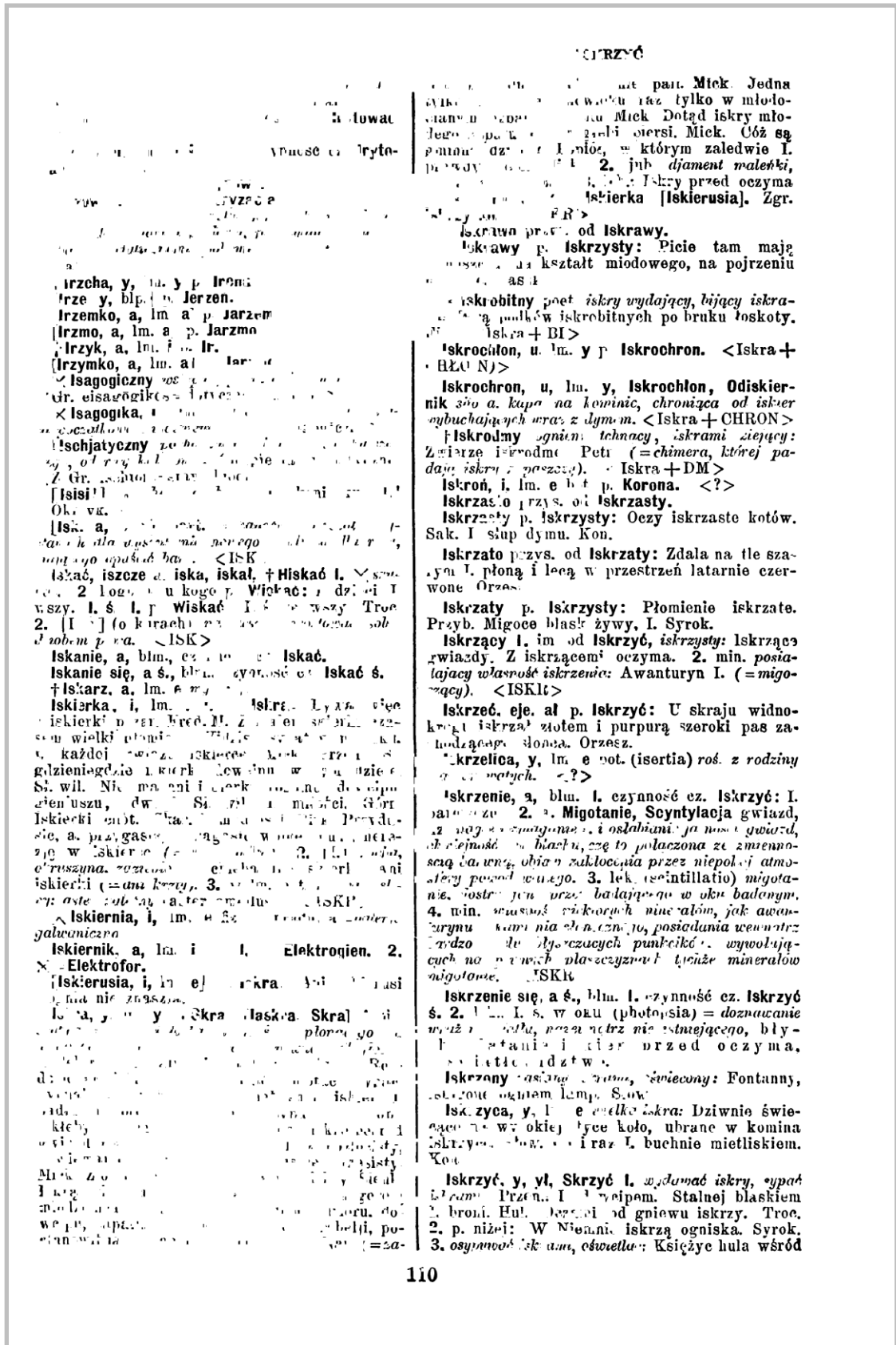
Iskrczenie, a, blm. I. czynność cz. Iskrzyć: I. parowozu. 2. a. Migotanie, Scyntylacja gwiazd, fiz. nagle wzmaganie s. i osłabianie ja-wności gwiazd, chwiejność ich blasku, często połączona ze zmiennością barwną, objaw zakłócenia przez niepokój atmosfery powodowanego. 3. lek. (scintillatio) migotanie, dostrzegane przez badającego w oku badanym. 4. min. własność niektórych mineralów, jak awanturynu i kamienia słonecznego, posiadania wewnątrz bardzo wielu błyszczących punkcików, wywołujących na pewnych płaszczyznach tychże mineralów migotanie. <ISKR>

Iskrczenie się, a s., blm. I. czynność cz. Iskrzyć s. 2. lek.: I. s. w oku (photopsia) = doznawanie wrażeń światła, nazewnątrz nie istniejącego, błyski, latanie iskier przed oczyma, × światłowidztwo.

Iskrzony zasiany iskrami, oświecony: Fontanny, iskrzone ogniem lamp. Słow.

Iskrzyca, y, lm. e wielka iskra: Dziwnie świecące na wysokiej tyce koło, ubrane w komina iskrzyce. Słow. Co i raz I. buchnie mieliskiem. Kon.

Iskrzyć, y, ył, Skrzyć I. wydawać iskry, sypać iskrami. Przen.: I. doweipem. Stalnej blaskiem I. broni. Hul. Oczami od gniewu iskrzy. Troc. 2. p. niżej: W Niemnie iskrzą ogniska. Syrok. 3. osypywać iskrami, oświecać: Księżyc hula wśród



OWANIE

czyn. Mat. I. *ś. gniewać ś., srożyć ś., być wzburzonym*: I. s. czym, czego. <Łć. irritare>

Irytowanie, a, blm., czynność cz. **Irytować, irytacja.**

Irytowanie się, a ś., blm., czynność cz. **Irytować ś.**, *irytacja.*

Iryzacja, i, blm. p. **Iryzowanie.**

Iryzowanie, a, blm., **Iryzacja fiz. mienienie ś. barwami tęczowymi, zjawisko, jakie okazują tkaniny jedwabne, macica perłowa, przedmioty sztucznie farbami metalicznymi pokryte.** <Z Gr. iris = tęcza>

† **Irzcha, y**, lm. **y p. Ircha.**

[**Irze, y**, bjp.] p. **Jerzeń.**

[**Irzemko, a**, lm. **a**] p. **Jarzemko.**

[**Irzmo, a**, lm. **a**] p. **Jarzmo.**

† **Irzyk, a**, lm. **i p. Ir.**

[**Irzymko, a**, lm. **a**] p. **Jarzmo.**

× **Isagogiczny** *wstępny, poprzednio wyjaśniający.* <Gr. eisagogikós = dotyczący wstępu>

× **Isagogika, i**, blm. *nauka wstępna, zawierająca początkowe wyjaśnienia.* <p. Isagogiczny>

! **Ischjatyyczny** *pochozący z bólu nerwu kulszowego, od rwy kulszowej; Cierpienia ischjatyyczne.* <Z Gr. ischion = staw biodrowy>

[**Isisi**] w. *hej! a hej!*: I, konieczu, I! <Okrzyk.>

[**Isk, a**, blm.] bart. *kilkanaście pszczoł, wysłanych dla wyszukania nowego siedliska dla roju, mającego opuścić barę.* <ISK>

Iskać, iszcze a. iska, iskał, † Hiskać I. × **szukać.** 2. kogo a. u kogo p. **Wiskać**: I. dzieci. I. wszy. I. ś. I. p. **Wiskać**: I. ś. ze wszy. Troc. 2. [I. ś.] (o kurach) *na deszcz prostować sobie dziobem pióra.* <ISK>

Iskanie, a, blm., *czynność cz. Iskać.*

Iskanie się, a ś., blm., *czynność cz. Iskać ś.*

† **Iskarz, a**, lm. **e myśliwy.**

Iskierka, i, lm. **i I. p. Iskra**: Bywa więc z iskierki pożar. Fred. M. Z malej iskierki czasem wielki płomień. Widzisz świat w prosku, w każdej gwiazd iskierce. Mick. Przen.: Są gdzieś gdzie iskierki dowiecipu w tym dziele. Sł. wil. Nie ma ani iskierki rozumu, dowiecipu, gienjuszu, odwagi. Sł. wil. I. miłości. Gór. Iskierki enót. Skar. I. mądrości. Birk. Przyduć, a. przygasić, a. zagasić wojnę, bunt, herezję w iskierce (= w zarodku). 2. [I.] *trocha, okruszyna, ździebło*: I. chleba. Ino iskierkę. Ani iskierki (= ani krzty). 3. w lm. bot. *nazwa rośliny: aster zalesny (aster amellus).* <ISKR>

× **Iskiernia, i**, lm. **e fiz.**: I. prądowa = *bateria galwaniczna.*

Iskiernik, a, lm. **i fiz. I. p. Elektrogien. 2.** × = **Elektrofor.**

[**Iskierusia, i**, lm. **e**] p. **Iskra**: Ani iskierusi w ognia nie znaszła.

Iskra, y, lm. **y, † Skra, [Jaskra, Skra] I. fiz.** *drobna cząstka ciała żarzącego ś. a. płonącego, od niego oderwana i padająca w pewnej odległości*: I. elektryczna, galwaniczna. Krzesać iskry. Rozdmuchać iskry. Iskry rzucać, miotać, sypać. Sypnął ś. rój iskier. Buchnął snop iskier. I. padła na prochy. Z komina sypały ś. snopy i kłęby złotych iskier. Przen.: Nową iskry pośród gwiazd rozwiecą. Syrok. Waehlarz pozłocisty, powiewając, rozlewał deszcz iskier rzęsisty. Mick. Z oczu wojowników sypały ś. iskry. Sienk. Iskry zdrowia oko strzela. Mick. Iskry goreją młodzieńca oczy. Syrok. I. rozumu, humoru, dowiecipu, zapału. Uwiadomiony król o rebelji, postanowił ją przygasić w samej iskry. Nar. (= za-

ISKRZYĆ

rodku). I. duszy już ś. nie pali. Mick. Jedna tylko I. jest w człowieku, raz tylko w młodocianym rozpala ś. wieku. Mick. Dotąd iskry młodego zapału tlą w głębi piersi. Mick. Cóż są gminne dzieje? Popiół, w którym zaledwie I. prawdy tleje. Mick. 2. jub. *djament maleńki, oddzielnie osadzony.* 3. lek.: Iskry przed oczyma p. **Iskrozenie ś.** Zdr. **Iskierka, [Iskierusia]. Zgr. Iskrzysko.** <ISKR>

× **Iskrawo** przys. od **Iskrawy.**

× **Iskrawy p. Iskrzysty**: Picie tam mają z muszcza na kształt miodowego, na pojrzeniu iskrawe. Paszk.

× **Iskrobitny** poet. *iskry wydający, bijący iskrami*: Wraż podków iskrobitnych po bruku łaskoty. Ejs. <Iskra + BI>

Iskrochlön, u, lm. **y p. Iskrochron.** <Iskra + CHŁON>

Iskrochron, u, lm. **y, Iskrochlön, Odiskiernik** *silo a. kapa na kominie, chroniąca od iskier wybuchających wraz z dymem.* <Iskra + CHRON>

† **Iskrodmy** *ogniem tchnący, iskrami ziejący*: Zwierzę iskrodme. Petr. (= chimera, której padają iskry z paszczy). <Iskra + DM>

Iskron, i, lm. **e bot. p. Korona.** <?>

Iskrzasto przys. od **Iskrzasty.**

Iskrzasty p. Iskrzysty: Oczy iskrzaste kotów. Sak. I. słup dymu. Kon.

Iskrzato przys. od **Iskrzaty**: Zdala na tle szarym I. płoną i lecą w przestrzeń latarnie czerwone. Orzesz.

Iskrzaty p. Iskrzysty: Płomień iskrzate. Przyb. Migocę blask żywy, I. Syrok.

Iskrzący I. im. od **Iskrzyć, iskrzysty**: Iskrzące gwiazdy. Z iskrzącymi oczyma. 2. min. *posiadający własność iskrzenia*: Awanturyn I. (= migoczący). <ISKR>

Iskrzeć, eje, ał p. Iskrzyć: U skraju widnokręgu iskrzał złotem i purpurą szeroki pas zachodzącego słońca. Orzesz.

Iskrzelica, y, lm. **e bot. (isertia) roś. z rodziny marzanowatych.** <?>

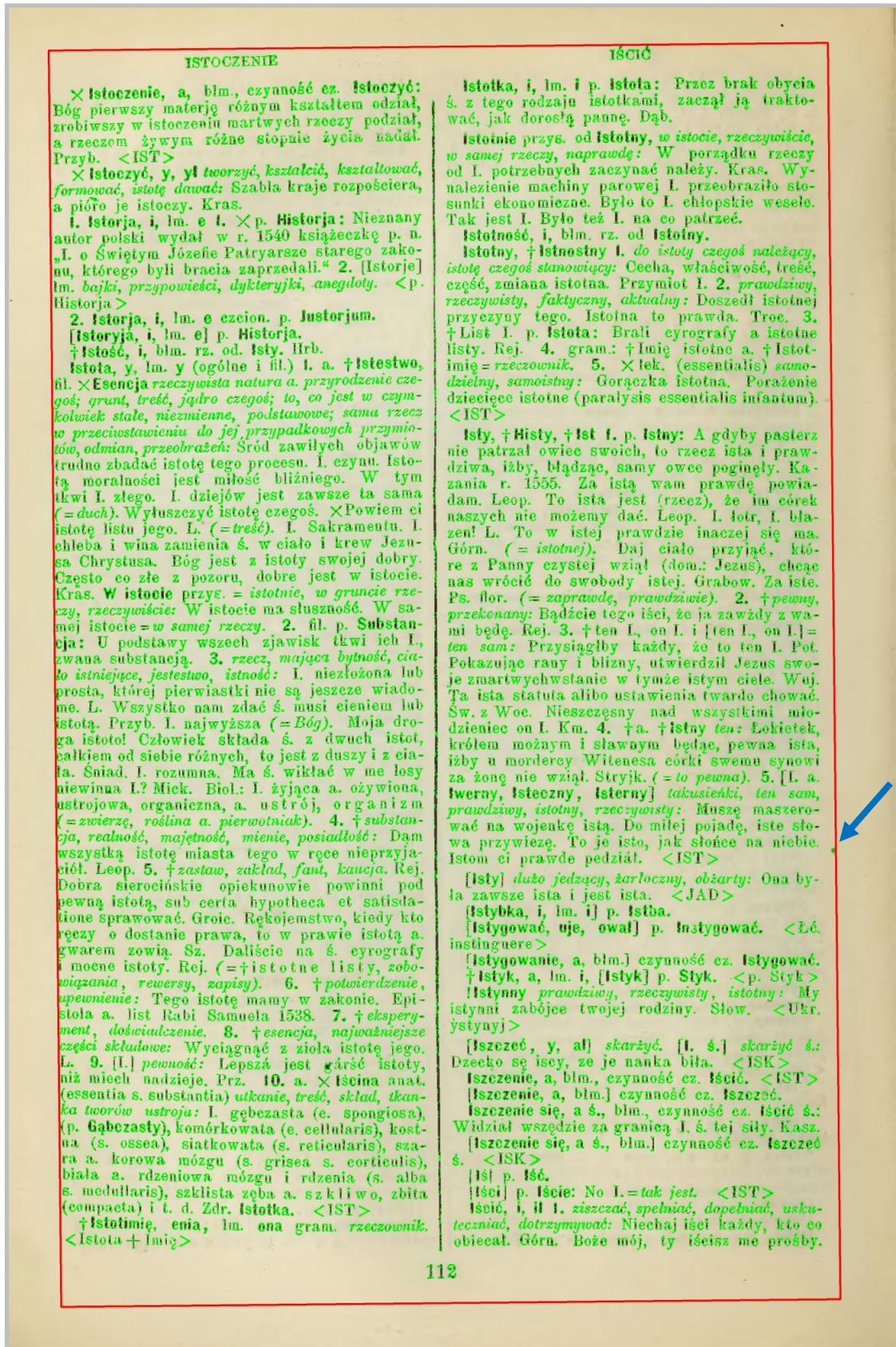
Iskrozenie, a, blm. **I. czynność cz. Iskrzyć**: I. parowozu. 2. a. **Migotanie, Scyntylacja** *gwiazd, fiz. nagłe wzmaganie ś. i osłabianie jasności gwiazd, chwilejność ich blasku, często połączona ze zmiennością barwną, objaw zakłócenia przez niepokój atmosfery powodowanego.* 3. lek. (scintillatio) *migotanie, dostrzegane przez badającego w oku badanym.* 4. min. *własność niektórych mineralów, jak awanturynu i kamienia słonecznego, posiadania wewnątrz bardzo wielu błyszczących punkcików, wywołujących na pewnych płaszczyznach tychże mineralów migotanie.* <ISKR>

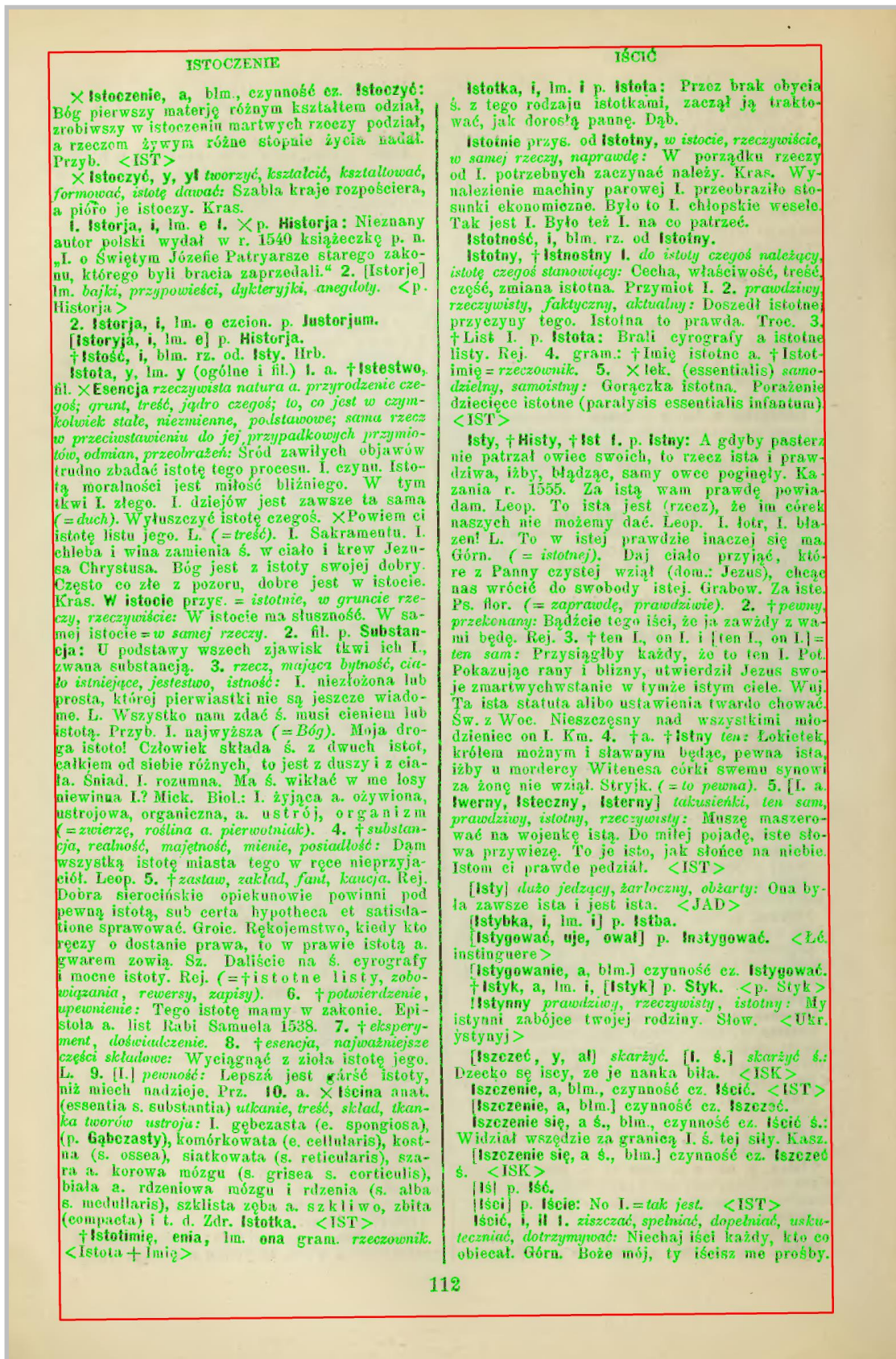
Iskrozenie się, a ś., blm. **I. czynność cz. Iskrzyć ś.** 2. lek.: I. ś. w oku (photopsia) = *doznawanie wrażenia światła, nazewnątrz nie istniejącego, błyski, latanie iskier przed oczyma.* × **światłowidztwo.**

Iskrzony *zasiany iskrami, oświecony*: Fontanny, iskrzone ogniem lamp. Słow.

Iskrzyca, y, lm. **e wielka iskra**: Dziwnie świecące na wysokiej tyce koło, ubrane w komina iskrzyce. Słow. Co i raz I. buchnie mietliskiem. Kon.

Iskrzyć, y, ył, Skrzyć I. *wydawać iskry, sypać iskrami.* Przen.: I. dowiecipem. Stalnej blaskiem I. broni. Hul. Oczami od gniewu iskrzy. Troc. 2. p. niżej: W Niemnie iskrzą ogniska. Syrok. 3. *osypywać iskrami, oświecać*: Księżyc hula wśród





Oczywiście tekst taki nigdy nie jest idealnie równy, dodawany jest więc niewielki margines dla bezpieczeństwa – tak zwane białe spacje. Dodatkowo, aby zredukować ryzyko potraktowania większych artefaktów na marginesach jako części tekstu, program ignoruje wszystko regiony znajdujące się zbyt blisko krawędzi obrazu. Po znalezieniu współrzędnych tekstu program przycina obraz, co daje rezultat taki jak przedstawiony na **ilustracji 32**.

Należy tutaj wspomnieć, że w przypadku niektórych stron mających przekosy (ang. *skew*), zniekształconą z perspektywą lub innymi zniekształceniami geometrycznymi, metoda ta może nieprawidłowo skorygować ich orientację. To z kolei może doprowadzić do złego przycięcia obszaru tekstowego. Jako że na chwilę obecną nawet wiodące komercyjne programy graficzne oraz do rozpoznawania tekstu nie zawsze są w stanie automatycznie rozwiązać tego rodzaju problemy, na najbardziej zniekształconych stronach dokonano ręcznej korekcji.

8.6.4. Etap czwarty

Po uzyskaniu odpowiednio przyciętego obrazu IndexCardSplitter przystępuje do podzielenia go na kolumny. Aby znaleźć granicę kolumny, wyszukuje najdłuższej pionowej linii w centralnej części obrazu.

Początkowo rozważane było wykorzystanie transformacji Hough'a, które służy do wykrywania wszystkich linii na stronie. Zaletą tej metody jest to, że znajduje linie nawet jeśli są one przekrzywione czy nieregularne. Zrezygnowano jednak z tej metody z dwóch powodów. Po pierwsze transformacja ta jest bardzo obciążająca dla procesora – do tego stopnia, że etap ten wymagał wielokrotnie więcej obliczeń niż wszystkie pozostałe czynności łącznie. Po drugie się piszącemu te słowa było trudno sparametryzować ją w taki sposób, aby po odfiltrowaniu pozostawała tylko linia dzieląca kolumny.

W związku z tym zastosowano mniej złożoną, ale skuteczną metodę: program ponownie oblicza średnią jasność dla każdej kolumny pikseli znajdującej się w centralnej części obrazu, i wybiera z nich tę o najmniejszej jasności. Koordynaty tej linii są wykorzystywane do podzielenia strony na lewą i prawą kolumnę, ponownie uwzględniając marginesy buforowe, aby nie przyciąć fragmentu tekstu. **Ilustracje 33 i 34** reprezentują kolejno lewą i prawą kolumnę powstałe w wyniku takiego podziału.

ISTOCZENIE

× **Istoczenie**, a, blm., czynność cz. **Istoczyć**: Bóg pierwszy materję różnym kształtem odział, zrobiwszy w istoczeniu martwych rzeczy podział, a rzeczom żywym różne stopnie życia nadał. Przyb. <IST>

× **Istoczyć**, y, yl tworzyć, kształcić, kształtować, formować, istotę dawać: Szabla kraje rozpościera, a pióro je istoczy. Kras.

1. **Istoria**, i, lm. e l. × p. **Historja**: Nieznany autor polski wydał w r. 1540 książeczkę p. n. „I. o Świętym Józefie Patryarsze starego zakonu, którego byli bracia zaprzędali.“ 2. [Istorje] lm. bajki, przypowieści, dykteryjki, anegdoty. <p. Historja>

2. **Istoria**, i, lm. e czcion. p. **Iustorjum**.

[**Istoria**], i, lm. e] p. **Historja**.

† **Istosc**, i, blm. rz. od. **Isty**. Ilrb.

Istota, y, lm. y (ogólne i fil.) l. a. † **Istestwo**,

fil. × **Esencja** rzeczywista natura a. przyrodzenie czegoś; gruntu, treść, jądro czegoś; to, co jest w czymkolwiek stale, niezmiennie, poistawowe; sama rzecz w przeciwstawieniu do jej przypadkowych przymiotów, odmian, przeobrażeń: Śród zawilych objawów trudno zbadać istotę tego procesu. I. czynu. Istotą moralności jest miłość bliźniego. W tym tkwi I. złego. I. dziejów jest zawsze ta sama (= duch). Wyłuszczyć istotę czegoś. × Powiem ci istotę listu jego. L. (= treść). I. Sakramentu. I. chleba i wina zamienia ś. w ciało i krew Jezusa Chrystusa. Bóg jest z istoty swojej dobry. Często co złe z pozoru, dobre jest w istocie. Kras. W istocie przye. = istotnie, w gruncie rzeczy, rzeczywistości: W istocie ma słuszność. W samej istocie = w samej rzeczy. 2. fil. p. **Substancja**: U podstawy wszech zjawisk tkwi ieh I., zwana substancją. 3. rzecz, mająca bytność, ciało istniejące, jestestwo, istność: I. niezłożona lub prosta, której pierwiastki nie są jeszcze wiadome. L. Wszystko nam zdać ś. musi cieniem lub istotą. Przyb. I. najwyższa (= Bóg). Moja droga istota! Człowiek składa ś. z dwóch istot, całkiem od siebie różnych, to jest z duszy i z ciała. Śniad. I. rozumna. Ma ś. widać w me losy niewinna I.? Mick. Biol.: I. żyjąca a. ożywiona, ustrojowa, organiczna, a. ustrój, organizm (= zwierze, roślina a. pierwotniak). 4. † **substancja**, realność, majątność, mienie, posiadłość: Dám wszystkim istotę miasta tego w ręce nieprzyjaciół. Leop. 5. † **zastaw**, zakład, fan, kaucja. Rej. Dobra sierocińskie opiekunowie powinni pod pewną istotą, sub certa hypotheca et satisfactione sprawować. Groic. Rękojemstwo, kiedy kto ręczy o dostanie prawa, to w prawie istotą a. gwarem zowią. Sz. Daliście na ś. cyrografy i moene istoty. Rej. (= † istotne listy, zobowiązania, rewery, zapisy). 6. † **potwierdzenie**, upewnienie: Tego istotę mamy w zakonie. Epistola a. list Rabi Samuela 1538. 7. † **eksperyment**, doświadczenie. 8. † **esencja**, najważniejsze części składowe: Wyciągnąć z ziola istotę jego. L. 9. [I.] **pewność**: Lepszą jest górsz istoty, niż miech nadzieje. Prz. 10. a. × **Iścina** anat. (essentia s. substantia) **utkanie**, treść, skład, tkan-ka tworów ustroja: I. gębezasta (e. spongiosa), (p. **Gąbezasty**), komórkowata (e. cellularis), kostna (s. ossea), siatkowata (s. reticularis), szara a. korowa mózgu (s. grisea s. corticulis), biała a. rdzeniowa mózgu i rdzenia (s. alba s. medullaris), szklista zęba a. szklivo, zbita (compacta) i t. d. Zdr. **Istotka**. <IST>

† **Istotimie**, enia, lm. ona gram. **rzeczownik**. <Istota + Imię>

IŚCIĆ

Istotka, i, lm. i p. **Istota**: Przez brak obycia ś. z tego rodzaju istotkami, zaczął ją traktować, jak dorosłą pannę. Dąb.

Istotnie przys. od **Istotny**, w istocie, rzeczywistości, w samej rzeczy, naprawdę: W porządku rzeczy od I. potrzebnych zaczynać należy. Kras. Wyznalezienie maszyny parowej I. przeobraziło stosunki ekonomiczne. Było to I. chłopskie wesele. Tak jest I. Było też I. na co patrzeć.

Istotność, i, blm. rz. od **Istotny**.

Istotny, † **Istotny** l. do istoty czegoś należący, istotę czegoś stanowiący: Cecha, właściwość, treść, część, zmiana istotna. Przymiot 1. 2. **prawdziwy**, **rzeczywisty**, **faktyczny**, **aktualny**: Doszedł istotnej przyczyny tego. Istotna to prawda. Troc. 3. † **List** l. p. **Istota**: Brali cyrografy a. istotne listy. Rej. 4. gram.: † **Imię** istotne a. † **Istotimie** = **rzeczownik**. 5. × **lek**. (essentialis) **samodzielny**, **samoistny**: Gorączka istotna. Porażenie dziecięce istotne (paralysis essentialis infantum). <IST>

Isty, † **Histy**, † **Ist** l. p. **Istny**: A gdyby pasterz nie patrzył owiec swoich, to rzecz ista i prawdziwa, iżby, błędząc, same owce poginęły. Kazania r. 1555. Za istą wam prawdę powiadam. Leop. To ista jest (rzecz), że im córek naszych nie możemy dać. Leop. I. łotr, I. błazen! L. To w istej prawdzie inaczej się ma. Górn. (= istotnej). Daj ciało przyjąć, które z Panny czystej wziął (dom.: Jezus), chcąc nas wrócić do swobody istej. Grabow. Za iste. Ps. flor. (= **zaprawdę**, **prawdziwie**). 2. † **pewny**, **przeknany**: Bądźcie tego iści, że ja zawsze z wami będę. Rej. 3. † **ten** I., **on** I. i [ten I., **on** I.] = **ten sam**: Przysięgłby każdy, że to ten I. Pot. Pokazując rany i blizny, utwierdził Jezus swoje zmartwychwstanie w tymże istym cielem. Wuj. Ta ista statuta albo ustawienia twardo chować. Św. z Woc. Nieszczęsny nad wszystkimi młodzieniec **on** I. Km. 4. † a. † **Istny** **ten**: Łokietek, królem możnym i sławnym będąc, pewna ista, iżby u mordercy Witenesa córki swemu synowi za żonę nie wziął. Stryjk. (= **to pewna**). 5. [I. a. **Iwerny**, **Isteczny**, **Isterny**] **takusienki**, **ten sam**, **prawdziwy**, **istotny**, **rzeczywisty**: Muszę maszerować na wojenkę istą. Do milej pojedę, iste słowa przywiezę. To je isto, jak słońce na niebie. Istom ci prawdę pedziął. <IST>

[**Isty**] **dużo jedzący**, **żarłoczny**, **obżarty**: Ona była zawsze ista i jest ista. <JAD>

[**Istybka**, i, lm. i] p. **Istba**.

[**Istygować**, **uje**, **ował**] p. **Instygować**. <Łé. instinguere>

[**Istygowanie**, a, blm.] czynność cz. **Istygować**.

† **Istyk**, a, lm. i, [**Istyk**] p. **Styk**. <p. **Styk**>

† **Istynny** **prawdziwy**, **rzeczywisty**, **istotny**: My istynni zabójce twojej rodziny. Słow. <Ukr. ystyny>

[**Iszczec**, y, a] **skarżyć**. [I. ś.] **skarżyć** ś.: Dzecko się isey, że je nanka biła. <ISK>

Iszczenie, a, blm., czynność cz. **Iścić**. <IST>

[**Iszczenie**, a, blm.] czynność cz. **Iszczec**.

Iszczenie się, a ś., blm., czynność cz. **Iścić** ś.: Widział wszędzie za granicą I. ś. tej siły. Kasz.

[**Iszczenie się**, a ś., blm.] czynność cz. **Iszczec** ś. <ISK>

[Iś] p. **Iśc**.

[Iści] p. **Iście**: No I. = **tak jest**. <IST>

Iścić, i, il l. **ziszczac**, **spełniać**, **dopełniać**, **usku-teczniać**, **dotrzymywać**: Niechaj iści każdy, kto co obiecał. Górn. Boże mój, ty iścisz me prośby.

× **Istoczenie**, a, blm., czynność cz. **Istoczyć**: Bóg pierwszy materję różnym kształtem odział, zrobiwszy w istoczeniu martwych rzeczy podział, a rzeczem żywym różne stopnie życia nadał. Przyb. <IST>

× **Istoczyć**, y, ył **tworzyć, kształcić, kształtować, formować, istotę dawać**: Szabla kraje rozpościera, a pióro je istoczy. Kras.

I. Istorja, i, lm. e I. × p. **Historja**: Nieznany autor polski wydał w r. 1540 książeczkę p. n. „I. o Świętym Józefie Patryarsze starego zakonu, którego byli bracia zaprzędali.“ **2.** [Istorje] lm. *bakki, przypowieści, dykteryjki, anegdoly.* <p. Historja >

2. Istorja, I, lm. e czcion. p. **Iustorjum**.

[Istoryja, i, lm. e] p. **Historja**.

† **Istość**, i, blm. rz. od. **Isty**. Hrb.

Istota, y, lm. y (ogólne i fil.) **1. a.** † **Istestwo**, fil. × **Esencja rzeczywista natura a. przyrodzenie czegoś; grunt, treść, jądro czegoś; to, co jest w czymkolwiek stałe, niezmiennie, podstawowe; sama rzecz w przeciwstawieniu do jej przypadkowych przymiotów, odmian, przeobrażeń**: Śród zawilych objawów trudno zbadać istotę tego procesu. I. czynu. Istotą moralności jest miłość bliźniego. W tym tkwi I. zlego. I. dziejów jest zawsze ta sama (= *duch*). Wyluszczyć istotę czegoś. × Powiem ci istotę listu jego. L. (= *treść*). I. Sakramentu. I. chleba i wina zamienia ś. w ciało i krew Jezusa Chrystusa. Bóg jest z istoty swojej dobry. Często co złe z pozoru, dobre jest w istocie. Kras. **W istocie** przyś. = *istotnie, w gruncie rzeczy, rzeczywistość*: W istocie ma słusność. W samej istocie = *w samej rzeczy*. **2. fil. p. Substancja**: U podstawy wszech zjawisk tkwi ich I., zwana substancją. **3. rzecz, mająca bytność, ciało istniejące, jestestwo, istność**: I. niezłożona lub prosta, której pierwiastki nie są jeszcze wiadome. L. Wszystko nam zdać ś. musi cieniem lub istotą. Przyb. I. najwyższa (= *Bóg*). Moja droga istot! Człowiek składa ś. z dwóch istot, całkiem od siebie różnych, to jest z duszy i z ciała. Śniad. I. rozumna. Ma ś. wkląć w me losy niewinna I.? Mick. Biol.: I. żyjąca a. ożywiona, ustrojowa, organiczna, a. ustrój, organizm (= *zwierzę, roślina a. pierwotniak*). **4. † substancja, realność, majątność, mienie, posiadłość**: Daj wszystkim istotę miasta tego w ręce nieprzyjaciół. Leop. **5. † zastaw, zakład, fanł, kaucja**. Rej. Dobra sierocińskie opiekunowie powinni pod pewną istotą, sub certa hypotheca et satisfactione sprawować. Groic. Rękojemstwo, kiedy kto ręczy o dostanie prawa, to w prawie istotą a. gwarem zowią. Sz. Daliście na ś. cyrografy i moene istoty. Rej. (= † *istotne listy, zobowiązania, rewery, zapisy*). **6. † potwierdzenie, upewnienie**: Tego istotę mamy w zakonie. Epistola a. list Rabi Samuela 1538. **7. † eksperyment, doświadczanie**. **8. † esencja, najważniejsze części składowe**: Wyciągnąć z zioła istotę jego. L. **9. [I.] pewność**: Lepszą jest gąrsć istoty, niż miech nadzieje. Prz. **10. a.** × **Iścina** anat. (essentia s. substantia) *utkanie, treść, skład, tkanka tworów ustroju*: I. gębczasta (e. spongiosa), (p. **Gąbczasty**), komórkowata (e. cellularis), kostna (s. ossea), siatkowata (s. reticularis), szara a. korowa mózgu (s. grisea s. corticulis), biała a. rdzeniowa mózgu i rdzenia (s. alba s. medullaris), szklista zęba a. szklisko, zbita (compacta) i t. d. Zdr. **Istotka**. <IST>

† **Istotimie, enia**, lm. ona gram. *rzeczownik*. <Istota + Imię>

Istotka, i, lm. i p. **Istota**: Przez brak obycia ś. z tego rodzaju istotkami, zaczął ją traktować, jak dorosłą pannę. Dąb.

Istotnie przyś. od **Istotny**, w *istocie, rzeczywistości, w samej rzeczy, naprawdę*: W porządku rzeczy od I. potrzebnych zaczynać należy. Kras. Wy-nalezienie maszyny parowej I. przeobraziło stosunki ekonomiczne. Było to I. chłopskie wesele. Tak jest I. Było też I. na co patrzeć.

Istotność, i, blm. rz. od **Istotny**.

Istotny, † **Istnotny** I. *do istoty czegoś należący, istotę czegoś stanowiący*: Cecha, właściwość, treść, część, zmiana istotna. Przymiot I. **2. prawdziwy, rzeczywisty, faktyczny, aktualny**: Doszedł istotnej przyczyny tego. Istotna to prawda. Troc. **3. † list** I. p. **Istota**: Brali cyrografy a istotne listy. Rej. **4. gram.**: † **Imię** istotne a. † **Istotimie** = *rzeczownik*. **5. × lek. (essentialis) samodzielny, samoistny**: Gorączka istotna. Porażenie dziecięce istotne (paralysis essentialis infantum). <IST>

Isty, † **Histy**, † **Ist** I. p. **Istny**: A gdyby pasterz nie patrzył owiec swoich, to rzecz ista i prawdziwa, iżby, błędząc, samy owce poginęły. Kazania r. 1555. Za istą wam prawdę powiadam. Leop. To ista jest (rzecz), że im córek naszych nie możemy dać. Leop. I. łotr, I. błazen! L. To w istej prawdzie inaczej się ma. Górn. (= *istotnej*). Daj ciało przyjąć, które z Panny czystej wziął (dom.: Jezus), chcąc nas wrócić do swobody istej. Grabow. Za iste. Ps. flor. (= *zaprawdę, prawdziwie*). **2. † pewny, przeknany**: Bądźcie tego iści, że ja zawżdy z wami będę. Rej. **3. † ten** I., on I. i [ten I., on I.] = *ten sam*: Przysięgłby każdy, że to ten I. Pot. Pokazując rany i blizny, utwierdził Jezus swoje zmartwychwstanie w tymże istym ciele. Wuj. Ta ista statuta albo ustawienia twardo chować. Św. z Woc. Nieszczęsny nad wszystkimi młodzieniec on I. Km. **4. † a. † Istny ten**: Łokietek, królem możliwym i sławnym będąc, pewna ista, iżby u mordercy Witenesa córki swemu synowi za żonę nie wziął. Stryjk. (= *to pewna*). **5. [I. a. lwerny, Isteczny, Isterny]** *takusieński, ten sam, prawdziwy, istotny, rzeczywisty*: Muszę maszerować na wojenkę istą. Do miłej pojedę, iste słowa przywiezę. To je isto, jak słońce na niebie. Istom ci prawdę pedział. <IST>

[Isty] *dużo jedzący, żarłoczny, obżarty*: Oua była zawsze ista i jest ista. <JAD>

[Istybka, i, lm. i] p. **Istba**.

[Istygować, uje, ował] p. **Instygować**. <Łć. instinguere>

[Istygowanie, a, blm.] czynność cz. **Istygować**.

† **Istyk**, a, lm. i, [Istyk] p. **Styk**. <p. Styk>

† **Istynny** *prawdziwy, rzeczywisty, istotny*: My istynni zabójcy twojej rodziny. Słow. <Ukr. ýstynnyj>

[Iszczec, y, ał] *skarżyć*. [I. ś.] *skarżyć* ś.: Dzecko się isey, że je nanka biła. <ISK>

Iszczenie, a, blm., czynność cz. **Iścić**. <IST>

[Iszczenie, a, blm.] czynność cz. **Iszczec**.

Iszczenie się, a ś., blm., czynność cz. **Iścić** ś.: Widział wszędzie za granicą I. ś. tej siły. Kasz.

[Iszczenie się, a ś., blm.] czynność cz. **Iszczec** ś. <ISK>

[Iś] p. **Iść**.

[Iści] p. **Iście**: No I. = *tak jest*. <IST>

Iścić, i, II I. *ziszczać, spełniać, dopełniać, uskuteniać, dotrzymywać*: Niechaj iści każdy, kto co obiecał. Górn. Boże mój, ty iścisz me próoby.

Ilustracja 33. Strona 112. – kolumna lewa:
po podziale.

× **Istoczenie**, a, blm., czynność cz. **Istoczyć**: Bóg pierwszy materję różnym kształtem odział, zrobivszy w istoczeniu martwych rzeczy podział, a rzeczem żywym różne stopnie życia nadał. Przyb. <IST>

× **Istoczyć**, y, yl tworzyć, kształcić, kształtować, formować, istotę dawać: Szabla kraje rozpościera, a pióro je istoczy. Kras.

I. Istorja, i, lm. e I. × p. **Historja**: Nieznany autor polski wydał w r. 1540 książeczkę p. n. „I. o Świętym Józefie Patryarsze starego zakonu, którego byli bracia zaprzędali.“ 2. [Istorje] lm. bajki, przypowieści, dykteryjki, anegdoly. <p. Historja>

2. **Istorja**, i, lm. o ezcion. p. **Justorjum**.

[**Istoryja**, i, lm. o] p. **Historja**.

† **Istość**, i, blm. rz. od **Isty**. Hrb.

Istota, y, lm. y (ogólne i fil.) I. a. † **Istestwo**,

fil. × **Esencja rzeczywista natura a. przyrodzenie czegoś**;

grunt, treść, jądro czegoś; to, co jest w czymkolwiek stale, niezmiennie, podstawowe; sama rzecz w przeciwstawieniu do jej przypadkowych przymiotów, odmian, przeobrażeń: Śród zawilęch objawów trudno zbadać istotę tego procesu. I. czynu. Istotą moralności jest miłość bliźniego. W tym tkwi I. złego. I. dziejów jest zawsze ta sama (= *duch*). Wyluszczyć istotę czegoś. × Powiem ci istotę listu jego. L. (= *treść*). I. Sakramentu. I. chleba i wina zamienia ś. w ciało i krew Jezusa Chrystusa. Bóg jest z istoty swojej dobry. Często co złe z pozoru, dobre jest w istocie. Kras. W **istocie** przys. = *istotnie*, w gruncie rzeczy, *rzeczywiście*: W istocie ma słusność. W samej istocie = w samej rzeczy. 2. fil. p. **Substancja**: U podstawy wszech zjawisk tkwi ich I., zwana substancją. 3. **rzecz, mająca bytność, cięło istniejące, jestestwo, istność**: I. niezłożona lub prosta, której pierwiastki nie są jeszcze wiadome. L. Wszystko nam zdać ś. musi cieniem lub istotą. Przyb. I. najwyższa (= *Bóg*). Moja droga istot! Człowiek składa ś. z dwóch istot, całkiem od siebie różnych, to jest z duszy i z ciała. Śniad. I. rozumna. Ma ś. widać w me losy niewinna I.? Miek. Biol.: I. żyjąca a. ożywiona, ustrojowa, organiczna, a. ustrój, organizm (= *zwierzę, roślina a. pierwotniak*). 4. † **substancja, realność, majątność, mienie, posiadłość**: Dąm wszystką istotę miasta tego w ręce nieprzyjaciół. Leop. 5. † **zastaw, zakład, fan, kaucja**. Rej. Dobra sierocińskie opiekunowie powinni pod pewną istotą, sub certa hypotheca et satisfactione sprawować. Groic. Rękojemstwo, kiedy kto ręczy o dostanie prawa, to w prawie istotą a. gwarem zowią. Sz. Dalście na ś. cyrografy i moene istoty. Rej. (= † istotne listy, *zobowiązania, rewersy, zapisy*). 6. † **potwierdzenie, upewnienie**: Tego istotę mamy w zakonie. Epistola a. list Rabi Samuela 1538. 7. † **eksperyment, doświadczenie**. 8. † **esencja, najważniejsze części składowe**: Wyciągnąć z ziola istotę jego. L. 9. [I.] **pewność**: Lepszą jest gąrsć istoty, niż miech nadzieje. Prz. 10. a. × **Iścina anat.** (essentia s. substantia) **utkanie, treść, skład, tkanka tworów ustroju**: I. gębczasta (e. spongiosa), (p. **Gąbczasty**), komórkowata (e. cellularis), kostna (s. ossea), siatkowata (s. reticularis), szara a. korowa mózgu (s. grisea s. corticulis), biała a. rdzeniowa mózgu i rdzenia (s. alba s. medullaris), szklista zęba a. szklivo, zbita (compacta) i t. d. Zdr. **Istotka**. <IST>

† **Istotimie, enia**, lm. ona gram. **rzeczownik**. <Istota + Imię>

Ilustracja 34. Strona 112. – kolumna prawa:
po podziale.

Istotka, i, lm. I p. **Istota**: Przez brak obycia ś. z tego rodzaju istotkami, zaczął ją traktować, jak dorosłą pannę. Dąb.

Istotnie przys. od **Istotny**, w *istocie, rzeczywistości, w samej rzeczy, naprawdę*: W porządku rzeczy od I. potrzebnych zaczynać należy. Kras. Wnalezienie maszyny parowej I. przeobraziło stosunki ekonomiczne. Było to I. chłopskie wesele. Tak jest I. Było też I. na co patrzeć.

Istotność, i, blm. rz. od **Istotny**.

Istotny, † **Istnotny** I. *do istoty czegoś należący, istotę czegoś stanowiący*: Cecha, właściwość, treść, część, zmiana istotna. Przymiot I. 2. **prawdziwy, rzeczywisty, faktyczny, aktualny**: Doszedł istotnej przyczyny tego. Istotna to prawda. Troc. 3. † **List** I. p. **Istota**: Brali cyrografy a istotne listy. Rej. 4. gram.: † **Imię** istotne a. † **Istotimie** = *rzeczownik*. 5. × **lek.** (essentialis) **samodzielny, samoistny**: Gorączka istotna. Porażenie dziecięce istotne (paralysis essentialis infantum). <IST>

Isty, † **Histy**, † **Ist** I. p. **Istny**: A gdyby pasterz nie patrzył owiec swoich, to rzecz ista i prawdziwa, iżby, błędząc, same owce poginęły. Kazania r. 1555. Za istą wam prawdę powiadam. Leop. To ista jest (rzecz), że im córek naszych nie możemy dać. Leop. I. łotr, I. blaizen! L. To w istej prawdzie inaczej się ma. Górn. (= *istotnej*). Daj ciało przyjąć, które z Panny czystej wziął (dom.: Jezus), chcąc nas wrócić do swobody istej. Grabow. Za iste. Ps. flor. (= *zaprawdę, prawdziwie*). 2. † **pewny, przeknany**: Bądźcie tego iści, że ja zawždy z wami będę. Rej. 3. † **ten** I., on I. i [ten I., on I.] = *ten sam*: Przysięgłby każdy, że to ten I. Pot. Pokazując rany i blizny, utwierdził Jezus swoje zmartwychwstanie w tymże istym cielesie. Wuj. Ta ista statuta albo ustawienia twardo chowac. Św. z Woc. Nieszczęsny nad wszystkimi młodzieniec on I. Km. 4. † a. † **Istny ten**: Łokietek, królem możnym i sławnym będąc, pewna ista, iżby u mordercy Witenesa córki swemu synowi za żonę nie wziął. Stryjk. (= *to pewna*). 5. [I. a. **Iwerny, Isteczny, Isterny**] *takusielki, ten sam, prawdziwy, istotny, rzeczywisty*: Muszę maszerować na wojenkę istą. Do miłej pojadę, iste słowa przywiezę. To je isto, jak słońce na niebie. Istom ci prawdę pedziat. <IST>

[**Isty**] *dużo jedzący, żarłoczny, obżarty*: Ona była zawsze ista i jest ista. <JAD>

[**Istybka**, i, lm. i] p. **Istba**.

[**Istygować**, uje, owal] p. **Istygować**. <Łé. instinguere>

[**Istygowanie**, a, blm.] czynność cz. **Istygować**.

† **Istyk**, a, lm. i, [**Istyk**] p. **Styk**. <p. **Styk**>

Istynny *prawdziwy, rzeczywisty, istotny*: My istynni zabójce twojej rodziny. Słow. <Ukr. ystynny>

[**Iszczec**, y, al] *skarżyć*. [I. ś.] *skarżyć* ś.: Dzecko się iscy, że je nanka biła. <ISK>

Iszczenie, a, blm., czynność cz. **Iścić**. <IST>

[**Iszczenie**, a, blm.] czynność cz. **Iszczec**.

Iszczenie się, a ś., blm., czynność cz. **Iścić** ś.:

Widział wszędzie za granicą I. ś. tej siły. Kasz.

[**Iszczenie się**, a ś., blm.] czynność cz. **Iszczec** ś.

<ISK>

[Iś] p. **Iść**.

[Iści] p. **Iście**: No I. = *tak jest*. <IST>

Iścić, i, II I. *ziszczac, spełniać, dopełniać, uskutecznić, dotrzymywać*: Niechaj iści każdy, kto co obiecał. Górn. Boże mój, ty iścis me prośby.

8.6.5. Etap piąty

Mając do dyspozycji obrazy z pojedynczymi kolumnami tekstu, pozbawione artefaktów, nadmiaru białych spacji oraz ze skorygowaną geometrią tekstu, program przechodzi do dzielenia obrazu na poszczególne akapity.

Aby do tego doprowadzić, w pierwszej kolejności tworzona jest kopia obrazu każdej kolumny, na której wielokrotnie przeprowadza operację otwierania. Operacja ta to sekwencyjne naprzemienne wykonanie erozji i dylatacji, czego efektem jest redukcja lub zupełne usunięcie mniejszych regionów (artefaktów), a wzmocnienia większych (w tym przypadku liter). W wyniku tego regiony czarnych pikseli odpowiadające pojedynczym glifom stają się większe, ich kształty się upraszczają. Wówczas, ponownie posługując się funkcją `cv2.getContours()` program znajduje kontury tekstu, a każdy kontur obrysowuje wypełnionym prostokątem. W ten sposób sąsiadujące regiony łączą się ze sobą, a kształt tekstu jeszcze bardziej się upraszcza.

Aby tak powstałe obrazy kolumny móc podzielić na akapity, program stara się znaleźć wcięcia poszczególnych akapitów. Przez wcięcie rozumie się tutaj białe światło na początku wiersza, czyli niezawierający czarnych pikseli prostokątny region, którego górny i dolny bok wyznaczają kolejno poprzedni i następny wiersz lub krawędzie obrazu, lewy bok wyznacza lewa krawędź obrazu, a prawy bok wyznacza początek tekstu.

Pomocniczo program podejmuje problem znalezienia światła separującego akapity, czyli poziomych linii białych pikseli. Wówczas za górną granicę wcięcia uznaje się właśnie owe linie pomocnicze. Światło to nie zawsze jednak jest obecne, wówczas program stara się wpisać prostokąty w wolną przestrzeń na początku wierszu, zachowując odstęp między kolejnymi wcięciami występującymi bezpośrednio po sobie.

Dzięki uprzednio uproszczonym kształtom maleje ryzyko, że wcięcie zostanie błędnie wykryte między wierszami w przypadku, gdy światło interlinii jest za duże. W przeciwnym razie ICS mógłby błędnie wykryć wcięcie pomiędzy wierszami, np. w sytuacji, gdy jednym wierszu jest wielka litera (np. „W”), a tuż pod nią, w wierszu następnym, litera mała (np. „a”).

Wizualizację elementów, na podstawie których ICS określa współrzędne akapitów ze strony 112, można zobaczyć na **ilustracjach 35** oraz **36**, na których kolorem **niebieskim** program wypełnił wykryty uproszczony kształt tekstu, kolorem **zielonym** – interlinii, a kolorem **pomarańczowym** – wcięcia. Na podstawie w ten sposób określonych koordynat program wycina poszczególne akapity z oryginalnej wersji obrazów z kolumnami.

Każdy taki obraz z pojedynczym segmentem tekstu, ze względu na swoje podobieństwo do niewielkich kart, na których zapisywane były poszczególne hasła przy tradycyjnej kompilacji słowników, będzie nazywany „fiszka”.

8.6.6. Etap szósty

Po wyekstrahowaniu haseł do osobnych fiszek uzyskiwany jest efekt taki, jak widoczny poniżej dla strony 112. na **ilustracjach 37–41**:

Ilustracja 37. Fiszka nr 1 („Istoczenie”) ze strony 112.

× Istoczenie, a, blm., czynność cz. Istoczyć:
Bóg pierwszy materję różnym kształtem odział,
zrobiwszy w istoczeniu martwych rzeczy podział,
a rzeczem żywym różne stopnie życia nadał.
Przyb. <IST>

Ilustracja 38. Fiszka nr 3 („Istorja”) ze strony 112.

Í. Istorja, i, lm. e l. × p. Historja: Nieznany autor polski wydał w r. 1540 książeczkę p. n. „I. o Świętym Józefie Patryarsze starego zakonu, którego byli bracia zapprzedali.“ **2. [Istorje]** lm. *bajki, przypowieści, dykteryjki, anegdoty.* <p. Historja >

Ilustracja 39. Fiszka nr 6 („Istość”) ze strony 112.

† Istość, i, blm. rz. od. Isty. Hrb.

Ilustracja 40. Fiszka nr 8 („Istotimię”) ze strony 112.

† Istotimię, enia, lm. ona gram. rzeczownik.
<Istota + Imię >

Ilustracja 41. Fiszka nr 12 („Iszczeń”) ze strony 112.

[Iszczeń, y, al] skarżyć. [l. ś.] skarżyć ś.:
Dzecko się iscy, ze je nanka biła. <ISK >

Można na ich podstawie dostrzec trzy rodzaje problemów:

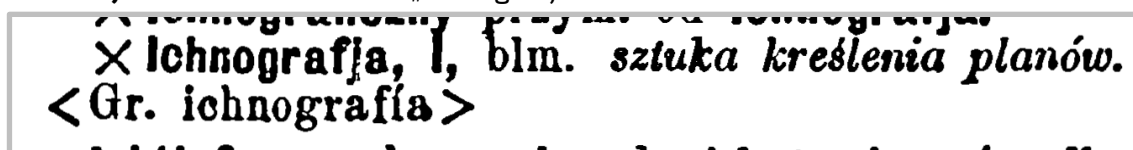
1. Na części fiszek obecne są fragmenty liter z haseł sąsiadujących;
2. Na części fiszek tekst dotyka górnego lub dolnego marginesu;
3. Na części fiszek jeden z marginesów jest większy.

Błędy te – choć w różnej liczbie i w różnym stopniu – pojawiają się na każdej z przetworzonych stron i związane są z brakiem lub nadmiarem białych spacji u góry lub dołu hasła. Wynika to zarówno z jakości druku samego SW, zniekształceń obrazów wejściowych, jak i błędnych współrzędnych dla poszczególnych akapitów.

W przypadku druku problem wynika z nieregularnej wielkości interlinii, w tym nawet jej negatywnej wartości (przez co dolna krawędź jednego wiersza nachodzi na górną krawędź kolejnego). Problemy wynikające z działania programu ograniczają do tych, które zostały uwzględnione w opisie poprzednich etapów działania programu IndexCardSplitter.

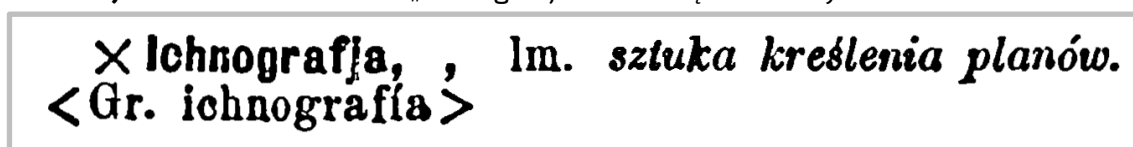
Przykład obrazu wejściowego, na którym program dokonuje korekcji, widoczny jest na **ilustracji 42**.

Ilustracja 42. Fiszka z hasłem „Ichnografja”.



Aby zniwelować wymienione wcześniej problemy, program ponownie dokonuje wyszukiwania konturów metodą `cv2.getContours`, ale tym razem ograniczając się tylko do tych, których co najmniej jedna koordynata równoznaczna jest z górną lub dolną krawędzią obrazu; tak znalezione regiony są usuwane. Metoda ta może powodować usunięcie pożądanego treści, co jest widoczne na **ilustracji 43**.

Ilustracja 43. Fiszka z hasłem „Ichnografja” – usunięte kontury.



Jak widać, metoda ta usunęła również litery hasła „Ichnografja”, które graniczyły z literami hasła poprzedniego. W związku z tym dodano ograniczenie wyszukiwania konturów do takich, które mają nie więcej niż 20 pikseli w pionie, czego rezultat jest przedstawiony na **ilustracji 44**.

Ilustracja 44. Fiszka z hasłem „Ichnografja” – kontury usunięte Metodą II.

× Ichnografja, I, bīm. sztuka kreślenia planów.
<Gr. ichnografía>

Po tejże modyfikacji nie zostały usunięte żadne elementy, które na obrazie powinny zostać, jednak na obrazie wyjściowym pozostało aż sześć artefaktów. W wyniku modyfikacji wartości granicznych udało się jednak zredukować liczbę artefaktów do zaledwie dwóch, które można dostrzec na **ilustracji 45**.

Ilustracja 45. Fiszka z hasłem „Ichnografja” – kontury usunięte Metodą III.

× Ichnografja, I, bīm. sztuka kreślenia planów.
<Gr. ichnografía>

Piszącemu te słowa nie udało się opracować algorytmu, który nie pozostawiałby żadnych artefaktów jednocześnie zachowując cały pożądaný tekst, ich łączna liczba jest jednak na tyle mała, że metoda nie była już dalej modyfikowana.

Po usunięciu artefaktów na brzegach IndexCardSplitter ponownie wykorzystuje rozwiązanie polegające na mierzeniu średniej jasności każdego wiersza, aby sprawdzić, czy tekst nie znajduje się za blisko lub za daleko krawędzi obrazu. Jeśli tak jest, program odpowiednio zwiększa lub białe spacje. Wybrane przykłady fiszek przed i po korekcie można zobaczyć poniżej (**ilustracje 46–53**).

Ilustracja 46. Hasło „Ibungsmarś” – przed korektą.

[Ibungsmarś, a, Im. e] ćwiczenia marszowe: Na
I. idę. <Nm. Uebungsmarsch>

Ilustracja 47. Hasło „Ibungsmarś” – po korekcie.

[Ibungsmarś, a, Im. e] ćwiczenia marszowe: Na
I. idę. <Nm. Uebungsmarsch>

Ilustracja 48. Hasło „Ichni” – przed korektą.

[Ichni] ich (z. dzierżawczy 3 os. 1m.). <Od Ieh >

Ilustracja 49. Hasło „Ichni” – po korekcie.

[Ichni] ich (z. dzierżawczy 3 os. 1m.). <Od Ieh >

Ilustracja 50. Hasło „Ichnograficzny” – przed korektą.

× Ichnograficzny przym. od Ichnografja.

Ilustracja 51. Hasło „Ichnograficzny” – po korekcie.

× Ichnograficzny przym. od Ichnografja.

Ilustracja 52. Hasło „Ichtjograficzny” – przed korektą.

× Ichtjograficzny przym. od Ichtjografja.

Ilustracja 53. Hasło „Ichtjograficzny” – po korekcie.

× Ichtjograficzny przym. od Ichtjografja.

Na powyższych przykładach widać jednak przede wszystkim obecność nadmiarowych elementów i zbyt duże marginesy. Może się jednak zdarzyć, że program usunie pożądany glif (np. znak diakrytyczny), jeśli ten będzie przylegał do górnego brzegu strony. Na szczęście został odnotowany tylko jeden taki przypadek w całym materiale testowym, zilustrowany poniżej:

Ilustracja 54. Hasło „Icek” – przed korektą.

Icek, icka, 1m. ickowie, Icyk Żyd. < Hb. Iechāk = Izaak >

Ilustracja 55. Hasło „Icek” – po korekcie.

Icek, icka, Im. ickowie, loyk Zyd. < Hb.
[cehāk = [zaak >

8.6.7. Znane problemy

Jeszcze w trakcie tworzenia programu zostały przez autora tych słów dostrzeżone pewne problemy, a także i potencjalnie rozwiązania, których implementacji można by było się podjąć w przypadku wznowienia prac nad SW. Poniżej wymienionych jest kilka z nich, wraz z rozważanymi rozwiązaniami.

8.6.7.1. Korekcja zniekształceń

Jak wspomniano, program w obecnej wersji nie jest w stanie automatycznie wyróżnić zniekształceń perspektywicznych i geometrycznych. Co prawda biblioteka OpenCV posiada funkcje do ich korekcji, konieczne jest im jednak podanie parametrów zniekształcenia. Potencjalnie sposobem na prawidłowe rozpoznanie części zniekształceń byłoby odejście od opisywania na tekście najmniejszego możliwego prostokąta, tylko zamiast tego ustalenie większej liczby punktów na krawędziach tekstu. Przed implementacją należałoby jednak uwzględnić, że zarówno kształt liter, jak i akapitów, jest inherentnie nieregularny, przez co znalezienie konturu tekstu nie jest trywialne.

8.6.7.2. Dynamiczne ustalanie jasności

Ze względu na niejednorodną jakość druku poszczególnych stron SW oraz degradację danych wynikających z utworzenia kopii cyfrowej, niektóre strony wymagały zmiany parametrów binaryzacji, aby usunąć szumy lub zapobiec utracie pożądanych informacji. Remedium na konieczność takiej manipulacji byłaby implementacja funkcji, która najpierw określałaby średnią jasność obszarów tekstowych i dopiero na jej podstawie automatycznie parametryzowała zarówno progowanie, jak i intensywność erozji oraz dyatacji.

8.6.7.3. Usuwanie samotnych wysp pikseli

Pojęcie to odnosi się do rodzaju szumu w postaci niewielkich zgrupowań czarnych pikseli, które są relatywnie oddalone od innych regionów z czarnymi pikselami na obrazie binarnym. Jest to jeden z dwóch głównych powodów błędnego rozpoznawania orientacji strony w badanym materiale. Dla opisaną wcześniej metody, która usuwa tego rodzaju szumy – otwierania – niektóre skupiska pikseli są jednakże na tyle duże, że ich usunięcie tą metodą pociągnęłoby za sobą degradację tekstu. Możliwym rozwiązaniem mogłoby być dodanie etapu, w którym program znajduje po znalezieniu konturów wszystkich regionów z czarnymi pikselami na obrazie usuwa te, które są blisko krawędzi obrazu, a współrzędne konturów są w znacznie większej odległości od najbliższych konturów, niż wynosi średni na obrazie.

8.6.7.4. Dynamiczne mierzenie wcięć

W niektórych sytuacjach, aby osiągnąć najlepszy jakościowo efekt, parametry dla niektórych obrazów lub grup obrazów były modyfikowane; pośród nich najczęściej były to wymiary wcięć. Sposobem, który mógłby zniwelować potrzebę ludzkiej ingerencji w ten parametr byłoby dynamicznie ustalanie tychże wymiarów dla poszczególnych obrazów kolumn. Wówczas mierzona by była średnia wysokość wiersza jak i średnia szerokość światła na początku akapitu. Wstępnie podjęte próby implementacji takiego rozwiązania wykazały jednak, że konieczne byłoby najpierw dużo skuteczniejsze manipulowanie geometrią.

8.6.7.5. Ulepszone usuwanie artefaktów

W chwili obecnej na obrazach wyjściowych z pojedynczymi akapitami pojawia się stosunkowo duża liczba artefaktów, które przeważnie są fragmentami tekstu z sąsiednich akapitów. Autorowi nie udało się obrać parametrów w taki sposób, żeby problemu tego zupełnie uniknąć. Nawet częściowe rozwiązanie problemu istotnie zmniejszyłoby czas obróbki manualnej obrazów wyjściowych.

8.6.7.6. Wielowątkowość

O ile operacje, które program wykonuje na pojedynczej stronie mają linearny charakter, to jednym ze sposobów na zwiększenie wydajności programu jest przetwarzanie wielu stron jednocześnie. Implementacją tego rodzaju rozwiązań zajmuje się programowanie współbieżne (ang. *concurrent programming*). Python daje możliwość zastosowania dwóch metod na osiągnięcie równoczesnych obliczeń: wielowątkowości (ang. *multithreading*) oraz wieloprocusowości (ang. *multiprocessing*)⁶². Zastosowanie tych metod nie było jednak zasadne ze względu na niewielki zakres materiału badawczego, ale być może byłoby wskazane w przypadku pracy z większym materiałem.

8.6.8. Wydajność

W czasie prac rozwojowych nad programem IndexCardSplitter kilkakrotnie ulepszano sprzęt, jednak wszystkie konfiguracje sprzętowe mieściły się w normie komputerów osobistych.

Specyfikacja pierwszej stacji roboczej to: Intel® Core™ i5-3210M CPU @ 2.50 GHz, 8 GB pamięci RAM, natomiast specyfikacja ostatniej stacji roboczej to: Intel® Core™ i7-9750 CPU @ 2.60 GHz, 32 GB RAM; oba komputery były wyposażone w dyski SSD i system operacyjny Windows 10. Wydajność podana poniżej jest w oparciu o tę ostatnią konfigurację, na której program działał niemal 4,5 szybciej niż na konfiguracji początkowej.

⁶² Koncepty te nie będą tutaj zgłębiane – przystępny zarys ich problematyki jest jednak dostępny w artykule Marcusa McCurdy'ego zatytułowanego „Python Multithreading and Multiprocessing Tutorial”: <https://www.toptal.com/python/beginners-guide-to-concurrency-and-parallelism-in-python>.

Głównym czynnikiem wpływającym na czas przetworzenia pojedynczej strony był stopień kompresji obrazów wyjściowych. W przypadku plików Portable Network Graphics (.png) dostępne jest dziewięć poziomów bezstratnej kompresji – różnica między nimi odczuwalna jest tylko podczas zapisywania obrazu.

W przypadku badanego materiału badawczego czas potrzebny na przetworzenie pojedynczej strony na komputerze osobistym wynosi około 1 sekundy przy najniższym stopniu kompresji, a średnio około 26 sekund przy najwyższym. Kompresja ta zredukowała jednak łączny rozmiar fiszek o zaledwie kilka procent, ponieważ obrazy w czerni i bieli inherentnie mają niewielkie rozmiary ze względu na bardzo małą przestrzeń barw. Różnica w czasie zapisu plików wynikała przede wszystkim z tego, że program zapisuje też pośrednie etapy przetwarzania obrazów, na których obecna jest większa paleta barw. Obrazy te jednak potrzebne są tylko w fazie testów, dlatego stosowanie kompresji daje niewielki zysk.

Oznacza to, że program potrzebuje około dwóch godzin na przetworzenie całej treści zasadniczej SW, co też zostało przetestowane i dało rezultat zgodny z przewidywaniami.

Chociaż wydajność programu ICS można by było w kilku miejscach usprawnić, według piszącego te słowa zysk byłby niewspółmierny do nakładu pracy i potencjalnej utraty czytelności kodu.

8.7. Rezultaty

Czas potrzebny na opracowanie, wdrożenie i przetestowanie programu IndexCardSplitter na wybranym materiale badawczym zamknął się w granicy 100 roboczogodzin. Jest to czas niemal dwukrotnie większy, niż potrzebny na przetworzenie stron z hasłami na literę F metodą manualną. Niemniej, jak to wcześniej zostało obliczone, opracowanie manualne całego słownika wymagałoby ok. 4000 roboczogodzin.

Oczywiście różnica czasu pracy ludzkiej i czasu opracowania ICS, czyli 3900 godzin, nie jest miarodajnym wynikiem, ponieważ program ten wymaga ingerencji człowieka zarówno w dane wyjściowe, jak i wyjściowe, lub włożenia dodatkowej pracy w jego usprawnienie. Trudno jest określić dokładny czas sumaryczny potrzebny do obróbki graficznej całego materiału w sposób zadowalający metodą hybrydową, jednak przetworzenie całego zakresu SW w ICS przy domyślnych parametrach wygenerowało fiszki, z których ok. 70% nie wymagało ingerencji.

Biorąc jednocześnie pod uwagę, że artefakty i zniekształcenia wpływające na rozpoznanie orientacji strony występują w przybliżeniu na 10–20% stron, piszący te słowa zakłada, że oszczędność czasu z opracowania programu IndexCardSplitter to kilkadziesiąt procent, a wynik ten prawdopodobnie można poprawić dzięki wprowadzeniu przynajmniej części rozwiązań zaproponowanych w **rozdziale 8.6.7**.

8.8. Alternatywne rozwiązania

Autor niniejszej tezy pragnie w tym miejscu wspomnieć o dwóch alternatywnych metodach. Pierwszą z nich jest odrzucona wcześniej metoda wykorzystująca uczenie maszynowe i sieci neuronowe. Warunkiem pozwalającym na skuteczne zastosowanie tej metody byłoby albo powstanie metod wymagających mniejszego zbioru danych, albo poszerzenie zakresu danych badawczych do rozmiarów, które uzasadniałyby implementację tego rozwiązania.

Drugą metodą jest wykorzystanie oprogramowania do OCR, które podczas rozpoznawania tekstu tworzy pliki robocze składające się zasadniczo z trzech warstw danych: oryginalnego obrazu, rozpoznanego tekstu, oraz koordynat, które opisują relację między poszczególnymi znakami a regionami na obrazie. Fakt wykorzystał m.in. Łukasz Borchmann w swoim narzędziu opracowanym na potrzeby projektu grantowego zatytułowanego *Narodowy Fotokorpus Języka Polskiego*⁶³, które przy pomocy interfejsu programistycznego aplikacji ABBYY FineReader dokonuje ekstrakcji współrzędnych poszczególnych akapitów; podobne podejście można znaleźć także u Bienia (2008b). O ile zastosowanie to bardzo dobrze sprawdziło się na tradycyjnych dziełach książkowych, to w przypadku słownikowego układu haseł rozpoznawanie akapitów okazało się niekonsekwentne, co skłoniło autora tych słów do znalezienia innego rozwiązania. Nie oznacza to jednak, że adaptacja takiego narzędzia jest wykluczona.

⁶³ Strona grantu, w którym i piszący te słowa miał swój udział, jest dostępna jest pod adresem: <http://nfjp.pl/>.

9. Konwersja warstwy graficznej na tekstową

W chwili rozpoczęcia pisania niniejszej pracy treść *Słownika warszawskiego* była dostępna w dwóch formach: graficznej oraz tekstowej w postaci dwóch warstw danych w pojedynczym pliku DjVu. Analiza składniowa na obrazach jest jednak możliwa jedynie na tekście, a nie na obrazach. Wynika to z faktu, że programy komputerowe traktują tekst inaczej niż człowiek; dane te nie mają dla nich charakteru analogowego czy graficznego, tylko dyskretny.

O ile człowiek może w przypadku niedokładnego zapisu czy druku – lub cyfrowej reprezentacji graficznej – pomylić ze sobą dwie podobne do siebie litery (np. małą literę **l** z wielką literą **L**, literę **l** z literą **ł** lub cyfrą **1**, literę **c** z literą **o** itd.), to komputery do przechowywania tekstu używają systemu, w którym (w przypadku kodowania w systemie Unikon) danej wartości szesnastkowej odpowiada dokładnie jeden znak. Oznacza to, iż każdy znak cyfrowy (lub ich kombinacja) posiada jedną interpretację. Zatem mapowanie treści analogowej, o charakterze ciągłym, na dane cyfrowe, o charakterze dyskretnym stanowi pewien problem.

Jeśli tekst nie powstał natywnie w formie cyfrowej, można go wprowadzić na komputer w ramach digitalizacji, która może przybrać jedną z dwóch postaci:

1. automatycznej konwersji obrazu na tekst;
2. transkrypcji.

Ten pierwszy przypadek oznacza zastosowanie technologii OCR (ang. *optical character recognition* ‘optyczne rozpoznawanie znaków’). Polega ona na próbie przyporządkowania fragmentów obrazu kształtom odpowiadającym poszczególnym znakom cyfrowym. Jedyną alternatywą dla OCR to ręczne wprowadzanie tekstu (Piotrowski 2012: 48–49) i jego przepisywanie wspomagane komputerowo (Piotrowski 2012: 49–51) – przykładem narzędzia, które do tego służy, jest program Transcribus⁶⁴. Ręczne przepisywanie tekstu jest jednak dużo powolniejszym procesem, przez co ograniczone jest niemal wyłącznie do digitalizacji pisma odręcznego, które jest dużym wyzwaniem dla komputerów.

Mimo kilku dekad rozwoju technologii OCR rozpoznawanie tekstu nadal stanowi problem i jego rezultat często wymaga dokonania korekty przez człowieka. Bierze się to stąd, że interpretacja graficznego znaku na dyskretny symbol, choć przeważnie prosta i intuicyjna dla człowieka, sprawia trudności przy opisie algorytmicznym.

Problemy mogą wynikać z degradacji papieru w czasie, ograniczeń techniki drukarskiej, nietypowej grafii czy obecności znaków współcześnie nieużywanych, lub ambiwalentnych w danym kontekście; nawet w sytuacji konwersji obrazu z tekstem o wysokiej rozdzielczości i jakości rezultat przetwarzania oprogramowania OCR wciąż generuje błędy. Ze względu na tę nieprzewidywalność rezultatów stosuje się pojęcie „brudnego OCR” (ang. *dirty OCR*), które Bień (2009a: 162) definiuje następująco:

⁶⁴ Strona projektu: <https://transcribus.eu/lite/>.

If the process is unattended and the results are neither verified nor corrected by humans, it is called “dirty OCR”.

Jeśli proces odbywa się bez nadzoru, a wyniki nie są weryfikowane ani korygowane przez człowieka, nazywa się to „brudnym OCR”.

Radzenie sobie z brudnym OCR jest dość skomplikowane (Bień 2011c). Jednym ze sposobów mitygacji problemów z odczytem i przeszukiwaniem takiego tekstu jest udostępnianie obrazów cyfrowych (skanów) wraz z automatycznie rozpoznanyim tekstem i konstruowanie złożonych kwerend przy pomocy wyrażeń regularnych, co pozwala na znalezienie fragmentów tekstu o niewielkiej dewiacji (Bień 2012b: 25). Choć rozwiązanie to jest przydatne, to stanowi ono jednak półśrodek wobec niskiej jakości danych.

9.1. Przebieg procesu OCR

Proces OCR może się nieznacznie różnić w między poszczególnymi programami, jednak można w nim wyróżnić kilka zasadniczych etapów wspólnych dla wszystkich lub niemal wszystkich dotychczasowych komercyjnych rozwiązań.

9.1.1. Przetwarzanie wstępne

W ramach tego etapu na początku dokonywana jest obróbka graficzna, która ułatwia (a czasem wręcz umożliwia) poprawne rozpoznanie tekstu. Obejmuje ona najczęściej:

- redukcję kolorów do czerni i bieli (binaryzacja);
- usuwanie szumu i samotnych wyps;
- rotację;
- redukcję przekosów i zniekształceń geometrycznych;
- wyprostowanie linii tekstu;
- normalizacja wielkości.

Niektóre programy oferują użytkownikowi ingerencję w powyższe czynności, jednak domyślnie wykonywane one są automatycznie.

Następnie oprogramowanie do OCR przeprowadza wstępne przetwarzanie tekstu, które obejmuje m.in.:

- usunięcie linii nietekstowych;
- rozpoznanie układu strony;
- rozpoznanie fontu tekstu;
- rozpoznanie języka tekstu;
- wyodrębnienie linii tekstu;
- wyodrębnienie granic słów;
- wyodrębnienie granic znaków.

Kiedy poszczególne znaki są wyizolowane, można przejść do ich odpowiedniego rozpoznania.

9.1.2. Rozpoznawanie znaków

Rozpoznawanie znaków we współczesnym oprogramowaniu OCR najczęściej wykorzystuje połączenie trzech technik:

1. porównania wektorowego (ang. *pattern matching* lub *feature extraction*) – polegającego na sprowadzeniu poszczególnych znaków do uproszczonych kształtów i próbie znalezienia znaku ze słownika o najbardziej zbliżonym kształcie;
2. porównywanie rastrowe (ang. *pixel by pixel matching* lub *matrix matching*) – polegające na porównaniu pikseli składających się na obraz danego wystąpienia znaku z pikselami prototypowego obrazu znaku;
3. porównywanie słownikowe (ang. *near-neighbor analysis*) – polegające na wykorzystaniu słownika do rozpoznania, czy dany wyraz istnieje, oraz czy wyrazy sąsiadujące często z nim współwystępują.

Wyjątkowo istotny jest szczególnie ten trzeci komponent, który nierzadko ma charakter decyzyjny; wcześniejsze etapy bowiem generują przeważnie rankingi możliwych znaków uporządkowane pod względem trafności, natomiast wykorzystanie informacji o danym języku, zwłaszcza o słownik, pozwala podjąć decyzję opartą o kontekst (Piotrowski 2012: 31).

9.1.3. Przetwarzanie końcowe

Po rozpoznaniu znaków następuje ostatni etap, który polega na nadaniu tekstowi formatowania i układu zgodnego ze źródłem, a także optymalizacje danych wyjściowych zgodne z ustawieniami i przeznaczeniem danego oprogramowania. W programach o przeznaczeniu ogólnym wiąże się to z nadaniem konkretnego formatu pliku.

9.2. Rozpoznawanie tekstu a SW

Rozpoznawanie tekstu zawartego w *Słowniku warszawskim* stanowi wyzwanie z kilku powodów:

1. Zeskanowane odbitki *SW* posiadają wiele artefaktów.
2. W tekście *SW* występuje wiele nietypowych znaków, wykraczających poza zakres polskiego alfabetu, interpunkcji i cyfr arabskich.
3. Język treści *SW* pokrywa się ze współczesnym tylko częściowo, przez co nie ma niepełne wsparcie słownikowe w oprogramowaniu OCR.
4. Programy do OCR są zoptymalizowane pod tekst w formie zwartej, a nie w formie typowym dla leksykonu.

Wszystko to sprawia, że jakość tekstu wyjściowego jest stosunkowo niska. Co prawda w *SW* nie występują takie problemy z digitalizacją jak np. w przypadku *SL* – który typograficznie jest znacznie trudniejszy ze względu na fragmenty zapisane frakturą, dwoma rodzajami cyrylicy, alfabetem hebrajskim i alfabetem greckim (Bilińska 2013: 97) – ale nadal wymaga istotnego nakładu pracy korektorskiej.

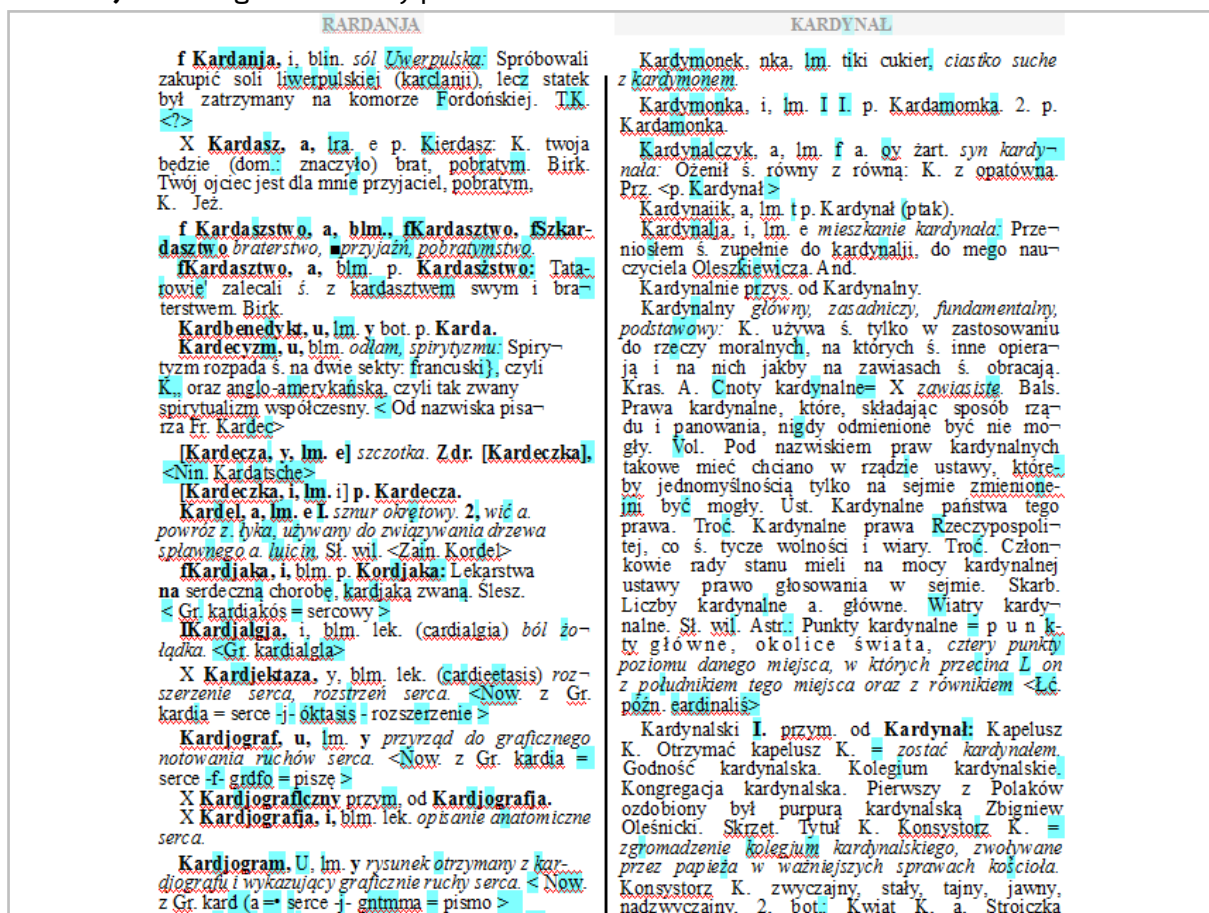
Poniżej przedstawiony jest oryginalny obraz strony (**ilustracja 56**) oraz wynik przetwarzania tejże strony przez program ABBYY FineReader 14 bez poprawek ze strony użytkownika

(ilustracja 57 – jasnoniebieskie tło wskazuje na znaki rozpoznane przez program do OCR jako wątpliwe).

Ilustracja 56. Fragment obrazu strony.

KARDANJA	KARDYNAŁ
<p>†Kardanja, i, blm. <i>sól liwepulska</i>: Spróbowali zakupić soli liwepulskiej (kardanji), lecz statek był zatrzymany na komorze Fordońskiej. T. K. <?></p>	<p>Kardymonek, nka, lm. nki cukier. <i>ciastko suche z kardymonem</i>.</p>
<p>×Kardasz, a, lm. e p. Kierdasz: K. twoja będzie (dom.: znaczyło) brat, pobratym. Birk. Twój ojciec jest dla mnie przyjacielem, pobratym, K. Jez.</p>	<p>Kardymonka, i, lm. i f. p. Kardamonka. 2. p. Kardamonka.</p>
<p>†Kardaszstwo, a, blm. p. Kardaszstwo: Tata- rowie zalecali ś. z kardaszstwem swym i bra- terstwem. Birk.</p>	<p>Kardynalczyk, a, lm. i a. oy żart. <i>syn kardyna- la</i>: Ożenił ś. równy z równą: K. z opatówną. Prz. <p. Kardynał></p>
<p>†Kardaszstwo, a, blm. p. Kardaszstwo: Tata- rowie zalecali ś. z kardaszstwem swym i bra- terstwem. Birk.</p>	<p>Kardynał, a, lm. i p. Kardynał (ptak). Kardynał, i, lm. e <i>mieszkanie kardynała</i>: Prze- niósł ś. zupełnie do kardynalji, do mego nau- czyciela Oleszkiewicza. Aud.</p>
<p>Kardbenedykt, u, lm. y bot. p. Karda. Kardecyzm, u, blm. <i>odłam spirytyzmu</i>: Spiry- tyzm rozpada ś. na dwie sekty: francuską czyli K., oraz anglo-amerykańską, czyli tak zwany spirytyzizm współczesny. <Od nazwiska pisa- rza Fr. Kardec></p>	<p>Kardynałnie przys. od Kardynałny. Kardynałny <i>główny, zasadniczy, fundamentalny, podstawowy</i>: K. używa ś. tylko w zastosowaniu do rzeczy moralnych, na których ś. inne opiera- ją i na nich jakby na zawiasach ś. obracałg. Kras. A. Onoty kardynałne = × <i>zawiasiste</i>. Bals. Prawa kardynałne, które, składając sposób rzą- du i panowania, nigdy odmienione być nie mo- gły. Vol. Pod nazwiskiem praw kardynałnych takowe mieć chciało w rządzie ustawy, które- by jednomyślnością tylko na sejmie zmienione- mi być mogły. Ust. Kardynałne państwa tego prawa. Troe. Kardynałne prawa Rzeczypospoli- tej, eo ś. tycze wolności i wiary. Troe. Człon- kowie rady stanu mieli na mocy kardynałnej ustawy prawo głosowania w sejmie. Skarb. Liczby kardynałne a. główne. Wiatry kardyna- łne. Sł. wil. Astr.: Punkty kardynałne = punk- ty główne, okolice świata, cztery punkty poziomu danego miejsca, w których przecina ś. on z południkiem tego miejsca oraz z równikiem. <Łć. późn. cardinalis></p>
<p>[Kardecza, y, lm. e] <i>szczołka</i>. Zdr. [Kardeczka].</p>	<p>Kardynałski i. przym. od Kardynał: Kapelusz K. Otrzymać kapelusz K. = <i>zostać kardynałem</i>. Godność kardynałska. Kolegium kardynałskie. Kongregacja kardynałska. Pierwszy z Polaków ozdobiony był purpurą kardynałską Zbigniew Oleśnicki. Skrzet. Tytuł K. Konsystorz K. = <i>zyromadzenie kolegium kardynałskiego, zwoływane przez papieża w ważniejszych sprawach kościoła</i>. Konsystorz K. zwyczajny, stały, tajny, jawny, nadzwyczajny. 2. bot.: Kwiat K. a. Stroiczka</p>
<p><Nm. Kardätsche></p>	
<p>[Kardeczka, i, lm. i] p. Kardecza.</p>	
<p>Kardel, a, lm. e i. <i>sznur okrętowy</i>. 2. <i>wió a. powróż z łyka, używany do związywania drzewa splewanego a. wicin</i>. Sł. wil. <Zam. Kordel></p>	
<p>†Kardjaka, i, blm. p. Kordjaka: Lekarstwa na serdeczną chorobę, kardjaką zwaną. Slesz. <Gr. kardiakós = sercowy></p>	
<p>!Kardjalgia, i, blm. lek. (cardialgia) <i>ból zo- łądka</i>. <Gr. kardialgia></p>	
<p>×Kardjektaza, y, blm. lek. (cardiectasis) <i>roz- szerzenie serca, rozstrzeń serca</i>. <Now. z Gr. kardía = serce + éktasis = rozszerzenie></p>	
<p>Kardjograf, u, lm. y <i>przyrząd do graficznego notowania ruchów serca</i>. <Now. z Gr. kardía = serce + gráfō = pisać></p>	
<p>×Kardjograficzny przym. od Kardjografja.</p>	
<p>×Kardjografja, i, blm. lek. <i>opisanie anatomiczne serca</i>.</p>	
<p>Kardjogram, u, lm. y <i>rysunek otrzymany z kar- djografu i wykazujący graficznie ruchy serca</i>. <Now. z Gr. kardía = serce + grámma = pismo></p>	

Ilustracja 57. Fragment strony po OCR.



Jak widać, program ten – choć mający silną pozycję na rynku – nie rozpoznaje poprawnie sporej części tekstu, momentami do tego stopnia, że utrudnia jego odczytanie i rekonstrukcję bez odwołania się do obrazu źródłowego. O ile „zwykłe” fragmenty tekstu jest łatwo poprawić, to zastosowanie znaków co prawda objętych systemem Unikod, ale rzadko stosowanych lub pochodzących z innych języków wymaga dodatkowych działań.

Aby znaki takie wprowadzić, został wykorzystany autorski układ klawiatury, opisany w **rozdziale 7.9**. Cały tekst został kilkakrotnie dokładnie przeczytany, a odpowiednie poprawki zostały naniesione. Miejsce też miały „celowe i przypadkowe podmiany znaków” (Biń & Biłńska & Sarnecki 2014: 7b), które wynikały z potrzeby standaryzacji tekstu na potrzeby analizy składniowej oraz poprawnego wyświetlania tekstu na komputerze.

Tekst cyfrowy odpowiadający treści oryginału – bez ingerencji w błędy autorów – został zapisany w formacie czystego pliku tekstowego `.txt` i posłużył jako dane wejściowe dla parsera, co jest opisane w rozdziale następnym.

Do rozwiązań zastosowanych w niniejszej pracy doktorskiej wykorzystano język programowania Ruby w wersjach od 1.9 do 3.0.

10. Przetwarzanie danych SW

Do analizy i parsowania treści *Słownika warszawskiego* w formie tekstowej został napisany program SWParser. Program ten jest w stanie nie tylko przetworzyć nieustrukturyzowany tekst w strukturę XML odpowiadającej budowie hasła – czyli dokonać jego ubazodanowienia – ale także odfiltrować poszczególne rodzaje informacji czy wygenerować nowe rodzaje danych.

Program ten został napisany w języku Ruby, ponieważ nadaje się on do zastosowań językoznawczych, a także z powodu preferencji autora do składni tego języka w opozycji do języków Python czy Perl.

Kod programu, po odrzuceniu komentarzy, osiąga długość ponad 7 000 linii w języku Ruby oraz ok. 1 000 linii kodu mieszanego w postaci schematów Embedded Ruby (ERB), przy pomocy których generowany jest wielokrotnie dłuższy kod w języku L^AT_EX.

Czas poświęcony na jego napisanie, choć nie był ściśle mierzony, najprawdopodobniej jest porównywalny do tego, który byłby konieczny do ręcznego otagowania treści SW. Podobnie jednak jak w przypadku programu IndexCardSplitter rozwiązanie to jest skalowalne, co będzie omówione w **podrozdziale 10.8.6**. Poza tym wykorzystanie metod automatycznych gwarantuje wysoką jakość danych – to właśnie dzięki tworzeniu i sprawdzaniu precyzyjnych reguł algorytmicznych, które z nieubłaganą precyzją były wykonywane przez komputer, autor odkrywał zawiłości w strukturze hasłowej i dewiacje od niej, co w innym wypadku mogłyby nie zostać dostrzeżone.

Oczywiście istnieją też głosy takie jak Dummer et al. (1998: 207), według których „Ein großer Teil der wünschenswerten Markierungen konnte also nur manuell erfolgen.” (‘Dużą część pożądaných oznaczeń można osiągnąć także ręcznie’). Autor tejże pracy doktorskiej przyjął jednak strategię, w ramach której żadne znaczniki nie są wprowadzone ręcznie – co najwyżej dodawane są wskazówki, które pomagają algorytmowi poradzić sobie z trudniejszymi przypadkami, ma to jednak miejsce w ograniczonej liczbie przypadków.

Funkcjonowanie programu ma charakter w pełni charakterystyczny i deterministyczny, w opozycji do popularnych dzisiaj metod uczenia maszynowego – strategia ta została przyjęta z czterech powodów:

1. **stosunkowo nietypowa charakterystyka projektu** – struktura logiczna artykułu hasłowego w dużym stopniu opiera się o ludzką percepcję i wiedzę pozasłownikową. Trudno przez to znaleźć analogiczne projektu, z których można by było czerpać rozwiązania;
2. **trudno o dostateczny materiał treningowy** – inne słowniki zbyt różnią i nie mają opisu strukturalnego, na podstawie którego można by było wytrenować system. Konieczne byłoby więc trenowanie uczenia maszynowego na bardzo dużym fragmencie samego SW, nieproporcjonalnie większym od tego, jaki został tu opracowany;
3. **celem jest maksymalna precyzja** – metody statystyczne niemal zawsze są kompromisem w próbie optymalizacji między nakładem pracy a jakością, tutaj jednak wysoka jakość była celem nadrzędnym;

4. **niskiej wiedzy autora** – próba zastosowania rozwiązań automatycznych wiązałyby się z dużym nakładem pracy na naukę i istotnym ryzykiem fiaska, którego piszący te słowa nie mógł

Na potrzeby programu SWParser i niniejszej tezy doktorskiej w pełni opracowano dane obejmujące część zasadniczą słownika z hasłami rozpoczynającymi się na litery **F**, częściowo także **Ó, V, X** oraz **Y**; wstępnie przygotowano i przeanalizowano również hasła na literę **I**.

Aby móc opracować materiał źródłowy, konieczne było postawienie sobie kilku istotnych pytań:

1. Jakimi dyrektywami kierowali się autorzy redagując artykułu hasłowe?
2. W jaki sposób dyrektywy te zostały identyfikowane?
3. Jakie istotne kategorie strukturalne można wyróżnić?
4. W jaki sposób powyższe można wykorzystać w analizie składniowej?

Na pytania te można udzielić następujących odpowiedzi:

- Ad. 1. Dyrektywy, którymi kierowali się autorzy, znane są tylko częściowo (co zostało wyjaśnione w rozdziale 3). Wiele dyrektyw udało się jednak zrekonstruować na podstawie analizy budowy haseł.
- Ad. 2. Sposób identyfikacji reguł dokonywany był w sposób hybrydowy – obejmuje ona analizę i interpretację rezultatów pracy analizatora składniowego oraz ich weryfikację.
- Ad. 3. Istotne kategorie zostały wyróżnione wtórnie, poprzez analizę strukturalno-funkcjonalną i wskazówki typograficzne.
- Ad. 4. Kategorie wyróżnione w ramach analizy zostały zaimplementowane w ramach opisywanego w tym rozdziale narzędzia, które zapewnia zgodność zapewnia odtworzenie reguł.

Działanie programu SWParser różni się od podobnych rozwiązań pod wieloma względami. Bilińska (2011: 47) w kontekście *SL* pisze o wyszukiwaniu po fontach (krojach), opisywane tu rozwiązanie – jak i prawie całą informację o formatowaniu tekstu – ignoruje. Wyróżnienie typograficzne jednak z reguły stosowane jest do oznaczenia elementów strukturalnych w słowniku.

Nie oznacza to, że rozwiązania zbliżone choć pod pewnymi względami nie miały miejsca – należy tutaj wskazać algorytm rozpoznawania cytatów obcojęzycznych w *SL* (Bilińska 2006: 30–31). Badaczka ta (idem 2013: 97–99) wypisuje znaki interpunkcyjne, które są znaczące z punktu widzenia mikrostruktury *SL* – te są bardzo istotne dla działania programu SWParser, ale rozpatrywane są one dopiero w kontekście poszczególnych elementów hasła. Podobna różnica w strategiach między piszącym te słowa a Bilińską widoczna jest również w przypadku cyfr oraz zabiegów interpunkcyjnych (Bilińska 2013: 100–101).

Większość programów do przetwarzania języka nie służy jednak do parsowania struktury leksykograficznej, ale do analizy morfologicznej:

Przez *analizę morfologiczną* rozumiemy pewną operację (program lub algorytm ją realizujący nazwiemy *analizatorem morfologicznym*), która dla każdego słowa stanowiącego dane wejściowe produkuje pewien jego opis. (Bień & Szafran 2001: 1)

Najważniejszym zadaniem analizatorów morfologicznym jest przypisanie wyrazowi formy hasłowej (Bień & Szafran 2001: 2). Jak ujął to sam Szafran (1998: 26):

Punktem wyjściowym tradycyjnej analizy morfologicznej jest skojarzenie słowa tekstowego z określonym leksemem. Umożliwia to nadanie mu określonej struktury i kategoryzację elementów tej struktury.

Nad stworzeniem indeksu morfologicznego dla języka polskiego pracował Jan Tokarski, dzieło to jednak zostało dokończzone i uzupełnione dopiero przez Saloniego; na podstawie tego indeksu Szafran utworzył wersję elektroniczną (Bień & Szafran 2001: 2–4).

Wyniki analizy morfologicznej wraz z systemem L^AT_EX można wykorzystać m.in. do tworzenia automatycznych skorowidzów (Bień & Szafran 2001: 8). Poza tym głębokie parsowanie morfologiczne i semantyczne ma realne zastosowania w analizie tekstu – znaczenie słów w kontekście wpływa na interpretację wieloznacznych form gramatycznych, jednak analiza na poziomie zdaniowym to trudne zadanie, dlatego też – jak pisze Szafran (1998: 27):

[P]unktem wyjściowym komputerowej analizy morfologicznej tekstu polskiego powinno być słowo tekstowe rozpatrywane samodzielnie, bez towarzyszącego mu kontekstu. (...)

Ceną za rezygnację z kontekstu jest pewna niejednoznaczność wyniku.

Analiza morfologiczna jest jednak analizą czysto językową – zadaniem opisywanego tu oprogramowania nie jest jednak praca nad językiem jako takim, a nad jego reprezentacją w słowniku. Niemniej, część prezentowanych w dalszej części tego rozdziału rezultatów ma bezpośrednie zastosowanie w systemach, które do analizy morfologicznej mogą służyć. Znaczenie szerzej wykorzystany będzie tutaj system L^AT_EX, który posłuży do automatycznego wygenerowania wszystkich aneksów.

10.1. Standardy w dziedzinie

Należy na początku wspomnieć, że istnieje standard opisu i przechowywania danych w tej dziedzinie, wypracowany przez European Network of e-Lexicography⁶⁵, inicjatywę, w ramach której zachęca się badaczy do ucyfrowienia zasobów przy pomocy standardu TEI (*Text Encoding Initiative* – ‘Inicjatywa Kodowania Tekstu’)⁶⁶ opartego o format XML. Piotrowski (2012: 61) definiuje standard anotacji TEI:

The TEI Guidelines define a markup language for annotating texts with respect to their structural, visual, and conceptual properties.

Wytyczne TEI definiują język znaczników do anotacji tekstów z uwzględnieniem ich strukturalnych, wizualnych, i konceptualnych właściwości.

TEI pozwala na anotowanie różnych typów tekstów, w tym słowników. Standard ten nie jest jednak sztywny, o czym pisze Piotrowski (2012: 64):

⁶⁵ Strona projektu: <https://www.elexicography.eu/>.

⁶⁶ Strona projektu: <https://tei-c.org/>.

TEI is not a mechanical process, but depends in many cases on the paleographical, linguistic, and historical interpretation of the text. Thus, many different encodings are possible.

TEI nie jest mechanicznym procesem, ale w wielu przypadkach zależy od paleograficznej, lingwistycznej i historycznej interpretacji tekstu. W związku z tym możliwych jest wiele różnych kodowań.

Przykład tego, jak wygląda korpus zakodowany przy pomocy XML z rekomendacjami TEI oraz XCES (*XML Corpus Encoding Standard*), można znaleźć u Bienia (2006a: 22ff).

O ile sama inicjatywa opiera się na dobrych założeniach, to jej wykonanie spotyka się z głosami krytyki, które przytacza Szafran (2004: 93):

W publikacjach na temat słowników elektronicznych można jednak znaleźć opinie, że w wielu przypadkach standard TEI nie jest wystarczający dla opisu struktury artykułu hasłowego. Dlatego też często twórcy słowników nie stosują się do niego ściśle. Natomiast większość z nich korzysta jednak z jakiegoś formalizmu znacznikowego. Najczęściej jest to XML/SGML.

Innymi słowy TEI ani nie odzwierciedla logicznej struktury słowników (a przynajmniej ich bardzo obszernego podzbioru), ani nie systematyzuje struktur w sposób uniwersalny (do wielu słowników potrzebne są adaptacje, co jest uwzględnione w samym standardzie).

Okazało się to również prawdziwe w kontekście niniejszej pracy. Piszący te słowa początkowo podejmował próby opisanie *Słownika warszawskiego* przy pomocy omawianego standardu (także ze wsparciem jego twórców), jednak okazało się, że przewidziana w TEI struktura ramowa nie odpowiada strukturze SW i wymagane są istotne modyfikacje znaczników. Oznaczało to więc konieczność pójścia na liczne kompromisy („naginanie” formatu TEI do struktury SW) przy wątpliwych zyskach (rezultat był zarówno dosyć odległy od standardów TEI, jak i właściwej reprezentacji struktury logicznej opisywanego słownika).

Zasadniczo brak jest alternatyw względem TEI. W związku z tym na potrzeby niniejszej tezy opracowano odautorski system znaczników oparty na języku XML, który wzoruje się na niektórych rozwiązaniach TEI, a który – w ocenie autora – pozwala znacznie lepiej oddać strukturę hasłową *Słownika warszawskiego*.

10.2. Dane wejściowe

Dane wejściowe, które są wczytywane a następnie parsowane przez program SWParser, zapisane są – z jednym wyjątkiem – w formacie tekstowym z kodowaniem UTF-8. Ich struktura i rozszerzenia się jednak od siebie różnią, co wynika to bezpośrednio z funkcji, jaką pełnią.

Biorąc pod uwagę cechy takie jak funkcja, pochodzenie i zastosowanie danych, można wyróżnić ich trzy kategorie:

1. Dane podstawowe;
2. Dane pomocnicze;
3. Dane predefiniowane.

Podział ten można przedstawić w poniższej tabeli:

Tabela 4. Typy danych wejściowych

	Niezbędne do działania programu?	Wspólne dla wszystkich liter?
Dane podstawowe	✓	✗
Dane pomocnicze	✗	✗
Dane predefiniowane	✓	✓

Dane te zostaną omówione szczegółowo w kolejnych podrozdziałach.

10.2.1. Dane podstawowe

Przez dane podstawowe rozumie się tutaj treść *SW*, która ma formę pliku tekstowego zawierającego listę haseł *Słownika warszawskiego*, w którym jednemu wierszowi zakończonym znakiem nowej linii odpowiada jeden artykuł hasłowy – przy czym artykułów zagnieżdżonych nie traktuje się tutaj osobno, ale jako część hasła głównego, przez co nie są umieszczane w odrębnych wierszach.

Jednocześnie każdy plik wejściowy jest równoznaczny z segmentem słownika, który obejmuje hasła rozpoczynające się na jedną i tę samą literę alfabetu – powiązane jest to z logiką głęboką programu, który traktuje słownik segmentowo. Powody takiego rozwiązania wyjaśnione są w **rozdziale 10.3**.

Podczas opracowywania treści *SW* w warstwie tekstowej znaleziono liczne błędy – obejmujące przede wszystkim literówki i niespójności w strukturze hasłowej. Aby nie tracić informacji o oryginalnym zapisie, błędy te nie zostały poprawione w pliku wejściowym – zamiast tego zastosowano mechanizm, który podczas wczytywania oryginalnej treści uwzględnia korekty umieszczone w pliku podmian, który opisane jest w **rozdziale 10.2.2.1**.

Format danych wejściowych zaprezentowano w **listingu 11**, gdzie dodano znaki końca akapitu (¶), aby wskazać, gdzie są faktyczne końce wierszy z punktu widzenia programu.

Listing 1. Pierwsze 10 wierszy z pliku dla litery F.

```
F. 1. spółgłoska wargowo-zębowa bezdźwięczna, czyli cicha (mocna),
powiewna, przeciągła. W piśmie występuje przeważnie w wyrazach
przyswojonych, w polskich zaś rzadko i głównie w wyrazach
dźwiękonaśladowczych, jak fruwać, fiukać i t. p.; w wymawianiu zaś ukazuje
się często, chociaż się pisze w, mianowicie na końcu wyrazów i w środku po
innych cichych, np. krew, łów, ... kwiat, twój, trwały, ... które brzmią: kref,
łuf, ... kfiat, tfuj, trfały... 2. a. Fa muz. = a) czwarty stopień tonacji c
major: F leży pomiędzy e a g. b) F dur a. major, F moll a. minor = nazwy
tonacji a. gam majorowej (twardej) i minorowej (miękkiej): Gama f minor ma
cztery bemole. Symfonia f major Beethovena. Piosnka (z) f major.¶
Fa nieod. muz. p. F.¶
†Fa! p. Fe!: Nie mogąc wstrzymać, zawołali: fa, fa! L. <Por. łć. fafael>¶
†Fabin, u, blm. jakaś tkanina.¶
[Fabjan, a, blm.] 1. gra dziecinna. 2. jeden z uczestników tej gry. <Zap.
od imienia Fabjan, łć. Fabianus>¶
[Fabor, a, lm. y] p. Fawor. Zdr. [Faborek]. <p. Fawor>¶
```

```
[Faborek, rka, lm. rki] p. Fabor.¶  
[Fabrek, brka, lm. brki] p. Chaber. <p. Chaber>¶  
[Fabrycarz, a, lm. e, Filanc, Chabrycarz, Habrycarz] urzędnik podatkowy.  
<p. Fabryka>¶  
Fabryczka, i, lm. i p. Fabryka.¶
```

Choć zasadniczo plik tekstowy nie zawiera informacji o formatowaniu, to umieszczone zostały w nim tyldy ~, które ograniczają fragmenty tekstu w oryginalne zapisane rozstrzelonym fontem – dzieje się tak, ponieważ rozstrzelenie to, obok funkcji typograficznej, pełni także funkcję strukturalną (co zostało opisane w **rozdziale 10.3.10**), co jest wykorzystywane w procesie parsowania.

10.2.2. Dane pomocnicze

Program SWParser wczytywane są wraz z plikiem głównym zawierającym hasła, o ile takie zostaną mu podane – nie są one niezbędne do działania programu, a jedynie rozszerzają wygenerowane dane. Poniżej zostały omówione trzy typy plików pomocniczych, które opracowano dla haseł na literę *F*.

10.2.2.1. Plik podmian

Dane w tym pliku służą do zastąpienia fragmentów oryginalnej treści zawartej w pliku hasłowym przez ich poprawione wersje. Zasadniczo plik ten nie jest konieczny do działania programu, jednak niektóre błędy w treści *SW* mogą spowodować problemy w parsowaniu konkretnych haseł – zwłaszcza jeśli są to niespójności w budowanie hasła. Część z tych problemów została opisana w **rozdziale 4**.

Dwa rodzaje zmian względem treści nie są jednak uwzględniane.

Po pierwsze są to zmiany regularne typograficzne wprowadzone podczas korekty wyników OCR, które nie odzwierciedlają treści *SW*, ale które ułatwiają parsowanie – przykładem takiej modyfikacji jest np. dodawanie spacji między kwalifikatorem symbolicznym a kwalifikatorem wyrazowym – por.:

- **Farba, y, lm. y, [Farwa]** 1. *koLor, barwa, maść*: (...) Farbą wstydu zarumieniła swe lice. Węg. × Przen.: *cera, pteć*: Farbę ma na twarzy ceglaną. Rej. (...)

Po drugie ignorowane są wszelkie potencjalne błędy w druku, które występują tylko w jednej kopii *SW* – błędy wynikające z robienia odbitek lub niedoskonałości w druku konkretnego egzemplarza nie są uwzględniane jako reprezentatywne dla treści.

Plik ten ma rozszerzenie `.tab`, stosowanym dla plików tekstowych o tabelarycznym układzie, w którym granice kolumn wyznaczone są znakami tabulacji, a wierszy – znakami nowej linii. Poza tym pierwszy wiersz zawiera nagłówki kolumn. Pliki tego rodzaju są zatem szczególnym przypadkiem pliku `.csv`, które przeważnie oddzielone są przecinkami lub średnikami, a rozszerzenie `.tab` jest jedynie konwencją. Dodatkowym powodem stosowania tego rozszerzenia jest fakt, że występują liczne problemy z kodowaniem plików `.csv` w programie Excel, których piszący te słowa pragnął uniknąć.

Plik podmian zawiera sześć kolumn, a każdy wiersz ma następujące pola:

- **Find** – pole z fragmentem tekstu, który ma zostać zastąpiony;
- **Replace** – pole z fragmentem tekstu, który jest poprawiony;
- **Type** – pole z nazwą kategorii podmiany;
- **Comment** – pole z komentarzem objaśniającym i uzasadniającym podmianę;
- **Reference** – pole z przykładami haseł, które uzasadniają podmianę;
- **ConcernedEntry** – pole z hasłem, w którym ma nastąpić podmiana.

W praktyce tylko dwa pierwsze pola są wykorzystywane do wykonania podmiany i uwzględniane w parsowaniu – pozostałe natomiast mają swoje zastosowanie dopiero podczas generowania **Aneksu L** oraz **Aneksu N**.

10.2.2.2. Indeks lokalizacji

Plik ten – również w formacie **.tab** – zawiera informacje, które pozwalają ustalić lokalizację hasła w druku. Zamiast jednak podawać lokalizację przy wszystkich hasłach, wskazane są tylko cztery separatory – hasła wyznaczające granice każdej kolumny na stronie, czyli kolejno:

1. Pierwsze hasło w kolumnie lewej;
2. Ostatnie hasło w kolumnie lewej;
3. Pierwsze hasło w kolumnie prawej;
4. Ostatnie hasło w kolumnie prawej.

Wszystkie hasła podane są wraz dezambiguatorami oraz kwalifikatorami symbolicznymi, ponieważ w przypadku niektórych ich grup jest to jedyna ich cecha dystynktywna.

Dla wszystkich pozostałych haseł lokalizacja jest ustalana dynamicznie, odliczając kolejne hasła między separatorami. Nierzadko pojedyncze hasło obejmuje więcej niż jedną kolumnę (por. hasła: **Fabryka**, **Fachowy**, **Falisty** czy **Fałdzisty**), a nawet stronę (por. hasła **Fakt** i **Fałd**) – wówczas na jego kontynuację wskazuje podanie go jako kolejnych separatorów.

Rozwiązanie to zmniejsza nakład pracy konieczny do stworzenia indeksu, co odbywa się ręcznie, a także pozwala na kompresję danych. Wadą takiego rozwiązania jest niska odporność na błędy – błąd w indeksowaniu pojedynczego hasła może spowodować niepoprawne zlokalizowanie kolejnych haseł aż do następnego separatora kolumny, co jest szczególnie istotne w przypadku generowania **Aneksu F**.

Struktura pierwszych dziesięciu wierszy pliku zawierającego indeks lokalizacji przedstawiona jest w **listingu 2**.

Listing 2. Pierwsze dziesięć wierszy indeksu lokalizacji

Page	LeftFirst	LeftLast	RightFirst	RightLast
708	F.	Fabryka	Fabryka	×Facecista
709	Facecja	Fachowy	Fachowy	×Facylitowanie się
710	Facyt	Fagas	Fagasostwo	Fajczany
711	Fajczany	Fajerwerkarz	Fajerwerkier	×Fajnguld
712	×Fajngult	Fajtnać	Fajtnięcie	Fakt
713	Fakt	Faktyczny	×Fakultatywny	Falansterjanin
714	Falansterjum	Falcowanie	Falczpany	Falisty
715	Falisty	Falowy	Falrurka	Fałd
716	Fałd	Fałdzisty	Fałdzisty	Fałszerstwo
...

10.2.2.3. Karty hasłowe

Karty te, wygenerowane przez program IndexCardSplitter, to dane graficzne w formacie `.png` i tym samym jedyne dane fundamentalnie nietekstowe, które są wykorzystywane w programie SWParser.

Zasadniczo jeden obraz odpowiada jednemu akapitowi tekstu i pojedynczemu artykułowi hasłowemu; zasadniczo, ponieważ można napotkać dwa rodzaje wyjątków:

1. Jeśli artykuł hasłowy podzielony jest między dwie lub więcej kolumn, wówczas fragmenty haseł z poszczególnych kolumn traktowane są jak odrębne akapity – w rezultacie jest tyle kart, na ile kolumn podzielony jest artykuł hasłowy.
2. Jeśli jedno hasło jest zagnieżdżone wewnątrz innego, to oba mają wspólną kartę hasłową (lub karty, jeśli występuje sytuacja z pkt. 1).

Karty hasłowe, wraz indeksem lokalizacji, wykorzystywane są przy generowaniu Aneksu F.

10.2.3. Dane predefiniowane

SWParser wykorzystuje podczas swojego działania także zdane predefiniowane, tj. takie, które są przygotowane ręcznie, ponieważ są „zewnątrzne” względem *SW*, i – w przeciwieństwie do danych opisanych w poprzednich dwóch rozdziałach – wykorzystywane przy parsowaniu wszystkich haseł, niezależnie od tomu czy litery, na którą się zaczynają.

Tylko lista skrótów jest niezbędna do prawidłowego przetworzenia treści *SW*, pozostałe rodzaje danych wykorzystywane są natomiast wyłącznie podczas generowania aneksów.

10.2.3.1. Lista skrótów

Lista skrótów zapisana jest w formacie `.tab` i posiada układ tabelaryczny, w którym jednemu wierszowi (poza wierszem nagłówka) odpowiada pojedyncze skrócenie, a kolejne komórki w wierszu reprezentują obecność, kolejność, formy skrócone i ich rozwinięcia w poszczególnych tomach (osobno w *Spisie skrótów* i *Dalszych skróceniach*), a także zapis ustandaryzowany i informacje o ich klasyfikacji (opisanej w **podrozdziale 5.2**).

Lista ta wykorzystywana jest przez program SWParser w procesie analizy składniowej, a także do wygenerowania aneksu poświęconego skrótom, który jest opisane w **podrozdziale 14.13**.

10.2.3.2. Wzory gramatyczne

Plik ten, również w formacie `.tab`, posiada wzory gramatyczne, przy pomocy których generowane są **Aneksy G** oraz **H**. Mowa tutaj nie o paradygmatach odmiany, ale strukturach informacji o gramatyce zawartych w samym *SW*, dzięki czemu parser jest w stanie zinterpretować znaczenie każdego elementu, a także przypisać część mowy; oryginalnie planowano wykorzystanie go do generowania pełnych paradygmatów, co jednak zarzucono⁶⁷.

Reprezentacja wybranych wzorów gramatycznych w formie tabelarycznej dostępna jest w **tabeli 5**.

Tabela 5. Tabelaryczna reprezentacja wzorów gramatycznych.

PatternName	Pattern	PoS	Args	Description
Noun 1	„nieod.” + gender	N	...	Rzeczownik nieodmienny.
Noun 2	„w lp.” + „nieod.” + „lm.” + suffix + suffix	N	...	Rzeczownik nieodmienny w liczbie pojedynczej.
...
Noun 29	„tylko w 1-ym przypadku”	N	...	Rzeczownik.
Noun 30	suffix + „lm.” + suffix + „zb.”	N	...	Rzeczownik o znaczeniu zbiorowym.
Noun 31	„nieod.” + „rz.”	N	...	Rzeczownik nieodmienny.
Gerund 1	„czynność cz.”	N:GER	...	Odsłownik.
Blend 1	multi-word + „lm.” + multi-word	N:BLEND	...	Zrost z zachowaną fleksją obu członów.
Ordinal 1	„licz. porządkowy”	NUM:ORD	...	Liczebnik porządkowy.
Adjective 1	„przym.”	ADJ	...	Przymiotnik odmienny.
Adjective 2	„nieod.” + „przym.”	ADJ	...	Przymiotnik nieodmienny.
...

10.2.3.3. Pary derywacyjne końcówek męskich i żeńskich

Plik ten zawiera, jak sama nazwa wskazuje, pary końcówek wyrazów męskich i derywowanych od nich feminatywów, które obecne są w hasłach *SW*. Lista ta została opracowana ręcznie na podstawie znalezionych przez parser danych w czasie prac rozwojowych, które ten następnie wykorzystuje do przypisania typu derywacji każdej z takich par w sposób automatyczny, czego rezultatem jest **Aneks P**.

⁶⁷ Wyjaśnienie tej decyzji zostało opisane w rozdziale **10.7.2**.

Treść pliku i jego struktura przedstawione są w **listingu 3**, gdzie puste wartości oznaczają końcówki zerowe).

Listing 3. Pary derywacyjne końcówek męskich i żeńskich

Masculine	Feminine
ak	aczka
anin	anka
ant	antka
arz	arka
at	atka
ator	atorka
a	ka
ący	ąca
er	erka
erz	erka
ista	istka
nik	nica
ski	ska
ysta	ystka
yk	yczka
y	a
	ówka
	ka
	a

10.2.3.4. Siatka pojęciowa słowników referencyjnych

Dane te zawierają listy haseł wyekscerpowane z 54 czterech słowników referencyjnych różnych rodzajów. Hasła każdego słownika z hasłami rozpoczynającymi się na daną literę umieszczone są w osobnym pliku tekstowym (.txt), co pozwala na swobodne dodawanie siatek pojęciowych kolejnych słowników dla kolejnych liter.

Pliki te zostały opracowane ręcznie poprzez automatyczne wykstrahowanie lub też ręczne wypisanie nagłówków haseł zaczynających się na odpowiednią literę z poszczególnych słowników referencyjnych. Wśród tych znalazło się wiele uprzednio opracowanych przez piszącego te słowa, ale zostały one uzupełnione także o ważniejsze dzieła polskiej leksykografii współczesnej i dawnej.

Poza plikami z hasłami z poszczególnych słowników jest też obecny plik z listą słowników w formacie .tab, który oprócz skróconych nazw słowników zawiera ich pełne dane bibliograficzne, a także przynależność do jednej lub więcej z dziewięciu kategorii, które określają, czy jest to słownik:

1. definicyjny;
2. przekładowy;
3. poprawnościowo-ortograficzny;
4. wyrazów obcych;

5. gramatyczny;
6. historyczny;
7. etymologiczny;
8. gwarowy;
9. specjalistyczny.

Dane te są wykorzystywane przez program SWParser do wygenerowania **Aneksu K**, który zawiera zestawienie siatki pojęciowej *Słownika warszawskiego* ze słownikami referencyjnymi.

10.2.3.5. Szablony ERB

System ERB (ang. *Embedded Ruby*)⁶⁸ pozwala na stosowanie szablonów łączących kod w języku Ruby ze zwykłym tekstem lub kodem napisanym w innych językach, co pozwala na łatwe i dynamiczne łączenie kodu.

Na potrzeby niniejszej pracy system szablonów ERB został wykorzystany w celu dynamicznego wygenerowania odpowiednich fragmentów kodu w języku \LaTeX dla poszczególnych zestawów danych, co z kolei pozwala na wygenerowanie aneksów w formie plików PDF.

Pliki szablonów zapisane są w formacie `.tex`, który jest natywny dla skryptów \LaTeX , a w których tylko fragmenty zmienne (zależne od danych wejściowych) zawierają fragmenty kodu. Każdemu plikowi `.tex` odpowiada wyjściowy plik `.pdf`.

10.3. Analiza składniowa

W niniejszym podrozdziale zostanie opisany proces parsowania (czyli przetwarzania, analizy składniowej) dokonywany przez program SWParser. W dużym uproszczeniu parsowanie można opisać jako zamianę danych tekstowych na wejściu na odpowiadającą im strukturę obiektową na wyjściu (tu: plik XML).

Nie zostanie tu jednak opisany cały kod programu, a jedynie główne elementy jego reprezentacji algorytmicznej – czyli główne kroki, które muszą zostać wykonane do otrzymania pożądanego rezultatu, a ignorując problemy czysto techniczne związane z konkretną realizacją kodu.

10.3.1. Podział na litery

Program SWParser napisany jest w taki sposób, aby przetwarzać treść *Słownika warszawskiego* partiami, z których każda zawiera hasła rozpoczynające się od tej samej litery alfabetu. Powody takiej decyzji były dwa.

Pierwszym powodem jest fakt, że grupy haseł rozpoczynających się tą samą literą często różnią się od siebie pod względem niektórych detali strukturalnych, podczas gdy hasła w obrębie danej grupy są z reguły homogeniczne. Obejmuje to nie tylko niewielkie wariacje w strukturze

⁶⁸ Por. oficjalna strona dokumentacji: <https://docs.ruby-lang.org/en/2.3.0/ERB.html>.

czy niespójne stosowanie oznaczeń źródłowych, ale także strategię stosowania pewnych elementów – dla przykładu, antonimy wprowadzane skróceniem **Przec.** są obecne w hasłach na literę **D** (np. **Dyletant**) i literę **G** (np. **Gnostycyzm**), za to nieobecne w przypadku hasel na literę **F**, mimo że ich część posiada znaczenia przeciwne (np. **Fałszywy**). Co zaskakujące, wszystkie trzy grupy hasel znajdują się jednak w tym samym tomie. Może to wskazywać, że podział słownika na litery miał swoje odzwierciedlenie również w procesie kompilacji.

Drugim powodem jest duża objętość danych słownika i związana z tym konieczność ograniczenia materiału badawczego do niewielkiego jego wyciągu. Wybór hasel w sekwencji w ramach jednej litery alfabetu podyktowany był założeniem, że hasła te będą z jednej strony wystarczająco reprezentatywne dla całego słownika, a z drugiej wystarczająco homogeniczne jako względem siebie.

Podział taki jest jednak z wielu względów nienaturalny i problematyczny, co zostanie szerzej omówione w **podrozdziale 10.8.7**.

10.3.2. Wczytanie i korekcja danych

Pierwszym etapem działania programu SWParser jest wczytanie i przygotowanie danych wejściowych, opisanych w rozdziale poprzednim. Najważniejsze są tutaj dane podstawowe, czyli treść słownika, jednak jednocześnie wczytywane są również pozostałe typy danych (z wyjątkiem kart hasłowych).

Nim zostanie rozpoczęte parsowanie, dane muszą zostać poddane wstępnej obróbce, na którą składa się zastąpienie fragmentów tekstu z błędami (na podstawie pliku podmian), konwersję znaków pełniących funkcje strukturalne w pliku XML (znaczników w postaci nawiasów trójkątnych `< i >` na `{ i }`), oraz automatyczne dodanie znaków ucieczki (nieobecnych w pliku podmian punktorów `•`). Obróbka ta jest o tyle istotna, że jej brak może spowodować nieprawidłową interpretację struktury danych.

10.3.3. Deklaracje wyrażeń regularnych

Większość elementów artykułu hasłowego jest identyfikowanych i modyfikowanych przy pomocy opisanych w **podrozdziale 7.12** wyrażeń regularnych. W ich skład, oprócz literałów i dzikich kart, mogą wchodzić także inne uprzednio zdefiniowane wyrażenia regularne, które można zagnieździć z wykorzystaniem interpolacji zmiennych typu string, które w przypadku języka Ruby zapisywane są w postaci `#{zmienna}`.

Wykorzystanie tej metody pozwalającej na uzyskanie wielopoziomowej struktury można zilustrować na przykładzie pierwszego z dziesięciu różnych wyrażeń regularnych wykorzystywanych do rozpoznania elementów segmentu inicjalnego (**listing 5**), które samo nie wchodzi w skład innych wyrażeń regularnych (poziom 1), ale w skład którego wchodzi wyrażenia składowe (poziom 2), które same zbudowane są z mniejszych jednostek (poziom 3), itd.:

Listing 4. Wyrażenie regularne Rx_header1 (poziom 1).

```
Rx_header1 = /^(#{Rx_Lemma_comp})(?:,  
(#{Rx_gram})\])?#{Rx_pronun}(#{Rx_syn})?(?: (#{Rx_abbr_combo_ch}))?(?:  
(#{Rx_pos}))?$/
```

W skład powyższego wchodzi sześć innych wyrażeń regularnych lub ich fragmentów o różnym stopniu złożoności. Pierwsze z nich to wyrażenie przechwytyjące wyraz hasłowy wraz z elementami powiązаныmi (**listing 7**), drugie natomiast przechwytyuje segment z informacją gramatyczną (**listing 8**), trzecie – informację o wymowie (**listing 9**), czwarte – segment z synonimami (**listing 10**), piąte – kwalifikatory wyrazowe przyłączone do lemmy (**listing 11**), a szóste – informację o części mowy (**listing 12**).

Listing 5. Wyrażenie regularne Rx_lemma_comp (poziom 2).

```
Rx_lemma_comp = /(?: (?:dok\.|nied\. ) )?(?: (?:#{Rx_abbr_combo}|zdr\. ) )  
)?(?: \d\. (?: i \d\. ) )?#{Rx_sym}? \ (?: \-  
? \ (?:#{Ltr_pl_c}) \ )? (?: (?:#{Ltr_pl_msc} | [ \ ( \ ) ] ) * ? | \ . ) | \ p { L } + \-  
|#{Ltr_pl_c} \ . ) (?: się | ś \ . )? (?: pod \ d \ d ? )? \ ]? (?: #{Rx_Lemma_scope} )? /
```

Listing 6. Wyrażenie regularne Rx_gram (poziom 2).

```
Rx_gram = /(?: w lp \ . nieod \ . , lm \ . #{Rx_ending},  
#{Rx_ending} | (?: #{Rx_ending}, )? (?: a \ . w )? lm \ . #{Rx_ending} (?: ,  
#{Rx_ending} )? (?: , część \ . w lm \ . )? (?: rz \ . (?!  
od) )? | przysłówkowo | (?: #{Rx_ending}, )? (?: bl [mp] \ . (?: zb \ . )? | (?: część \ .  
| zwykle )? (?! [^\ . ] , ) w l [mp] \ . ) (?: \ ( (?: zwykle )? tylko w (?: \ d \ -ym  
przypadku \ d pp \ . ) \ ) )? |#{Rx_gen} | (?: przy [ms] | rz ) \ . (?! od  
) | w \ . | zb \ . | \ ( dok \ . i nied \ . \ ) | 2 pp \ . #{Rx_ending} | nieod \ .  
(?: przy [ms] \ . | rz \ . ) ) (?: #{Rx_gen} )? (?: \ ( część \ . w lm \ . \ ) )? \ ]? | \ ( tylko w  
\ d \ - [yi] m przypadku \ ) /
```

Listing 7. Wyrażenie regularne Rx_pronun (poziom 2).

```
Rx_pronun = /(?: \ ( (?:#{Ltr_pl_c}#{Ltr_pl_s}+ \ -#{Ltr_pl_s}+ ) \ ) )? /
```

Listing 8. Wyrażenie regularne Rx_syn (poziom 2).

```
Rx_syn = /(?: a \ . | , ? a \ . | , | i | ( ? < = [ ; ; + ! ] ) ) (?: (?: dok \ . | nied \ . | st \ .  
w \ . | zgr \ . | zgr \ . ) )? (?: (?: #{Rx_abbr_combo} | niek \ . | część \ . )  
)? (?: #{Rx_lemma_scope} )? [ + x ! \ [ ]? (?: #{Ltr_pl_c}#{Ltr_pl_msc} + (?: się |  
ś \ . ) )? (?: #{Ltr_pl_s} + ! )? |#{Ltr_pl_c} \ . (?: #{Ltr_pl_msc} + | ś \ . ) ) (?:  
(?: #{Rx_gen} | lm \ . ) (?: ( ? = , ) | ( ? = \ n ? ) ) )? \ ]? (?: , (?: (?: dok \ . | nied \ . | st \ . w \ . )  
)? (?: (?: #{Rx_abbr_combo} | niek \ . | część \ . ) ) )? (?: #{Rx_lemma_scope}  
)? [ + x ! \ [ ]? #{Ltr_pl_c}#{Ltr_pl_msc} + (?: się | ś \ . ) )? (?: #{Ltr_pl_s} + (?!  
) )? (?: (?: #{Rx_gen} | lm \ . ) ( ? = , ) )? \ ]? ) * ? /
```

Listing 9. Wyrażenie regularne Rx_abbr_combo_ch (poziom 2).

```
Rx_abbr_combo_ch = / \ ( (?: zwykle ) )? (?: #{Rx_abbr_combo} ) (?: i ?  
? (?: #{Rx_abbr_combo} ) )? \ )? | \ ( i (?: #{Rx_abbr_combo} )? \ ) /
```

Listing 10. Wyrażenie regularne `Rx_pos` (poziom 2).

```
Rx_pos = /(?:przy[ms]\.|rz\.|w\.|im\.)?! od )/
```

Wiele z powyższych składa z kolejnych zadeklarowanych elementów. Dla przykładu w skład `Rx_gram` wchodzi wyrażenia regularne takie jak `Rx_ending` (**listing 14**) czy `Rx_gen` (**listing 15**); dodatkowo w tym pierwszym zawiera się zdefiniowany fragment obejmujący małe litery polskiego alfabetu `Ltr_pl_s` (**listing 17**).

Listing 11. Wyrażenie regularne `Rx_ending` (poziom 3).

```
Rx_ending = /(?:[+x!]?#{Ltr_pl_s}+?(?: ś\.| się)?(?: a\.[+x!]?#{Ltr_pl_s}+?(?: ś\.| się)?\)?\)?(=,| |$)/
```

Listing 12. Wyrażenie regularne `Rx_gen` (poziom 3).

```
Rx_gen = /(?: (?:m\.|ż\.|ni\.) (?: i (?:m\.|ż\.|ni\.) ) )?
```

Listing 13. Element składowy wyrażenia regularnego `Ltr_pl_s` (poziom 4).

```
Ltr_pl_s = '[a-ząęśćźńółáé]'
```

Niektóre elementy są też definiowane dynamicznie, jak dzieje się to w przypadku wyrażen zawierających skrócenia, np. wchodzące w skład zarówno `Rx_abbr_combo_ch` (łącacego więcej niż jedno skrócenie w różnych kombinacjach), jak i `Rx_lemma_comp`, czy wielu innych większych struktur przechwytyjących wyrażenie `Rx_abbr_combo` (**listing 14**), które zawiera jednostki zaczerpnięte z opisanej wcześniej listy predefiniowanej, którym przypisana jest kategoria oznaczeń domenowych lub rejestrowych, a także skrócenie `niek..`

Listing 14. Wyrażenie regularne `Rx_abbr_combo`.

```
Rx_abbr_combo = Rx_abbr_dom + '|' + Rx_abbr_reg + '|niek.'
```

Wszystkie zagnieżdżenia i zmienne są interpretowane przez parser w momencie wczytania poszczególnych wyrażen – oznacza to, że wartość `Rx_abbr_combo` podczas wykonywania programu prezentuje się w sposób przedstawiony na **listingu 15**:

Listing 15. Wyrażenie regularne `Rx_abbr_combo` (po interpretacji).

```
elektrot\.|przyrod\.|zczion\.|szczot\.|cukier\.|fortep\.|gieol\.|rzeźb\.|przęd\.|fecht\.|piern\.|blach\.|patol\.|terat\.|piecz\.|wiośl\.|teatr\.|malar\.|chem\.|chir\.|cieś\.|ślus\.|cukr\.|szew\.|tech\.|dent\.|buch\.|dru c\.|tkac\.|ekon\.|biol\.|embr\.|farb\.|węgl\.|stol\.|siod\.|fizj\.|zool\.|bart\.|rzeź\.|garb\.|garn\.|rzem\.|bakt\.|pusz\.|astr\.|arch\.|gorz\.|pr aw\.|gram\.|powr\.|zoot\.|piek\.|hist\.|antr\.|intr\.|druk\.|pale\.|okul\.|myśl\.|karc\.|mier\.|mech\.|anat\.|leśn\.|kosz\.|kość\.|kotl\.|kuśn\.|k raw\.|kuch\.|kup\.|kow\.|lek\.|koń\.|lud\.|żyż\.|koź\.|mat\.|kol\.|men\.|met\.|kip\.|min\.|apt\.|mor\.|mul\.|muz\.|myd\.|kam\.|nar\.|noź\.|ogr\.|
```

```
jub\.|org\.|pal\.|jęz\.|inż\.|hut\.|her\.|han\.|piw\.|poł\.|pow\.|haf\.|g
ór\.|gis\.|gin\.|bed\.|gim\.|ret\.|ręk\.|rol\.|ryb\.|gie\.|fot\.|zeg\.|fi
z\.|fil\.|der\.|cyk\.|szk\.|szm\.|ceg\.|tap\.|bur\.|bud\.|tel\.|bot\.|zdń
\.|tok\.|wet\.|bil\.|bib\.|woj\.|den\.|zł\.|ps\.|mł\.|fl\.|bł\.|dziec\.|w
łaśc\.|przen\.|piesz\.|prow\.|lekc\.|posp\.|poet\.|późn\.|żart\.|pouf\.|l
ep\.|daw\.|rub\.|ucz\.|pog\.|gm\.|ir\.|niek.
```

W przypadku `Rx_abbr_combo_ch`, które łączy dwa ciągi skróceń ze sobą, wartość ta wyglądała natomiast w sposób następujący:

Listing 16. Wyrażenie regularne `Rx_abbr_combo_ch` (po interpretacji).

```
/\((?:zwykle
)?(?:elektrot\.|przyrod\.|czcion\.|szczot\.|cukier\.|fortep\.|gieol\.|rze
źb\.|przęd\.|fecht\.|piern\.|blach\.|patol\.|terat\.|piecz\.|wiośl\.|teat
r\.|malar\.|chem\.|chir\.|cieś\.|ślus\.|cukr\.|szew\.|tech\.|dent\.|buch\
.|druc\.|tkac\.|ekon\.|biol\.|embr\.|farb\.|węgl\.|stol\.|siod\.|fizj\.|z
ool\.|bart\.|rzeź\.|garb\.|garn\.|rzem\.|bakt\.|pusz\.|astr\.|arch\.|gorz
\.|praw\.|gram\.|powr\.|zoot\.|piek\.|hist\.|antr\.|intr\.|druk\.|pale\.|o
kul\.|myśl\.|karc\.|mier\.|mech\.|anat\.|leśn\.|kosz\.|kość\.|kotl\.|kuś
n\.|kraw\.|kuch\.|kup\.|kow\.|lek\.|koń\.|lud\.|żyż\.|koł\.|mat\.|kol\.|m
en\.|met\.|kip\.|min\.|apt\.|mor\.|mul\.|muz\.|myd\.|kam\.|nar\.|noż\.|ogr
r\.|jub\.|org\.|pal\.|jęz\.|inż\.|hut\.|her\.|han\.|piw\.|poł\.|pow\.|haf
\.|gór\.|gis\.|gin\.|bed\.|gim\.|ret\.|ręk\.|rol\.|ryb\.|gie\.|fot\.|zeg\
.|fiz\.|fil\.|der\.|cyk\.|szk\.|szm\.|ceg\.|tap\.|bur\.|bud\.|tel\.|bot\
.|zdń\.|tok\.|wet\.|bil\.|bib\.|woj\.|den\.|zł\.|ps\.|mł\.|fl\.|bł\.|dziec
\.|właśc\.|przen\.|piesz\.|prow\.|lekc\.|posp\.|poet\.|późn\.|żart\.|pouf
\.|lep\.|daw\.|rub\.|ucz\.|pog\.|gm\.|ir\.|niek.)(?: i?
)?(?:elektrot\.|przyrod\.|czcion\.|szczot\.|cukier\.|fortep\.|gieol\.|rzeź
b\.|przęd\.|fecht\.|piern\.|blach\.|patol\.|terat\.|piecz\.|wiośl\.|teatr
\.|malar\.|chem\.|chir\.|cieś\.|ślus\.|cukr\.|szew\.|tech\.|dent\.|buch\
.|druc\.|tkac\.|ekon\.|biol\.|embr\.|farb\.|węgl\.|stol\.|siod\.|fizj\.|zo
ol\.|bart\.|rzeź\.|garb\.|garn\.|rzem\.|bakt\.|pusz\.|astr\.|arch\.|gorz\
.|praw\.|gram\.|powr\.|zoot\.|piek\.|hist\.|antr\.|intr\.|druk\.|pale\.|o
kul\.|myśl\.|karc\.|mier\.|mech\.|anat\.|leśn\.|kosz\.|kość\.|kotl\.|kuś
n\.|kraw\.|kuch\.|kup\.|kow\.|lek\.|koń\.|lud\.|żyż\.|koł\.|mat\.|kol\.|me
n\.|met\.|kip\.|min\.|apt\.|mor\.|mul\.|muz\.|myd\.|kam\.|nar\.|noż\.|ogr
r\.|jub\.|org\.|pal\.|jęz\.|inż\.|hut\.|her\.|han\.|piw\.|poł\.|pow\.|haf
\.|gór\.|gis\.|gin\.|bed\.|gim\.|ret\.|ręk\.|rol\.|ryb\.|gie\.|fot\.|zeg\
.|fiz\.|fil\.|der\.|cyk\.|szk\.|szm\.|ceg\.|tap\.|bur\.|bud\.|tel\.|bot\
.|zdń\.|tok\.|wet\.|bil\.|bib\.|woj\.|den\.|zł\.|ps\.|mł\.|fl\.|bł\.|dziec
\.|właśc\.|przen\.|piesz\.|prow\.|lekc\.|posp\.|poet\.|późn\.|żart\.|pouf
\.|lep\.|daw\.|rub\.|ucz\.|pog\.|gm\.|ir\.|niek.))?)?(i
(?:elektrot\.|przyrod\.|czcion\.|szczot\.|cukier\.|fortep\.|gieol\.|rzeźb
\.|przęd\.|fecht\.|piern\.|blach\.|patol\.|terat\.|piecz\.|wiośl\.|teatr\
.|malar\.|chem\.|chir\.|cieś\.|ślus\.|cukr\.|szew\.|tech\.|dent\.|buch\
.|druc\.|tkac\.|ekon\.|biol\.|embr\.|farb\.|węgl\.|stol\.|siod\.|fizj\.|zoo
l\.|bart\.|rzeź\.|garb\.|garn\.|rzem\.|bakt\.|pusz\.|astr\.|arch\.|gorz\.
```

```
|praw\.|gram\.|powr\.|zoot\.|piek\.|hist\.|antr\.|intr\.|druk\.|pale\.|ok
ul\.|myśl\.|karc\.|mier\.|mech\.|anat\.|leśn\.|kosz\.|kość\.|kotl\.|kuśn\
.|kraw\.|kuch\.|kup\.|kow\.|lek\.|koń\.|lud\.|żyż\.|koż\.|mat\.|kol\.|men
\.|met\.|kip\.|min\.|apt\.|mor\.|mul\.|muz\.|myd\.|kam\.|nar\.|noż\.|ogr\
.|jub\.|org\.|pal\.|jęz\.|inż\.|hut\.|her\.|han\.|piw\.|poż\.|pow\.|haf\
.|gór\.|gis\.|gin\.|bed\.|gim\.|ret\.|ręk\.|rol\.|ryb\.|gie\.|fot\.|zeg\.|
fiz\.|fil\.|der\.|cyk\.|szk\.|szm\.|ceg\.|tap\.|bur\.|bud\.|tel\.|bot\.|z
dń\.|tok\.|wet\.|bil\.|bib\.|woj\.|den\.|zł\.|ps\.|mż\.|fl\.|bł\.|dziec\
.|właśc\.|przen\.|piesz\.|prow\.|lekc\.|posp\.|poet\.|późn\.|żart\.|pouf\
|lep\.|daw\.|rub\.|ucz\.|pog\.|gm\.|ir\.|niek.)?)\|/
```

Możliwość dynamicznego budowania wyrażeń regularnych przynosi kilka istotnych korzyści:

1. czas potrzebny na ich budowanie wyrażeń staje się krótszy;
2. modyfikacje we wczytywanych zmiennych (np. z pliku skrótów) samoistnie się propagują, niwelując tym samym potrzebę ręcznej aktualizacji kodu;
3. ryzyko błędów spada, ponieważ mniejszy fragment kodu podlega ingerencji człowieka.

10.3.4. Wykorzystanie skrótów

Program SWParser wykorzystuje skrócenia ze spisów w *Słowniku warszawskim* jako składnik licznych wyrażeń regularnych: część z nich pozwala na prawidłową segmentację hasła, inne zaś przechwytyują jego elementy i przesyłają je do dalszej obróbki. Skrócenia te są umieszczone w predefiniowanym pliku, wczytywanym przez program podczas uruchamiania.

Oprócz procesu standaryzacji zapisu opisanego w **rozdziale 5.5** była konieczna jeszcze jedna modyfikacja. Stało się tak, ponieważ w *Spisie* pojawiają się takie skrócenia wieloczłonowe – takie, które zawierające spacje – w przypadku których pierwszy człon może być mylnie zakwalifikowany jako odrębne skrócenie, jak działo się to choćby w przypadku grupy poniższych oznaczeń źródłowych:

1. Jak. = Jan Walenty Jakubowski (†1582)
2. Jak. J. = Józef Jakubowski (w. XIX)
3. Jak. W. = Wojciech Jakubowski (†1784)

Jako że program sekwencyjnie próbuje dopasować każde skrócenie, natrafiając na fragment tekstu z ciągiem znaków Jak. J. lub Jak. W. rozpoznaje źródło Jak. i drugi człon traktuje jako odrębną jednostkę. Dlatego konieczne była zmiana kolejności niektórych elementów tak, żeby najpierw podejmowana była próba wyszukania skrótów złożonych przed skrótami prostymi – oznacza to, że w przypadku elementów cytowanych ich kolejność zmieniała się na następującą:

2. Jak. J. = Józef Jakubowski (w. XIX)
3. Jak. W. = Wojciech Jakubowski (†1784)
1. Jak. = Jan Walenty Jakubowski (†1582)

Poprawne rozpoznanie skrótów jest kluczowe, bowiem można je znaleźć we wszystkich głównych segmentach artykułu hasłowego, i to w ramach wielu różnych segmentów, ponieważ

mogą opisywać nie tylko wyraz hasłowy, ale także poszczególne jego znaczenia, czy inne jednostki przytoczone w ramach artykułu hasłowego.

Obrazuje to choćby fakt, że samo dodanie do listy predefiniowanej kwalifikatora `kosz.`, który w *Spisie skrótów* znajduje się dopiero od tomu drugiego, pozwoliło programowi zacząć poprawnie parsować hasło `Flektowany`.

Należy też wspomnieć, iż od strony programistycznej oznaczenia źródłowe chronologiczne również zostały potraktowane na równi ze skrótami, co umożliwiło poprawną analizę składniową wszystkich tych przykładów użycia, dla których została podana informacja o roku lub wieku pochodzenia. Przechwytywanie tej informacji osiągnięto przy pomocy prostego wyrażenia regularnego, przedstawionego w **listingu 17**.

Listing 17. Wyrażenie regularne do przechwytywania informacji temporalnej.

```
(?:\((?[Ww]\. [•][XVI]+\)?|\((?:z )?r\.[•]\d{3,4}\))
```

10.3.5. Supersegmentacja artykułu hasłowego

Pierwszym etapem związanym z przetworzeniem poszczególnych artykułów hasłowych jest ich podział na segmenty najwyższego poziomu, co powiązane jest to z logiką głęboką programu: ze względu na mnogość struktur i czysto deterministyczne przetwarzanie konieczne jest operowanie na ograniczonych jednostkach, co pozwala wykluczyć wiele typów treści podczas parsowania.

Zgodnie z opisem w **podozdiale 4.3**, każdy artykuł hasłowy składa się z trzech segmentów najwyższego poziomu: inicjalnego, centralnego i terminalnego, które z kolei składają się z dalszych podsegmentów. Do parsowania każdego z supersegmentów poświęcony jest odrębny moduł, ponieważ różnią się one od siebie typograficznie, strukturalnie i funkcjonalnie. Aby jednak każdy z segmentów poddać odrębnej analizie składniowej, konieczne jest ich zidentyfikowanie, do czego wykorzystywane są wyrażenia regularne.

10.3.5.1. Wyodrębnianie haseł zagnieżdżonych

Hasła zagnieżdżone wykrywane są przy pomocy wyspecjalizowanych reguł, jako że przyjąć mogą jedną z ograniczonej liczby postaci, choć zasadniczo ich struktura jest podobna do haseł głównych. W materiale badawczym poświadczono pięć kategorii takich lemm:

1. Lemma główna w postaci zwrotnej, zapisanej w postaci `F. ś.` (a także wraz z kwalifikatorami: `†F. ś.`, `×F. ś.` i `[F. ś.]`, również w postaci `†F. ś. i [F. ś.]`, `†Fremarczyć ś.`, `[Frymaczyć ś.]`, `Fyrmaczyć ś.`);
2. Podhasła o postaci wyrażen przysłówkowych z przyimkiem *po* i zakończonych na *-u* (np. `Po faryzeuszowsku`, `×Po filozofsku`, `Po fuszersku`);
3. Podhasła imiesłowowe (pojedyncze poświadczenie: `†Figurowany im.`);
4. Podhasła pochodzenia przymiotnikowego, używane także rzeczownikowo (dwa poświadczenia: `×Folwarczny`, `ego`, `lm. i rz.`, `×Folwarkowy` oraz `Frontowy`, `ego`, `lm. i rz.`);

5. Podhasła o postaci przysłówków o pochodzeniu rzeczownikowym (dwa poświadczenia: [Fuksem] przys. oraz †Fużą przys.).

Wyizolowane w ten sposób hasła zagnieżdżone (podhasła) przetwarzane są następnie w sposób identyczny do haseł głównych.

10.3.5.2. Wyodrębnienie segmentu terminalnego

W pierwszej kolejności program podejmuje próbę znalezienia segmentu terminalnego, ponieważ ten jest łatwy w identyfikacji i wspólny dla haseł głównych i zagnieżdżonych: może składać się z czterech elementów, które zawsze wskazane są eksplicitnie, jednoznacznie i w ustalonej kolejności – segmenty derywacyjne wprowadzone są poprzez skrócenia, a segment etymologiczny ograniczony jest nawiasem trójkątnym.

10.3.5.3. Wyodrębnienie haseł zagnieżdżonych

Po przechwyceniu segmentu terminalnego program przystępuje do detekcji haseł zagnieżdżonych, które znajdują się w segmencie centralnym hasła głównego, najczęściej na jego końcu. Program stara się przechwycić jeden z pięciu rodzajów haseł zagnieżdżonych, z których każdy powiązany jest z konkretną częścią mowy i kategorią morfologiczną. Ich identyfikacja na tym etapie jest kluczowa, ponieważ kolejne kroki dotyczą haseł głównym i zagnieżdżonych w sposób symetryczny.

10.3.5.4. Wyodrębnienie segmentu inicjalnego i centralnego

Rozdzielenie tych segmentu inicjalnego od segmentu centralnego dokonywane jest z wykorzystaniem jednego z dwunastu wyrażeń regularnych. Do wykrycia granicy obu części wykorzystywane są takie elementy strukturalne jak informacja, ciągi synonimiczne, numeracja znaczeń, fonetyka, a także elementy typograficzne: interpunkcja i wielkość liter.

10.3.6. Parsowanie segmentu inicjalnego

Budowa segmentu inicjalnego skonsolidowana jest wokół pojedynczego elementu, wyrazu hasłowego, który stanowi jego rdzeń; wszystkie pozostałe elementy są opcjonalne.

Sama lemma może być poprzedzona przez dwa elementy: dezambiguator i kwalifikator symboliczny; po niej natomiast mogą pojawić się (w kolejności) informacja o gramatyce, informacja o wymowie, ciąg synonimiczny, kwalifikator wyrazowy i informacja o zakresie.

10.3.6.1. Dezambiguator

Dezambiguatory interpretowane są przy pomocy prostej reguły, która przechwytyje cyfrę rzymską wraz kropką, jeśli są to pierwsze znaki artykułu hasłowego, bez oddzielającej je spacji.

10.3.6.2. Kwalifikator symboliczny

Kwalifikatory symboliczne łatwo przechwycić, ponieważ są pojedynczymi znakami poprzedzającymi wyraz hasłowy (†, ×, ! oraz *), z jednym wyjątkiem: nawiasu kwadratowego [], który obejmuje nie samą lemmę, ale zbiorczo i wszystkie inne elementy, których dotyczy – takie jak informacja gramatyczna czy ciąg synonimiczny, np.:

- **[Fabrycarz, a, lm. e, Filanc, Chabrycarz, Habrycarz]** *urzędnik podatkowy. (...)*

W przypadku takich haseł, jak i wszelkich innych elementów zawartych w nawiasie kwadratowych, uruchamiany jest moduł, który ekstrahuje wszystkich elementy i umieszcza kwalifikator [] przed nimi. Przytoczony uprzednio artykuł hasłowy uzyskuje zatem na tym etapie analizy składniowej następującą formę:

- **[]Fabrycarz, a, lm. e, []Filanc, []Chabrycarz, []Habrycarz** *urzędnik podatkowy. (...)*

Dzięki temu możliwe jest symetryczne parsowanie wszystkich kwalifikatorów symbolicznych przy pomocy wspólnych reguł.

10.3.6.3. Lemma

Wykrycie lemmy jest stosunkowo prostym zadaniem: każdy nagłówek artykułu hasłowego rozpoczyna się wielką literą, po której występują małe litery alfabetu polskiego (także historyczne), np.:

- **Fertanie, a, blm., czynność cz. Fertać.**
- **[Ferezyjá, i, lm. e] p. Ferezja.**

Lemmy mogą zawierać także spacje, zwłaszcza w przypadku wyrazów zwrotnych, po których występuje partykuła „się” i przecinek:

- **Fertanie się, a ś., blm., czynność cz. Fertać ś.:** *Trzeba nam na tym baliku młodzieży dla fertania ś. L.*
- **Fiokować się, uje ś., ował ś. ubierać ś. w fioki; stroić ś., elegantować ś. śmiesznie, przesadnie.**

Bardzo rzadko spacje występują też w jednostkach niezwrotnych – wówczas drugi człon jest zapisywany małą literą, a kolejne jego formy również zawierają spację:

- **[Facjender jegier, facjendera jegra, lm. facjendery jegry]** *Lokaj a. strzelec dworski, tułający ś. bez służby. (...)*

Wystąpić mogą także niektóre znaki interpunkcyjne – wykrzykniki (!) oznaczające wykrzyknienia, apostrofy (') w funkcji fonetycznej (zmiękczającej), a także dywizy (-):

- **Fe! 1. a. Pfe!, Fi!, Fuj!, Pfuj!, xFa!, tPfa!, tPfy!, tPi!, tP'!** *w. obrzydzenia, wstrętu, nagany = tju! be! bwe! bla!: (...)*
- **Fik-miki** *p. Fik-mik.*
- **Fligiel-adjutant, a, lm. ci oficer na służbie adjutania przy panującym.** *<Nm. Flügeladjutant>*

Program wyszukuje ciągi opisanych wyżej znaków na początku artykułu hasłowego (lub po dezambiguatorze lub kwalifikatorze symbolicznym) aż do kolejnego elementu strukturalnego.

10.3.6.4. Informacja gramatyczna

Segment gramatyczny znajduje się bezpośrednio po lemnie i składa się z czterech typów elementów: (1) końcówek fleksyjnych, (2) pełnych form fleksyjnych, (3) oznaczeń gramatycznych

dotyczących fleksji, (4) klas wyrazów. Elementy te mogą być oddzielone spacjami, przecinkami lub spójnikami w ustalonych konfiguracjach np.:

- **Fakturowanie**, a, blm., czynność cz. **Fakturować**.
- **×Falcgrabia**, i a. **ego**, lm. **owie** p. **Falcgraf**.
- **[Fleja**, i, lm. **e]** m. i ż. p. **Flejtuch**.
- **Fraj** nieod. przym. **1. a. ×Frej**, **×Frejowy**, **[Frajny]** *swobodny, wolny, bezpłatny, gratisowy: (...)*

10.3.6.5. Walencja

Informacja o walencji realizowana jest przy pomocy dwóch zamkniętych klas elementów: (1) zaimków oraz (2) przymków, które oddzielone są przecinkiem albo spójnikiem. Algorytm przechwytyjący zakłada istnienie sześciu różnych pól, z których pola **1.**, **3.**, **4.** i **6.** mogą zawierać zaimek, a pola **2.** i **5.** przimek. Jedynym polem obowiązkowym jest pole **3.** – wszystkie pozostałe są fakultatywne. Struktura ta wraz z pięcioma typami realizacji i ich przykładami jest zilustrowana przy pomocy **tabeli 6.**

Tabela 6. Struktura informacji o walencji

#	Pola w strukturze						Przykład		
	1.	2.	3.	4.	5.	6.	Hasło	Miejsce	Realizacja
1.	–	–	PRN	–	–	–	2. Futrować	nagłówek	kogo
2.	–	PRP	PRN	–	–	–	Fundować	znaczenie 3.	w co
3.	–	–	PRN	a. PRN	–	–	†Fortunić	nagłówek	komu a. kogo, co
4.	–	PRP	PRN	a. PRN	–	–	Fuknąć	znaczenie 1.	na kogo a. ×kogo
5.	PRN	PRP	PRN	a. PRN	PRP	PRN	Frymarczyć	znaczenie 1.	co za co a. czym na co

10.3.6.6. Ciąg synonimiczny

Ciągi synonimiczne wychwytywane są na podstawie wielkich liter, od których się zaczynają, oraz i lokalizacji (po informacji gramatycznej). Pojedyncze leksemy w ciągu mogą mieć własne kwalifikatory oraz informację gramatyczną, co również jest przechwytywane:

- **[Facambuł**, a, lm. **y**, **Pacambuł]** *człowiek opasty a głupi. (...)*
- **†Facelet**, u, lm. **y**, **†Facelit**, **†Facilet**, **†Facolet** *chustka, szczeg. do nosa: (...)*
- **Filharmoniczny**, **Filarmoniczny**, **niek. Filharmonicy** *muzyczny, miłujący, uprawiający muzykę: (...)*
- **[Forknąć**, **nie**, **nął**, **Furknąć**, **nied. Forkać**, **Furkać]** *(o zwierzętach) parsknąć, czmychnąć: (...)*

Wykrycie jednostek zawierających spacje jest uproszczone dzięki właściwego odseparowania segmentu inicjalnego od centralnego:

- **Ficlaus**, a, lm. **y**, **Ficlauz**, **[Bekas kurka**, **Bekas mały]** *zool. (limnocryptes gallinula) ptak podkasaty z rodziny bekasów. (...)*

- **Flaszowiec**, wca, lm. wce, **Flaszodrzew**, **Flaszka**, **Flaszkwiec**, **Drzewo flaszowe**, **Siężyśnia** bot. (anona) roś. z rodziny *flaszowcowatych*.

Poszczególne leksemy są wyodrębniane z ciągów synonimicznych przy pomocy odgraniczających je przecinków.

10.3.6.7. Wymowa

Kolejnym elementem segmentu jest wymowa. Ta podana jest zawsze w nawiasie, podobnie jak informacja o zakresie, łatwo ją zidentyfikować, ponieważ jest to powtórzona lemma z nagłówka z dodatkiem dywizu – ten rozdziela dwuznaki takie jak *ch* czy *rz* i oddziela *i* zmiękczone od kolejnej samogłoski wskazując, że odseparowane powinny być wymawiane osobno, a nie zgodnie ze zwykłymi regułami ortograficznymi, np.:

- **Felchubel**, bla, lm. ble (**Felc-hubel**) *hebel do robienia felców*. <Nm. Falzhobel>
- **Fitobiologja**, i, blm. (**Fitobi-ologja**) bot. *biologja roślin, nauka o życiu roślin*. <Now. z Gr. fytón = roślina + Biologja>
- **Forzec**, u, lm. e (**For-zec**) intr. *papier do podklejania wewnętrznej strony okładki*. <Nm. Vorsatz>

10.3.6.8. Kwalifikator wyrazowy

Przechwytywanie kwalifikatorów wyrazowych opiera się o prosty algorytm, który identyfikuje ich wystąpienia przy pomocy listy, które jest dynamicznie generowana na podstawie wybranych klas oznaczeń po standaryzacji opisanej w **rozdziale 5**. Algorytm wychwytuje także ciągi kwalifikatorów oddzielone przecinkiem lub spójnikiem (jeśli należą do tej samej klasy) lub spacją (jeśli należą do różnych klas).

10.3.6.9. Informacja o literze

Segment ten nie posiada realizacji w przypadku materiału badawczego, jednak występuje w sekcjach poświęconych hasłom rozpoczynającym się literą **I**, **X** oraz **Y**. Parsowanie informacji o specyfice ortograficznej haseł rozpoczynających się na te litery jest możliwe dzięki znakom ucieczki, które wskazują na rodzaj informacji niehasłowej.

10.3.7. Parsowanie segmentu centralnego

Segment centralny jest obligatoryjny, natomiast jego najprostszą realizację stanowi pojedyncza definicja lub pojedynczy odsyłacz. Często jego konstrukcja jest jednak znacznie bardziej złożona.

Problem z parsowaniem struktur nadrzędnych w tej części hasła polega w pierwszej kolejności na znalezieniu granic wielu elementów równorzędnych, czyli opisów poszczególnych znaczeń. W drugiej kolejności należy właściwie zinterpretować ich strukturę wewnętrzną – ta nie tylko jest wielopoziomowa, ale może być także rekurencyjna; oznacza to, że wewnątrz segmentu o określonej strukturze może być zawarty subsegment o strukturze analogicznej.

10.3.7.1. Definicja

Definicja wyodrębniana jest nie wprost, tzn. poprzez rozpoznanie najpierw wszystkiego tego, co nią nie jest, ze względu na to, że jej treść może być bardzo różna. Część skrótów jest automatycznie zabezpieczana przez niepożądaną segmentacji przy pomocy znaków ucieczki.

10.3.7.2. Numeracja

Rozpoznawanie numeracji jest konieczne do wyodrębnienia poszczególnych znaczeń. Jest to zasadniczo bardzo prosta implementacja reguł opisanych w **podrozdziale 4.3.4.2**, wyszukująca liczb lub liter w funkcji porządkowej wraz z odpowiednim znakiem interpunkcyjnym wskazującym na taki ich charakter. Kontekst jest dodatkowo wykorzystywany do odróżnienia przypadków bardziej złożonych, m.in. do odróżnienia I. w funkcji numeracji od I. jako integralnego elementu treści słownika.

10.3.7.3. Kwalifikator symboliczny

Kwalifikatory symboliczne występują bezpośrednio po numeracji i oddzielone są spacjami. Szczególnie należy potraktować przypadek kwalifikatora [], wewnątrz którego należy ponownie pierwszą literę lemmy, np.:

- **Fasować, uje, ował (...)** 5. † i [F.] *bić, ćwiczyć, prac*: Dobrze go fasuje po gębie. (...)

10.3.7.4. Kwalifikator wyrazowy

Do detekcji kwalifikatorów wyrazowych wykorzystywany jest spis oznaczeń. Kwalifikatory mogą występować w ciągach – wówczas oddzielone są przecinkami lub spójnikiem „i”; jeśli należą do różnych klas, mogą być także oddzielone spacją. Aby wykryć kwalifikatory w użyciu takim, jak zostało to opisane w **podrozdziale 10.3.7.3**, program tworzy dynamiczne ich wersje w zapisie wielką literą.

10.3.7.5. Przykład użycia

Przykłady użycia znajdują się bezpośrednio po (1) definicji, (2) odsyłaczu, a niekiedy także po (3) kwalifikatorze (także w przypadku znaczeń niewyróżnionych numeracją) czy (4) po segmentie inicjalnym i zawsze wprowadzane są przez dwukropek, co ułatwia ich detekcję – por.:

1. przykład użycia po definicji:
 - **Fanfaron, a, lm. y, [Fenfaron] (...)** 2. *elegant, modniś, dandys, frant*: **F. pudrowany a bosy. Prz.** (...)
 - **†Fortelny, †Fortylny, st. w. †Fortelniejszy (...)** 2. *podstępny, chytry, peten fortelów*: **Człowiek chytry i F. Skar. Nie ten wygra, kto ma sprawiedliwą, ale ten, kto ma więcej pieniędzy świadkom na zapłatę i fortelniejsze świadki. Górn.**
2. przykład użycia po odsyłaczu:
 - **Feniczek, czka, lm. czki p. Fenik**: **Feniczka nie znajdują. Ani feniczka nie wart.**
 - **Filozoficznie przys. od Filozoficzny**: **Zapatruje ś. na to wszystko F. F. biorąc, jest to fałsz.**

3. przykład użycia po kwalifikatorze:

- **Femiczny:** (...) *Żart.: W usposobieniu umysłu femicznym, kiedy głos przeczuć bardzo głośno gada trzykrotnym głosem: Biada! biada! biada! Słow. (= podobnym do tego, jakie mieli skazani na śmierć przez sąd femiczny). <Nm. Fe(h)me dosł. = kara, męki; Femgericht = sąd tajny>*
- **Flektowany** kosz.: *Koszyk F. = ażurowy (np. do papierów). (...)*
- **Frymark, u, lm. i 1. † zamiana, wymiana, handel zamienny; szacherka:** *Na frymarku jeden traci. Prz. Pan Bóg wynalazł jarmarki, a djabeł frymarki. Prz. Przen.: Dni swoje w żalości, nocy tracę łzami, a gdyby ś. F. ten stał między nami, że od tego czasu miałbym żyć w radości, snaby jej mniej było. Grabow. (= zamiana, układ). (...)*

4. przykład użycia bezpośrednio po segmencie inicjalnym:

- **Femiczny:** *Sąd F. = średniowieczny trybunał tajemny, karzący śmiercią winowajców bogatych i możnych: Sąd tajemny nazywał ś. jeszcze trybunałem femicznym. Mick. (...)*

Obsługa przykładów użycia jest dodatkowo skomplikowana przez fakt, że przykład użycia może mieć swoją definicję, która z kolei może mieć własny przykład użycia

- **Fortuna, y, lm. y, [Fortona, Fertuna, Fartuna] 1. Los, dola, przeznaczenie:** (...) *†Lewa F. = nieszczęście: Hetman tam pracował lewą bardzo fortuną. Tw. (...)*

Detekcja takich przypadków struktury rekurencyjnej w programie SWParser odróżnia też sytuacje, w których dwukropek jest częścią samego przykładu – por.:

- **Fabryczność, i, blm. ogół przedmiotów i zjawisk, dotyczących fabryki, życie fabryczne, świat fabryczny:** *Przybył mu do studjowania żywioł jeden więcej: F. Jeż.*
- **1. Fora! 1. w. żądający powtórzenia jakiegokolwiek ustępu na scenie a. estradzie = bis, drugi raz, powtórnie, nanowo:** (...) *I gdy aktor na scenie zabijał aktora, urznąłem huczne brawo i krzyknąłem: F.! Dmusz. (...)*
- **Fuknąć, nie, nął, nied. Fukać 1. a. F. ś., nied. Fukać ś. na kogo a. *kogo = krzyknąć, ofuknąć ś., obwisnąć ś., obruszyć ś., wybuchnąć gniewem, oburzyć ś., rzucić ś. z gniewu, zgromić, złażać:** *Książę fuknął ś.: A co mi u licha! Pol.*

10.3.7.6. Źródło

Dzięki ustandaryzowaniu i ujednoznacznieniu zapisu dokonanego w **rozdziale 5.5**, identyfikacja źródeł dokonywana jest poprzez sprawdzenie ich obecności na liście – podobnie jak w przypadku kwalifikatorów wyrazowych – i otagowaniu ich znakami pomocniczymi (do danego przykładu lub leksemu może podane być więcej niż jedno źródło). Problematiczne są jedynie kategorie dynamiczne, czyli zagadki i nienazwane źródła określone datą (rokiem lub wiekiem). Przechwytywanie takich przypadków dokonywane jest poprzez wyrażenia regularne, które obejmują konkretne wzory opisów i dat (liczb wraz ze skrótem).

10.3.7.7. Ciąg synonimiczny

Ciągi synonimiczne w segmencie tym są przetwarzane podobnie do tego, jak ma to miejsce w segmencie inicjalnym. Różnicą jest problem z wykryciem granicy ciągu w przypadku obecności jednostek wielowyrazowych – zadanie to jest proste w przypadku obecności kwalifikatorów lub numeracji znaczeń, np.:

- Fluoryt, u, lm. y, Fluspat, †Szpat fluoryczny min. fluorek wapnia, tworzący kryształy równoosiowe (...)
- Fiu! a. Fiu fiu! (...) 2. a. Fiu bździu przysłówkowo pusto, pstro: (...)

Niekiedy jednak algorytmy są w stanie wykryć takie przypadki dynamicznie, posiłkując się podobieństwem drugiego członu do lemmy, np.:

- Faktura, y, lm. y 1. (...) b) a. Faktur-księga a. lep. Księga fakturowa księga kupiecka ze spisem należności za towary. (...)

Dzięki takiemu rozwiązaniu program nie przechwytuje błędnie przypadków takich jak tu:

- Frędzlowanie, a (...) 2. lm. a, a. Fręzlowanie, ×Frاندzlowanie frędzla w bramowaniu.

W niektórych przypadkach jednak konieczne jest użycie znaków pomocniczych i pliku podmian, aby odpowiednio wykryć granicę ciągu synonimicznego.

10.3.7.8. Informacja gramatyczna

Informacja gramatyczna w segmencie centralnym jest zbliżona do tej w segmencie poprzednim, jednak przeważnie ma charakter częściowy (uzupełniający względem informacji wspólnej dla całego hasła), czasem podając także pełne formy zapisywane wielką literą:

- Fum, u, lm. y (...) 4. i w lm. Fumy, ów, a. Fuma pretensjonalne uroszczenie, próżność, zarozumiałość, nadętość, grymasy, duma, pycha: (...)

10.3.7.9. Walencja

Walencja w segmencie centralnym jest parsowana analogicznie do tego, jak ma to miejsce w segmencie inicjalnym, jednak w tym przypadku analizator składniowy oczekuje, informacja ta będzie oddzielona od definicji znakiem równości lub będzie umieszczona w nawiasie.

10.3.7.10. Pojęcia zagnieżdżone

Pojęcia zagnieżdżone, które obecne są w różnych podsegmentach artykułu hasłowego, identyfikowane są wyjątkowo na podstawie formatowania tekstu (rozstrzelenia), które zaznaczona jest w danych wejściowych przy pomocy odpowiednich znaków pomocniczych, co jest opisane w **podrozdziale 10.3.9**. Ich dalsze przetwarzanie jest analogiczne do tego, jak ma to miejsce w przypadku lemm. Poza tym jednostki takie niekiedy mają formę zagregowaną – przetwarzanie tego rodzaju sytuacji opisane jest w **podrozdziale 10.3.10**.

10.3.7.11. Podział typologiczny

Jednostki tego rodzaju identyfikowane są przy pomocy przygotowanej ręcznie listy konkretnych elementów, po których jest dwukropek. Lista ta zawiera jednostki:

- „gatunek”
- „gatunki”;
- „odmiany”;
- „odmiana”;
- „w wyrażeniach”.

10.3.7.12. Zakres

Informacja o zakresie identyfikowana jest na podstawie lokalizacji (po konkretnych typach jednostek), nawiasie, w którym jest umieszczona, oraz tego, że jej treści rozpoczyna się od jednego z zamkniętej listy przyimków.

10.3.7.13. Komentarz i glosa

Komentarze i glosy są identyfikowane analogicznie do zakresu, jednak zlokalizowane są w innym kontekście strukturalnym (przed definicją wprowadzoną przez znak równości).

10.3.7.14. objaśnienie

Objaśnienia, podobnie jak dwa poprzednie rodzaje subsegmentów, znajdują się w nawiasie, jednak ich wykrycie odbywa się dzięki znakowi równości, który jest pierwszym znakiem wewnątrz tego nawiasu.

10.3.7.15. Odsyłacz

Odsyłacze wszystkich rodzajów rozpoznawane są przy pomocy przygotowanej ręcznie listy skrótów, po których znajduje się leksem lub ich ciąg zakończony kropką lub średnikiem (kiedy podana jest definicja synonimiczna do hasła w danym artykule).

10.3.8. Parsowanie segmentu terminalnego

Segment terminalny jest jedyną częścią hasła najwyższego poziomu, która jest opcjonalna. Detekcja jej komponentów jest bardzo prosta dzięki regularnej i jednoznacznej strukturze.

10.3.8.1. Segmenty derywacyjne

Na segmenty derywacyjne składają się: (1) zdrobnienia, (2) zgrubienia, i (3) odsyłacze referencyjne. Wszystkie trzy rozpoczynają się zawsze tymi samymi skrótami – kolejno **Zgr.**, **Zdr.** i **Por.** – dzięki czemu łatwo je wyodrębnić. Ich wewnętrzna kompozycja również jest prosta: jej struktura jest zbliżona do ciągów synonimicznych, tj. szeregu leksemów wraz z elementami je opisującymi, które są oddzielone przecinkami i niekiedy spójnikami.

10.3.8.2. Segment etymologiczny

Segment etymologiczny zawsze znajduje się w nawiasie trójkątnym **< >**, dzięki czemu łatwo zidentyfikować jego początek i koniec. Jego wewnętrzna budowa jest jednak znacznie bardziej złożona i miał jej być poświęcony odrębny moduł, co zostało opisane w **podrozdziale 10.7.1.**

Elementów, które mogą wystąpić jest bardzo wiele, zasadniczo są to jednak wyrazy obce wprowadzane przez skróty nazw języków wraz z ekwiwalentami, derywacje i kompozycje słowotwórcze, rodzime słoworody, nazwy własne, a także elementy opisowe. Wymagało to opracowania 32 dwóch reguł przechwytyjących kilkanaście rodzajów elementów o różnym układzie.

10.3.9. Wykorzystanie znaków pomocniczych

W procesie parsowania *Słownika warszawskiego* wykorzystano znaki pomocnicze, które umieszczane były w treści tylko na potrzeby analizy składniowej, a następnie usuwane w pliku wyjściowym. Znaki te zostały dobrane arbitralnie spośród tych, które obecne są w czcionce Consolas i nie są użyte w oryginalnej treści *SW*.

Znak punktora • użyty jest zamiast spacji, które w przeciwnym razie mogłyby zostać rozpoznane przez parser jako granicę segmentów. Wiele punktów jest wstawianych automatycznie w pewnych klasach przypadków, część jednak jest wprowadzana ręcznie z wykorzystaniem pliku podmian.

W sposób zbliżony stosowany jest czarny kwadrat o połowie szerokości ■, którego zadaniem jest zastąpienie spacji wprowadzającej jednostki niższe w hierarchii w bardzo specyficznych kontekstach.

Pary tyld ~ ~ umieszczane są wokół tekstu, który typograficznie został wyróżniony rozstrzeleniem liter. Rozstrzelenie to jest wykorzystywane do oznaczania terminów ukrytych, ponieważ te nie są w żaden sposób wyróżnione strukturalnie. Znaki te również umieszczane są w pliku podmian wszędzie tak, gdzie fragment tekstu był wyróżniony w oryginale.

Pary małp @ @ umieszczane są wokół oznaczeń źródeł, pary znaków paragrafu § § umieszczane są wokół treści przykładu cytowanego, natomiast pary znaków karo ◆ ◆ umieszczane są całe jednostki (cytowania ze źródłem) – w obu przypadkach program robi to automatycznie, a celem tego jest poprawna segmentacja kolejnych przykładów cytowanych.

10.3.10. Zagregowane związki wyrazowe

Nazywane w ten sposób są tu jednostki, które spotykane są w segmencie zwłaszcza centralnym (najczęściej w przykładach), ale także i w segmencie terminalnych (w części etymologicznej), a które *de facto* są zbiorem jednostek leksykalnych zapisanych wspólnie w sposób skrótowy dla zaoszczędzenia miejsca.

Ich detekcja zależy od ich rodzaju, program zasadniczo jednak stara się (skutecznie) na różne sposoby przewidzieć, gdzie leżą granice między jednostkami i które fragmenty należy rozwinąć przy pomocy lemmy. Stosowana jest do tego analiza strukturalna, weryfikacja wielkości liter, a także porównanie z samą lemmą.

Niepełna lista klas jednostek zagregowanych rozwijanych automatycznie przez program SWParser to m.in.:

1. Pierwszy z dwóch wyrazów jest stały, drugi wyraz się zmienia:
 - F moll a. Minor
→ F moll, F minor

- Fajeru dać a. Zadać
→ Fajeru dać, Fajeru zadać
 - Fijołkami marcowemi a. Brunatnemi
→ Fijołkami marcowemi, Fijołkami brunatnemi (zwane)
 - Formacje paleozoiczne a. Pierwszorządowe
→ Formacje paleozoiczne, Formacje pierwszorządowe
 - [fonfry robić a. stroić]
→ [Fonfry robić, [Fonfry stroić
 - Groch turecki a. włoski a. Szablasy
→ Groch turecki, Groch włoski, Groch szablasy
2. Pierwszy wyraz jest opcjonalny:
 - W fałdy a. fałdy układać
→ W fałdy układać, Fałdy układać
 3. Dwa równorzędne segmenty, pierwszy wyraz jest w obu taki sam:
 - Mieć w głowie fijołki a. mieć fijołka
→ Mieć w głowie fijołki, Mieć fijołka
 4. Dwa równorzędne segmenty, drugi wyraz jest taki sam:
 - Niemoc francowata a. choroba francowata
→ Niemoc francowata, Choroba francowata
 5. Dwa równorzędne segmenty, pierwszy element jest w obu podobny (posiada formy odmienione):
 - Fałszywy stosunek a. fałszywa relacja
→ Fałszywy stosunek, Fałszywa relacja
 6. Drugi element pierwszego segmentu jest równy pierwszemu elementowi segmentu drugiego (w postaci skrócenia lub całego wyrazu):
 - F. różyczka a. różyczka Panny Marji
→ Firletka różyczka, Różyczka Panny Marji
 7. Pierwszy element (skrócenie) wszystkich segmentów jest wspólny, drugi i kolejne elementy (wyrazy) są różne:
 - F. celny, niemiecki a. metryczny
→ Funt celny, Funt niemiecki, Funt metryczny
 8. Ostatni element jest wspólny (wyraz), lista elementów wchodzących w skład poprzedza
 - Góry, doliny fałdowe
→ Góry fałdowe, Doliny fałdowe
 - Mieć a. cierpieć fiksum-dyrdum
→ Mieć fiksum-dyrdum, Cierpieć fiksum-dyrdum
 - Pokazać a. xukazać komu figę
→ Pokazać komu figę, xUkazać komu figę
 - Sprawić a. dać komu F.
→ Sprawić komu fryko, Dać komu fryko
 9. Ostatni element jest wspólny (skrócenie), lista elementów wchodzących w skład poprzedza
 - Alkohol, xkwas F.
→ Alkohol fenilowy, xKwas fenilowy

- Sprawić a. dać komu F.
→ Sprawić komu fryko, Dać komu fryko
 - Śpiew, kontrapunkt F.
→ Śpiew figularny, Kontrapunkt figularny
10. Pierwszy element jest wspólny dla obu segmentów, drugi element się zmienia (skrótowania i wyrazy):
- Kwas F., fosforowy, †fosforyczny
→ Kwas fosforowy, Kwas fosforowy, †Kwas fosforyczny
11. Dwa równorzędne elementy jednowyrazowe:
- Metaftalowy a. Izoftalowy
12. Skrócenie i dalsze wyrazy są wspólne w obu segmentach, ostatni wyraz jest inny:
- F. w futrze od drzwi a. Okna
→ Felc w futrze od drzwi, Felc w futrze od okna
13. Pierwszy wyraz jest wspólny, drugi element jest wspólny lub skrócony (w innej formie):
- †Obie fortuny a. †oboja F.
→ †Obie fortuny, †Oboja fortuna
 - Choroba francuska a. gościec F. a. kołnierzyki francuskie a. niemoc francuska a. ospa francuska a. słabość francuska a. świerzba francuska
→ Choroba francuska, Gościec francuski, Kołnierzyki francuskie, Niemoc francuska, Ospa francuska, Słabość francuska, Świerzba francuska
14. Fragment wyrazu jest w nawiasie:
- fer(r)adże, feredże
→ feradże, ferradże, feredże

Wszystkie przytoczone powyżej reguły konwersji w pełni radzą sobie parsowaniem materiału badawczego, z wyjątkiem ostatniej z nich, która może doprowadzić do sytuacji takiej, jak w przypadku hasła **×Fizjonotyp**, co zostało opisane w **podrozdziale 10.8.5**.

10.4. Generowanie danych pochodnych

Po przeprowadzonej analizie składniowej danych wejściowych i wygenerowaniu wyjściowego pliku XML program SWParser przystępuje do sekwencyjnego generowania danych pochodnych. Bezpośrednie podpięcie modułów w ten sposób pozwala na zautomatyzowanie procesu obróbki danych, chociaż podatne jest również na kaskadową propagację błędów parsera – nie stanowi to jednak problemu, gdy błędy te zostały wyeliminowane, a dane wyjściowe przeszły pomyślną walidację.

Opis funkcjonowania poszczególnych modułów wchodzących w skład SWParser został opisany poniżej.

10.4.1. Indeks artykułów hasłowych

Moduł ten ekstrahuje z pliku XML informacje o wyrazach hasłowych wraz z dezambiguatorami i kwalifikatorami symbolicznymi, a także informacją o lokalizacji hasła. Lista lemm jest

następnie uporządkowywana alfabetycznie w dwóch wariantach: *a fronte* i *a tergo*. Indeks ten jest otrzymywany dzięki odfiltrowaniu odpowiednich węzłów w pliku XML i zastosowaniu na nim własnoręcznie wykonanego algorytmu sortującego.

Dane wygenerowane przez ten moduł mają format `.csv` pozwalają na indeksowanie wyrazów hasłowych i ich łatwe wyszukanie i mają zastosowanie w badaniach leksykologicznych i słowotwórczych (zwłaszcza wariant *a tergo*). Dane te można wykorzystać także m.in. do porównania siatki hasłowej *SW* z siatkami innych słowników.

10.4.2. Indeks statych połączeń wyrazowych

Moduł ten ekstrahuje z pliku XML informacje o wszystkich rozpoznanych jednostkach wielowrazowych, a także o hasła, w którym się one znajdują. Jednostki te są następnie porządkowane w kolejności alfabetycznej.

Dane wygenerowane przez ten moduł mają format `.csv` i mają zastosowanie szczególnie w badaniach frazeologicznych oraz składniowych.

10.4.3. Indeks form wyrazowych

Moduł ten ekstrahuje z pliku XML wszelkie fragmenty tekstu znajdujące się w definicjach, wyrażeniach, glosach i komentarzach, dzieli je na wyrazy, sprowadza je do zapisu małymi literami i zlicza wystąpienia każdej unikalnej grafoformy, a następnie sortuje rezultaty: w pierwszej kolejności pod względem liczby wystąpień, a w drugiej zaś alfabetycznie – to dzieje się zaś w dwóch wariantach: *a fronte* i *a tergo*.

Dane wygenerowane przez ten moduł mają format `.csv` zawierają sporadycznie zapis pseudofonetyczny. Po odfiltrowaniu takich przypadków można wykorzystać do uzupełnienia siatki hasłowej *SW* (opracowania autosuplementu dzięki porównaniu znalezionych form wyrazowych z siatką hasłową) czy opracowywaniu paradygmatów fleksyjnych. Dane te mają także zastosowanie w badaniach leksykologicznych.

10.4.4. Korpus zdań i wyrażeń przykładowych

Moduł ten ekstrahuje z pliku XML wszystkie zidentyfikowane przykłady i cytowania, ich źródło (jeśli jest podane) oraz hasło, z którego pochodzą. Następnie skróty rzadko stosowane w tekście, takie jak `a.` (= albo) czy `ś.` (= się), są automatycznie rozwijane; skróty typowe, takie jak `św.`, `ks.`, `i` `t.` `d.` pozostają bez zmian. Zapisywane w pełnej formie są też wszelkie wystąpienia lemmy, które w *Słowniku* skracane są do postaci inicjału, co ma miejsce tylko gdy lemma występuje w tekście w formie słownikowej.

Niektóre z przykładów można zinterpretować jako terminy wielowrazowe, mimo że nie pełnią takiej funkcji w *Słowniku* – nie udało się jednak opracować metody automatycznej ich identyfikacji. Cytowania są też bardzo zróżnicowane regionalnie, stylistycznie i ortograficznie, co jednak nie jest w żaden sposób oznaczone w słowniku.

Dane wygenerowane przez ten moduł mają format `.csv` i mogą stanowić mini-korpus do badań składniowych, fleksyjnych czy semantycznych.

10.4.5. Wykaz znaków w systemie Unikod

Moduł ten opracowuje na podstawie poprawionego pliku wejściowego listę wszystkich użytych znaków systemie Unikod, dla których podaje współrzędne kodowe a także oficjalną nazwę znaku w języku angielskim z wykorzystaniem rozszerzenia `unicode-name` dla języka Ruby. Znaki te są uporządkowane malejąco pod względem liczby wystąpień.

Dane wygenerowane przez ten moduł mają format `.csv` i mają zastosowanie przede wszystkim podczas opracowywania narzędzi operujących na tekście wejściowym, a także podczas opracowywania prezentacji treści *SW*.

10.4.6. Fotokorpus haseł

Moduł ten łączy wygenerowaną uprzednio siatkę hasłową z informacją o lokalizacji każdego artykułu na nośniku fizycznym z pliku lokalizacji opisanego w **podrozdziale 10.2.2.2**, oraz z kartami hasłowymi wygenerowanymi przez program IndexCardSplitter.

W przypadku, w którym ta sama lemma ma podane wielokrotne lokalizacje, brane są skrajne z nich, co pozwala na wygenerowanie zakresu stron i kolumn, na których znajduje się dane hasło.

Dane wygenerowane przez ten moduł wykorzystywane są dopiero przy generowaniu aneksów i pełnią przede wszystkim funkcję zasobu fotodokumentacyjnego.

10.4.7. Informacja gramatyczna

Moduł ten ekstrahuje z pliku XML informację gramatyczną dla poszczególnych haseł, którą scala (jeśli informacje o paradygmacie dotyczą poszczególnych znaczeń) i normalizuje z użyciem pliku pomocniczego opisanego w **podrozdziale 10.2.3.2**. To uczyniwszy tworzony jest spis wszystkich unikalnych paradygmatów, które następnie są w grupowane ze względu na część mowy, jakiej dotyczą i sortowane o najczęstszego do najrzadszego, a na tej podstawie nadawane im są sekwencyjnie nazwy – w rezultacie `N-001` oznacza najczęstszy paradygmat rzeczownikowy, `N-002` drugi najczęstszy i tak dalej.

Wygenerowane powyżej danej użyte są do stworzenia dwóch zasobów. Pierwszym z nich jest słowniczek, w którym obok każdej lemmy umieszczana jest informacja gramatyczna oraz klasa paradygmatyczna. Drugim natomiast jest zbiór paradygmatów, gdzie obok każdego z nich podawana jest pełna informacja gramatyczna, którą reprezentują, a wraz z liczbą jednostek, które do danej klasy należą.

Opis gramatyczny uzyskany tą metodą jest pod wieloma względami zbliżony do tego, jaki przedstawiła Bilińska (2013: 115–118) na potrzeby opisu *SL*.

Dane wygenerowane przez ten moduł mają format `.csv` i można je wykorzystać w badaniach nad gramatyką historyczną, a także do rekonstrukcji tablic fleksyjnych, co z kolei pozwoliłoby na stworzenie słownika sprawdzania pisowni dla języka nowopolskiego.

10.4.8. Słowniczek walencyjny

Moduł ten ekstrahuje z pliku XML wszelkie hasła, które mają podaną informację o walencji, a także zakres znaczeń lemmy, jakiego informacja ta dotyczy. Lemmy są następnie sortowane alfabetycznie.

Dane wygenerowane przez ten moduł mają format `.csv` i mogą zostać wykorzystane do badań nad walencją i składnią.

10.4.9. Słowniczki przekładowe

Moduł ten ekstrahuje z części etymologicznej pliku XML wszelkie informacje o zapożyczeniach z języków obcych, które następnie grupuje na podstawie skrótu języka. Skróty te są następnie rozwijane przy pomocy *Spisu skrótów* i prostego algorytmu, który wykorzystuje rozwinięcia do wygenerowania odpowiedniej nazwy słowniczka, np.:

Nm. → niemiecki → niemiecko-polski → Mały słownik niemiecko-polski

Zgrupowane hasła wraz z podanymi w *SW* polskimi ekwiwalentami są następnie umieszczane w osobnych plikach, z których każdy odpowiada parom przekładowym między danym językiem a polszczyzną.

Dane wygenerowane przez ten moduł mają format wielu plików `.csv` i mogą posłużyć do badania semantyki wyrazów polskich i obcych, zwłaszcza diachronicznych.

10.4.10. Matryca siatek hasłowych SW i wybranych słowników

Moduł ten porównuje uzyskaną wcześniej siatkę pojęciową i porównuje ją z siatkami hasłowymi słowników referencyjnych (diagnostycznych), które zostały wyekstrahowane metodami ręcznymi i automatycznymi, a następnie umieszcza w plikach predefiniowanych, opisanych w **podrozdziale 10.2.3.4**. Ze względu na duże zróżnicowanie typograficzne dzieł przyjęto ogólną zasadę ekscerpcji jednostek, które w danym układzie typograficznym stanowiły główny element hasła lub podhasła. W przypadku zasobów elektronicznych, w tym hasłownika *SL*⁶⁹ oraz *Elektronicznego Słownika Języka Polskiego XVII i XVIII wieku*⁷⁰, przyjęto hasła wyodrębnione przez autorów.

Jednostki hasłowe są następnie normalizowane, co obejmuje przede wszystkim rozwinięcie skrótów i zapisanie wariantów jako osobnych jednostek, sprowadzenie wielkich liter w funkcji typograficznej do małych, a także usunięcie komentarzy, symboli naukowych i wskaźników fonetycznych (zwłaszcza dywizów). Przykład normalizacji widoczny jest w **tabeli 1**.

⁶⁹ Źródło danych: <https://github.com/jsbien/iLindeCSV>.

⁷⁰ Strona słownika: <https://sxvii.pl/>.

Tabela 7. Przykłady normalizacji haseł.

Hasło oryginalne	Hasło znormalizowane	Źródło
iść prawym chwytem; prawym wiatrem (prawym halsem)	iść prawym chwytem iść prawym wiatrem iść prawym halsem	SMor
ilość pocisków gazowych potrzebnych do skażenia, do zagazowania	ilość pocisków gazowych potrzebnych do skażenia ilość pocisków gazowych potrzebnych do zagazowania	SBrChem
impregnit (amerykański)	impregnit impregnit amerykański	SBrChem
fale przerywane (tonowane) fale przerywane, tonowane	fale przerywane fale tonowane	STey
tniak (na trzonku) tniak (kowalski)	tniak tniak na trzonku tniak kowalski	ŚSlus
farbownik lek.	farbownik lekarski	SGier
faj(l)kluba	fajkluba fajlkluba	SRzem
fersegung, fersenkung, fer-zegung, fer-zenkung	fersegung, fersenkung, ferzegung, ferzenkung	SRzem
inwersor odległościowy (uzgadniający odległości a, b, f, przetwornika)	inwersor odległościowy	SSłużGeogr
indol C ₈ H ₇ N	indol	STow
frankowany, list ~, nie-~	frankowany list frankowany list nie-frankowany	SHandl
inwentarzowy, księga ~a	inwentarzowy księga inwentarzowa	SHandl
fajans szlachetny z kaolinu t. zw. niemiecki	fajans szlachetny z kaolinu fajans szlachetny niemiecki	SHandl
fosforan wapnia jednozasadowy albo kwaśny	fosforan wapnia jednozasadowy fosforan wapnia kwaśny	SHandl

Abscessus — ropień. ~ alveolaris — r. zębodołowy. ~ apicalis — r. wierzchołkowy.	Abscessus — ropień. Abscessus alveolaris — ropień zębodołowy. Abscessus apicalis — ropień wierzchołkowy.	SDent
Atonia — wątłość, zwątlenie. ~ gingivae — dziąseł.	Atonia — wątłość, zwątlenie. Atonia gingivae — wątłość dziąseł, zwątlenie dziąseł.	SDent

Oprócz tego ekscerpowane były również lemmy wewnątrz, które posiadały formę słownikową i były wyróżnione typograficznie względem sąsiadującego tekstu (np. w *SKarł*). Pomijano za to nazwy osobowe (np. *SBrk*, *SPedag*) czy warianty rekonstruowane (np. *SStarop*). W przypadku słowników wielojęzycznych posiadających więcej niż jeden kierunek przekładowy, brano pod uwagę jednostki polskie znajdujące się we wszystkich sekcjach (np. *SAut*, *STkac*).

Znormalizowane siatki hasłowe są następnie łączone w jedną listę uporządkowaną alfabetycznie bez powtórzeń, po czym program sprawdza obecność lemmy w poszczególnych słownikach stosując pełne dopasowanie. Na tej podstawie opracowywana jest matryca wystąpienia haseł.

Dane wygenerowane przez ten moduł mają format `.csv` i mają zastosowanie w badaniach leksykologicznych oraz metaleksykograficznych.

10.4.11. Lista podmian w treści

Moduł ten wczytuje plik podmian opisany w **podrozdziale 10.2.2.1** i dokonuje porównania oryginalnych fragmentów treści *SW* z ich poprawionymi wersjami, które program wykorzystuje podczas parsowania. Celem tej czynności jest znalezienie koordynat w łańcuchu znaków, które wskażą podłańcuch różnicujący oba fragmenty.

Problem ten jest znany w informatyce jako problem najdłuższego wspólnego podłańcucha (ang. *longest common substring problem*) i nie posiada jednego optymalnego rozwiązania; piszącemu te słowa nie udało się również znaleźć gotowego rozwiązania, które można by było zastosować w przypadku programu *SWParser*, dlatego napisał własną implementację opartą o założenie, że dla danej pary łańcuchów zawsze będzie obecna dokładnie jedna różnica.

Program najpierw dokonuje kontrolnego binarnego porównania obu łańcuchów, oczekując, że będą od siebie różne. Wówczas przystępuje do pierwszej metody znalezienia koordynat sprawdzając, czy któryś z łańcuchów zawiera się w drugim. Jeśli tak się stanie, to pozostała część dłuższego z nich (na jego początku lub końcu) stanowi fragment różnicujący: oznacza to odpowiednio dodanie znaków względem treści oryginału, jeśli drugi łańcuch jest dłuższy, albo ich odjęcie, jeśli dłuższy jest łańcuch pierwszy. Zapisywane są wówczas koordynaty części niewspólnej.

Jeśli natomiast żaden łańcuch znaków nie zawiera się w całości w drugim, porównanie dokonywane jest metodą drugą, która sprawdza kolejne znaki obu łańcuchów aż do momentu

napotkania na rozbieżność. Następnie robi to samo idąc od końca obu łańcuchów. Znalezione koordynaty wyznaczają fragment stanowiący różnicę obu łańcuchów.

W miejscu koordynat dla każdego z pary łańcuchów umieszczana jest para znaków ►◄, które wskazują na lokalizację rozbieżności między nimi. W **tabeli 8** przedstawiono niewielki fragment pliku podmian po umieszczeniu znaków wskazujących wprowadzone zmiany.

Tabela 8. Fragment przetworzonego pliku podmian.

Pole „Znajdź”	Pole „Zamień”	...
p. Gryzetk►i◄	p. Gryzetk►a◄	...
p. Gdu►ł◄a	p. Gdu►l◄a	...
Fulminacja, i, lm. e 1. (o złocie, srebrze a. prochu) ►1. ◄łoskot, szczęk, brzęk; trzask.	Fulminacja, i, lm. e 1. (o złocie, srebrze a. prochu) ►◄łoskot, szczęk, brzęk; trzask.	...
F. ś. 1. ►1. ◄troszczyć ś.	F. ś. 1. ►◄troszczyć ś.	...
Kobyła ma większy łeb, to niech ś. frasuje. Prz. ►◄F. ś. o kogo a. [za kogo]:	Kobyła ma większy łeb, to niech ś. frasuje. Prz. ►2. ◄F. ś. o kogo a. [za kogo]:	...
<Nm. [vormiedel]►◄	<Nm. [vormiedel]►>◄	...
...

Dane wygenerowane przez ten moduł mają format `.csv` i pozwalają zweryfikować poprawność modyfikacji wprowadzonych do treści *SW*, a także dokonać analizy błędów redaktorskich i ich kategorii, które w dziele tym się pojawiły.

10.4.12. Skonsolidowany spis skrótów

Moduł ten wczytuje listę skrótów opisaną w **rozdziale 10.2.3.1** i uzupełnia ją o liczbę wystąpień każdego oznaczenia w materiale badawczym.

Dane wygenerowane przez ten moduł mają format `.csv` i mają zastosowanie w badaniach metaleksykograficznych.

10.4.13. Dane liczbowe

Moduł ten generuje dane liczbowe na podstawie pliku XML (a konkretnie: podliczenia znaczników oraz analizy zawartej w nich treści) oraz wygenerowanych danych pochodnych. Dane posiadają kategorię, nazwę (opis) wartości oraz przypisaną jej wartość liczbową. Warto zaznaczyć, że wszystkie mediany są obliczane po odrzuceniu wartości zerowych, a liczba definicji uwzględnia także poddefinicje.

Dane wygenerowane przez ten moduł mają format `.csv` i dostarczają nie tylko informacji o samym słowniku, ale także o procesie jego przetwarzania i wyekstrahowanych danych.

10.4.14. Słowniczek par derywacyjnych

Moduł ten generuje na podstawie pliku XML pary derywacyjne znajdujące się w treści *SW* w trzech kategoriach: (1) pary przymiotników i pochodzących od nich przysłówków, (2) pary

czasowników niezwrrotnych i zwrotnych, a także (3) listę czasowników i pochodzących od nich rzeczowników odczasownikowych (odsłowników). Dane te ekstrahowane z odsyłaczy derywacyjnych w artykułach hasłowych, a także – w przypadku wariantów zwrotnych – dodatkowo poprzez wyszukiwanie wariantów artykułów hasłowych różniących się tylko partykułą „się”.

Dane wygenerowane przez ten moduł mają format odrębnych plików `.csv` i mogą posłużyć do badań słowotwórstwa.

10.4.15. Słowotwórstwo rzeczowników żeńskich (feminatywów)

Moduł ten generuje na podstawie pliku XML pary rzeczowników męskich i pochodzących od nich rzeczowników żeńskich (feminatywów). Dane te ekstrahowane z odsyłaczy derywacyjnych w artykułach hasłowych.

Dane wygenerowane przez ten moduł mają format `.csv` i mogą posłużyć do badań leksykologicznych poświęconym symetrii leksykalnej w polszczyźnie.

10.5. Dane wyjściowe

Rezultatem działania programu SWParser jest wygenerowanie różnego rodzaju danych, które następnie są zapisywane na dysku. Poniżej opisana jest specyfika ich cyfrowego formatu.

10.5.1. Format i prezentacja danych

Pliki wyjściowe programu SWParser zapisywane są w trzech generacjach odzwierciedlających trzy etapy działania programu, każda stosująca inny format danych – wszystkie łączy jednak zastosowanie kodowania UTF-8, które jest jedną z implementacji standardu Unikod. Formy wyjściowe są następujące:

- sparsowana treść słownika ma format `.swp`, będący formatem wewnętrznym (przejściowym) programu SWParser;
- na podstawie powyższego generowany jest plik `.xml`, które jest właściwą reprezentacją struktury słownika;
- dane pochodne są umieszczane w osobnych plikach `.csv`, które mają układ tabelaryczny.

Plik XML oraz pliki pomocnicze, a także dane wyjściowe programu IndexCardSplitter są następnie – z pomocą systemu przygotowanych ręcznie szablonów ERB dla języka Ruby – do wygenerowania kodu w języku \LaTeX , który następnie jest wykonywany, czego końcowym rezultatem są pliki PDF, które stanowią aneksy do niniejszej tezy doktorskiej.

10.5.2. Wyświetlanie znaków

Jednym z problemów, jaki napotkano podczas obróbki danych wejściowych, a który istotny jest zwłaszcza z punktu widzenia danych wyjściowych, jest problem wyświetlania znaków. O ile użyto wyłącznie znaków użytych standardem Unikod, to nie każdy font posiada graficzne realizacje każdego znaku, przez co nie może ich wyświetlić, lub są one niespójne graficznie z innymi znakami w tej samej kategorii. Najczęściej tyczy się to znaków rzadko używanych,

takich, które weszły do standardu po stworzeniu fontu, lub pochodzących z innego systemu pisma.

W przypadku treści SW dzieje się tak zwłaszcza w przypadku znaków zapisanych w indeksie górnym lub dolnym: jako że dane wejściowe są w formie czystego tekstu (bez formatowania), znaki zapisane w indeksach zostały zakodowane przy pomocy odrębnych punktów kodowych, tj. znaki w indeksie górnym (np. ¹, ², ³, lub ⁿ) lub dolnym (np. ₁, ₂, ₂ lub _n) stanowią byty odrębne od swoich odpowiedników w tekście podstawowym (czyli kolejno znaków 1, 2, 3 oraz n).

Na **ilustracji 58** widoczne jest, jak wiersz z hasłem „Ferycyjan” wygląda na ekranie komputera, kiedy jest wyświetlane w foncie Courier New w programie Notepad++. W **listingu 18** widać natomiast, jak wygląda tekst w foncie Cambria Math. Pierwszy z wymienionych fontów cechuje się stałą szerokością znaku i jest powszechnie wykorzystywany w konsolach systemowych oraz oprogramowaniu do edycji czystego tekstu, podczas gdy drugi jest zaprojektowany specjalnie do zadań matematycznych, ale nie nadaje się do wykorzystania w konsoli – czyli środowisku, w którym dane te funkcjonują – ponieważ wyświetla znaki o zmiennej szerokości.

Ilustracja 58. Hasło „Ferycyjan” w foncie Courier New (Notepad++).

```
1309 Ferycyjan, u, blm., Nadżelazocyjan, (w Galicji) Nadżelasin
chem. rodnik złożony z żelaza trój (sześćcio-) wartościowego,
węgla i azotu, Fe(C□N□) I I I a. Fe□(C□□N□) I v: Ferycyjanowodór =
Fe(C□N□) H□ a. Fe□(C□□N□□) H□. <Now. z Łć. ferrum = żelazo +
Cyjan>
```

Listing 18. Hasło „Ferycyjan” w foncie Cambria Math.

```
Ferycyjan, u, blm., Nadżelazocyjan, (w Galicji) Nadżelasin chem. rodnik złożony z żelaza trój
(sześćcio-) wartościowego, węgla i azotu, Fe(C6N6)III a. Fe3(C13Nn)IV: Ferycyjanowodór =
Fe(C6N6)H3 a. Fe3(C12N13)H6. <Now. z Łć. ferrum = żelazo + Cyjan>
```

Poniżej, na **ilustracji 59** znajduje się prezentacja tekstu (1) w indeksie górnym, (2) głównego (zwykłego), oraz (3) w indeksie dolnym w foncie Courier New. Dla porównania ten sam zestaw znaków w foncie Cambria Math został przedstawiony w **listingu 19**. Można łatwo zauważyć, że Courier New nie posiada wszystkich potrzebnych znaków, a część z nich wygląda niespójnie.

Ilustracja 59. Wybrane znaki w foncie Courier New.

```
1 □1 □2 □3 □□□□□□ I v n
2 0123456789IVn
3 □□□□□□□□□□ i v □
```

Listing 19. Wybrane znaki w foncie Cambria Math.

```
0123456789IVn
0123456789IVn
0123456789ivn
```

Innym problemem jest niespójne wyświetlanie znaków złożonych, czyli takich, w skład których wchodzi znak podstawowy i znak lub znaki dostawne. W przypadku niektórych z nich dany font może renderować je jako pojedynczą jednostkę graficzną (zajmującą jedno pole w przypadku fontu o stałej szerokości), a czasem jako odrębne znaki. Fragment danych statystycznych wyświetlanych w konsoli, gdzie widoczny jest powyższy problem, przedstawia **ilustracja 60**.

Ilustracja 60. Wybrane znaki złożone wyświetlane w konsoli.

```
94.  é      LATIN SMALL LETTER E WITH MACRON AND ACUTE      U+1E17  27
95.  ...    HORIZONTAL ELLIPSIS      U+2026  22
96.  ȯ̂     LATIN SMALL LETTER O + COMBINING MACRON + COMBINING CIRCUMFLEX ACCENT  U+006F  21
```

Tego typu złożenia znaków są w *SW* niemal całkowicie ograniczone do części hasła z informacją etymologiczną. Poprawne wyświetlanie bywa silnie uzależnione od konkretnego znaku podstawowego i dostawnego – porównaj:

- <Now. z Gr. fōs, 2 pp. fōtós = światło + Chemja>
<Now. z Gr. fȯ̂s, 2 pp. fōtós = światło + Chemja>
- <Now. z nazwy litery Gr. êta>
<Now. z nazwy litery Gr. ē̂ta>
- <Now. z Gr. ísos = równy + anómalos = nieprawidłowy>
<Now. z Gr. ísos = równy + anómalos = nieprawidłowy>

Trudno znaleźć font, który spełnia oba wymagania, tj. wyświetla wszystkie potrzebne znaki i zachowuje ich stałą szerokość. Nie przeszkadza to jednak w operowaniu programistycznym na tekście, a na poszczególnych etapach można stosować różne sposoby wyświetlania tekstu.

10.6. Podsumowanie procesu

Wszystkie etapy działania programu SWParser są w pełni automatyczne – oznacza to, że wystarczy uruchomić program podając mu odpowiednie dane wejściowe i parametry, a rezultatem końcowym będą aneksy w formacie `.pdf`. Wygląd zakończonego wykonania programu w konsoli przedstawiony jest na **ilustracji 61**, natomiast poszczególne etapy działania programu oraz pliki generowane na każdym z nich opisuje schemat na **ilustracji 62**.

Ilustracja 61. Wygląd programu w konsoli.

```

RUBY - Warsawian Dictionary
Enter ruby command (or 'Q' to exit) [ruby swparser.rb f 0]:

>>>>>>>> PARSING BEGINS <<<<<<<<<<
12:56:36

NOW PARSING LETTER: F

===== PARSED LETTER: F =====
12:57:12

MAIN ENTRIES : 3880
NEST ENTRIES : 42
TOTAL ENTRIES : 3922

TIME LAPSED : 00:00:35
OUTPUT SIZE : 7 673 KB

SWAPPED LINES: 121 / 0

S-1v2 MAIN : 3880 / 0
S-1v2 NEST : 42 / 0
S-1 PARSED : 3922 / 0
S-2 PARSED : 5397 / 0
USG-BL PARSED : 2033 / 0
USG-EL PARSED : 2886 / 0
S-3 PARSED : 1889 / 0
ETYM PARSED : 1843 / 0
LOAN PARSED : 1541 / 0
GRAM PARSED : 8799 / 0
SYN PARSED : 1225 / 0

SWAP DIFFERENTIATION TABLE: DONE

ENTRYWORDS - AFRONTE: DONE
ENTRYWORDS - ATERGO: DONE

XML FILE: DONE

NON/REFLEXIVE DERIV. PAIRS: DONE
ADJ/ADV DERIV. PAIRS: DONE
VERB/GERUND DERIV. PAIRS: DONE

WORDFORMS - AFRONTE: DONE
WORDFORMS - ATERGO: DONE

GENDER SYMMETRY: DONE

CORPUS OF EXAMPLES: DONE

WORD COMBINATION LIST: DONE

HEADWORD MATRIX: DONE

PARADIGM PARSING: DONE

ABBREVIATION ANALYZER: DONE

DICTIONARIES: DONE

WORDGEN LIST: DONE

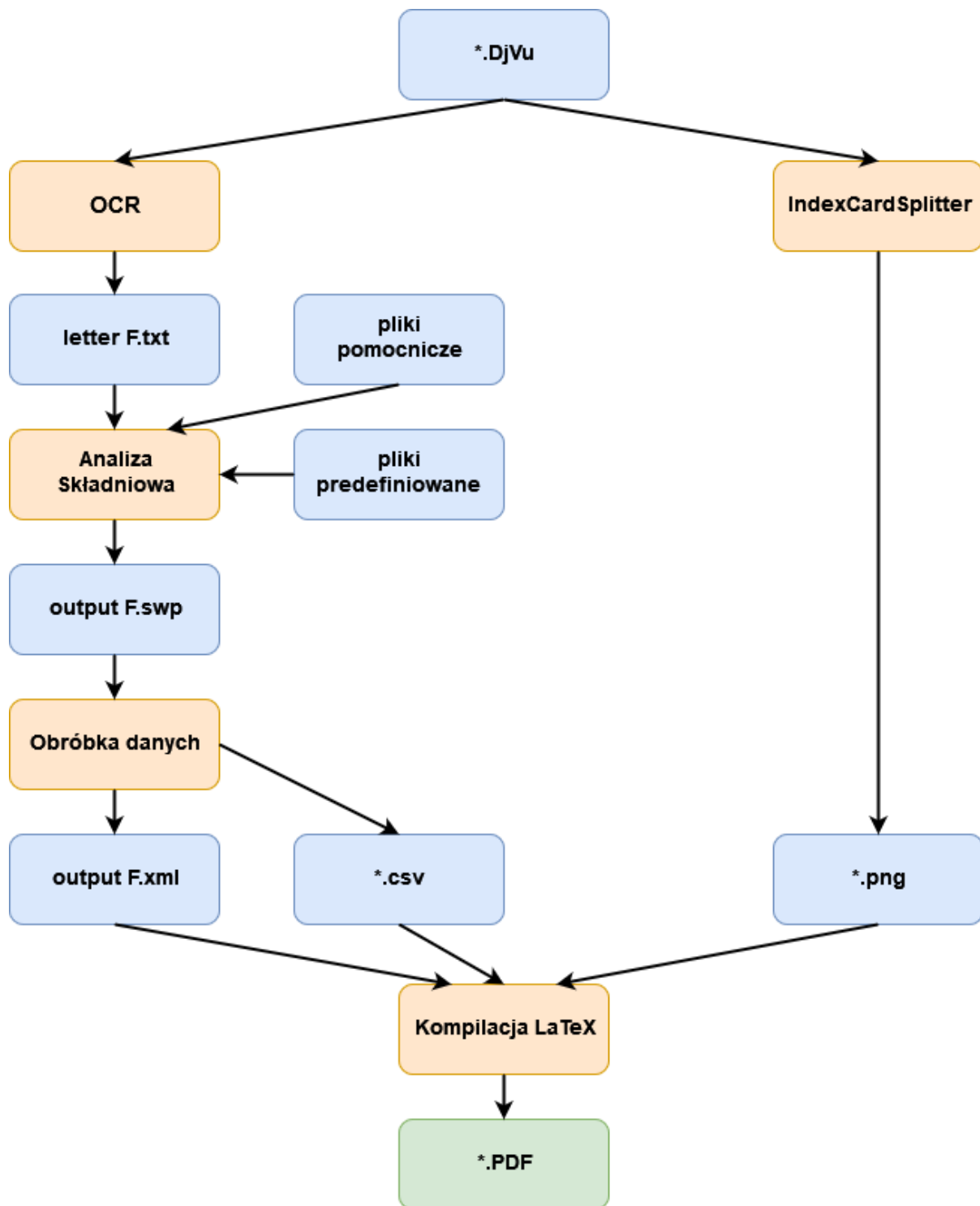
CHARACTER LIST: DONE

STATISTICS: DONE

ANNEX A: DONE
ANNEX B: DONE
ANNEX C: DONE
ANNEX D: DONE
ANNEX E: DONE
ANNEX F: DONE
ANNEX G: DONE
ANNEX H: DONE
ANNEX I: DONE
ANNEX J: DONE
ANNEX K: DONE
ANNEX L: DONE
ANNEX M: DONE
ANNEX N: DONE
ANNEX O: DONE

```

Ilustracja 62. Etapy przetwarzania danych.



10.7. Niezaimplementowane moduły

Oprócz modułów, które generowały wymienione powyżej aneksy, w planach było ich jeszcze kilka, a nad częścią z nich rozpoczęto nawet pracę. Ze względu jednak na ograniczenia czasowe oraz i tak obszerną część praktyczną tezy doktorskiej, piszący te słowa zmuszony był zarzucić pomysł włączenia ich do niniejszej pracy. Obrazują one jednak potencjał danych wyekstrahowanych ze *Słownika warszawskiego*, dlatego też zostaną pokrótce opisane.

10.7.1. Klasyfikacja informacji etymologicznej

Moduł ten miał pozwolić na automatyczną klasyfikację rodzajów informacji, które znajdują się w części etymologicznej hasła. Pozwoliłoby to na wyróżnienie haseł o pochodzeniu rodzimym od pożyczek językowych, reprezentujących różne strategie słowotwórcze, czy dźwiękonaśladowczych itd.

W ramach prac okazało się jednak, iż klas takich jest bardzo wiele, i wynikały one nie tylko z mnogości fenomenów językowych w materiale badawczym, ale także niskiego usztywnienia struktury rubryki etymologicznej; konieczne byłoby więc zaawansowane mapowanie fenomenów do struktur. Jednocześnie można było zaobserwować w nich rozkład Pareta⁷¹: mała liczba częstych struktur obejmowała wiele haseł, natomiast duża liczba rzadkich struktur obejmowała niewielką liczbę haseł. Wstępna analiza wskazywała też, że rozszerzenie materiału badawczego problem ten pogłębi, ponieważ hasła na literę **F** nie wyczerpują sposobów opisu informacji o pochodzeniu.

10.7.2. Generowanie tabel fleksyjnych

Planowane było również wykorzystanie informacji o paradygmatach – których ekstrakcja i automatyczna klasyfikacja zostały opisane w **podrozdziale 10.4.7** – wraz z wiedzą o języku wykraczającą poza *SW*, aby wygenerować pełne tablice fleksyjne dla języka nowopolskiego.

Pomysł ten jednak nie został zrealizowany. Pierwszym powodem jest duży nakład pracy konieczny, aby móc właściwie uzupełnić formy wraz z przegłosami i alternacjami spółgłoskowymi. Po drugie informacja, która znajduje się w *SW*, nie zawsze jest poprawna i pełna – aby móc więc zapewnić odpowiednią jakość danych, konieczna byłaby pogłębiona analiza w obrębie morfologii historycznej.

10.7.3. Słownik sprawdzania pisowni

Moduł ten miał wykorzystywać wygenerowane przed moduł opisywany powyżej do stworzenia słownika sprawdzania pisowni nowopolskiej (i nie tylko), co byłoby wsparciem dla wszelkiego rodzaju analiz tekstów dawniejszych. Przykładem formatu takiego słownika może być Hunspell, który ma postać plików tekstowych kodowanych w UTF-8, w których każdy wyraz znajduje się w osobnym wierszu⁷².

⁷¹ Działanie tego rozkładu, nazwanego w ten sposób od imienia jego odkrywcy, Vilfreda Pareto, wyjaśnione jest na stronie Wikipedii: https://pl.wikipedia.org/wiki/Rozk%C5%82ad_Pareta.

⁷² Por.: http://pwet.fr/man/linux/fichiers_speciaux/hunspell/.

Moduł ten nie został zrealizowany ze względu na brak implementacji modułu poprzedniego, od którego jest zależny, oraz ze względu na dziedziczenie opisanych wcześniej problemów merytorycznych.

10.7.4. Słownik gniazd derywacyjnych

Planowany był również moduł, którego zadaniem byłoby pogrupowanie elementów siatki hasłowej w gniazda derywacyjne. Możliwe byłoby to przy pomocy odsyłaczy użytych w artykułach hasłowych oraz słoworodów w części etymologicznej. Co prawda wykryto pewne problemy merytoryczne w tym podejściu, jednak są one marginalne, a o zarzuceniu implementacji modułu zdecydowały względy praktyczne, a mianowicie brak czasu.

10.7.5. Lista słoworodów

Moduł, znajdujący wszystkie instancje słoworodów w *SW*, został częściowo zaimplementowany, jednak na obecnym etapie jego wartość merytoryczna jest wątpliwa. Potencjalnie można by w przyszłych wersjach programu scalić go ze modulem generującym gniazda derywacyjne opisanym powyżej.

10.7.6. Mapa relacji semantycznych

Jednym z rozważanych modułów miał wykorzystywać informacje o synonimii (zarówno dla całego hasła, jak i poszczególnych znaczeń) i antonimii do zbudowania podstawowej mapy sieci semantycznej. Rozwiązanie to byłoby o wiele bardziej przydatne dopiero w przypadku opracowania całego materiału badawczego. Poza tym nierozwiązany pozostał problem ewentualnej demonstracji relacji pionowych między hasłami (hiperonimicznych i hiponimicznych).

10.7.7. Detekcja martwych odniesień

Jak zostało wcześniej zaznaczone, w *SW* pojawiać się mogą hasła główne wraz z synonimami, do których powinny prowadzić odsyłacze, ale tak się nie dzieje; możliwa jest też sytuacja odwrotna, w której istnieje odsyłacz do hasła, w którym nie ma informacji o jednostkach powiązanych. Planowany był moduł, który weryfikowałby obecność ewentualnych asymetrii w sposób definitywny – ich ręczne wykrycie jest bardzo pracochłonne – jednak ograniczenia czasowe wykluczyły jego realizację.

10.7.8. Makieta *SW* on-line

Zadaniem tego modułu miało być stworzenie próbnej wersji internetowej edycji *Słownika warszawskiego*. Od strony technicznej odbywałoby się to poprzez konwersję kodu XML na kod HTML – który jest kompatybilnym systemem znaczników – wraz ze wzbogaceniem niektórych rodzajów informacji oraz dodaniem stylów, które pozwoliłyby na estetyczną i czytelną prezentację treści *SW*. I tutaj na brak realizacji zamierzeń decydujący wpływ miał brak czasu.

10.8. Potencjalne usprawnienia

Zarówno stworzone na potrzeby niniejszej pracy narzędzia, jak i opisywany tu proces mogą być na wiele sposobów usprawnione. Poniżej omówione zostaną jednak te usprawnienia, które były oryginalnie zaplanowane, ale na których implementację nie starczyło czasu.

10.8.1. Ekstrakcja haseł ukrytych

W chwili obecnej program SWParser jest w stanie dokonać ekstrakcji haseł ukrytych z definicji i przykładów. Instancje takich haseł, z którymi program sobie nie radzi, należą raczej do przypadków granicznych i występują w nietypowych miejscach, co nie zmieni jednak faktu, że ich kompozycja strukturalna powinna pozwolić na ich właściwą ekstrakcję przy odpowiednim usprawnieniu algorytmu.

Przykładem takiego hasła jest `dno cedne`, które znajduje się wewnątrz glosy do kolokacji w hasle `Filtracyjny`:

- **Filtracyjny** przym. od **Filtracja**: Dno filtracyjne (= służące do oddzielania brzezki od masy zaciernej, in. dno cedne). Każdą filtracyjną p. **Każd**.

Funkcjonalność ta nie została zaimplementowana ze względu na rzadkie występowanie tego rodzaju struktur.

10.8.2. Detekcja glos łacińskich

Jak zostało to nakreślone w **pododziale 4.3.4.1**, niektóre definicje poprzedza ekwiwalent łaciński, który – choć wyróżniony typograficznie – nie jest wyodrębniany przez analizator składniowy, np.:

- **Fankiel**, **kla** a. **klu**, lm. **kle**, **Fankiel**, **Fankul** bot. (foeniculum) roś. z rodziny baldaszkowych. (...)
- **Figa**, **i**, lm. **i** (...) 5. w lm. a. **†Fiżki** lek. (condylomata in ano) *sadzele bolesne w siedzeniu, szyszkowiny*. Urzęd. (...)
- **Figlarek**, **rka**, lm. **rki** bot. 1. (leucocarpus) roś. z rodziny trędownikowatych. 2. p. **Figlarz**.
- **×Fosforzyca**, **y**, blm. lek. (phosphorismus) *cierpienie powstałe z przewlekłego trucia ś. fosforem*.

Dzieje się tak, ponieważ element ten strukturalnie nie różni się od definicji rozpoczętej nawiasem, por.:

- **Flejtuch**, **a**, lm. **y**, **†Flajtuch** (...) 4. a. [**Flej** m., **Fleja**, **Flufa**] (dla płci obojej) *człowiek nieporządny, niechluj, flądra, świńtuch, brudas, plucha*. (...)
- **Fluks**, **u** a. **a**, lm. **y** I. (...) 2. a. **†Flus** lek. = a) (według pojęć dawnych) *napływ nieprawidłowy soków do pewnej części ustroju: Wino z jabłek granatowych kwaśnych skuteczne jest przeciw płynieniu wewnętrznemu po członkach, to jest fluksom*. (...)

Są dwa potencjalne rozwiązania tej niespójności. Po pierwsze można dodać detekcję glos łacińskich opartą o słowniki lub cechy morfologiczne, ponieważ semantycznie glosy

łacińskie pełnią na tej pozycji inną funkcję niż komentarze w języku polskim – trudno jest jednak przewidzieć, jaki będzie stosunek nakładu pracy do rezultatów. Drugim rozwiązaniem jest potraktowanie obu sytuacji w ten sam sposób ze względu na wyróżnienie ich przez autorów taką samą typografią. Nie zostało to jednak dotychczas zrobione, ponieważ wymaga to stosunkowo dużej ingerencji w obecny kod programu.

10.8.3. Detekcja zakresu w wyrażeniach

Część wyrażen posiada rodzaj uwagi, która funkcjonuje podobnie do informacji o zakresie występującej w innych segmentach hasła, a która umieszczana jest w nawiasie, bez wyróżnienia jej w jakikolwiek inny sposób – por.:

- **Farsowy** przym. od **Farsa**: **Karmią nas farsową strawą (w teatrze).**
- **For, u, lm. y, Fora** 1. **przewaga, ustępstwo, ulga, dana komuś w początku gry a. wyścigów:** F. dawać. **Dał mu konia F. (w grze szachowej).** (...)
- **Fus, u, lm. y** (...) 2. w lm. **kawa rozgotowana, opadła na dno po ustaniu ś.:** **Przegotować fusy (dla otrzymania nanowo kawy).** (...)
- [**Futer, tru, lm. try, Fuder, Futro**] (...) 2. **podszewka, podszycie:** **Barani F. (u kapuzy).** (...)

W chwili obecnej informacja ta jest traktowana jako część wyrażenia ze względu na potencjalne interferencje reguł przechwytyjących tego rodzaju elementy z innymi elementami.

10.8.4. Kwalifikatory opcjonalne

Część kwalifikatorów, zarówno tekstowych jak i symbolicznych, jest wyróżnionych jako opcjonalne, to jest opisują hasło w sposób nieostry. W przypadku tych pierwszych stosuje się do tego nawias, w przypadku tych drugi – spójnik. Porównaj:

- **Fantazyjny (posp. i fil.)** przym. od **Fantazja** 1. **fantazji dotyczący:** (...)
- [**Fundytus**] 1. **i † z gruntu, zupełnie, całkiem, doszczętnie:** (...)

W chwili obecnej parser ignoruje informację o opcjonalności w takich przypadkach. W przyszłości planowana jest jednak jego modyfikacja, tak, aby informacja ta nie była tracona podczas analizy składniowej.

10.8.5. Pozostałe osobliwości hasłowe

Występuje także kilka przypadków, które charakteryzują się pewnymi anomaliami w strukturze i są źródłem problemów z interpretacją, przez co nie zostały właściwie oprogramowane.

Pierwszym przykładem może być pojedyncze poświadczenie użycia nawiasu w części etymologicznej w funkcji innej niż kompresja dwóch leksemów do jednej postaci:

- **xFizjonotyp** (...) <Fizjono(mja) + Gr. týpos = odbicie, typ>

W chwili obecnej parser błędnie interpretuje część **Fizjono(mja)** jako dwa leksemy: **Fizjono** oraz **Fizjonomja**, podczas gdy chodzi tu o zaczerpnięcie niesamodzielnej podstawy słowotwórczej wyrazu.

Drugim przykładem jest pojawiająca się w niektórych hasłach, wyróżniona typograficznie informacja o zmianie znaczenia wraz ze zmianą liczby – jej pozycja utrudnia jednak regularną ekstrakcję, por.:

- **Farfoce1** (...) 2. [F.] część. w lm. *szyszkowina*, w lm. *sadzele bolesne w siedzeniu*. (...)
- **Flores, u, lm. y** część. w lm. *figiel rysunkowy, figlas, zawity deseń, ozdoba w kształcie gałązki z liśćmi i kwiatami*, w lm. *esy-floresy*: (...)

Trzecim przykładem jest informacja o użyciu wyrazu w funkcji innej części mowy, niż jest przypisana (także przez pominięcie) hasłu głównego:

- **Filut, a, lm. ci, [Filurt]** (...) 2. jako przym. *chytry, przebiegły, szpakami karmiony, przemądrzały*: (...)

Osobliwości tego rodzaju można z pewnością w strukturze SW znaleźć więcej, jednak trudno tutaj wymienić wszystkie. Planowane jest jednak uwzględnienie możliwie największej liczby z nich.

10.8.6. Usprawnione parsowanie mikrostruktury

Mimo stosunkowo uporządkowanej mikrostruktury hasłowej SW, format wielu elementów nie jest całkowicie usztywniony; poza tym w treści słownika niewchodzącej w obręb materiału badawczego pojawiają się elementy obecnie przez parser nieobsługiwane, a także niekonsekwencje autorów w dalszych tomach. Dlatego w przypadku próby parsowania kolejnych hasel konieczne będą odpowiednie adaptacje.

Symulacja takich adaptacji została wykonana kolejno dla sekcji słownika obejmujących litery: **Ó**, **V**, **Y**, **X** (wraz z implementacją) oraz **I** (tylko analiza teoretyczna). Aby zapewnić poprawne parsowanie hasel na literę:

1. **Ó** konieczne były:
 - 1 modyfikacja w przechwytywaniu lemm;
 - 2 modyfikacje w separacji segmentu inicjalnego i centralnego;
 - 2 modyfikacje w parsowaniu segmentu etymologicznego.
2. **V** konieczne była 1 modyfikacja w przechwytywaniu lemm;
3. **Y** konieczna była 1 modyfikacja w separacji segmentu inicjalnego i centralnego;
4. **X** konieczne było dodanie kategorii opisu hasła
5. **I** konieczne byłoby wprowadzenie ponad 140 modyfikacji na różnym etapie parsowania i ponad 50 podmian treści na wejściu.

Na szczęście wymienione modyfikacje są kompatybilne z obecnymi rozwiązaniami. Wskazuje to jednak, że analiza składniowa całej treści SW wymagać będzie licznych dodatkowych usprawnień poza wymienionymi powyżej.

10.8.7. Rezygnacja z podziału na litery

W chwili obecnej program SWParser przetwarza dane w obrębie plików z hasłami rozpoczynającymi się na daną literę. Wynika to z przyjętej metodyki doboru materiału badawczego, i odzwierciedla to makrostrukturę samego SW (podział siatki hasłowej na sekcje literowe).

Rozwiązanie to ma jednak istotne wady, ponieważ wiele relacji między hasłami wykracza poza daną sekcję literową i nie powinny one być rozpatrywane w odłączeniu – program podczas przetwarzania nie może ich w pełni uwzględnić ze względu na ten podział, np. podczas próby ustalenia klasy wyrazu przy pomocy synonimy głównego, który rozpoczyna się na inną literę.

10.8.8. Walidacja i DTD

Jak to opisano w **rozdziale 7.3**, format XML można walidować przy pomocy kilku standardowych rozwiązań. Zamiarem piszącego te słowa było opracować odpowiedni dokument DTD lub XSD, który pozwoli na sprawdzenie poprawności generowanego pliku XML.

Nie udało się jednak tego zrealizować ze względu na ograniczenia czasowe, jednocześnie implementacja tegoż ma niski priorytet, ponieważ poprawne parsowanie wygenerowanych danych stanowi dowód na poprawność wyjściowego pliku XML.

11. Ewolucja słowników polszczyzny

Aby możliwe było w sposób koherentny mówić o miejscu *Słownika warszawskiego* wśród innych leksykonów, a także w szerszym kontekście o ich obecnej i przyszłej postaci, konieczna jest analiza zmian, które zaszły w leksykografii do momentu jego powstania. Badania w tym obszarze określa się mianem archeologii słownikowej, którą Hartmann & James (2002: 41) definiują w sposób następujący:

dictionary archaeology

The uncovering of links between different dictionaries by studying their contents, history and genetic affiliations.

archeologia słownikowa

Odkrywanie powiązań między różnymi słownikami poprzez badanie ich zawartości, historii i przynależności genetycznej.

O tym, dlaczego warto badać słowniki, wspomniano w poprzednich rozdziałach, jednakże warto powody te podkreślić przytaczając argumentację Kahane & Kahane (1992: 20):

The lexicon, with its many facets, is a mirror of its time, a document to be understood in sociolinguistic terms. It both describes and prescribes the lexis of a language according to (in Karl Jasper's phrasing) 'the intellectual situation of the time'. Lexicographic strategy, as even the small sample of the lexica outlines will indicate, reflects the attitudes of a society, as expressed in the **word**, toward the dominant problems of the ever changing here and now. His own milieu provides the specific motivations which guide the lexicographer.

Leksykon, z wieloma aspektami, jest lustrem swoich czasów, dokumentem, który należy rozumieć w kategoriach socjolingwistycznych. Zarówno opisuje, jak i przepisuje leksykę języka zgodnie z (według sformułowania Karla Jaspiera) „sytuacją intelektualną tamtych czasów”. Strategia leksykograficzna, na co wskazuje choćby niewielka próbka zarysów leksykalnych, odzwierciedla postawy społeczeństwa, tak jak są wyrażone w **słowie**, wobec dominujących problemów wiecznie zmieniającego się tu i teraz. Jego własne środowisko dostarcza konkretnych motywacji, którymi kieruje się leksykograf.

Ten obraz epoki w leksykonie jest pochodną faktu, iż leksykografia zmienia się „nie pod wpływem natury, ale społeczeństwa” (Tarp 2009: 43). Dlatego też tak istotny jest aspekt dokumentacyjno-historyczny badań leksykograficznych, na co nacisk kładzie Zgusta (1992: 14):

History as a chain of events can, however, also bring a person, society, or language to an end: then it is the task of the lexicographer, who also is in service of history to record what was or perhaps yet is, and what will not be.

Historia jako ciąg wydarzeń może jednak doprowadzić do kresu także człowieka, społeczeństwo lub język: wówczas zadaniem leksykografa, który również służy historii, jest zapisanie tego, co było, a może jeszcze jest, i czego nie będzie.

Zasadnicze jest tu założenie, że historia – rozumiana jak ciąg wydarzeń – ma wpływ na kształt leksykografii jako praktyki, ta z kolei jest formą rejestrowania historii jako szczególnej formy narracji, co w formie diagramu widocznego na **ilustracji 63** przedstawia Podhajecka (2018: 133):

Ilustracja 63. Możliwe relacje między leksykografią a historią.

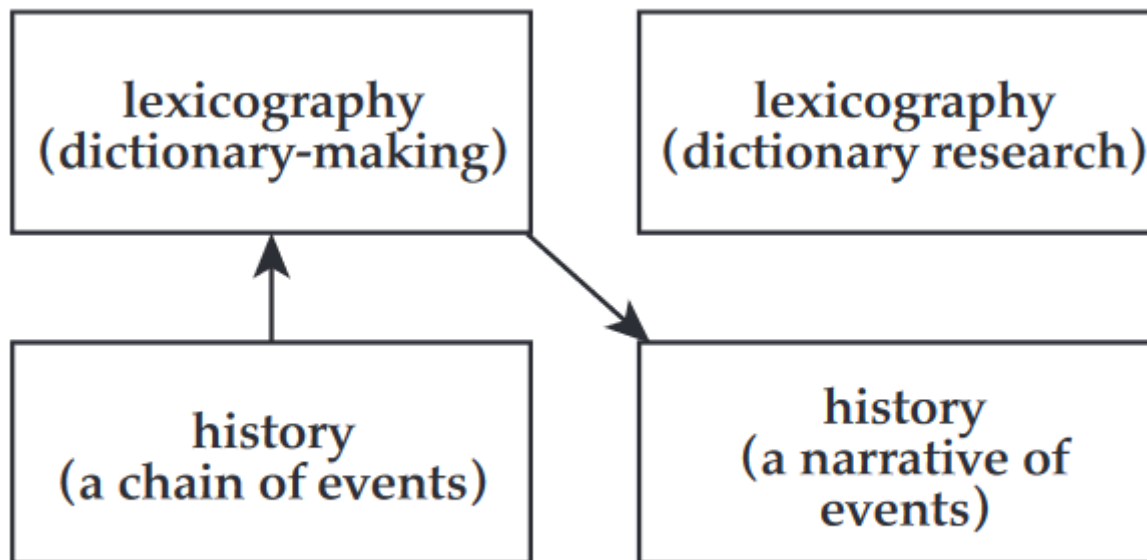


FIGURE 1 Possible Relationships between *lexicography* and *history*

Aby jednak badać przeszłość słowników, potrzebny jest odpowiedni aparat badawczy, który może przybrać różne formy. Badacze Yong & Peng (2008: 7–8) przedstawiają jednak uogólniony zarys metodologii badań historii leksykografii:

The basic methodology in the study of the history of lexicography should be a combination of sequence, evaluation, and generalization, following the diachronic sequence as the main thread of inquiry and exploration, making objective assessment of major lexicographical works, formulating theoretical generalizations, and eventually establishing a coherent framework to incorporate all the findings of such research.

Podstawową metodologią w badaniu historii leksykografii powinna być kombinacja sekwencji, wartościowania i uogólniania, podążanie za ciągiem diachronicznym jako głównym wątkiem dociekań i eksploracji, dokonywanie obiektywnej oceny najważniejszych dzieł leksykograficznych, formułowanie teoretycznych uogólnienia i ostatecznie ustanowienie spójnych ram, aby uwzględnić wszystkie wyniki takich badań.

Spośród wielu zagadnień, które bada współczesna metaleksykografia historyczna, Simpson (2009: 55) wskazuje trzy najważniejsze w jego ocenie pytania, na które dziedzina ta stara się odpowiedzieć:

1. Jak dostępność danych historycznych zmieni oblicze leksykografii historycznej?
2. Jak fakt, że leksykografia przestaje się skupiać na pojedynczych artykułach hasłowych, a zamiast tego skupia się na języku jako całości wpłynie na postrzeganie języka?

3. Jaki jest status słowników historycznych względem użytkownika języka?

Pytania te są pokrewne tym stawianym w niniejszej pracy. Zagadnień związanych z leksykografią i metaleksykografią historyczną jest jednak znacznie więcej – ich przegląd znajduje się w pracy autorstwa Colemana & McDermotta (2004), natomiast zarys historii leksykografii znaleźć można u Hanksa (2013). W niniejszym rozdziale autor skupi się jedynie na tych dziełach, które bezpośrednio dotyczą nakreślenia kontekstu dla materiału badawczego, jakim jest *Słownik warszawski*.

11.1.1. Okres przedstownikowy

Za początki leksykografii w Polsce uważa się średniowieczne glosy do tekstów – te objaśniały znaczenie pojedynczych wyrazów lub całych zdań (Doroszewski 1954: 9ff). Nie są one jednak odrębnymi dziełami, ale stanowią integralną część tekstów średniowiecznych – nie tylko poprzez samo glosowanie, ale także okazjonalne wciągania glos do kolejnych wersji tekstu. Jest to główny powód, dla którego teksty te cechuje wielowarstwowość genetyczna (Mika & Masłej 2020: 121ff).

Kolejnym etapem było wyodrębnienie się mamotrektów (inaczej mamotreptów), czyli pierwszych mini-słowniczków, które przeznaczone były dla początkujących, na co wskazuje sama nazwa – łacińskie słowo *mamotreptus* oznacza bowiem ‘piersią karmiony’. Były one niczym innym jak wypisanymi glosami – choć formalnie stanowiły osobny tekst, to nie były samodzielne, ponieważ ich kolejność była taka, w jakiej występowały w tekście źródłowym (Żmigrodzki 2009: 125).

Do czasów dzisiejszych dochowały się dwa polskie mamotrepty: tak zwany mamotrekt lubiński (pochodzący z biblioteki gminy ewangelickiej w Lubinie na Dolnym Śląsku) i mamotrekt kaliski (z biblioteki benedyktyńskiej w okolicach Kalisza), datowany na 1471 r. (Żmigrodzki 2009: 126).

11.1.2. Wybrane dawniejsze słowniki drukowane

Pierwsze słowniki języka polskiego, podobnie jak glosy i mamotrepty, miały charakter przekładowy. O ich funkcji w społeczeństwie pisze Bańko (2020: 24)

W okresie kształtowania się języków literackich nowożytnej Europy – na przykład w Polsce w XVI wieku – autorzy słowników dwujęzycznych byli nieraz zmuszeni proponować rodzimy odpowiednik jakiegoś wyrazu łacińskiego i czyniąc tak, nie ukrywali, że wypowiadają się tylko w swoim imieniu (...).

Słowniki te, w przeciwieństwie do glos i mamotrektów, są już w pełni niezależnymi dziełami, niepowiązanymi z żadnym konkretnym tekstem. Dzieł tego rodzaju było więcej, jednak autor tych słów pozwoli sobie wymienić tylko te jego zdaniem najważniejsze.

11.1.2.1. Słownik Murmeliusza

W roku 1526 w Krakowie został wydany trójjęzyczny, łacińsko-niemiecko-polski słownik zatytułowany *Dictionarius Ioannis Murellii variarum rerum tum pueris tum adultis utilissimus, cum Germanica atque Polonica interpretatione, adiecto etiam vocabulorum et capitulorum indice* (‘Słownik rozmaitości Jana Murmeliusza, niezwykle pożyteczny zarówno dla

młodzieży, jak i dla dorosłych, z przekładem niemieckim i polskim, opatrzone także spisem słów i rozdziałów’).

Jako autor jest wskazany autorstwa Jan Murmeliusz (Murmelius), faktycznym autorem części polskiej był jednak Hieronim z Wielunia (Hieronim Polyconius, zwany też Hieronimem Spiczyńskim) (Bańko et al. 2011; Gruszczyński 1997: 113).

Słownik ten ma układ rzeczowy i dzieli się na 47 rozdziałów, które zawierają łącznie 2444 hasła polskie, w skład których wchodzi 4366 wyrazów (Cygalska-Krupa 1979: 37). Przy hasłach bywa podawana informacja gramatyczna i dezambiguacja znaczeń.

Dzieło to miało wiele wydań – po 1533 r. łacińsko-niemiecko-polski układ haseł zastąpiony został łacińsko-polsko-niemieckim – według Gruszczyńskiego (1997: 120) wskazuje to „na wzrost znaczenia języka polskiego wśród wykształconych mieszkańców Krakowa”.

11.1.2.2. Słownik Mymera

Kolejnym, trójjęzycznym dziełem zawierającym język polski jest słownik Franciszka Mymera (Mymerusa) o tytule *Dictionarius trium linguarum: Latine, Teutonice et Polonice potiora vocabula continens nunc denuo pluribus in locis auctus in titulis per seriem alphabeti concinnatus tum peregrinantibus tum domi desidentibus et quibuslibet tribus loqui linguis cupientibus maxime necessarius et utilissimus* (‘Słownik trójjęzyczny, zawierający ważniejsze słowa po łacinie, po niemiecku i po polsku, teraz znowu poszerzony w wielu miejscach, zestawiony w rozdziałach w porządku alfabetycznym, niezwykle potrzebny i przydatny zarówno podróżującemu, jak i przebywającemu w domu oraz w ogóle wszystkim, którzy pragną mówić trzema językami.’).

Łączna liczba haseł wynosi 2400 – część z nich zaczerpnięta ze słownika Murmeliusza, przy czym nie podawane są informacje dodatkowe. Hasła te są podzielone na 40 rozdziałów tematycznych, wewnątrz których występuje porządek alfabetyczny (Bańko et al. 2011; Gruszczyński 1997: 125).

11.1.2.3. Słownik Mączyńskiego

Słownik *Lexicon Latino Polonicum ex optimis Latinae linguae Scriptoribus concinnatum* (‘Leksykon łacińsko-polski zestawiony z [tekstów] najlepszych pisarzy łacińskich’), którego autorem był Jan Mączyński (Ioannes Macziński) ukazał w 1564 r. w Królewcu. Dzieło to liczy ponad 1000 starannie opracowanych stron. O dziele tym Źmigrodzki (2009: 129) pisze w ten sposób:

Pod względem układu, a także wyboru materiału ilustracyjnego autor wzorował się przede wszystkim na łacińsko-niemieckim słowniku Piotra Dasypodiusza (wydanym w roku 1535). Leksykon miał — często wówczas stosowany — układ alfabetyczno-gniazdowy, tzn. we wspólnych hasłach grupowano wyrazy pokrewne słowotwórczo (np. wszystkie czasowniki, także prefiksalne, oparte na tym samym rdzeniu umieszczone były obok siebie).

Dzieło to zawiera około 30 000 wyrazów łacińskich i dokładnie 20 200 ich polskich odpowiedników podle indeksacji wykonaną przez Kuraszkiewicza (1962–1963). Czyni to z tego słownika największy zbiór łacińsko-polskich par przekładowych z XVI wieku (Bańko et al. 2011).

Lexicon... jest dziełem renesansowym i pod wieloma względami zrywającym ze średniowieczną tradycją leksykograficzną, uwzględniającym także wyrazy regionalne i potoczne. Jego autor przejawiał zresztą wysoką świadomość językową (Bańko et al. 2011).

11.1.2.4. Słownik Calepinusa

Istotnym dziełem wielojęzycznym, w którym uwzględniono język polski, jest także jedenastojęzyczny słownik autorstwa Ambrożego Calepinusa (Doroszewski 1954: 11):

Język polski uwzględniony został w opracowanym przez Calepinusa a wydanym w Bazylei w r. 1582 słowniku jedenastojęzycznym: *Dictionarium undecim linguarum*, oraz w wydanym w r. 1594 słowniku dziesięciojęzycznym tegoż autora: *Dictionarium decem linguarum*.

Oryginalny słownik zawierał tylko łacinę z dodatkami greki, które jednak było uzupełniane przez dalszych autorów o kolejne języki narodowe. Dzieło to było popularne, o czym świadczyć może fakt, że do 1779 r. było wznawiane przeszło 200 razy (Przywara 2017: 401). Gruszczyński (2011: 66) dostrzega jednak pewne wady tego dzieła:

Trzeba zaznaczyć, że polskie ekwiwalenty znajdowały się nie we wszystkich artykułach hasłowych, a poza tym zawierały wiele błędów, więc trudno uznać, że z polskiego punktu widzenia były pełnowartościowymi dziełami leksykograficznymi.

11.1.2.5. Słownik Knapiusza

W 1621 r. w Krakowie wydany został słownik polsko-łacińsko-grecki zatytułowany *Thesaurus polono-latino-graecus seu promptuarium lingua Latinae et Graece* ('Skarbiec polsko-łacińsko-grecki lub podręczne kompendium języka łacińskiego i greckiego'). Jego autorem był Grzegorz Knapski, znany także jako Knapiusz (Cnapius, Gregorius Cnapius). Ponad dwie dekady później, bo w 1643 r. ukazało się drugie wydanie słownika (Doroszewski 1954: 11–12).

Leksykon Knapskiego był największym, bo liczącym ok. 40 000–50 000 haseł, słownikiem polskim XVII w. i posiadał „liczne i obficie wykorzystywane kwalifikatory stylistyczne” (Żmigrodzki 2009: 130; Bańko et al. 2011). Stanowił on odpowiedź na zapotrzebowanie społeczne na słownik, którego część polska byłaby najistotniejsza (Puzynina 1961: 5), jednak jego warsztat leksykograficzny kontynuuje tradycję średniowieczną, językowo-encyklopedyczną i nadal poświęcającą wiele uwagi wyrazom łacińskim (w tym specjalistycznym), z drugiej strony posiadał nowoczesny układ alfabetyczny, a nie rzeczowy (Puzynina 1961: 14–15). Knapiusz jednocześnie starał się oczyścić i wzbogacić język polski, co jest odzwierciedlone w treści jego słownika (Puzynina 1961: 23).

Dzieło to jest też nietypowym słownikiem wielojęzycznym, ponieważ nie ograniczał się do podawania jedynie wyrazów polskich i odpowiedników. Poza tym autor uwzględnił różne znaczenia wyrazów, a wśród haseł znajdowały się także frazeologizmy i afiksy, dodatkowo opatrując je kwalifikatorami (Bańko et al. 2011).

11.1.2.6. Słownik Dasypodiusza

W Gdańsku w 1642 r. przez Andrzej Hünefelda wydany został słownik *Dictionarium latino-germanico-polonicum, germanico-latinum et polono-latino-germanicum*, który był spolszczoną

wersją słownika łacińsko-niemieckiego Piotra Dasypodiusa (Dasypodiusza), który ukazał się ponad pół wieku wcześniej (Doroszewski 1954: 12; Gruszczyński 2011: 58).

Na dzieło to składają się trzy części: łacińsko-niemiecko-polska, niemiecko-łacińska i polsko-łacińsko-niemiecka (Gruszczyński 2011: 58). Podstawa tego słownika wprowadziła układ alfabetyczny oraz gniazdowanie haseł wieloelementowych pod hasłem głównym (Lenartowicz-Zagrodna 2018: 231).

11.1.3. Słowniki narodowe przed *SW*

Polska leksykografia pierwszej poł. XX w. była nienormatywna i włączała wszystkie zaświadczone wyrazy, co ją wyróżniało na tle leksykografii europejskiej. Jednocześnie w strukturze haseł nie odzwierciedlano chronologii znaczeń, co również odróżniało praktykę polską od europejskiej. Novum było też pokazywanie powiązań systemowych w etymologii, co współcześnie zostało zrobione dopiero w latach 50-tych, jak zaznacza Piotrowski (1993: 585). Badacz ten wskazuje też na podobieństwa względem leksykografii rosyjskiej (idem 1993: 586):

Zasady metodologiczne słowników polskich pierwszej połowy XX w. najbardziej zbliżone były do założeń bardzo interesującego *Słownika rosyjskiego języka* pod redakcją Jakowa Grota i Aleksieja Szachmatowa, który wprowadził zasadę nienormatywności w leksykografii rosyjskiej. Pracowali nad tym słownikiem najwybitniejsi językoznawcy rosyjscy, w większości uczniowie Polaka Jana Boudouina de Courtenay.

Jednocześnie polska leksykografia specjalistyczna zaczęła się formować w XIX w. w odpowiedzi na potrzebę uporządkowania polskiego słownictwa specjalistycznego, pełniącego głównie funkcję narzędzia terminologicznego (Łukasik 2019: 179–183). To jednak, co zdefiniowało polskie słownictwo to według Horodyskiej (1986: 150) trzy dzieła wyznaczające etapy polskiej leksykografii i jednocześnie odzwierciedlające trzy epoki narodu polskiego: *SL*, *SW* i *SJPDor*.

11.1.3.1. Słownik Lindego

Na początku XIX w., a dokładniej w latach 1807–1814 został w Warszawie opublikowany *Słownik języka polskiego* Samuela Bogumiła Lindego, zwanego powszechnie od nazwiska autora *Słownikiem Lindego* (*SL*). Doroszewski (1954: 16) pisze o jego wydaniu tymi słowami: „Ukazanie się tego pomnikowego dzieła było w swoim czasie wydarzeniem naukowym w skali europejskiej. Świadczy o tym choćby fakt, że pierwsze recenzje słownika Lindego ukazały się nic w wydawnictwach polskich, ale w prasie zagranicznej”.

SL zawiera blisko 60 000 haseł obejmujących leksykę od XVI do początków XIX w. W artykułach hasłowych znajdują się też takie elementy jak kwalifikatory, warianty, ekwiwalenty w innych językach słowiańskich oraz niemieckim, a na końcu polska definicja wraz z przykładami (Piotrowski 1993: 573).

Z założenia dzieło to jest nienormatywne – *SL* miał odzwierciedlać polszczyznę, aby ją uratować, a nie ją zmieniać (Piotrowski 1993: 573). Linde chciał też zachować wyłącznie faktycznie użyte słowa, a nie potencjalizmy (Doroszewski 1954: 20), jednak Lewaszkiewicz (1980: 61–62) wskazuje, że autor ten sztucznie wymyślał słowa na podstawie innych języków słowiańskich.

Podstawą słownika jest materiał empiryczny, z którego zaczerpnięto przykłady z 400 autorów (Piotrowski 1993: 573). Linde celowo wybrał tylko część zebranego materiału do ekscerpcji, a z niego wybrał tylko podzbiór jednostek, które utworzyły siatkę hasłową jego słownika (Pudzikowski & Wawrzyńczyk 2017: 282).

Opis semantyki przestrzega porządku logicznego od znaczeń konkretnych do abstrakcyjnych (Piotrowski 1993: 573). Linde chciał oddzielić znaczenie ogólne wyrazu samodzielnego od tego, który wynikało z kontekstu – dlatego celowo podawał po kilka przykładów do znaczenia, żeby pokazać wszystkiego jego odcienie (Doroszewski 1954: 16, 22). *SL* jest dobrym źródłem informacji o poczuciu językowym użytkowników języka polskiego w przeszłości, a więc ówczesnego językowego obrazu świata. (Kurzowa 1985: 307f).

Warto też wspomnieć, że jest to pierwszy słownik porównawczy języków słowiańskich i zarazem i jeden z pierwszych naukowych dzieł leksykograficznych poświęconych językom słowiańskim w historii (Piotrowski 1993: 573).

Do dzieła przedostało się także trochę błędów i braków: Linde niekonsekwentnie stosował kwalifikatory i pomijał zapożyczenia (Żmigrodzki 2009: 139). Nie zmienia to faktu, że *SL* jest źródłem informacji o wielu językach (Bilińska 2014b); badaczka ta zademonstrowała też (Bilińska 2014a), jak można ze słownika wyekstrahować informację o konkretnym języku – w tym przypadku słoweńskim – i podaje także statystyki dotyczące informacji o poszczególnych językach słowiańskich w *SL* (idem 2014a: 8).

Bilińska (2014b: 11; 15–23) zwraca też uwagę, że w *SL* panuje też niekonsekwencja w stosowaniu skrótów i oznaczeń, np.:

- dla włoskiego: Ital., It., Włosk., Wło.;
- dla węgierskiego: Weg., Hung., Hg., Hng., Weg., Hungar., Ungar., Ung.;
- dla łużyckiego: Surab., Sorab.;
- dla czeskiego: Bh., Boh., Cz., Czesk., czesk.;

Co więcej, w dziele tym stosowana jest kompresja informacji i podawanie kilku skrótów językowych dla jednego wyrazu, a nawet pomieszczenie ich ze źródłami (Bilińska 2014b: 14).

Problemy te jednak nie przesłaniają zasług *SL*, były spotkały się z uznaniem zarówno w przeszłości (Rykaczewski 1866: II):

Słownik jaki napisał Linde, i doskonalszy jakiego dziś naród żąda, a który Linde wywoływał życzeniem na końcu swej wielkiej pracy, taki słownik jest dziełem lat i poszukiwań mozolnych.

Jak i w czasach bliskich współczesności (Doroszewski 1966: 124):

Oceniając ogólnie słownik Lindego, nie można go określić inaczej niż jako dzieło w swoim czasie imponujące i wyjątkowe, a i dziś jeszcze zachowujące swą wartość.

11.1.3.2. Słownik wileński

Drugim istotnym dziełem leksykograficznym poprzedzającym *SW* jest staraniem i kosztem Maurycego Orgelbranda *Słownik języka polskiego, obejmujący oprócz zbioru właściwie polskich, znaczną liczbę wyrazów z obcych języków polskiemu przyswojonych* autorstwa Aleksandra Zdanowicza, Michała Bohusza-Szyszki, Januarego Filipowicza, Waleriana Tomaszewicza, Floriana Czeplińskiego i Wincentego Korotyńskiego, z udziałem Bronisława Trentowskiego,

który w 1861 r. został opublikowany w Wilnie. Stąd też jego popularna nazwa – *Słownik wileński* (*SWil*), od wydawcy zwany niekiedy także *Słownikiem Orgelbranda*.

Warto podkreślić, iż jest to pierwsze leksykograficzne dzieło zbiorowe w Polsce (Żmigrodzki 2009: 141). Zdanowicz et al. (1861: „Doniesienia literackie”) wyjaśniają, że chcieli uzupełnić Lindego, dlatego prosili o pomoc szeroką publiczność – spontaniczne gromadzenie kart hasłowych pod pewnymi względami przywodzi na myśl dzisiejszy crowd-sourcing:

Wiadomo nam, że wielu literatów zwykło przy czytaniu książek, notować nowo-odkryte lub w pomnikowém dziele Lindego ominięte, a nigdzie jeszcze zbiorowo nieogłoszone wyrazy (...)

Słownik wileński jest oparty o *SL*, ale rozbudowany względem niego nie tylko liczbą haseł, ale i zakresem – poprzez włączenie słownictwa regionalnego, potocznego i specjalistycznego: na blisko 2300 stronach znajduje się bowiem ok. 110 000 haseł. Autorzy piszą o swoich staraniach ku temu, aby stworzyć dzieło możliwie najbardziej kompletne (Zdanowicz et al. 1861: 1):

Najglówniejszém przeto było założeniem naszego przedsięwzięcia, jak najskrzętniej się zająć zbieraniem wszelkich, posługujących kutemu materiałóm i sumienném ich opracowywaniem.

Oprócz obfitego materiału wyrazowego słownik ma częściowo charakter encyklopedyczny, ponieważ na jego końcu znajduje się dodatek, który obejmuje nazwy osobowe i miejscowe, nazwy miar oraz wyrazy nieregularne (Zdanowicz et al. 1861: 1–2; Doroszewski 1954: 37; Piotrowski 1993: 574).

SWil wprowadza także wiele nowych rozwiązań metodologicznych, w tym stosowanie obfitego systemu kwalifikatorów (Piotrowski 1993: 574). Wśród wad tego dzieła należy wymienić bezkrytyczne i skrótowe wykorzystanie cytowań. „Jako typ słownika jest słownikiem niewątpliwie pożytecznym” (Doroszewski 1954: 37). Dziełu temu całą monografię poświęcił też Walczak (1991).

11.1.4. Słowniki narodowe po *SW*

11.1.4.1. Słownik Doroszewskiego

Pierwszym istotnym powojennym dziełem leksykograficznym w Polsce był wydawany w Warszawie w latach 1958–1969 *Słownik języka polskiego* pod redakcją Witolda Doroszewskiego, od nazwiska którego wzięła się nazwa zwyczajowa leksykonu – *Słownik Doroszewskiego* (*SJPDor*).

Słownik ten obejmuje zwłaszcza leksykę XX w., ale uwzględnia też częściowo XIX i XVIII w. (Bartnicka 2000: 18). Doroszewski wiedział o niekompletności zasobu słownictwa, nie tylko dawniejszego, ale współczesnego – jednak jak pisze Bartnicka (2000: 23):

[D]ziś wielki Słownik języka polskiego jest dla nas jedynym źródłem prezentującym słownictwo XIX-wieczne, które w ciągu XX wieku całkowicie wyszło z użycia lub zmieniło swoje znaczenie, z pełną dokumentacją kontekstową.

Słownik Doroszewskiego zawiera też wiele przykładów użycia zaczerpniętych z literatury. Wawrzyńczyk (2010: 9) pisząc o *SJPDor* zwraca uwagę na to, że jego „unikatowa wartość

polega na tym, że daje on w wielkiej obfitości cytaty z literatury źródłowej dokładnie zlokalizowane, zatem łatwo kontrolowalne co do wierności”.

Sam materiał ekscerpcyjny również jest pokaźny. W samym dziele widnieje informacja o wykorzystaniu ok. 3 000 źródeł – jednak faktycznie jest ich ponad 3 600 (Bartnicka 2000: 18). Sam autor o kompozycji materiału źródłowego pisze w ten sposób:

Mniej więcej połowę tej liczby stanowią utwory literackie. Oprócz utworów literackich wyzyskiwane były pamiętniki, dzieła historyczne, prace naukowe z zakresu wszystkich działów humanistyki, prace matematyczne, przyrodnicze i techniczne, literatura popularnonaukowa, publicystyka, prasa, encyklopedie, słowniki (...) (Doroszewski 1968: s. XLI).

Wymaga dużego nakładu pracy, aby móc stwierdzić, z którego konkretnie wieku są niektóre źródła, ale według Bartnickiej (2000: 19):

Najogólniej można szacować, że liczba utworów ekscerpowanych, które powstały w XIX w., wynosi prawie 900 pozycji, czyli około 1/4 ogólnej liczby.

SJPDor stosuje także rozbudowany system kwalifikatorów, które są stosowane zarówno do opisu słów, jak i znaczeń. Jest w nich jednak trochę niespójności, także kategoryalnych, zwłaszcza w opisie chronologii – nie tylko poszczególne kwalifikatory mają nieostre granice, ale także część tego rodzaju informacji jest podawana w osobnej rubryce lub bezpośrednio w definicji (Bartnicka 2000: 19–22).

Dzieło Doroszewskiego jest pod wieloma względami innowacyjne i nowoczesne. *SJPDor* charakteryzują przede wszystkim koncepcje (1) behawioryzmu, (2) racjonalizmu i scjentyzmu, a także (3) encyklopedyzmu, (4) perswazyjności, i wreszcie (5) normatywizmu (Żmigrodzki 2009: 34–36). Koncepcja leksykograficzna Doroszewskiego jest kompatybilna ze współczesnym kognitywizmem, m.in. przez badanie procesów myślowych w języku i ich zakotwiczenie w biologii (Grzegorzczkova 2000: 8–9)

Przy redakcji tegoż słownika udało się także zrealizować to, co zamierzał zrobić Karłowicz, czyli informować o postępach w redakcji – organem, który to czynił, był „Poradnik Językowy” (Doroszewski 1966: 111).

Słownik Doroszewskiego posiada także wersję elektroniczną opracowaną przez PWN⁷³, jednak bardzo prostej: składa się ona z siatki haseł, pod które podpięte są obrazy z odpowiednimi artykułami – wyszukiwanie poprzez rubryki inne niż nagłówki nie jest więc możliwe.

11.1.4.2. Słownik Szymczaka

Drugim istotnym dziełem leksykograficznym wydanym po ukazaniu się *SW* jest *Słownik języka polskiego* pod redakcją Mieczysława Szymczaka, wydanym w trzech tomach w Warszawie w latach 1978–1981. Podobnie jak w przypadku *SJPDor*, nazywany jest zwyczajowo *Słownikiem Szymczaka* (*SJPSZym*) od imienia redaktora.

Pierwsze wydanie tego dzieła zostało rozprowadzone w przeszło 300 000 egzemplarzy (Bańko 2014: 96); słownik ten ze względu na swoje znaczenie i przydatność był wielokrotnie wznawiany (Bańko 2014: 97):

⁷³ Strona www: <https://doroszewski.pwn.pl/>.

Wznawiany w okresie ćwierćwiecza, stał się SJPSzym najbardziej znanym i najpopularniej używanym słownikiem polszczyzny powojennej.

*SJPS*zym cechowało nowe rozumienie słownictwa i podejście popularne – był to pierwszy słownik większych rozmiarów dla masowego odbiorcy (Bańko 2014: 97). Sam autor pisze o tym we wstępie do swojego dzieła w ten sposób (Szymczak 1978–1981):

Sposób myślenia i sposób wypowiedzania myśli wpływa bezpośrednio na jakość naszego indywidualnego życia, określa to, co nazywamy stylem życia człowieka...

Słownik Szymczaka ma 30 000 nowych haseł względem *SJPDor* i również uwzględnia wyrazy gwarowe, jednak w przeciwieństwie do poprzednika – co można uznać za regres – jest brak przykładów zdaniowych; w słowniku tym obecne są tylko kolokacje (Markowski 1986: 534, 542).

W r. 1996 opublikowano go łącznie ze *Słownikiem wyrazów obcych* PWN z 1995 r. na dysku CD w formie programu komputerowego; w latach 2001–2006 był też dostępny w wersji internetowej.

11.1.5. Słowniki nowego milenium

W latach osiemdziesiątych w Polsce zaczynała pojawiać się świadomość stosowania komputerów do leksykografii (Saloni 1986: 362), która w nowym stuleciu doprowadziła do powstania nowych dzieł, które wykorzystywały nowe medium, jakim jest postać cyfrowa, w znacznie większym stopniu, niż czyniły to dotychczasowe digitalizacje dzieł wcześniejszych.

Pruvost (2000: 188) już dwie dekady temu uprzedzał o nadejściu nowej epoki pod względem tego, jakie projekty leksykograficzne są realizowane cyfrowo:

[V]ient maintenant une autre nouvelle étape consistant à ne plus traduire informatiquement un dictionnaire papier, mais au contraire à concevoir d'emblée le dictionnaire comme un produit informatique, en prenant en compte dès l'élaboration du produit les avantages du support électronique, hypertextes et multimédia.

[N]adchodzi teraz kolejny nowy etap polegający na tym, że nie należy już tłumaczyć papierowego słownika za pomocą komputera, a wręcz przeciwnie, projektować słownik od początku jako produkt komputerowy, uwzględniając w rozwoju produktu zalety wsparcia elektronicznego, hipertekstów i multimediiów.

Zmiany te wywołały refleksje takie jak ta wyrażona przez Hanksa (2010a: 596) w odniesieniu do problemu trwałości informacji;

Is the Internet as secure and durable a medium as the printed page? Will future readers, in five hundred years' time, be able to consult an electronic dictionary on the Internet in the same way that a present-day reader can consult an old book in a library?

Czy Internet jest tak bezpiecznym i trwałym nośnikiem jak drukowana strona? Czy przyszli czytelnicy za pięćset lat będą mogli korzystać ze słownika elektronicznego w Internecie w taki sam sposób, w jaki współczesny czytelnik może zapoznać się ze starą książką w bibliotece?

Poniżej omówione będą dwa najważniejsze w ocenie piszącego te słowa dzieła, które odnalazły się w nowym medium.

11.1.5.1. Uniwersalny słownik języka polskiego

Prace nad *Uniwersalnym słownikiem języka polskiego (USJP)* pod redakcją Stanisława Dubisza rozpoczęły się w 1994 r. i trwały aż do r. 2001, natomiast sama publikacja miała miejsce dopiero dwa lata później. Początkowo przedsięwzięcie to miało być rozszerzeniem *SJPSzym*, jednak zmieniono koncepcję słownika i uwzględniono nowoczesny warsztat leksykograficzny (Latko 2015: 109)

USJP zawiera około 100 000 haseł, co przekłada się na około 150 000 jednostek leksykalnych – wraz z *SJPDor* są to dwa największe powojenne jednojęzyczne dzieła leksykograficzne poświęcone polszczyźnie ogólnej (Latko 2015: 108–109). Dzieło to zawiera słownictwo różnych rejestrów polszczyzny ogólnej – autorzy korzystali z *SJPDor*, Kartoteka PWN oraz Korpus tekstów PWN – przy czym łączy w sobie charakter synchroniczny i diachroniczny (Latko 2015: 109):

Zanotowane w *USJP* jednostki należą zasadniczo do zasobu leksykalnego polszczyzny drugiej połowy XX wieku. Opiszem objęto jednak także wybrane historyzmy, pozostające w biernym zasobie słownictwa i – wyjątkowo – archaizmy leksykalne.

Słownik ten cechuje się poza tym szeregiem innowacji warsztatowych – Latko (2015: 108–110) wymienia następujące z nich:

- włączenie większej liczby wyrazów obcych i neologizmów;
- uwzględnienie ważniejszych nazw własnych;
- odnotowanie częstszych skrótów i skrótowców;
- zrównoważoną reprezentację poszczególnych odmian języka.
- podawanie informacji z więcej niż jednego typu słownika: wyrazów obcych, ortograficznego, poprawnej polszczyzny itd.;
- objaśnianie ograniczeń frazeologicznych i pragmatycznych;
- obfity system kwalifikatorów;
- ograniczenie encyklopedyzmu i scjentyzmu w definicjach znaczeniowych;
- przytoczenie licznych i zróżnicowanych przykładów;
- informację etymologiczną o bezpośrednich źródłach zapożyczeń;
- rozbudowane tabele fleksyjne.

Piszący te słowa powyższą listę uzupełniłby o wskazanie relacji semantycznych między jednostkami oraz wyjątkowo konsekwentne traktowanie wyrazów funkcyjnych, zwłaszcza przyimków.

USJP ma rozbudowaną mikrostrukturę, w ramach której przy pomocy różnej typografii wyraźnie zaznaczone są różne typy informacji w artykule hasłowym, w tym rubryki zawierające: definicje, przykłady, frazeologia, kwalifikatory, informację gramatyczną, informację etymologiczną, pisownię, wymowę itd.

Kolejne znaczenia w haśle zasadniczo uszeregowane są od znaczeń podstawowych (mających największy zakres występowania) do znaczeń specjalistycznych i przestarzałych (Latko 2015: 111). W *USJP* wykorzystywane są definicje realnoznaczeniowe, strukturalnoznaczeniowe, strukturalne, zakresowe, synonimiczne oraz gramatyczne.

W wersji elektronicznej hasła są dodatkowo uhipertekstowane, co oferuje nowe ścieżki dostępu; wersja ta jest dodatkowo uzupełniona i rozbudowana względem papierowej.

11.1.5.2. Wielki słownik języka polskiego

Wielki słownik języka polskiego (WSJP) pod redakcją Piotra Źmigrodzkiego i przy współpracy czołowych polskich językoznawców jest jedynym omawianym tutaj słownikiem, który od samego początku powstawał w formie cyfrowej (Źmigrodzki 2017a: 358):

Jest to jedyna forma istnienia słownika, już bowiem we wstępnej fazie prac zrezygnowano z perspektywy jego publikacji książkowej, nie brano również pod uwagę upowszechniania w formie aplikacji zapisanej na nośniku trwałym. Dlatego też propozycje dotyczące metainformacji w nim umieszczonej z natury rzeczy mieć będą charakter w pewnym sensie pionierski i eksperymentalny, a w stosunku do dorobku metaleksykografii, który opiera się na analizie słowników książkowych – nowatorski.

Jest to też najnowocześniejszy z omawianych słowników, zarówno pod względem formy, jak i treści. Źmigrodzki (2018a: 17–18) wyjaśnia, że z założenia *WSJP* ma być:

- słownikiem dokumentacyjnym – wykorzystuje autentyczne materiały;
- słownikiem polszczyzny współczesnej – uwzględnia słownictwo od 1945 r. do chwili obecnej;
- słownikiem deskryptywnym – nie narzuca normy (rejestruje jednostki niepoprawne);
- słownikiem akademickim – implementuje najnowsze osiągnięcia warsztatowe;
- słownikiem „born digital” – powstał bezpośrednio w formie cyfrowej.

Słownik ten odnotowuje nie tylko pojedyncze wyrazy, ale także jednostki nieciągłe (w tym przysłowia), skróty, skrótowce i najważniejsze nazwy własne (Źmigrodzki 2018a: 22).

Od strony strukturalnej artykuły hasłowe w *WSJP* są złożone – obejmują one następujące rubryki, zależnie od opisywanej jednostki hasłowej (Źmigrodzki 2018b: 29):

- formę hasłową wyrazu;
- wariant lub warianty leksemu;
- informację o chronologizacji;
- informację o pochodzeniu;
- definicję
- kwalifikatory;
- klasyfikację tematyczną;
- relacje znaczeniowe;
- odmianę wyrazu;
- informacje syntaktyczne;
- istotne połączenia wyrazowe;
- cytaty egzemplifikujące hasło;
- informację normatywną.

Powyższa struktura odzwierciedla nowoczesną i rygorystyczną metodologię leksykograficzną, która została zastosowana w słowniku (Źmigrodzki et al. 2018). Pod względem typów zawartej informacji jest to najbardziej rozbudowany słownik polszczyzny, a także jeden z najbardziej zaawansowanych na świecie. Obecna liczba haseł tego słownika nie jest znana,

ale w 2018 r. osiągnął on liczbę 70 000 haseł⁷⁴, co czyni go dziełem średniej wielkości pod względem gęstości siatki hasłowej.

11.1.6. Miejsce *SW* wśród słowników polszczyzny

Słownik warszawski nie jest dziełem zawieszonym w próżni – po pierwsze dlatego, że czerpał z tradycji słowników dawniejszych – przede wszystkim *SL* i *SWil* – a po drugie dlatego, że ukształtował on wiele standardów w praktyce leksykograficznej, które miały silny wpływ zwłaszcza na *SJPDor* (także poprzez świadome odejście strategii autorów *SW*), a w mniejszym stopniu na słowniki późniejsze, które z tej tradycji czerpały pośrednio.

SW od *SL* i *SWil* przejął nie tylko wiele rozwiązań warsztatowych, ale także i poszczególne hasła – wraz z oryginalnymi błędami. Zarówno *SWil*, jak i *SL*, a nawet *SW* konsekwentnie cytują neologizmy ze słownika Knapiusza, które nie były używane (Walczak 1995: 85). Należy tu też pamiętać, że do utworzenia *SW* wykorzystany został materiał *SKarł*, *SWOKarł* oraz kartoteka nieukończonego słownika etnolingwistycznego, choć nie zawsze w całości – pominięte zostały przykłady użycia i źródła (Walczak 1995: 84).

Największym skokiem jakościowym było powstanie *SL*, które zyskało rangę europejską. Linde miał też duży wpływ na rozwój leksykografii słowiańskiej, zwłaszcza dzięki uwzględnieniu ekwiwalentów „innosłowiańskich” (Horodyska 2008: 257). *Słownik warszawski*, jak i zresztą inne omawiane słowniki, również miały swój wkład w tym obszarze.

Jednym z podstawowych sposobów badania rozwoju polskiego słownictwa jest analiza siatek hasłowych. Już *SWil* oferował powiększenie materiału względem *SL* – dodał wiele prowincjonalizmów i dialektyzmów, ponieważ zgodnie z duchem epoki romantyzmu jego autorzy chcieli odnowić język literacki. Słownik ten rejestrował też wiele wariantów, gdyż system terminologiczny był nieuporządkowany, a jego autorzy chcieli zbudować możliwie najbardziej reprezentatywną bazę dla ewentualnych prac standaryzacyjnych (Walczak 1995: 84–88).

SW poszedł jeszcze dalej i rejestrował całą rozciągłość stylistyczną i normatywną języka polskiego, w tym bardzo liczne zapożyczenia oraz leksykę gwarową. Podobną do *SW* jeśli chodzi o regionalizmy, koncepcję „ogólnopolskiego” słownika realizował Kazimierz Nitsch, którego uważa się za ojca polskiej dialektologii – krytykował on *SW* za niedociągnięcia w młodym jeszcze warsztacie gwaroznawstwa, natomiast sam chciał stworzyć dzieło ściśle naukowe (Okoniowa 2019: 307–310).

Powojenne słowniki wydawnictwa Artca niewiele wносиły (Piotrowski 1993: 577). Przed wojną zaczęto co prawda opracowywać słownik języka polskiego pod redakcją Tadeusza Lehra-Spławińskiego (w wydawnictwie Trzaski, Everta i Michalskiego), ale go nie dokończono; polską leksykografię zdominowała bowiem koncepcja Doroszewskiego oraz idea Suplementu do *SW*, do którego zgromadzono 60 000 haseł.

To właśnie model Doroszewskiego ukształtował dalsze dzieje leksykografii w Polsce – model, który idealizował osiągnięcia *SL*, a w dużym stopniu demonizował wady *SW*. Może to być zaskakujące, ponieważ *SJPDor* wziął swoje początki właśnie jako próba utworzeniu

⁷⁴ Informacja dostępna na stronie portalu *Nauka w Polsce*: <https://naukawpolsce.pl/aktualnosci/news-%2C31604%2Cwielki-sownik-jezyka-polskiego-pan-ma-juz-70-tys-hasel.html>.

suplementu do *SW* po utracie oryginalnej kartoteki kompilowanej do tego dzieła; *SJPDor* posiada zresztą wiele definicji inspirowanych tym słownikiem (Piotrowski 1993: 577–581).

SJPDor uważa się pierwszy nowoczesny słownik w Polsce (Miodunka 1989: 218). Leksykon ten niechętnie notował słownictwo nowe, a wyrazy poddawał zasadniczo ostrej selekcji. Jak pisze Piotrowski (1993: 578), według Doroszewskiego funkcja słowników to „doskonale nie stosunków między ludźmi dzięki właściwej interpretacji rzeczy, do których odnoszą się wyrazy” – a więc mają charakter perswazyjny (poprawnościowy) i starają się odzwierciedlić rzeczywistość obiektywną (a przynajmniej naukową), a nie społeczno-kulturową. Jak zauważa na kolejnej stronie:

Według teorii Doroszewskiego, najwłaściwszym opisem znaczenia jest definicja realnoznaczeniowa, która opisuje nie tyle znaczenie wyrazu, ile świat zewnętrzny. (Piotrowski 1993: 579)

Doroszewski (1952: 6) miał też „wielkie sny o słowniku wszechogarniającym” (Bańko 2020: 48), które wyraził m.in. w ten sposób:

W ścisłym sensie słowniki nie są opisem języka, lecz tylko słownictwa i nawet jeśli uwzględniają np. dyspozycje składniowe poszczególnych wyrazów, nie obejmują zasad ich łączenia w zdania, czyli reguł gramatyki. Niewątpliwie jednak słowniki są źródłem informacji o języku.

Dzisiaj naukowe słowniki odbiegają od metody Doroszewskiego (Piotrowski 1993: 581). Zmieniły się zwłaszcza sposoby definiowania, ale także słowniki coraz częściej tworzy się jako opis cech języka wykraczających poza leksykę. Najlepiej egzemplifikuje to *WSJP*, który ma usystematyzowany sposób definiowania jednostek i który podaje wiele informacji o fleksji i składni.

Zastosowanie w *SW* ujednoliconej grafii również było istotnym zabiegiem, który miał swój udział w kształtowaniu się standardu narodowego. W leksykografii dialektologicznej dość szybko zaczęto jednak od tego rozwiązania odchodzić – zarówno *Atlas gwar polskich* Karola Dejny, jak i *Mały atlas gwar polskich* zbierały wyrazy w ściśle określonym zapisie fonetycznym (Zdaniukiewicz 2006: 214).

Zarówno w *SL* jak i *SW* feminatywom poświęcone są odrębne artykuły hasłowe (chyba że forma wyrazowa jest w obu przypadkach identyczna), z reguły nie mają jednak własnej definicji. W słownikach tych widać spory przyrost nazw żeńskich, co odzwierciedla ruch emancypacyjny (Krysiak 2016: 85–86) – tendencja ta utrzymała się do dziś.

Co interesujące, *Słownik warszawski* do dziś pozostaje największym słownikiem ogólnym dla języka polskiego (Piotrowski 1993: 575), a do niedawna był największym słownikiem polszczyzny w ogóle. Kolejne po nim są dwutomowy *Indeks alfabetyczny wyrazów z kartoteki „Słownika gwar polskich”* pod redakcją Jerzego Reichana z 1999 r. oraz *Wielki słownik ortograficzno-fleksyjny* z 2001 r., którego autorami są Barbara Janik-Płocińska, Małgorzata Sas oraz Ryszard Turczyn (Wawrzyńczyk 2015: 8).

SW został jednak zdeklasowany przez *Słownik gramatyczny języka polskiego (SGJP)* – choć jego pierwsze wydanie zawierało zaledwie 245 000 jednostek, to wersja elektroniczna z października 2021. haseł tych notuje już ponad 345 000, w tym nazwy własne, wybrane związki wyrazowe, ale także sporo „mechanicznych” derywatów, takich jak np. „niezmiennobarwistość” (Woliński & Kieraś 2017: 85). Należy tu zaznaczyć, że słownik ten czerpał

słownictwo także z *SW*. Nowoczesne słowniki ogólne stronią jednak od stosowania potencjalizmów i form niepoświadczonych empirycznie.

Jeśli chodzi o większe jednostki, to *SW* nie jest nowatorski w kwestii traktowania związków frazeologicznych i „przynależy jeszcze do tak zwanego praktycznego okresu wyodrębniania frazeologii” (Piela 2018: 145–146). Z omawianych słowników pierwszym, które frazeologizmy traktował w miarę systematycznie jest *USJP*, jednak i tutaj stanowią one część hasła, podobnie jak w *SW*, jednak posiadają bardziej rozbudowany opis.

SW był też silnie krytykowany za niewystarczającą egzemplifikację; widoczne to było zwłaszcza w przypadku wyrazów gwarowych, ale nie tylko; siatka hasłowa *SJPDor* również była słabo wsparta cytatami i użyciem. Z problemem tym systematycznie poradził sobie dopiero *WSJP*, który czerpie przykłady z Korpusu Języka Polskiego PWN.

Jednym z elementów, który zmienił się najbardziej w kolejnych słownikach polszczyzny, są kwalifikatory. Według Jankowiak (2013: 609) najistotniejszą rolę w ustaleniu się polskich kwalifikatorów miały trzy słowniki: *Knapiusza*, *Troca* i *Słownik wileński*.

Co prawda *Lexicon Latino-Polonicum* Jana Mączyńskiego jako pierwszy zawierał kwalifikatory, choć te były bardzo ograniczone – postęp natomiast był w przypadku *Thesaurusa... Knapiusza*, wprowadzającego liczne kwalifikatory. Te rozwinął Troc w *Nowym dykcyonarzu...* sprowadzając kwalifikatory do pojedynczych wyrazów w formie skrótów, które były umieszczane po hasłach lub przed znaczeniami, wyróżniając je typograficznie, a także wprowadzając nowe kwalifikatory specjalistyczne.

Linde natomiast, niewspółmiernie do rangi swojego dzieła, reprezentował regres w stosunku do słownika Troca. Głównym problemem był brak konsekwencji w ich stosowaniu – Linde wymyślał kwalifikatory ad hoc, przez co wiele z nich ma liczne alternatywne formy (Bilińska 2013: 129). Największy przełom miał miejsce w przypadku *SWil*, który wprowadził rozbudowany i nowoczesny system kwalifikatorów, i choć jest ich mniej, to on również posiada warianty w „odautorskim” systemie skrótów (2020: 331–335) przypisała także do poszczególnych ich klas. System ten uściślił *SW* (Walczak 1988: 415–421), jednak system ten był poprawiany przez kolejne słowniki (*SJPDor* i późniejsze), które pozbyły się wariantów i uprecyzowały poszczególne kwalifikatory.

Systematyczne porównanie kwalifikatorów w *SL*, *SWil*, *SW* i *SJPDor* znajduje się w artykule Pielu (2020: 60). Również Mariak podaje w swojej pracy statystyki dotyczące wyrazów opatrzonych kwalifikatorem x oraz ich pogłębioną analizę, także względem siatek pojęciowych słowników wcześniejszych (Mariak 1991/1992: 147–158). Zasadniczo kompatybilna systematyka została zaproponowana przez piszącego te słowa w **podrozdziale 5.2.3**, jednak w pracy tej kwalifikatory środowiskowe i stylistyczne były potraktowane zbiorczo.

Omawiane słowniki – oprócz *SL* – posiadają także bogatą informację fleksyjną. *SW* nie wniósł żadnych zmian w prezentowaniu informacji fleksyjnej względem *SWil*, to Karłowicz chciał wprowadzić wzory, co jednak zaimplementował w swoim słowniku Doroszewski (Walczak 1995: 82), co jest – z pewnymi modyfikacjami – kontynuowane w *USJP* i *WSJP*.

Wiele zmieniło się – chociaż w sposób nieoczywisty – w kwestii normatywizmu w słownikach. *SW* był dziełem typowo pozytywistycznym – kładł nacisk na cele dydaktyczne, kompletność materiału i zbieranie całego materiału bez oceniania go (Piotrowski 1993: 576). Późniejsze słowniki, zwłaszcza *SJPDor* i *SJPSzym*, były już normatywne, co mogłoby

wskazywać na osiągnięcie pewnego etapu rozwoju języka i świadomości narodowej. Trend do krytycznej oceny leksemów się rozluźnił wraz z powstaniem *USJP* i *WSJP*, gdzie słownictwo notowane głównie na podstawie frekwencji, i są co najwyżej opatrzone uwagami co do kwestii normatywnych.

Nie jest to jednak jedyna taka sytuacja – encyklopedyzm, od którego odchodzono w *SW*, a który częściowo powrócił w słownikach późniejszych: *SJPDor*, *USJP* i *WSJP* notują niektóre nazwy własne. Nie inaczej było ze scjentyzmem, który był w *SW* krytykowany przez Doroszewskiego, jednak definicje scjentyistyczne zaczęły się znów pojawiać w *USJP*.

Słownik warszawski z pewnością zostawił swój ślad na kartach polskiej leksykografii – nie jako dzieło idealne ani apogeum warsztatu słownikarskiego, ale jako dzieło unikalne i wprowadzające wiele innowacji warsztatowych, który wyznaczyły standardy w dziedzinie w dużym stopniu aktualne i dziś. Leksykon ten zawiera nie tylko odbicie języka wieków ubiegłych, lecz odzwierciedla także świadomość językową odradzającego się narodu. Jednocześnie dzieło to jest relatywnie słabo zbadane w porównaniu ze słownikami zarówno dawniejszymi, jak i bliższymi współczesności.

12. Praktyczna aplikacja wersji cyfrowej

Przeszło dwie i pół dekady temu Bailey (1986: 123) przewidywał, że na rozwój leksykografii będą miały wpływ dwa fenomeny: (1) rozwój samej leksykografii oraz dziedzin pokrewnych, oraz (2) zastosowanie rozwiązań komputerowych.

W niniejszej pracy skupiono się na tym drugim, w ramach którego dokonano transformacji treści *Słownika warszawskiego* do postaci cyfrowej, co wiąże się ze zmianą paradygmatu z leksykografii tradycyjnej i historycznej na leksykografię komputerową (elektroniczną, cyfrową) – to z kolei pozwala na reinterpretację tegoż dzieła i wykorzystanie go na nowe sposoby. Komputery pełnią w tym procesie dwie role: miejsca przechowania dużych baz danych (ang. *container of huge data banks*) oraz narzędzia do transformacji informacji (ang. *information-rearranging device*) (Zgusta 1986: 145; Bailey 1986: 126).

Zasób taki ma wielorakie zastosowania, ponieważ potencjał badawczy źródła, jakim jest *SW*, pozostał do tej pory niewyczerpany (Okoniowa 2013: 256). Cyfryzacja *Słownika warszawskiego* jest pod wieloma względami analogiczna względem elektronicznych wersji *SWil* i *SL*.

12.1. Słownik elektroniczny

Już trzy dekady temu Gruszczyński et al. (1981: 13) pisali o niemożności opracowania siatki haseł na potrzeby badawcze bez zastosowania komputera, natomiast w 1997 Schröder pisał o perspektywach dla leksykografii w dobie słowników elektronicznych, dziś natomiast niepodważalnym jest fakt, że słowniki „znalazły swoją idealną platformę w medium internetowym” (Rundell 2012: 15).

Słowniki elektroniczne mogą być przeznaczone dla ludzi lub do odczytu maszynowego (Dziemianko 2018: 663). W tym pierwszym przypadku definiuje się jako (De Schryver 2003: 146):

[C]ollections of structured electronic data that can be accessed with multiple tools, enhanced with a wide range of functionalities, and used in various environments.

Zbiory ustrukturyzowanych danych elektronicznych, do których można uzyskać dostęp za pomocą wielu narzędzi, wzbogaconych o szeroki zakres funkcji i używanych w różnych środowiskach.

Powstanie słowników elektronicznych jest wynikiem rewolucji informacyjnej. Bailey (1986: 129) wymienia dwa problemy, które początkowo towarzyszyły tejże rewolucji: wysoki koszt dostępu i trudność użytkowania. Oba te problemy jednak – zgodnie z jego przewidywaniami – zostały rozwiązane przez rozwój technologii: przechowywanie i przetwarzanie danych jest coraz tańsze, a interfejsy coraz prostsze w użyciu.

Cyfrowa postać słowników jest ważna nie tylko efektywniejszej dydaktyki, ale także z perspektywy nowych zastosowań naukowych – pozwala ona na przeprowadzenie różnego

rodzaju operacji na swojej treści, które teoretycznie można wykonać także w przypadku słowników drukowanych, jednak jest to często praktycznie nierealizowalne, np. w przypadku statystycznej klasyfikacji jednostek (Fiedoruszkow 2009: 28).

Zaletą słowników elektronicznych jest też ich trwałość, chociaż, jak zauważa Szafran (2004: 90):

Słownik elektroniczny jest też zapewne trwalszy niż tradycyjny, choć – ze względu na relatywnie krótki okres stosowania komputerów – nie można tego sądu formułować kategorycznie.

Nie oznacza to jednak, że leksykografia zakończyła proces swojej transformacji, na co zwraca uwagę Tarp (2018a: 30):

Today, lexicography is in a process of adapting to digital media. The introduction of new disruptive computer and information technologies have opened completely new horizons to this 4,000-year-old discipline.

Obecnie leksykografia jest w trakcie adaptacji do mediów cyfrowych. Wprowadzenie nowych przełomowych⁷⁵ technologii komputerowych i informatycznych otworzyło zupełnie nowe horyzonty dla tej liczącej 4000 lat dyscypliny.

Wyróżnić należy dwa zasadnicze rodzaje słowników elektronicznych, o których pisze (Karpinska 2020: 84):

Based on their origin, electronic dictionaries can also be divided in two types – dictionaries transferred from existent print dictionaries or digitized print dictionaries and dictionaries compiled for the electronic medium or purpose-built electronic dictionaries (Svensén, 2009: 438–439).

Ze względu na ich pochodzenie, słowniki elektroniczne można również podzielić na dwa rodzaje – słowniki przeniesione z istniejących słowników drukowanych lub zdigitalizowane słowniki drukowane, oraz słowniki skompilowane dla nośnika elektronicznego lub słowniki elektroniczne specjalnie skonstruowane (Svensén, 2009: 438–439).

To rozróżnienie niekiedy miewa swoje odzwierciedlenie w terminologii: Campos Martín (2009: 325) w języku francuskim stosuje na słowniki elektroniczne termin *dictionnaire informatique* ('słownik informatyczny') w odróżnieniu od *dictionnaire informatisé* ('słownik zinformatyizowany').

Leksykografia cyfrowa jest silnie związana z językoznawstwem komputerowym, na co wskazuje choćby liczba wariantów tego ostatniego pojęcia – Hartmann & James (2002) wymieniają następujące terminy: *computational lexicography* (idem 2002: 26), zwana także *computer-assisted lexicography*, *computer-based lexicography*, *computer-enhanced lexicography* (idem), *machine-aided lexicography*, *machine lexicography* (idem: 90), *automatic lexicography* (idem: 11) oraz *electronic lexicography* (idem: 48).

Relacja ta wynika z faktu, że leksykografia to dziedzina poświęcona przechowaniu informacji (a raczej dość specyficznym jej wycinkiem), zarządzania nią i tworzeniem

⁷⁵ Trudno tutaj oddać znaczenie angielskiego wyrazu *disruptive*, który można by było opisać jako „zmieniający *status quo*” czy „wprowadzający zamieszanie”.

odpowiednich struktur dostępowych do tejże informacji. W sposób następujący sformułował to Tarp (2009: 45):

The common nature of all lexicographically relevant needs is that they are needs for information which may subsequently be used for a huge variety of purposes.

Cechą wspólną wszystkich istotnych leksykograficznie potrzeb jest to, że stanowią one zapotrzebowanie na informacje, które mogą być następnie wykorzystane do bardzo różnych celów.

Właściwości słownika drukowanego, który został przystosowany do formy cyfrowej Debus-Gregor & Heid (2013: 1002) opisują jako mające cechy „gdzieś «pomiędzy» słownikiem papierowym a słownikiem powstałym wyłącznie jako narzędzie elektroniczne”. Karpinska (2020: 84–85) zauważa przy tym, że leksykony te mogą albo przybrać postać niemal identyczną z oryginałem, albo w pełni wykorzystywać zalety formy cyfrowej, w tym nawet – hipotetycznie – adaptować się do potrzeb poszczególnych użytkowników. Typologię tego rodzaju produktów w swojej pracy opisuje Tarp (2011).

Aby osiągnąć bardziej zaawansowane stopnie cyfryzacji – zwłaszcza umożliwiające lepsze rozumienie treści przez komputery – konieczne jest dokonanie konwersji struktury słownika. To jednak nie zawsze jest proste, na co uwagę zwraca Heid (2014: 26):

Not only formal conversions are necessary, but in particular a reinterpretation of those indications which serve to formulate equivalence conditions (...). For the use in NLP programmes, such indications must be mapped onto well-circumscribed classes of items.

Niezbędne są nie tylko konwersje formalne, ale w szczególności reinterpretacja tych oznaczeń, które służą do formułowania warunków równoważności (...). Do wykorzystania w programach NLP, takie wskazania muszą być mapowane na dobrze opisane klasy jednostek.

Konwersja może oznaczać także jednoczesną transformację z jednego rodzaju słownika na inny, np. poprzez odwrócenie kierunku przekładowego w słownikach bilingwalnych (Hartmann & James 2002: 29), a także różnego rodzaju modyfikacje treści.

Zalety tego rodzaju zabiegów oraz innowacje związane z leksykografią cyfrową podsumowuje Karpinska (2020: 86f):

1. zniesienie ograniczenia miejsca – a tym samym potrzeby kondensacji wiadomości, przez co można używać pełnych form i większego zakresu danych;
2. zwiększona dostępność informacji – nowe medium oferuje więcej równoległych ścieżek dostępu;
3. hipertekstowość – hiperłącza wewnętrzne i zewnętrzne pozwalają łączyć ze sobą różne treści lub ich fragmenty;
4. adaptacja interfejsu – dzięki oddzieleniu formy od treści można łatwo modyfikować interfejs do użytkowników;
5. powiązanie publikacji – możliwe jest połączenie różnego rodzaju zasobów językowych;
6. integracja korpusów – bezpośrednio do opisu leksykograficznego;

7. multimodalność – obejmująca nie tylko tekst i obrazy, ale także nagrania audio i video;
8. zaangażowanie użytkowników – poprzez zgłaszanie błędów i wkład merytoryczny.

Aby było to możliwe, konieczne jest właściwe przechowywanie danych: czy to w formie bazy danych, czy – jak jest to czynione w niniejszej tezie – w ramach struktury XML. Należy przy tym pamiętać, że oba te rozwiązania nie są tożsame ze słownikiem, a jedynie formą przechowania danych, a na podstawie pojedynczej bazy danych lub pliku ze znacznikami można wygenerować więcej niż jeden słownik (Bergenholtz & Nielsen 2013), co jest możliwe dzięki oddzieleniu informacji o słowniku od informacji o języku w nim zawartej. Typowe elementy składowe w słownikach opisują Hartmann & James (2002: XI):

The entry database includes the headwords, definitions and other information categories (e.g. pronunciation, syntactic functions, usage labels). This database also includes formatting codes, page layout instructions, cross-references and links across entries, which enable rapid updating of information, and easy extraction of specific entries or parts of entries from the database for, for example, the production of an abridged or compact edition of a larger dictionary.

Baza danych artykułów hasłowych zawiera wyrazy hasłowe, definicje i inne kategorie informacji (np. wymowę, funkcje składniowe, etykiety użytkownika). Zawiera ona również kody formatowania, instrukcje dotyczące układu strony, odsyłacze i łącza między wpisami, które umożliwiają szybką aktualizację informacji i łatwe wyodrębnianie określonych wpisów lub części wpisów z bazy danych, na przykład w celu stworzenia skróconego lub zwięzłego wydania większego słownika.

Jednym z najwcześniejszych przykładów informatyzacji słowników można wskazać opracowaną przez Lawrence'a Urdgana zaawansowaną bazę danych dla swojego słownika w 1959 r., która pozwalała na przechowanie złożonej struktury hasłowej, jednak maszyny wykorzystywane do składu tekstu nie potrafiły tego potencjału w pełni zrealizować (Bailey 1986: 125f).

Ważniejsze digitalizacje europejskich słowników papierowych zostały zrealizowane np. dla języka francuskiego w ramach projektów CNRS oraz Trésor de la langue française (Bailey 1986: 134) czy dla języka angielskiego w ramach procesu „elektronifikacji” Oxford English Dictionary, która nastąpiła w 1989 r. (Brewer 2004). Na rodzimym polu warto wymienić (1) elektroniczną wersję *SWi*⁷⁶, gdzie dostępny jest przeszukiwalny tekst oraz oryginalny obraz strony, (2) elektroniczne indeksy *SL*⁷⁷ (Bień 2015), czy (3) *Indeks alfabetyczny wyrazów z kartoteki Słownika gwar polskich* pod red. Jerzego Reichana (1999). Wśród słowników tych wciąż brakuje *Słownika warszawskiego*.

12.1.1. Układ i prezentacja

Medium cyfrowe nie tylko pozwala na nowy układ i prezentację treści słowników, ale wewnętrzna zmiana układu danych jest wręcz konieczna, o czym pisze Heid (2014: 25):

⁷⁶ Strona projektu: <https://eswil.ijp.pan.pl/?str=bibliografia>.

⁷⁷ Strona projektu: <https://githubmemory.com/repo/jsbien/iLindeCSV>.

[T]he use of computational methods has affected (i) the (internal) representation of lexical data for electronic dictionaries, (ii) the presentation of data in dictionaries, and (iii) the processes of dictionary production and of the preparation of raw material for dictionaries.

[W]ykorzystanie metod obliczeniowych wpłynęło na (i) (wewnętrzną) reprezentację danych leksykalnych dla słowników elektronicznych, (ii) prezentację danych w słownikach oraz (iii) procesy tworzenia słowników i przygotowywania surowych materiałów na słowniki.

Mimo nowego paradygmatu leksykografowie nadal przenoszą teorię budowy słownika drukowanego na platformę elektroniczną, która jest relewantna tylko częściowo (Gouws 2018b: 234). Badacz ten podkreśla:

The transition from printed to online dictionaries demands a re-assessment of and adaptations to dictionary structures. (Gouws 2018b: 225)

Przejście od słowników drukowanych do internetowych wymaga ponownej oceny struktur słownikowych i ich dostosowania do nich.

A także:

[M]etalexigraphers should re-assess dictionary structures in order to negotiate the options made possible by the new medium. (Gouws 2018b: 235)

[M]etaleksykografowie powinni dokonać ponownej oceny struktur słownikowych w celu wynegocjowania opcji, które umożliwia nowe medium.

Dzięki odpowiedniemu wyróżnieniu struktur głębokich słownika możliwe jest właściwe oddzielenie danych od formy publikacji, o czym pisze Richard (2005: 107–108). Autor ten opisuje to zjawisko w sposób następujący:

Układ słownika ma zasadnicze znaczenie dla jego funkcji. Czytelnik będzie oczekiwał, że znajdzie informacje, których szuka, bez konieczności czytania całego artykułu hasłowego. Z tego powodu nadal uważamy, że dla leksykografa istotne jest, aby miał ostatnie słowo w sprawie układu swojego słownika. Oznacza to, że leksykograf musi mieć przynajmniej elementarną znajomość typografii i układu. To, czego nie potrzebuje, to podejmowanie ostatecznych decyzji, w jakich mediach słownik jest publikowany lub jakie oznaczenia leksykograficzne są potrzebne dla produktu końcowego. Decyzje te są w rzeczywistości lepiej podejmowane pod koniec tworzenia słownika, gdy dostępna jest odpowiednia ilość danych. (...) Tworzenie słownika za pomocą bazy danych pozwala leksykografowi skoncentrować się na funkcji i treści słownika. Publikacja staje się wtedy kwestią naciśnięcia właściwego przycisku.

Dzięki odpowiedniej strukturze i wykorzystaniu generatora kodu układ słownika można modyfikować globalnie (Richard 2005: 103), w sposób zbliżony do wykorzystywania stylów do edycji tekstu. Badacz (Richard 2005: 101) ujmuje to tak: „In fact, it is possible to work on a dictionary without even having an idea of the final layout” (‘Właściwie możliwe jest podjęcie prac nad słownikiem nie mając nawet pomysłu na jego końcowy układ’). Kosztem zastosowania takiego rozwiązania jest jednak zwiększony poziom złożoności procesu decyzyjnego na początku projektu.

Słowniki elektroniczne współcześnie najczęściej przyjmują postać stron internetowych. Granger (2012: 2) opisuje słowniki internetowe jako:

[T]estimony that the innovations afforded by the electronic medium can radically transform every facet of dictionary design and use.

[D]owód na to, że innowacje oferowane przez medium elektroniczne mogą radykalnie zmienić każdy aspekt projektowania i użytkowania słownika.

Richard (2005: 105) opisuje, że wykorzystanie wszystkich danych redaktorskich *DanskOrdbogen* i zmiana ich prezentacji na potrzeby wydania internetowego zwiększyłaby jego objętość co najmniej dwukrotnie. Dzieje się tak, ponieważ treść słowników bywa w dużym stopniu kodowana ze względu na ograniczenia nośników fizycznych – to z kolei może części użytkowników utrudniać korzystanie z nich (Gasiglia 2013a: 9).

Głównym sposobem na pokonanie bariery, którą posiadają słowniki papierowe, jest atomizacja elementów treści zasadniczej podczas digitalizacji (Dummer & Michaelis & Schlaefel 1998: 199). Elementy takie można przedstawiać w różnych konfiguracjach. Prowadzi to w kierunku indywidualizacji formy zależnie od potrzeb danego użytkownika (Tarp 2009: 61; Dziemianko 2018: 666–667). Heid (2014: 28) objaśnia, że wiąże się to z odfiltrowaniem danych, które są w danym kontekście potrzebne, oraz wykorzystaniem odpowiednich narzędzi adaptacyjnych, które pozwalają na spersonalizowaną prezentację.

Choć w danym kontekście spersonalizowany słownik staje się jednofunkcyjny, to zasadniczo sam fakt możliwości jego indywidualizacji i adaptacji jest dowodem na jego wielofunkcyjność, którą Hartmann & James (2002: 96) definiują w sposób następujący:

multifunctional dictionary

A type of REFERENCE WORK designed to meet several needs of the user simultaneously. ELECTRONIC DICTIONARY systems allow the combination of information types so that several functions can be fulfilled, e.g. looking up definitions for decoding activities (reading), synonyms for encoding activities (writing), and factual knowledge for study and practice.

wielofunkcyjny słownik

Rodzaj PUBLIKACJI ENCYKLOPEDIYCZNEJ przeznaczony do jednoczesnego zaspokojenia kilku potrzeb użytkownika. Systemy SŁOWNIKÓW ELEKTRONICZNYCH pozwalają na łączenie typów informacji tak, aby spełniać kilka funkcji, m.in. wyszukiwanie definicji do dekodowania (czytanie), synonimów do kodowania (pisanie) oraz wiedzy faktograficznej do nauki i praktyki.

Słowniki elektroniczne, zwłaszcza internetowe, mają jeszcze jedną bardzo istotną zaletę: mają charakter dynamiczny, tj. mogą być na bieżąco aktualizowane, jako że – w przeciwieństwie do wydań tradycyjnych – trudno wyodrębnić obiektywny punkt ich publikacji w zamkniętej formie. Schröder (1997: 60) w związku z tym dzieli słowniki na dwie grupy ze względu na ich finalność: *Ausbauwörterbuch* ('słownik w budowie') oraz *Abschlusswörterbuch* ('słownik ukończony'). Rozwiązanie pierwsze pozwala zredukować opóźnienie między zredagowaniem poprawki a jej dostępnością dla użytkownika, który w przeciwnym razie musiałby czekać, aż zebranych zostanie wystarczająco dużo uzupełnień, aby zasadne było wydanie suplementu.

Dzięki opracowaniu *SW* jako ustrukturyzowanego słownika w formie XML z odłączeniem logiki informacji od jej postaci byłoby możliwe nadanie mu niemal dowolnej formy prezentacji, dopasowanej pod potrzeby konkretnego rodzaju użytkowników.

12.1.2. Nowe ścieżki dostępu

Już ćwierć wieku temu przewidywano, że słowniki XXI wieku będzie charakteryzować większa przeszukiwalność, a jej prędkość, trafność i wydajność będą nieporównywalne (Bailey 1986: 127). Stało się to możliwe dzięki odpowiednim strukturom danych i właściwie zaprojektowanym interfejsom, które pozwoliły na powstanie nowych ścieżek dostępu, czyli sposobów dotarcia do danej informacji.

Nowe ścieżki dostępu mogą zostać utworzone także z pomocą metod analogowych, czego przykład opisuje Doroszewski (1954: 11):

W pracowni Instytutu Badań Literackich zajętej obecnie przygotowaniem słownika języka polskiego wieku XVI słownik [Lexicon latino-polonicum] Mączyńskiego został rozkartkowany i „odwrócony”, to znaczy, że ułożony został odpowiadający łacińsko-polskiemu słownikowi Mączyńskiego słownik polsko-łaciński. Była to bardzo dobra myśl.

Jak widać, już samo odwrócenie kolejności par przekładowych sprawia, że powstają nowe ścieżki dostępu. Dzięki rozwiązaniom komputerowym można jednak osiągnąć znacznie więcej. Centralnym zagadnieniem jest tutaj zmiana paradygmatu od przeszukiwania słowników (ang. *lookup*, niem. *Nachschlag*) do ich zapytywania (ang. *query*, niem. *Abfrage*) jak bazy danych, co pozwala odejść od linearnego układu danych (Schröder 1997: 59).

Różnie sformułowane zapytania mogą być w swojej naturze semajologiczne lub onomajologiczne – de facto mowa tu o wyszukiwaniu po słowach lub po znaczeniach (a także i innych obiektach lub parametrach). Aby użytkownik mógł jednak łatwo znaleźć to, czego szuka, należy najpierw określić jego potrzeby, następnie wyodrębnić odpowiednie klasy informacji, a następnie zoptymalizować strukturę bazy danych pod kątem możliwych i prawdopodobnych zapytań (Schröder 1997: 60). Problem ten porusza także Heid (2014: 27):

Devices for search and filtering allow users to query their dictionary in ways that would require an unmanageable effort if they had to be done manually, on a paper dictionary. (...) [T]o be searchable, these properties must be identified as in dividual types of lexicographic data, and they must be represented internally in a way that corresponds to the attribute-value-based approach described above.

Urządzenia do wyszukiwania i filtrowania umożliwiają użytkownikom przeszukiwanie ich słownika w sposób, który wymagałby niemożliwego do opanowania wysiłku, gdyby musiały być wykonywane ręcznie, w słowniku papierowym. (...) [A]by można było je przeszukiwać, właściwości te muszą być identyfikowane jak w poszczególnych typach danych leksykograficznych i muszą być reprezentowane wewnętrznie w sposób, który odpowiada opisanemu powyżej podejściu opartemu na wartości atrybutów.

Te zasadnicze usprawnienia w korzystaniu ze słowników rzadko są jednak realizowane:

Many dictionary portals provide additional features which might be of interest to the user. These features generally do not assist the user in the basic task of dictionaries, which is finding the meaning or translation of words. (Boelhouwer & Dykstra & Sijens 2018: 756)

Wiele portali słownikowych udostępnia dodatkowe funkcje, które mogą zainteresować użytkownika. Cechy te na ogół nie wspierają użytkownika w podstawowym zadaniu słowników, jakim jest odnajdywanie znaczenia lub tłumaczenie słów.

Optymalne ścieżki dostępu do danych są również istotne dla zaawansowanych użytkowników, w tym językoznawców i samych leksykografów pracujących nad słownikami – czas zaoszczędzony na ręcznej konkordancji badacz może bowiem poświęcić na inne zadania (Heid 2014: 28).

Dzięki opracowaniu pełnej treści *SW* w formie ustrukturyzowanej możliwe by było konstruowanie zapytań przy pomocy wielu różnych parametrów – nie tylko treści nagłówków artykułów hasłowych, ale dowolnej właściwości wyróżnionej strukturalnie, jak właściwości morfosyntaktyczne, sposób derywacji, pochodzenie i etymologia, opis znaczeń, użyte przykłady, kwalifikatory i innych.

Możliwa by była również propozycja Piotra Wierzchoń (2010a: 58–61), polegająca na utworzeniu gniazd derywacyjnych nie w kontekście synchronicznym, jak przeważnie ma to miejsce, ale także diachronicznym. Wierzchoń w ramach dziedziny, którą określa lingwochronografią definiuje taką jednostkę opisu jako „gniazdo chronologiczne zbudowane z form pochodnych” (Wierzchoń 2010a: 61).

12.1.3. Hipertekstowość i integracja zasobów

Poszczególne artykuły hasłowe – a także ich elementy składowe – stanowią więc odrębne jednostki, co skłoniło Piotrowskiego (2010: 52) do stwierdzenia, że „słownik to zbiór dyskretnych elementów tekstu”. Na ten hierarchiczny charakter tekstowości słownika zwraca uwagę również Wiegand (2000), dla którego słownik to tekst, ale też nośnik tekstów, z których ów słownik się składa. Można mówić o intertekstualności w słowniku: powiązania są obecne nie tylko między elementami hasła, a także odnoszenie hasła do większych struktur, takich jak treści zewnętrzne (Wiegand et al. 2013: 57–58).

W przypadku zakodowania wewnętrznych i zewnętrznych powiązań między elementami słownika w sposób możliwy do odczytania przez komputery mówi się o hipertekstowości. Słownik hipertekstowy nie może automatycznie podpinać informacji, które nie są eksplicytnie zakodowane – chyba że zastosowane są odpowiednie algorytmy, których trafność jest jednak różna (Gasiglia 2013b: 183).

Do tej pory metaleksykografowie skupiali się jednak na opisie formy książkowej słownika, a nie jego struktury wyabstrahowanej od medium, dlatego też ten aspekt formy cyfrowej był nierzadko pomijany, na co uwagę zwraca Żmigrodzki (2017a: 355):

Jak łatwo dostrzec, podział ten dotyczy słownika książkowego, w dodatku uwzględnia zarówno elementy odautorskie, jak i pochodzące od wydawnictwa (np. karty tytułowe). W dobie rozwoju leksykografii elektronicznej traci on aktualność, gdyż

powszechne obecnie słowniki komputerowe, a zwłaszcza internetowe, nie mają struktury linearnej, tylko hipertekstową.

Zalet hipertekstowości jest wiele – funkcje hipertekstowych e-tekstów wyliczają badacze Гасов & Цыганенко (2001):

1. obsługa łączy (odwołanie do tekstu lub innego fragmentu, do którego jest podany link) zarówno w obrębie jakiegoś materiału (słownik elektroniczny, encyklopedia, katalog itp.), jak i w ramach określonego repozytorium elementów informacyjnych publikacji jako całości;
2. wyszukiwanie informacji za pomocą przeglądania w programach przeglądarek;
3. utrzymanie (przyrostowej) struktury wydania;
4. nawigacja po hiperłączach w celu studiowania lub przyswajania wiedzy z zakresu tematycznego objętego publikacją;
5. selekcja struktur „wirtualnych” podczas nawigacji, zbiorów bloków informacyjnych, które są nieco zbliżone (w odniesieniu do tekstu zbiory te są czasami nazywane „kontekstami”);
6. zapamiętywanie „ścieżek”, układanie tekstów liniowych podczas nawigacji;
7. automatyczne konstruowanie powiązanych tekstów z sieci hipertekstowej z wykorzystaniem algorytmów heurystycznych;
8. osadzenie w tekście innych komponentów wydania elektronicznego;
9. organizacja prezentacji wydania elektronicznego.

Słowniki można też łączyć z innymi zasobami – wówczas zyskują nowy potencjał, zwłaszcza jeśli są podlinkowane do materiałów źródłowych (Bonati 2018: 150). Szczególnie interesująca jest integracja zasobów leksykograficznych i korpusowych – o zagadnieniu tym pisze m.in. Wieczorek (2021), a także przytoczona poniżej Karpinska (2020: 85):

Another subtype of electronic dictionaries comprises products where dictionary is combined with raw data from corpus. Asmussen (2013: 1082–1083) mentions several examples of such combined dictionary-corpus products, but also notes that the degree of integration of the material from both sources may vary.

Kolejny podtyp słowników elektronicznych to produkty, w których słownik jest łączony z surowymi danymi z korpusu. Asmussen (2013: 1082–1083) wymienia kilka przykładów takich połączonych produktów słownikowo-korpusowych, ale zaznacza też, że stopień integracji materiału z obu źródeł może mieć różny stopień.

Na rodzimym polu warto odnotować propozycję uhipertekstowania *SJPDor*, którą wysunął Fiedoruszkow (2009: 32–33), brak jednak polskojęzycznych zasobów wykorzystujących hipertekstowość w większym stopniu. Opracowanie hipertekstowej wersji *SW* pozwoliłoby na powiązanie go z innymi słownikami elektronicznymi, nie tylko współczesnymi, co z kolei ułatwiłoby prześledzenie zmian leksykalnych (i nie tylko) w polszczyźnie, a także uprościło proces kompilacji korpusów języka nowopolskiego.

12.1.4. Słownik jako korpus

Lingwistyka korpusowa stoi w pewnym sensie w opozycji do szkoły Chomsky'ego i poszukiwania uniwersalnej struktury – zamiast tego badacz (prawie) niczego nie zakłada, a jedynie patrzy, co jest w korpusie (Grefenstette 1998: 27). Dane te są kluczowe do opisu znaczenia i łączliwości wyrazów:

Corpus evidence provides an essential source of information for collocations and syntagmatics, which need to be studied statistically in order to understand the relationship between word use and word meaning. This provides a structure or framework of a dictionary. (Hanks 2010a: 590)

Dowody korpusowe stanowią istotne źródło informacji dla kolokacji i syntagmatyki, które należy badać metodami statystycznymi, aby zrozumieć związek między użyciem słów a znaczeniem słów. Zapewnia to strukturę lub ramę słownika.

Nie są to jednak jedyne cechy języka, które można badać metodami korpusowymi – dzięki nim można badać też fenomeny (1) leksykalne, (2) morfologiczne, (3) gramatyczno-składniowe, (4) frazeologiczne, oraz (5) semantyczne (Davis 2015: 13). Oczywiście, korpus nie jest konieczny, aby analizować tego rodzaju fenomeny językowe, jednak posiada one szereg zalet w porównaniu z metodami tradycyjnymi, które przytacza Svartvik (2011: 7), postulując, że korpus:

- Jest bardziej obiektywny; mówcy często nie są w stanie zdać dokładnego sprawozdania o tym, co mówili.
- Jest weryfikowalny.
- Jest przydatny do badania wariacji językowych, dialektu, rejestru, stylu i dla badań historycznych.
- Określa częstotliwość użytkowania.
- Stanowi zasób teoretyczny.
- Jest przydatny do tłumaczenia maszynowego, rozpoznawania i syntezy mowy oraz oprogramowania do obsługi języka.
- Jest bardziej reprezentatywny dla użycia języka niż kolekcje cytatów.
- Można wykorzystać do różnych celów.
- Ma potencjał, aby być szeroko (czasem bez ograniczeń) dostępnym do badań naukowych.
- To jedyny sposób na zbadanie użycia wyrazów przez nierodzimych użytkowników – a jest ich sporo – ponieważ żadne inne techniki nie będą działać.

Choć ich przydatność w opisywaniu znaczenia czy zbieraniu danych niektórych rodzajów jest ograniczona, korpusy są bardzo ważne w wielu przedsięwzięciach leksykograficznych (Fuertes-Olivera & Tarp 2014; Bergenholtz 2015; Argebo 2018). Aby jednak w pełni wykorzystać ich możliwości, muszą one być czytelne nie tylko dla ludzi, ale także dla komputera, żeby mógł je odpytywać i parsować (Faaß 2018: 124). Zastosowanie rozwiązań cyfrowych pozwala rozwiązać wiele problemów, ale stawia także przed badaczami nowe wyzwania:

Given this enhanced availability of linguistic data, the question for lexicographers now is not whether to use digitally formatted language samples for research, but how to use them most wisely and effectively (...). (Hargraves 2018: 701)

Biorąc pod uwagę zwiększoną dostępność danych językowych, teraz leksykografowie nie zadają sobie pytania, czy używać do badań cyfrowo sformatowanych próbek językowych, ale jak wykorzystać je najmądrzej i najefektywniej (...).

Stworzenie korpusu wymaga jednak dużego nakładu pracy i sprowadza się do ręcznej anotacji tekstu. Co prawda były podejmowane automatycznego lub wspomaganego automatyzacją porządkowania korpusów, ale algorytmy okazały się do tej pory niewystarczające (Żmigrodzki 2009: 31–32; Węgrzynek 1995). Jak ujmuje to Boneti (2018: 157):

Annotation is indeed a cardinal part of the linguistic analysis of a corpus of texts, the method of describing, recording and analysing linguistic phenomena through computer-based text corpora that is better known as Corpus Linguistics.

Anotacja jest rzeczywiście główną częścią analizy lingwistycznej korpusu tekstów, metody opisywania, rejestrowania i analizowania zjawisk językowych za pomocą komputerowych korpusów tekstowych, lepiej znanej jako Corpus Linguistics.

Nowoczesne słowniki są przeważnie tworzone w oparciu o korpusy, bywają także z nimi bezpośrednio integrowane, jednak dzieła te same mogą być traktowane jako korpusy, ponieważ składają się z dyskretnych, opisanych lingwistycznie elementów tekstowych – artykułów hasłowych i ich komponentów składowych. Oznacza to, że są gotowym do użycia źródłem anotowanych danych.

W polskim językoznawstwie zagadnieniu temu wiele prac poświęcił Bień, co doprowadziło do powstania wyszukiwarki Poliqarp zintegrowanej ze słownikami historycznymi w formacie XML⁷⁸ (2009a: 165), w tym *SL*, które można przeszukiwać jak korpus (Bilińska & Bień 2010: 4ff). Zasoby tam zawarte posiadają jednak różną jakość – w przypadku *SW* jest to jedynie brudny OCR, co uniemożliwia spolegliwego wykorzystania jego treści w pełnym zakresie.

Opracowanie całej treści *SW* w formie przejrzanego i ustrukturyzowanego pliku XML pozwoliłoby na łatwe przeszukiwanie go jako zasobu korpusowego, a zapytania dodatkowo można by było zawęzić do dowolnych elementów strukturalnych, takich jak definicje, przykłady, cytowania, informacja gramatyczna itd.

12.2. Przetwarzanie komputerowe

Przetwarzanie komputerowe pozwala na automatyczną obróbkę danych w dużych ilościach do różnych zastosowań, a także wykonywać zadania, które mogą stanowić problem dla człowieka – niektóre z nich wymienia Knowles (1990: 301):

[L]exicographical computing or – for short – lexi-computing can make allowances for missing data, can accommodate highly individualistic approaches, can virtually avoid the hassle of code-conversions.

⁷⁸ Strona projektu: <https://szukajwsloownikach.uw.edu.pl/>.

- █ Leksykograficzne przetwarzanie komputer (...) może uwzględniać brakujące dane, może uwzględniać wysoce indywidualne podejścia, może praktycznie uniknąć kłopotów z konwersją kodu.

Metody komputerowe można także wykorzystać na potrzeby leksykografii, o czym już pół wieku temu pisał Zgusta (1971: 353). Umożliwiają one m.in. na przeszukiwanie makrokategorii – jak klasa wyrazu, dziedzina, czy źródło – oraz identyfikowanie wariantów gramatycznych (zdobnień, kognatów), fonetycznych i ortograficznych (Boneti 2018: 153–156). Trudno jednak jest w pełni automatycznie wygenerować słownik, co wyjaśnia Fiedlerová (1994: 114):

Za základní předpoklad možného solidního počítačového zpracování slovní zásoby pokládá pak jejich vědeckou analýzu provedenou lingvisty-lexikografy. Ta by se měla orientovat třemi směry: 1. na sestavení inventáře relevantních sémantických komponentů – sémů (ztvárněných zvl. morfémy a valencemi), 2. na stanovení kritérií pro objektivní určení různých charakteristik lexikálních jednotek (např. slovesného vidu), 3. na promyšlení struktury definice významu lexikální jednotky na základě struktury sémů dané její příslušností k jistým lexikálním subsystémům daného jazyka (např. se zřetelem k centralnosti a perifernosti sémů v struktuře lexikálního významu).

- █ Za podstawowy warunek solidnego komputerowego przetwarzania słownictwa uważa się jego naukową analizę przez lingwistów-leksykografów. Ta powinna obrać trzy kierunki: 1. sporządzenie inwentarza odpowiednich składników semantycznych – sememów (reprezentowanych głównie przez morfemy i walencję), 2. ustalenie kryteriów obiektywnego określania różnych cech jednostek leksykalnych (np. forma werbalna), 3. Przemyślenia o strukturach określających znaczenie jednostki leksykalnej na podstawie struktury sememów wynikającej z jego przynależności do pewnych podsystemów leksykalnych danego języka (np. ze względu na centralność i peryferyjność sememów w strukturze znaczenia leksykalnego). (Fiedler 1994: 114)

Jednym z powodów takiego stanu rzeczy jest fakt, który Hanks (2010a: 586) ujmuje w sposób następujący: „[C]omputers (...) do not exercise judgement” (‘[K]omputery nie dokonują oceny’). Z tego powodu wykorzystanie algorytmicznych rozwiązań w dziedzinach humanistycznych nie było do tej pory tak spektakularne, jak miało to miejsce w przypadku nauk ścisłych (McCarty 2015: 72).

Sprawia to, że poprawnie opracowane leksykony są nieocenionym zasobem. Dane słownikowe mogą być więc nie tylko produktem przetwarzania językowego, ale i narzędziem lub jego częścią – przeglądu tego rodzaju narzędzi dokonuje Piotrowski (2012: 85–100). Niektórzy badacze, jak Heid (2014: 25), dostrzegają tendencję do odchodzenia od słowników dla NLP i powrót do słowników przeznaczonych dla ludzi – wynikać to może z popularyzacji rozwiązań uczenia maszynowego, które rzadko posiłkują się słownikami.

Nie oznacza to wszak, że zasoby leksykograficzne nie mają zastosowania w rozwiązaniach algorytmicznych – mowa wówczas o słownikach do odczytu maszynowego, które Hartmann & James (2002: 91) definiują następująco:

machine-readable dictionary

A version of a dictionary which can be processed by or in a computer. Such dictionaries are used, e.g. in natural language processing, for purposes of semantic analysis and disambiguation, machine translation or the representation of knowledge.

słownik do odczytu maszynowego

Wersja słownika, która może być przetwarzana przez komputer lub w komputerze. Z takich słowników korzysta się m.in. w przetwarzaniu języka naturalnego, do celów analizy semantycznej i ujednoznacznienia, tłumaczenia maszynowego lub reprezentacji wiedzy.

Żmigrodzki (2017b: 6) podsumowuje rolę wzajemnego oddziaływania na siebie lingwistyki i technik komputerowych:

[W]szystko wskazuje na to, że jakiegokolwiek badanie języka w XXI wieku bez odwoływania się do metod czy technik związanych z lingwistyką komputerową, z drugiej strony jakiegokolwiek rozwój lingwistyki komputerowej i inżynierii językowej bez korzystania z wiedzy uzyskanej w drodze analiz czysto językoznawczych, nie będą, poza nielicznymi wyjątkami, możliwe (...)

W myśl tę, a także analogicznie do tego, co pisze to Kay w odniesieniu do „The Historical Thesaurus of English” (1983: 87), celem opisywanego w tej pracy rozwiązania jest zapewnienie narzędzia badawczego dla semantyków i badaczy historii języka. Gdyby opracować całość treści *SW*, można by wspomóc lub opracować szereg narzędzi pochodnych.

12.2.1. Słownik sprawdzania pisowni i OCR

Jednym z pierwszych zastosowań dla leksykografii elektronicznej było opracowanie sprawdzania pisowni – to wymagało stworzenia pełnych „indeksów morfologicznych”, czyli pełnych tabeli odmian, czego w leksykografii tradycyjnej się w zasadzie nie stosowało (Jost 2014). To pozwoliło na wygenerowanie list wszystkich akceptowalnych form wyrazowych – jeśli dany ciąg liter na nie jest na takiej liście odnajdywany, sprawdzanie pisowni traktuje go jako błędny.

Większość słowników sprawdzania pisowni jest jednak opracowana na podstawie współczesnych postaci języka, ponieważ to w nich produkuje się tekst wymagający korekty. Słowniki takie mają jednak też inne zastosowania – zasoby leksykalne są istotne dla digitalizacji tekstów, jako że większość programów do rozpoznawania tekstu posiłkuje się listami wyrazów do podjęcia ostatecznej decyzji w przypadkach ambiwalentnych. Dlatego też powstał projekt IMPACT20 (“Improving Access to Text”), którego jednym z zadań było usprawnienie OCR (Piotrowski 2012: 41).

Podczas digitalizacji *SW* nie można było spolegliwie wykorzystać współczesnych list słownikowych, jednak wyczyszczoną treść *SW* można wykorzystać do wygenerowania diachronicznego sprawdzania pisowni (Piotrowski 2012: 78–79). Jest to możliwe, ponieważ w dziele tym jest *de facto* wprowadzone rozwiązanie, które Piotrowski (2012: 70) nazywa kanonizacją pisowni (ang. *spelling canonicalization*), które obejmuje warianty fonetyczne i ortograficzne. To jest z jednej strony ułatwieniem – listy takiej nie trzeba dodatkowo standaryzować. Potencjalnie może to być jednakże źródłem problemów w mapowaniu jednostek hasłowych na treść korpusów, które nie były ujednoczone, przy czym autorzy *SW* ingerowali wyłącznie w zapis form poświadczonych jedynie ustnie, zawsze zachowując oryginalną postać wyrazu napotkanego w treści materiału referencyjnego.

Opracowanie całej treści *SW* pozwoliłoby na uzyskanie zasobu, które zwiększyłyby trafność programów do rozpoznawania tekstu w przypadku konwersji tekstów dawniejszych, a tym samym uprościłoby proces konstrukcji korpusów historycznych.

12.2.2. Autosuplement

W niektórych słownikach istnieją tak zwane jednostki latentne, czyli leksemy, które znajdują się w treści słownika (np. w definicjach, przykładach, glosach lub treściach zewnętrznych), jednak nie stanowią one odrębnych artykułów hasłowych ani przedmiotu opisu językoznawczego. Na potrzebę ekstrakcji takich jednostek i zgromadzenie ich w ramach tzw. autosuplementu zwracał uwagę przede wszystkim Wawrzyńczyk (1991: 173). W pierwszej kolejności postulował opracowanie *SJPDor* (Wawrzyńczyk 1995: 3):

Analiza zawartości SJP postrzeganego jako *tekst* obejmuje wiele zagadnień. Mamy tu m. in. kwestię występowania w tym specyficznym tekście haseł słownikowych potencjalnych, czyli takich wyrażen, które nie figurują w siatce haseł (hasłowniku) samego SJP, niemniej są w nim udokumentowane – i ukryte – głównie w cytatach ilustracyjno-dokumentacyjnych oraz w innych partiach tekstu SJP, nie stanowiących przytoczeń.

Podobny postulat wysunął także względem *SW*, a następnie podjął się pracy nad jego autosuplementem (Wawrzyńczyk 2009a). Rezultatem jego działań jest lista wyrazów niehasłowych, w tym sześciu na literę **F**, która stanowi materiał badawczy niniejszej pracy (Wawrzyńczyk 2009a: 8):

1. **Faetoński** (raz w przykładzie do hasła **Złotokoły**);
2. **fenilometyl** (raz w objaśnieniu do hasła **Benzyl**);
3. **Filadelfja** (dwukrotnie w przykładzie do hasła **Kongres**);
4. **frankoński** (dwukrotnie w przykładzie do hasła **Kapitularz**);
5. **frigotherapia** (raz w przykładzie do hasła **1. Zimno**), oraz:
6. **furfur** (raz w objaśnieniu do hasła **Otrąbowiałość**).

Badacz ten nie jest jednak odosobniony w opisie jednostek niehasłowych – Bień (2015: 11–13) poza hasłami głównymi opisuje też kilka innych rodzajów jednostek, które uwzględniono podczas tworzenia indeksu *a tergo* dla *SL*:

1. hasła wewnętrzne jawne i niejawne – które nie są w słowniku wyróżnione typograficznie, zapisywane w sposób pełny (jawne) i skrócony (niejawne);
2. hasła odtworzone – które jako hasło widnieją w formie odmienionej i zostają sprowadzone do formy słownikowej;
3. hasła uzupełniające – odsyłacze, których desygnat nie istnieje w indeksie głównym;
4. hasła pomocnicze – mające charakter odsyłacza lub komentarza, a których nie ma w indeksie głównym;
5. hasła nieindeksowane – takie, które są fleksyjnie pochodne, np. formy zwrotne czasownika, a także zestawienia;
6. hasła odesłane – które mogą zostać potraktowane jako hasła uzupełniające w przypadku, gdy nie ma hasła głównego, na które wskazuje dany odsyłacz.

Wierzchoń (2015) opisuje nową próbę stworzenia suplementu do *SW* i kroki, które zostały podjęte. I piszący te słowa miał swój udział w tym przedsięwzięciu, zarówno jako programista, jak i redaktor. Pominięty w artykule został jednak istotny problem porównania elementów wysekcerpowanych do hasłownika suplementu; problem ten polega na tym, że przeszukiwanie ręczne ponad 80 tys. Wytypowanych jednostek (także w treści haseł) jest z praktycznych względów awykonalna, a automatyczne przeszukiwanie brudnego OCR i nieustrukturyzowanego tekstu – niemiarodajne.

Autor niniejszej tezy rozwiązał ten problem poprzez napisanie własnego programu do tzw. wyszukiwania aproksymacyjnego (ang. *fuzzy match*), które znajduje wszystkie łańcuchy znaków najbardziej zbliżone do tego, który został zadany w wyszukiwaniu. Algorytm programu następnie ocenia ich podobieństwo na podstawie z góry określonych parametrów i przypisuje odpowiednią wartość punktową zamianie znaków w wyrazie miejscami, ich dodanie lub odjęcie itd., a następnie sortuje pod względem uzyskanej wartości do ręcznego przejrzania wraz z oryginalnym kontekstem. Na tej podstawie można było nie tylko znaleźć poświadczenia lemmy mimo błędów OCR w dowolnej części artykułu hasłowego, ale także znaleźć warianty ortograficzne i formantowe.

W przypadku tak obszernego dzieła, jakim jest *SW*, ekstrakcja form latentnych nie tylko jest pracochłonna, ale także rezultaty ręcznego ich opracowania nie dają gwarancji kompletności. Przy zastosowaniu opisywanego w niniejszej pracy narzędzia i mając do dyspozycji w pełni opracowaną treść *SW* możliwe by było automatyczne wytypowanie jednostek niehasłowych i ich wstępne przyporządkowanie przy pomocy wymienionej wyżej metody, co pozwoliłoby na uzupełnienie elektronicznej wersji tegoż dzieła.

12.2.3. Analiza składniowa i morfologiczna

Sparsowaną treść *SW* można wykorzystać nie tylko do badań obecności jednostek wyrazowych, ale także do ich opisu poprzez stworzenie odpowiednich narzędzi leksykograficznych – narzędzie takie Tarp (2008: 123) definiuje jako

[A] tool that can be used via consultation or passive searching by users with a specific type of communicative or cognitive need to gain access to lexicographical data, from which they can extract the type of information required to cover their specific needs.

Narzędzie, które może być wykorzystywane poprzez konsultację lub bierne wyszukiwanie przez użytkowników z określonym rodzajem potrzeby komunikacyjnej lub poznawczej, aby uzyskać dostęp do danych leksykograficznych, z których mogą wydobyć rodzaj informacji wymaganych do zaspokojenia ich konkretnych potrzeb.

Narzędzia tego rodzaju dzieli się na (1) słowniki oraz (2) leksymaty – ten drugi rodzaj narzędzia Tarp (2008: 123) opisuje jako:

[A] lexicographical tool consisting of a search engine with access to a database and or the internet, enabling users with a specific type of communicative or cognitive need to gain access via active or passive searching to lexicographical data, from which they can extract the type of information required to cover their specific needs.

Narzędzie leksykograficzne składające się z wyszukiwarki, która ma dostęp do bazy danych lub Internetu, pozwalająca użytkownikom ze szczególnym rodzajem potrzeby

komunikacyjnej lub kognitywnej uzyskać dostęp poprzez aktywne lub pasywne przeszukiwanie danych leksykograficznych, z których mogą pozyskać rodzaj informacji wymagany do zaspokojenia swoich konkretnych potrzeb.

Leksymaty można wykorzystać do osiągnięcia różnych poziomów abstrakcji językowej, które wymienia Grefenstette (1998: 28):

- tokenizacja – polegająca na rozgraniczeniu porównywanych obiektów;
- lematyzacja – polegająca na sprowadzaniu form fleksyjnych do znormalizowanych lemm;
- tagowanie części mowy – polegające na przypisaniu poszczególnym wyrazom odpowiednich klas (części mowy);
- parsowanie płytkie – polegające na uogólnieniu informacji o pozycji wyrazu do jego funkcji syntaktycznej;
- tagowanie semantyczne – polegające na uogólnieniu informacji o wyrazie do klas semantycznych.

Słownik elektroniczny może mieć zastosowanie w klasycznych (deterministycznych) rozwiązaniach tagujących oraz wspomagać systemy o partę o uczenie maszynowe. Działanie tagerów przybliżają Górski et al. (2019: 13):

Ujednoznacznienia form fleksyjnych w tekście dokonuje się przy pomocy programu zwanego tagerem. W rzeczywistości tager składa się z dwu programów: analizator morfologiczny rozpoznaje wszystkie potencjalnie istniejące interpretacje form homonimicznych w tekście (mamy interpretuje jako rzeczownik, jak i czasownik), następnie dezambiguator dokonuje ujednoznacznienia, tzn. ustala, z interpretacji jest w danym kontekście prawidłowa (...).

Automatyczna analiza składniowa i tagowanie na wyższym poziomie abstrakcji nadal stanowi pewne wyzwanie, ponieważ o ile ludzie są dobrzy w abstrahowaniu semantycznym, to komputery nadal nie (Grefenstette 1998: 38). Nie oznacza to, że nie ma zadań, w których metody automatyczne nie dają przydatnych rezultatów; lingwiści co prawda niekiedy zakładają, że dowolny kontrprzykład czy całe rozwiązanie niepoprawnym, a tym samym uniemożliwia stworzenie parsera – Grefenstette (1998: 27) jednak proponuje podejście aproksymacyjne w lingwistyce, która opiera się na idei, która brzmi:

[H]igher linguistic analyses such as those to be produced by perfect full-sentence parsing, can be represented more or less faithfully by imperfect systems.

[A]nalizy lingwistyczne wyższego rzędu, takie jak te, które mają być tworzone przez doskonałe parsowanie pełnych zdań, mogą być reprezentowane mniej lub bardziej wiernie przez niedoskonałe systemy.

Takie podejście zdaje się być optymalne, ponieważ „approximations to linguistic theory (...) are both incrementally perfectible as well as being robust and immediately useful.” (‘przybliżenia do teorii lingwistycznej (...) można stopniowo doskonalić, a także są natychmiastowo przydatne’) (Grefenstette 1998: 37).

Co prawda narzędzia takie dla języka polskiego istnieją, lecz nie są przystosowane do analizy tekstów dawniejszych:

Tym samym możliwości użycia do tekstów dawnych analizatora zbudowanego dla współczesnego języka są ograniczone, przy czym oczywiście im wcześniejsze teksty, tym bardziej skuteczność tagera maleje (Eder, Klapper i Kołodziej, 2015). (Górski et al. 2019: 13)

Opracowanie i przetworzenie pełnej treści *SW* pozwoliłoby na jego transformację w leksymat, który posłużyłby do automatycznej analizy morfologicznej języka nowopolskiego oraz identyfikacji nowych jednostek. Na wagę takiego przedsięwzięcia może wskazywać fakt, że próby o zbliżonych celach są podejmowane m.in. w ramach projektu NCN zatytułowanego „Automatyczna analiza fleksyjna tekstów polskich z lat 1830-1918 z uwzględnieniem zmian w odmianie i pisowni”⁷⁹. Narzędzie takie znalazłoby zastosowanie w szeroko pojętych badaniach historycznojęzykowych, a także mogło posłużyć do budowy bardziej zaawansowanych narzędzi.

12.2.4. Klasyfikacja maszynowa

Współcześnie można zaobserwować wzrost zainteresowania rozwiązaniami w obszarze klasyfikacji maszynowej, którą definiują Górski et al. (2019: 71):

Klasyfikacja maszynowa jest jedną z najgwałtowniej rozwijających się dziedzin współczesnej statystyki; pojęcie to stosuje się do całej grupy metod analizy danych ilościowych. Mowa tutaj zarówno o prostych technikach eksploracji danych, jak i metodach tzw. uczenia maszynowego (ang. *machine learning*), których z kolei podzbiorem są algorytmy sztucznej inteligencji (ang. *artificial intelligence*, w skrócie *AI*), wykorzystujące zaawansowane sieci neuronowe, zwane też czasami sieciami uczenia głębokiego (ang. *deep learning*). Wszystkie te techniki łączy zdolność do znajdowania podobieństw pomiędzy elementami zbioru danych i grupowanie ich we względnie jednorodne klasy.

Choć obecnie coraz popularniejsze są rozwiązania oparte o metody statystyczne, opisane dane językoznawcze wysokiej jakości nadal mogą mieć zastosowanie w klasyfikacji tekstów lub poszczególnych – pod względem chronologicznym, tematycznym lub emocjonalnym, a także przy identyfikacji klas obiektów typu *named entity*⁸⁰, które obejmują nazwy własne oraz oznaczenia czasowe. Spektakularnym przykładem tego ostatniego jest opracowana przez Błoch & Vasques Filho & Bojanowskiego (2020) mapa relacji stworzona na podstawie 170 000 portugalskich dokumentów.

Tego rodzaju rozwiązania mogłyby zostać wzbogacone poprzez odpowiedni materiał referencyjny. Pełne opracowanie *SW* w połączeniu z innymi metodami miałyby swoje zastosowanie w wymienionych wyżej obszarach.

12.3. Wartość badawcza

Cyfrowa wersja *Słownika warszawskiego* stanowi nie tylko samodzielne narzędzie czy element bardziej złożonych narzędzi, ale źródło wiedzy o języku i leksykografii. Źródło to dostarcza

⁷⁹ URL: https://projekty.ncn.gov.pl/index.php?projekt_id=199880.

⁸⁰ Piszącemu te słowa nie jest znany żaden polski odpowiednik tego pojęcia.

tych informacji dwojako: bezpośrednio – poprzez informacje eksplicytnie zakodowane w słowniku, jak i pośrednio – poprzez analizę tego, jak słownik ten jest zredagowany i jakim językiem został napisany.

Novum jest tutaj nowy interfejs pozwalający na algorytmiczną analizę zagadnień językowych i metaleksykograficznych zawartych eksplicytnie oraz implicytnie w *SW*. Wykorzystanie tegoż interfejsu wymaga jednak kompetencji w dwóch dziedzinach:

[P]latformą komunikacji między historykiem a programistą jest umiejętność konstruowania algorytmów – logicznych schematów dochodzenia do odpowiedzi na postawione pytanie badawcze. (Wilkowski 2013: 32, za: Prinke 1997: 50)

Johnson, autor ważnego angielskiego słownika, *A Dictionary of the English Language*, pisał o tym, że jego marzeniem jest, aby jego praca przydała się obcym narodom i w odległych wiekach (Johnson 1963: 27f); prawdopodobnie jest to ambicja wielu innych leksykografów. Dzięki zaawansowanej digitalizacji *SW* mógłby zyskać na wartości i relewantności jako zasób leksykograficzny w czasach współczesnych.

12.3.1. Porównanie siatek hasłowych

Słowniki, nawet mając na uwadze wszelkie ich ograniczenia, stanowią cenne źródło danych do badań leksykalnych. Jednym ze sposobów ich wykorzystania jest porównanie notowania jednostek przez poszczególne dzieła. Te mogą różnić się na przykład zakresem objętych jednostek czy algorytmem ich ekstrakcji – na przykład Piotrowska-Wojaczyk (2011: 53) wskazuje w swojej książce na związek między pochodzeniem autorów a stopniem objęcia leksykalizmów regionalnych, a także o problemach w warsztacie leksykograficznym autorów *SW* pod względem traktowania regionalizmów opisała (idem 2011: 152–162).

Możliwe jest także kontrastowanie jednostek z perspektywy chronologicznej – wówczas, jeśli wziąć stosowne poprawki na różnice w metodologii redaktorskiej poszczególnych dzieł, można prześledzić zmiany leksykalne w języku.

Byłoby to odpowiedzią na zapotrzebowanie na słownik polszczyzny XIX w., jako że wszystkie inne okresy ery słownikowej – zarówno wieki poprzedzające jak i współczesność – zostały opisane. Jednocześnie każdy z czterech najważniejszych słowników ogólnych (narodowych) polszczyzny – *SL*, *SWil*, *SW* i *SJPDor* – opisuje słownictwo więcej niż jednego okresu; jak zauważa Kwapien (2014: 259):

Słownik warszawski sięga wstecz aż do okresu staropolskiego, a zatem leksyka XIX stulecia na tle całości zgromadzonego materiału jest jedynie pewną częścią.

Jeśli jednak skonstruować siatki pojęciowe poszczególnych dzieł, to umożliwiłyby to kompleksowe badania nad słownictwem tego okresu. Świadomość roli słowników jako dzieł referencyjnych mieli sami redaktorzy – niektóre słowniki notują obecność wyrazu w innych słownikach, np. *SJPDor* robi tak względem *SL*, *SWil* i *SW* (Żmigrodzki 2009: 108–109), a *SW* robi tak względem *SL* i *SWil*.

O wykorzystaniu siatek pojęciowych do badań nad językiem w kontekście *SL* i *SJPDor* Kwapien (2006: 209) stawia następujące pytania:

- 1) czy można wyciągać wnioski o tym, co dziewiętnastowieczne na podstawie zestawienia informacji z tych dwu słowników;
- 2) czy rzeczywiście można zestawić dane z tych słowników w taki sposób, by można było je potraktować jako podstawę badania słownictwa XIX wieku;
- 3) co jest typowe dla XIX wieku, jeśli chodzi o zasób leksykalny.

Na pytania odpowiedź jest według badaczki twierdząca (Kwapien 2006: 217–218).

Zresztą nadal liczne prace na ten temat (Kwapien 2014: 257–262). Słowa „hapaksowe” w *SW*, tj. nieobecne w siatkach hasłowych innych głównych słowników języka polskiego, jeszcze niedawno były przedmiotem ręcznego opracowania przez Wawrzyńczyka (2015: 9). Kwapien (2010: 241–246) z kolei zestawiła aneks porównujący etymologie różnych słowników. W pracy Piotrowskiej-Wojaczyk (2011: 57–70) natomiast znajduje się opis statystycznego porównania siatek hasłowych czterech słowników: *SL*, *SWil*, *SW* i *SJPDor*, na podstawie którego autorka skompilowała słownik regionalizmów (idem 2011: 193–531).

Porównanie zasobu leksykalnego słowników w sposób idealny byłoby „maksymalistyczne”, czyli przebiegało hasło po hasle, fragment po fragmencie, ale w praktyce nierealizowalne tradycyjnymi metodami (Kaszewski 2014: 12). Opracowanie *SW* jako zasobu byłoby istotnym krokiem w tym kierunku, a wykorzystanie opisywanych w części inżynierskiej tej pracy algorytmów pozwoliłoby na generowanie tego rodzaju danych w sposób niemal całkowicie automatyczny.

12.3.2. Test hipotezy lingwochronologicznej

W językoznawstwie diachronicznym opis języka historycznego polega na zawężeniu się do tego, co kontrastuje ze współczesnością lub innym obranym punktem przeszłości. Sytuację z jednej strony ułatwia fakt, że język historyczny jest „skończony” i zamknięty (Dubois & Dubois 1971: 105–106), z drugiej jednak dostęp do dawniejszych źródeł pisanych jest ograniczony, a do żywego języka mówionego – niemożliwy inaczej niż przez pryzmat tego, co zostało zarejestrowane.

Nie oznacza to jednak, że badania zmian w języku jest niemożliwe – ich teorię ramową przedstawiają Górski et al. (2019: 16):

W językoznawstwie diachronicznym mamy do czynienia z ustaleniami typu: „Forma/struktura X zastępuje formę/strukturę Y w czasie T”. Znany jest punkt wyjścia i punkt dojścia oraz – w mniejszym czy większym przybliżeniu – czas, kiedy dana zmiana zaszła, a przynajmniej chronologia względna zmiany (zjawisko A nastąpiło po zjawisku B, choć moment przejścia jednego zjawiska w drugie jest nieznan).

Zbadanie wystarczająco wielu takich punktowych zmian własności języka w czasie – o ile możliwe jest wyizolowanie sygnału chronologicznego z korpusu (Górski et al. 2019: 82) – pozwala na dokonanie generalizacji, które dostarczają informacji o systemie językowym jako całości lub przynajmniej o istotnym jego segmencie. Badacze zwracają jednak uwagę na ograniczenia operowania w skali pojedynczych słów (idem 2019: 82):

Trudno śledzić zmiany pojedynczych słów (...) Znacznie sensowniej będzie przyjąć hipotezę, że za ewolucją systemu językowego stoi bardzo duża liczba ledwie zauważalnych zmian leksykalnych, być może da się zaobserwować *en masse*.

Zgromadzenie większej ilości materiału badawczego pozwala zatem na sformułowanie dokładniejszych wniosków. Aby jednak materiał taki właściwie przebadano, konieczne są odpowiednie narzędzia – tych dostarcza teoria lingwochronologizacji (TLCH) Wierzchoń (2008), czyli chronologizacji językoznawczej. Dziedzina ta jest zbliżona do etymologii, ale różni się od niej w kilku zasadniczych kwestiach, ponieważ (Wierzchoń 2010b: 39):

1. Nie ma charakteru opisowego, ale kwantyfikuje informacje o leksyce;
2. Przypisuje poszczególnym jednostkom leksykalnym informacji o dacie poświadczenia (najczęściej z dokładnością do roku);
3. Nie skupia się na zmianach, ale na przyroście rocznym słownictwa (i jego dalszej obecności);
4. Ogranicza się do zarejestrowanego tekstu;
5. Nie sięga do tak odległych okresów jak etymologia (co wynika z innych metod badawczych);
6. W jej przypadku znacznie łatwiej zaimplementować metody komputerowe;
7. Oparta jest wyłącznie na materiale badawczym, a nigdy na rekonstrukcjach.

Wszystko powyższe sprawia, że dziedzina ta jest silnie osadzona w empirii. Nie oznacza to, że datacja i redatacja udzielają na pytania badawcze niepodważalnych odpowiedzi – Wawrzyńczyk (2011: 7–8) podważa takie twierdzenie pisząc, że „[w] procesie datowania jednostek posługujemy się zatem (...) nie kategorią pewności, lecz kategorią prawdopodobieństwa”. Datowanie jednostek leksykalnych nigdy bowiem nie mówi, że nie było wcześniejszych ich użycie, a jedynie wskazuje na najwcześniejsze znane; datacja jest więc procesem, który można ukończyć dopiero po wyczerpaniu źródeł, i to tylko w odniesieniu do języka pisanego.

Badacz ten (idem 2011: 8) wysuwa również hipotezę, że gdyby autorzy *SJPDor* mieli dostęp do zaginionej kartoteki suplementu do *SW*, to najprawdopodobniej wywołałoby to masową redatację artykułów hasłowych wyrażaną kwalifikatorami. Te są stosowane, ponieważ nie zawsze jest możliwy metodyczny i rygorystyczny opis chronologii wyrazów (tak jak robią to Bogusławski, Wawrzyńczyk i Wierzchoń); choć zbyt niedokładne dla celów badań diachronicznych, kwalifikatory chronologiczne są przydatne stylistycznie (Walczak 2011: 323–324).

Izolacja informacji chronologicznej nie należy do prostych zadań: na podstawie leksyki łatwiej wykryć gatunek tekstu niż okres jego pochodzenia. Badania ewolucji języka dodatkowo komplikuje fakt, że tempo zmian nie jest stałe (te mogą mieć charakter ciągły lub skokowy) i potrafi zmienić się na przestrzeni zaledwie kilku lat, z wychwyceniem czego słowniki miały do tej pory problem. Mogą one jednak być źródłem wiedzy w większej skali, jeśli rejestrują znaczenia w kolejności ich wejścia do języka – Walczak (2011: 328) zgrabnie opisuje chronologiczny układ znaczeń w artykule hasłowym jako „warstwy stanowiące przyrost leksykalny kolejnych wieków historii języka polskiego”.

Zasadniczo więc lingwochronologizacja oparta jest o kryterium tekstowe oraz koncepcja ekscerpcyjnego badania języka orzeka o istnieniu jednostki, jeśli jest poświadczona w tekście źródłowym, w opozycji do badania siatek słownikowych (Wierzchoń 2010b: 41). Jest to zaleta,

ponieważ poświadczenia wyrazów w tekstach niejęzykoznawczych cechuje „autentyzm poświadczenia filologicznego”, w czym mają przewagę nad słownikami (Wierchoń 2009: 64–65).

Badacz wskazuje też na relację między tekstami źródłowym a leksykonami: „Natomiast zasób słownikowy (leksykograficzny) jest pewną reprezentacją tych tekstów (...)” (Wierchoń 2010b: 48), przez co lingwochronologizacja dzieli niektóre problemy z leksykografią – te, które obie dziedziny dziedziczą po leksykologii (Wierchoń 2010b: 66). Jednocześnie chronologizacja językoznawcza pozwala weryfikować treść słowników:

Centralnym zagadnieniem, które metoda fotodokumentacyjna weryfikuje, jest tzw. kryterium leksykograficzne (...) w zakresie orzekania o nowości danych faktów językowych. (Wierchoń 2009: 66)

Jest to szczególnie istotne w przypadkach wyrazów nienotowanych przez słowniki, ale poświadczonych w tekstach (Wierchoń 2012: 8). Nie zawsze jest jednak możliwe stwierdzić, czy dana forma słowa jest zawiera błąd literowy, czy fakt językowy z danego okresu (Wierchoń 2012: 67). Poza tym jednostki języka (takie jak wyrazy, ale też syntagmy, morfemy czy większe struktury) datuje się na podstawie momentu ukazania się tekstu; to jednak oznacza tylko tyle, że w danym momencie wyrazu już (lub jeszcze) istniał, ale nie oznacza to, że nie ma poświadczeń wcześniejszych (lub późniejszych) (Wierchoń 2010b: 31).

Przykładem wykorzystania takiego mechanizmu jest *Depozytorium leksykalne języka polskiego*, które zawiera hasła poświadczone w zebranych korpusie tekstów, ale odfiltrowane dzięki siatce hasłowej *SW*. W pełni opracowaną kopię cyfrową tegoż słownika można by było wykorzystać do automatycznej ekscerpcji, a w połączeniu z rozwiązaniem pozwalającym na wyszukiwanie aproksymacyjne mogłaby ona posłużyć do wspomaganiej weryfikacji zebranego materiału leksykalnego i jednoczesnej weryfikacji hipotezy lingwochronologizacyjnej z wysoką precyzją.

12.3.3. Badania metaleksykograficzne

Zaprezentowana w części inżynierskiej niniejszej tezy struktura danych, choć oderwana od układu strony, jest odbiciem logicznej struktury słownika; oba te poziomy w przypadku słowników drukowanych są ze sobą powiązane.

Zmiany w językoznawstwie, technice drukarskiej, jak i praktyce leksykograficznej i redakcyjnej odbiły się na kształcie słowników wydawanych w poszczególnych okresach, reprezentujące różne typy i poświęcone danym językom. Różnice między poszczególnymi dziełami do tej pory badano zwłaszcza opisowo, ponieważ większość informacji o treści słownika trudno skwantyfikować, jednak istnieją przykłady operowania na liczbach.

Sławiński (1873) na przykład policzył wyrazy dla trzech słowników w sposób ręczny – były to *SL*, *SWil* i *SRyk*; nie był on jednak pierwszą osobą, która posłużyła się taką metodą (Sławiński 1873: 1–2). Z kolei przeprowadzona już w czasach współczesnych na sześciu tomach *SL* kwerenda wykazała, że zawiera on ok. 60 000, haseł, ok. 200 000 cytatów, oraz ponad 800 źródeł od ok. 400 autorów (Strzyżewski 1994: 148); dwie dekady później statystykom dotyczącym liczby i charakterystyki zapożyczeń w *SWil* poświęciła swój artykuł Majewska (2014). Warto też wskazać na pracę Walczaka (1991: 91–92), w której porównuje statystyki różnych słowników języka polskiego, gdzie *SW* wiedzie niekwestionowany prymat.

Badania te jednak skupiają się jednak głównie na jednostkach języka, a przez zastosowanie próbkowania i metod statystycznych podają jedynie przybliżone wartości. Wykorzystanie cyfrowej wersji *SW* pozwala nie tylko na automatyczne i dokładne zliczenie jednostek, ale także rozszerzenie danych liczbowych o dowolne kategorie wyróżnione strukturalne, a nawet treściowo, co zostało zademonstrowane w **Aneksie N**.

Gdyby zestawić strukturę *SW* w formacie XML z reprezentacjami innych słowników w tym samym formacie, to można by wyciągnąć wnioski na temat zmienności tego gatunku w czasie oraz jego zróżnicowania synchronicznego zależnie od przyjętych kryteriów.

Aby to jednak było możliwe, konieczne jest ustanowienie struktur na tyle uniwersalnych, aby mogły reprezentować niuanse strukturalne porównywanych dzieł. Najprostszym przykładem jest porównanie zbioru znaczników (tagów) wykorzystanych w pliku XML. Aby to osiągnąć, konieczne by było zastosowanie standardu TEI, którego piszący te słowa jest przeciwnikiem ze względu na trudność opisu struktury *SW*, albo modyfikacji wypracowanego w tej tezie doktorskiej systemu na potrzeby innych słowników.

12.3.4. Pozostałe obszary

Powyższe przykłady nie wyczerpują możliwości wykorzystania elektronicznej kopii *SW* na potrzeby badawcze, zwłaszcza z obszaru leksykologii. Jednym z planowanych wcześniej, ale ostatecznie w pracy tej niezrealizowanych założeń było automatyczne wygenerowanie tablic odmiian w sposób zbliżony do tego, co proponował Bień (1991: 37–45).

Trudno jednak skompilować wyczerpującą listę – piszący te słowa ma podstawy sądzić, że w pełni opracowany *SW* stanowiłby zasób, po który sięgali by językoznawcy badający poszczególne fenomeny językowe, a dane będące rezultatem pełnej digitalizacji same się wypowiedzą (ang. *the data will speak for themselves*) (Górski et al. 2019: 77).

13. Falsyfikacja tez badawczych

Na przestrzeni wieków można zaobserwować w słownikach stopniową zmianę nie tylko ich treści, ale także ich postaci. Słowniki polszczyzny nie są tutaj wyjątkiem. Przemiany te mają liczne źródła – wśród nich można wymienić:

1. zmiany w samym języku;
2. zwiększenie się świadomości językowej użytkowników;
3. zmianę poziomu prestiżu języka;
4. zmianę funkcji i percepcji słowników w społeczeństwie;
5. rozwój warsztatu językoznawczego i leksykograficznego;
6. rozwój nauki i techniki;
7. sytuację socjoekonomiczną.

Wśród cech, które się zmieniały, należy podkreślić takie jak:

1. objętość siatki hasłowej;
2. zakres stylistyczny;
3. liczbę ścieżek dostępu;
4. poziom merytoryczny warsztatu;
5. sposób pozyskiwania materiału źródłowego;
6. liczbę rodzajów informacji;
7. struktura logiczna;
8. prezentacja (układ i typografia);
9. dostępność słownika.

Powyższe ogólne zmiany mają swoje konkretne realizacje w słownikach ogólnych polszczyzny – są to przede wszystkim następujące tendencje:

1. rosnąca liczba jednostek hasłowych;
2. rozszerzanie zakresu słownictwa;
3. rosnąca liczba rubryk;
4. usztywnienie formy rubryk;
5. systematyzacja elementów (kwalifikatorów, odsyłaczy, tabel odmian);
6. coraz większa rola materiału empirycznego;
7. rosnąca liczba ścieżek dostępu;
8. rosnąca liczba rodzajów informacji;
9. malejąca liczba błędów technicznych (literówek, nieściśłości, odsyłaczy);
10. wzrost złożoności typograficznej;
11. fluktuacje w podejściu normatywnym.

Cechy słowników zasadniczo trudno jest porównać w sposób kwantytatywny, a tam, gdzie jest to możliwe – choćby w przypadku liczby stron, haseł lub znaczeń – dane liczbowe niewiele mówią o jego jakości. Można jednak stwierdzić obecność pewnych własności, co też zostało uczynione w **tabeli 9**.

Tabela 9. Porównanie cech słowników tradycyjnych i elektronicznych.

Cechy słownika	Słowniki tradycyjne				Słowniki elektroniczne		
	SL	SWiI	SJPDor	SW	USJP	WSJP	Cyfrowy SW
W słowniku brak jest realnych ograniczeń miejsca	X	X	X	X	✓	✓	✓
Informacja w hasłach słownika ma postać pełną	X	X	X	X	✓	✓	✓
Porządek haseł słownika jest dynamiczny (nielinearny)	X	X	X	X	✓	✓	✓
Słownik można przeszukiwać (<i>lookup</i>)	✓	✓	✓	✓	✓	✓	✓
Słownik można zapytywać (<i>query</i>)	X	X	X	X	✓	✓	✓
Układ słownika niezależny od jego logiki	X	X	X	X	✓	✓	✓
Interfejs słownika dostosowany do użytkownika	X	X	X	X	✓	✓	✓
W słowniku można zmienić style i typografię	X	X	X	X	✓	✓	✓
W słowniku można zmienić wyświetlane rubryki	X	X	X	X	✓	✓	✓
W słowniku można dynamicznie zmienić treść	X	X	X	X	✓	✓	✓
Słownik posiada wiele ścieżek dostępu	X	X	X	X	✓	✓	✓
Słownik można przeszukać po lemmie	X	X	X	X	✓	✓	✓
Słownik można przeszukać po znaczeniach	X	X	X	X	✓	✓	✓
Słownik można przeszukać po kategorii tematycznej	X	X	X	X	X	✓	X
Słownik można przeszukać po etymologii	X	X	X	X	X	✓	✓
Słownik można przeszukać po kwalifikatorach	X	X	X	X	X	✓	✓
Słownik można przeszukać po informacji fleksyjnej	X	X	X	X	X	✓	✓
Słownik można przeszukać po części mowy	X	X	X	X	X	X	✓
Słownik można przeszukać po źródle	X	X	X	X	X	X	✓
Słownik można przeszukać po innych rubrykach	X	X	X	X	X	X	✓
W słowniku występują powiązania hipertekstowe	X	X	X	X	✓	✓	✓
Słownik może być zintegrowany z innymi słownikami	X	X	X	X	✓	✓	✓
Słownik może być zintegrowany z korpusami	X	X	X	X	✓	✓	✓
Słownik można aktualizować w czasie rzeczywistym	X	X	X	X	✓	✓	✓
Proces redakcji słownika gwarantuje koherentność	X	X	X	X	✓	✓	✓
Medium słownika jest odporne na degradację danych	X	X	X	X	✓	✓	✓

Poniżej, także w oparciu o przytoczone tutaj własności, zostanie podjęta próba obu wysuniętych wcześniej hipotez badawczych.

13.1. Weryfikacja hipotezy I

Hipoteza I przedstawiona w niniejszej pracy doktorskiej brzmi: **Rozwój technologii jest głównym czynnikiem zmian postaci fizycznej i organizacji słowników.**

W sposób pośredni w części praktycznej pracy starano się odpowiedzieć na pytanie, czy oddzielenie treści słownika od jego formy daje nową jakość. Poniżej znajduje się próba potwierdzenia tej hipotezy.

W przypadku ewolucji słowników od stron technologicznej można wyróżnić trzy okresy:

1. okres słowników rękopiśmiennych;
2. okres słowników drukowanych;
3. okres słowników elektronicznych.

Granice tych okresów odpowiadały przeskokom technologicznym.

Poszczególne przeskoki miały wpływ zarówno na medium, jak i sam proces kompilacji słowników.

Pierwszy przeskok, czyli odejście od słowników rękopiśmiennych na rzecz słowników drukowanych, zmienił przede wszystkim ich dostępność i cenę. Łatwo, szybko i tanio można było je powielać. W rezultacie stały się one dostępne dla stosunkowo dużej grupy użytkowników.

Nie wprowadziło to jednak zasadniczych zmian jeśli idzie o narzędzia stosowane przez ówczesnych leksykografów: Ani struktura słowników ani ich typografia nie uległy znaczącej zmianie. Nie zmieniła się także metoda ich tworzenia, która nadal była ręczna i polegała na tworzeniu osobnych kart hasłowych, które dopiero na końcu były porządkowane i kompilowane do zwartej formy.

Drugi przeskok, czyli zastosowanie komputerów, pogłębił te zmiany, jednocześnie wprowadzając zupełnie nowe medium, oferując tym samym szereg nowych narzędzi zarówno dla redaktorów, jak i dla użytkowników.

Początkowo komputery stosowane niemal wyłącznie do sporządzania słowników, a nie korzystania z nich, co jednak na przestrzeni ostatnich trzech dekad się zmieniło. Warto tu zaznaczyć, że choć dziś prawie wszystkie dzieła leksykograficzne powstają w formie elektronicznej, to publikacje w postaci cyfrowej współistnieją z drukowanymi.

Nadejście leksykografii komputerowej nie tylko pozwoliło na tworzenie obszerniejszych i dokładniejszych słowników, ale także przekształciło je w publikacje wygodniejsze w użyciu. Dzięki postaci cyfrowej znacznemu skróceniu uległy operacje korzystania ze słowników, wyszukiwanie w nich informacji stało się efektywniejsze. Informatyzacja leksykonów otworzyła także zupełnie inne możliwości.

Oczywiście technologia nie jest tu jedynym czynnikiem – bo przecież istotny jest również poziom rozwoju danego języka, jego status, sytuacja normatywna, świadomość twórców

słowników i odbiorców, poziom warsztatu leksykograficznego, stan wiedzy językoznawczej, czy w końcu potrzeby użytkowników. Jednak wszystkie wymienione wyżej czynniki wpływały raczej na dobór haseł i ich treść, a nie na ich budowę, na którą decydujący wpływ miała użyta technologia.

Układ słowników w przypadku języka polskiego był dotąd alfabetyczny, i jeśli decydowano się dodanie kolejnej ścieżki dostępu, to odbywało się to kosztem ogromnego nakładu pracy i znacznym powiększeniem objętości słownika.

Technologia znosi ograniczenia objętościowe, eliminuje skrótowość zapisu, nie wymaga rezygnacji z niektórych haseł z powodów technicznych, i – co najważniejsze – uwalnia czytelnika od ograniczeń systemu alfabetycznego. Stwarza liczne możliwości docierania do nie tylko do szukanych haseł, ale również uzyskiwania ich zbiorów w różnych konfiguracjach.

Znaczenie traci typografia i układ tekstu, które w słownikach papierowych były sposobem na reprezentację struktury danych. W przypadku słowników cyfrowych pełnią one zaś funkcję wyłącznie estetyczną, a poszczególne ich realizacje są oderwane od tego, w jaki sposób dane są przechowywane.

Przedstawiony w części inżynierskiej pracy program SWParser pozwolił przekształcić formę materiału badawczego – czyli wybranego fragmentu *Słownika warszawskiego*. W rezultacie ta sama treść mogła zostać zaprezentowana równoległe na wiele nowych sposobów. Za egzemplifikację służą aneksy dołączone do niniejszej pracy.

13.2. Weryfikacja hipotezy II

Hipoteza II postawiona w niniejszej pracy doktorskiej brzmi: **Pojawienie się nowych funkcji słowników jest pochodną zmian technologicznych.** Słowniki są narzędziem, i jak wszystkie narzędzia – dostosowywane są do celu, jakiemu mają służyć. Objawiać się to może w wyborze treści, układzie podporządkowanym szukaniu i sposobie prezentacji danych.

I tak słowniki normatywne będą zawierać inną siatkę haseł niż słowniki deskryptywne – a przynajmniej będą je oznaczać inaczej. Słowniki słowotwórcze będą miały przeważnie układ gniazdowy lub *a tergo*. Słowniki ogólne będą miały inny zestaw rubryk niż słowniki specjalistyczne itd.

Dzięki uniezależnieniu formy od treści możliwe jest wykorzystanie tego samego zestawu informacji do różnych celów i przy pomocy różnych sposobów ich prezentacji: te same hasła uporządkowane w inny sposób mogą posłużyć do wyszukiwania różnych typów informacji, np. układ *a fronte* do wyszukania znaczeń wyrazu, natomiast *a tergo* do znalezienia wyrazów o tym samym formancie słowotwórczym.

Dzięki możliwości przeszukania dowolnej rubryki można wykorzystać każdy rodzaj informacji w słowniku o charakterze tekstowym, przez to każdy słownik może stać się wielofunkcyjny i dostarczać wielu rodzajów treści równoległe.

Co więcej, technologia pozwala wykorzystywać słowniki w procesach tzw. przetwarzania maszynowego. O ile zasadniczo słowniki przeznaczone tylko do korzystania przez ludzi i te przygotowane do głównie do wykonywania operacji informatycznych się różnią, to część

informacji jest wspólna – odpowiednio sformalizowany opis morfologiczny jest przydatny nie tylko dla osób uczących się języka obcego, ale także dla różnego rodzaju narzędzi NLP.

Im prostsza budowa słownika, tym łatwiej przenieść jego strukturę informacji do baz danych, przy czym im bardziej złożona struktura danych, tym łatwiej wykorzystać zasób informacji na różne sposoby.

Szczególnie potwierdza się to w przypadku ucyfrowienia i transformacji słowników papierowych. Np. Majewska (2020: 325–330) opisuje model artykułu hasłowego *SWil*, który pod wieloma względami jest zbliżony do *SW*, jednak jego struktura jest sztywniejsza, przez co łatwe było przeniesienie go do postaci bazy danych (Majewska 2020: 330ff).

Sam proces digitalizacji słownika również pełni istotną rolę. Po pierwsze spełnia funkcję informacyjną – dostarcza nowych rodzajów wiadomości o samej publikacji, zwłaszcza o jej strukturze głębokiej. Po drugie pełni funkcję normalizacyjną: wszelkie niespójności i osobliwości są wychwytywane i harmonizowane ze względu na charakterystykę baz danych i języków znaczników opartych o XML.

W przypadku retrodigitalizacji dochodzi jeszcze jedna perspektywa funkcjonalna: otóż dotyczy to konwersji słowników dawnych które jako papierowe należały do leksykografii tradycyjnej (klasycznej, historycznej). natomiast po konwersji zalicza się je do leksykografii komputerowej (elektronicznej, cyfrowej).

W postaci pierwotnej leksykony takie były przeznaczone do powszechnego użytku. W nowej postaci stały się użyteczne głównie dla językoznawców (przede wszystkim historyków języka). Z procesem tym jest więc nierozdzielnie związana ewolucja funkcji słownika.

Najprawdopodobniej nie wszystkie zmiany funkcjonalne zostały już odkryte, i analogicznie do tego, jak współczesna digitalizacja oraz integracja z innymi zasobami nadała nowy kontekst dawniejszym źródłom, podobne zjawisko może mieć miejsce w przyszłości. Szukanie tych nowych kontekstów jest właściwie humanistyce cyfrowej trzeciej fali (Maryl 2017; za: Berry 2011: 12):

Jest też coś, co David Berry nazwał „trzecią falą” humanistyki cyfrowej, czyli krytyczny namysł nad epistemologią narzędzi cyfrowych czy wizji świata, którą owe technologie pośrednio promują. Berry postuluje „przyjrzeć się cyfrowemu komponentowi humanistyki cyfrowej w świetle specyfiki jego medium, by analizować, jak przemiany mediów wytwarzają przemiany epistemologiczne”.

W przypadku retrodigitalizacji *Słownika warszawskiego* uzyskano nowe funkcje dzięki analizie składniowej i przekształceniu treści tegoż dzieła na formę ustrukturyzowaną przy pomocy znaczników. Wspomniane funkcje wynikają z:

1. Nowego sposobu prezentacji treści *SW*;
2. Nowych ścieżek dostępu do elementów treści *SW*.

Potencjalne metody wykorzystania treści *SW* są egzemplifikowane praktycznie w sposób obszerny, choć niewyczerpujący potencjału uzyskanych danych, w aneksach do niniejszej pracy, których funkcje zostały opisane w **podrozdziale 10.4** – te obejmują derywowane mini-dzieła leksykograficzne (m.in. słownik gramatyczny w **Aneksie G** słowniczki par przekładowych w **Aneksie J**), a także opisy metaleksykograficzne (m.in. dane liczbowe w **Aneksie N**

czy porównanie siatek hasłowych różnych słowników w **Aneksie K**) oraz dokumenty językoznawcze niemające charakteru leksykonu (m.in. mini-korpus w **Aneksie D**).

13.3. Podsumowanie

Niniejsza praca dowodzi tego, jak wielkie możliwości – w kontekście rozszerzenia funkcjonalności – kryje w sobie proces cyfryzacji w dziedzinie leksykografii.

W niniejszej pracy rzecz dotyczy względnie niewielkiego fragmentu treści *Słownika warszawskiego* w postaci cyfrowej. Mimo to digitalizacja okazała się czynnikiem, który istotnie zmienia zakres i sposoby wykorzystania słownika. Digitalizacja całego słownika czy wręcz zbiorów słowników stwarza zatem możliwości, które dopiero są badane i odkrywane.

Co więcej, samo przewidywanie potencjału digitalizacji w tym zakresie – zarówno jeśli idzie o naukę, edukację, jak i codzienne korzystanie ze słownika – stanowi dzisiaj bardzo ważny obszar badań, który obejmuje nie tylko językoznawstwo, ale i antropologię, informatologię, kulturoznawstwo i nauki filologiczne. Rola ucyfrowienia staje się jednocześnie coraz istotniejsza, także dlatego, że informatyka podsuwa badaczom kolejne, coraz bardziej zaawansowane narzędzia.

14. Prezentacja danych

Program SWParser po przetworzeniu treści *Słownika warszawskiego* i wygenerowaniu wyjściowego pliku XML generuje także na jego podstawie dane pochodne, z których część to dane oryginalne zaprezentowane w nowy sposób, inne natomiast są wynikiem wieloetapowego przetworzenia.

SWParser następnie wykorzystuje te dane, aby przy pomocy silnika szablonów języka Ruby oraz technologii L^AT_EX w pełni automatycznie wygenerować aneksy do pracy doktorskiej – program sam przetwarza dane w odpowiedniej kolejności, a następnie generuje ich ludzko-brzmiące opisy i interpretacje w języku naturalnym przy pomocy autorskich skryptów bez ingerencji człowieka. Na końcu dane są formatowane w taki sposób, aby układ graficzny plików wyjściowych w formacie PDF był zgodny z częścią zasadniczą tezy doktorskiej.

Rezultatem opisywanego w niniejszej tezie przetwarzania słowników są wszystkie trzy sposoby reprezentacji treści wyróżniane przez TEI⁸¹, które w swojej pracy przytacza Bień (2010b: 132):

1. Perspektywa typograficzna — dwuwymiarowy obraz strony, z zachowaniem informacji o podziale na wiersze i strony.
2. Perspektywa edytorska — jednowymiarowy ciąg znaków, który może stanowić dane wejściowe dla systemu składania tekstów.
3. Perspektywa leksykalna — reprezentacja zawartej w słowniku informacji abstrahująca od jej formy tekstowej.

Bień (2010b: 134–135) odróżnia pliki także **poligraficzne** (przeznaczone do druku) od **redakcyjnych** (zawierających dodatkową informację, nieprzeznaczoną dla użytkownika końcowego). Pliki redakcyjne – zarówno zawierające dane wejściowe opisywane w **rozdziale 10.1**, jak i wszelkie pliki pośrednie generowane na poszczególnych etapach wykonywania programów – nie są jednak w niniejszej pracy prezentowane ze względu na ich niewielką wartość badawczą.

Poniżej opisany został układ graficzny poszczególnych aneksów, a także objaśnienia, jak należy interpretować zwarte w nich dane – proces opracowania samych danych znajduje się natomiast w **rozdziale 10.4** i nie jest tutaj opisywany ponownie.

14.1. Aneks A: Indeks artykułów hasłowych

W aneksie A wypisane są wszystkie hasła ze sparsowanej części *Słownika*. Aneks ten na układ kolumnowy i składa się z dwóch części: indeksu alfabetycznego *a fronte* oraz indeksu alfabetycznego *a tergo*.

⁸¹ Punkt 9.5 na stronie www: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html>.

Układ danych w części *a fronte* ma format:

1. hasło;
2. dezambiguator;
3. kwalifikator.

Natomiast w części *a tergo* kolejność elementów to:

1. dezambiguator;
2. kwalifikator;
3. hasło.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.1**.

14.2. Aneks B: Indeks stałych połączeń wyrazowych

W aneksie B zostały uwzględnione wszystkie połączenia wyrazowe, które zostały wyróżnione przez autorów *Słownika* – obejmują one wszystkie pojęcia zagnieżdżone opisane w **rozdziale 4.3.4.10**, jeśli składają się one z więcej niż jednego wyrazu.

Jednostki w tym aneksie ułożone są w kolejności alfabetycznej, na podstawie której każdemu przyporządkowany jest numer porządkowy; poza tym przy każdym z nim jest wskazane hasło, z którego zostały wyekstrahowane.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.2**.

14.3. Aneks C: Indeks form wyrazowych

W aneksie C umieszczony jest spis wszystkich form wyrazowych, rozumianych jako unikalne ciągi znaków reprezentujących wyraz lub skróty języka polskiego, które ekstrahowane są z:

1. definicji;
2. przykładów użycia;
3. objaśnień;
4. komentarzy (glos).

Jednostki leksykalne zawarte w ramach innych struktur, takich jak kwalifikatory czy elementy informacji gramatycznej, nie są uwzględniane.

Aneks ten ma układ kolumnowy i składa się z dwóch części: indeksu alfabetycznego *a fronte* oraz indeksu alfabetycznego *a tergo*. W obu z nich przy każdej formie podana jest liczba jej wystąpień w tekście, na podstawie której formy te są uporządkowane na liście w kolejności od najwyższej do najniższej wartości; w przypadku jednostek o tej samej liczbie wystąpień decyduje kolejność alfabetyczna – odpowiednio *a tergo* i *a fronte*. Oprócz tego każdej formie przypisana jest jej liczba porządkowa na podstawie frekwencji.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.3**.

14.4. Aneks D: Korpus zdań i wyrażen przykładowych

W aneksie D znajduje się lista wszystkich wyrażen, zdań oraz większych fragmentów tekstu ilustrujących użycie wyrazów hasłowych. Przy każdym z nich podane jest hasło, z którego zostały zaczerpnięte, a na podstawie którego zostały pogrupowane; posortowane w ramach grup zostały natomiast według kolejności wystąpienia w artykule. Każdy fragment jest też opatrzony numerem porządkowym, który kolejność tę odzwierciedla, a także – w przypadku cytatów – skrócenie ze źródłem.

Wszystkie formy skrócone hasła zostały rozwinięte – tak samo jest też ze skrótami: **ś.** (= *się*), **a.** (= *albo*) oraz **wog.** (= *wogóle*), np.:

- Na bokach każdej deszczki dwoje F. będzie, przez które deszczka z deszczką **ś.** spoi.

zostało rozwinięte do:

- Na bokach każdej deszczki dwoje fugowanie będzie, przez które deszczka z deszczką **się** spoi.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.4**.

14.5. Aneks E: Wykaz znaków w systemie Unikod

W aneksie E zawarty jest wykaz wszystkich unikodowych znaków, które znajdują się w treści zasadniczej (hasłowej) *Słownika*. Pod uwagę brany jest cały tekst po zastosowaniu poprawek z pliku podmian i po usunięciu dodanych znaków ucieczki. Znaki posortowane są na podstawie częstotliwości względem pozostałych znaków, uszeregowanych od najczęstszego do najrzadszego, a w przypadku takiej samej częstotliwości – na podstawie kodu szesnastkowego w kolejności rosnącej.

Wykaz ma układ tabelaryczny i zawiera następujące kolumny:

- **ID** – będące liczbą porządkową nadaną zgodnie z sortowaniem;
- **Użyć** – czyli liczba wystąpień znaku;
- **Znak** – czyli graficzną reprezentację znaku;
- **Kod** – zapisany w formacie szesnastkowym;
- **Nazwa** – oficjalnie nadana w systemie Unikod (w języku angielskim);
- **Rodzaj** – opisujący, czy znak składa się z jednego (prosty) czy wielu (złożony) punktów kodowych.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.5**.

14.6. Aneks F: Fotokorpus haseł SW

Aneks F jest rezultatem połączenia trzech typów informacji zawartych w SW:

1. wygenerowanej siatki hasłowej;
2. pliku z informacją o lokalizacji haseł w druku, co obejmuje kolejno:

- tom;
- literę alfabetu (uzyskaną na podstawie numeru tomu);
- stronę;
- kolumnę;
- numer akapitu w kolumnie.

3. cyfrowych kart hasłowych uzyskanych dzięki programowi IndexCardSplitter.

SWParser podczas generowania tego aneksu odlicza kolejne hasła od ostatniego znalezione w pliku lokalizacji i na tej podstawie dobiera odpowiednie pliki graficzne, których lokalizacja opisana jest w ich nazwie. Jeśli hasło obejmuje więcej kolumn lub stron, to lokalizacja jest prezentowana w formie zakresu kart hasłowych. Informacja ta jest następnie zamieniana na opis zrozumiały dla człowieka.

W wygenerowanym pliku widnieje zatem wyraz hasłowy (wraz z dezambiguatorami i kwalifikatorami symbolicznymi), obok którego jest umieszczony opis lokalizacji, a następnie karta lub karty hasłowe, które zawierają treść artykułu hasłowego. Hasła zagnieżdżone zostały zgrupowane z hasłami głównymi.

Obrazy umieszczone w fotokorpusie różnią się od tych wygenerowanych przez IndexCardSplitter – zostały one ręcznie poprawione, jednak naniesione zmiany mają charakter niemal wyłącznie kosmetyczne – poprawiają zwłaszcza estetykę, a tylko niekiedy czytelność.

Metoda uzyskania danych do aneksu została opisana w **rozdziale 10.4.6**.

14.7. Aneks G: Słowniczek gramatyczny

Aneks G zawiera listę form hasłowych wraz z informacją o części mowy i odmianie. Treść aneksu jest wygenerowana na podstawie informacji dostępnej w *SW* oraz pozasłownikowej wiedzy o języku.

W przypadku, gdy dane hasło posiada więcej niż jeden paradygmat, traktowane jest jako odrębna jednostka leksykalna. Cały aneks ten ma układ tabelaryczny i jest obrócony o 90° względem orientacji strony. W tabeli umieszczone są następujące kolumny:

- **Lemma** – czyli wyraz hasłowy słownika;
- **Inf. gramatyczna** – czyli oryginalna informacja gramatyczna ze słownika;
- **Cz. mowy** – część mowy przypisana na podstawie informacji gramatycznej;
- **Zakres** – czyli informacja o tym, których znaczeń wyrazu hasłowego dotyczy dany paradygmat;
- **Detekcja** – czyli sposób rozpoznania części mowy:
 - **merge** – na podstawie pełnej lub scalonej (z fragmentów paradygmatu dla poszczególnych znaczeń) informacji gramatycznej w słowniku;
 - **derivation** – na podstawie informacji o derywacji wyrazu w słowniku;
 - **heuristic** – na podstawie pozasłownikowej wiedzy o języku;
- **Wzór** – na podstawie którego rozpoznano część mowy;
- **Klasa** – wygenerowana automatycznie na podstawie części mowy oraz informacji gramatycznej.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.7**.

14.8. Aneks H: Klasyfikacja paradygmatów gramatycznych

Aneks H zawiera wykaz wszystkich klas paradygmatów, które zostały przypisane wyrazom hasłowym w aneksie G, uporządkowane w kolejności alfabetycznej. Dane zaprezentowane są w układzie tabelarycznym, a obok każdej klasy przytoczona jest kolejno informacja o części mowy, zapisie paradygmatu, a także jego liczbie wystąpień w materiale badawczym.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.7**

14.9. Aneks I: Słowniczek walencyjny

Aneks I zawiera wykaz wszystkich odrębnych informacji o walencji – ta ogranicza się w *Słowniku* do informacji o łączliwości z przyimkami i przypadkami. Dane zaprezentowane są w układzie tabelarycznym, gdzie przy każdym wyrazie hasłowym podana jest jego walencja oraz zakres znaczeń, których dotyczy; jeśli jeden wyraz posiada wiele walencji, wówczas jest powielony na liście, tak jak zostało to przedstawione w **tabeli 10**:

Tabela 10. Fragment aneksu I.

Lemma:	Walencja:	Zakres:
...
† foldrować	na kogo	znaczenie 1.
† foldrować	kogo	znaczenia 2. i 4.
† foldrować	kogo/co	znaczenie 3.
...

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.8**.

14.10. Aneks J: Słowniczki przekładowe

Aneks J zawiera wyrazy obce wraz z polskimi tłumaczeniami, zaczerpnięte z części etymologicznej artykułów hasłowych *SW*; wyrazy obce nieposiadające tłumaczeń są ignorowane oraz nazwy własne nie są uwzględnione.

Struktura aneksu ma układ kolumnowy, a hasła zaprezentowane są w układzie alfabetycznym. Sam aneks podzielony jest na automatycznie generowane sekcje, z których każda obejmuje jednostki leksykalne z danego języka; tytuły sekcji generowane są automatycznie na podstawie rozwinięć skrótów językowych.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.9**.

14.11. Aneks K: Matryca siatek hasłowych SW i wybranych słowników

Aneks K jest rezultatem porównania opracowanej części siatki hasłowej *Słownika warszawskiego* z 53 słownikami referencyjnymi. Dobór słowników jak i sposób ich porównania został opisany w **podrozdziale 10.4.10**.

Cały aneks ma układ tabelaryczny, a strony mają układ poziomy. Na jego początku znajduje się tabela z listą wszystkich słowników referencyjnych wraz z informacją bibliograficzną oraz przypisaną kategorią, która wskazuje, czy dany słownik można opisać jako:

1. Definicyjny;
2. Przekładowy;
3. Poprawnościowo-ortograficzny;
4. Wyrazów obcych;
5. Gramatyczny;
6. Historyczny;
7. Etymologiczny;
8. Gwarowy;
9. Specjalistyczny.

Każdemu słownikowi przypisana jest jedna lub więcej kategorii, na podstawie których są one grupowane, co ułatwia analizę.

Treść właściwa aneksu, ze względu na swoją objętość, jest zaprezentowana w trybie dwustronnym, tj. pierwsza połowa szerokości tabeli dla danego zakresu wierszy znajduje się na stronie parzystej, a druga na stronie nieparzystej.

W pierwszej kolumnie tabeli widnieje informacja o numerze porządkowym jednostki leksykalnej (w tym jednostek wielowyrazowych, a także i morfemów), w drugiej zaś umieszczone są kolejne jednostki z połączonych siatek pojęciowych wszystkich słowników, uporządkowane w kolejności alfabetycznej. Kolejne wiersze natomiast zawierają informację o obecności jednostki leksykalnej w poszczególnych słownikach.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.10**.

14.12. Aneks L: Lista podmian w treści

Aneks L jest reprezentacją jednego z plików pomocniczych i zawiera informację na temat ingerencji w oryginalną treść *Słownika* dokonanych przez program SWParser. Każdej modyfikacji odpowiada pojedyncza tabela, tabele te są natomiast pogrupowane są w sekcje, zależnie od tego, do jakiej kategorii należą zebrane w niej podmiany.

W nagłówku każdej tabeli widnieje informacja o liczbie porządkowej kolejnych zastosowanych podmian, natomiast jej komórki zawierają kolejno:

1. Oryginalny tekst, który podlegał modyfikacji;
2. Poprawiony tekst, który zastąpił tekst oryginalny;
3. Uzasadnienie wprowadzonej zmiany;
4. Hasło, w którym dokonano podmiany.

Fragment, który różni się między tekstem oryginalnym a poprawionym, jest także – dzięki modułowi porównującemu programu SWParser – podświetlony: w przypadku tego pierwszego kolorem zielonym, a w przypadku tego drugiego – czerwonym. Pusty kwadrat na pół wysokości wiersza dodatkowo oznacza odpowiednik zerowy (w przypadku dodania lub usunięcia ciągu znaków), natomiast pusty prostokąt na całą wysokość wiersza należy interpretować jako spację. Przykład wyniku takiego porównania widać na przykładzie hasła **Fulminacja** (ilustracja 64):

Ilustracja 64. Modyfikacja hasła „Fulminacja”.

MODYFIKACJA 1.	
Oryginalny tekst:	Fulminacja, i, lm. e 1. (o złocie, srebrze a. prochu) 1. łoskot, szczęk, brzęk; trzask.
Zamieniono na:	Fulminacja, i, lm. e 1. (o złocie, srebrze a. prochu) □ łoskot, szczęk, brzęk; trzask.
Uzasadnienie:	<i>zduplikowany numerator „1.”</i>
Dotyczy hasła:	Fulminacja

Metoda uzyskania danych do aneksu została opisana w **rozdziale 0**.

14.13. Aneks M: Skonsolidowany spis skrótów

Aneks M zawiera listę skrótów (i innych oznaczeń) obecnych w pliku predefiniowanym. Jego treść podzielona jest na sekcje odpowiadające kategoriom skrótów opisanych w **rozdziale 5.2**. Każdemu skrótowi w sekcji odpowiada jeden akapit.

Na początku każdego akapitu umieszczona jest liczba porządkowa danego skrócenia w kolejności alfabetycznej (choć elementy dodane ręcznie są poza tą kolejnością) – liczby w ramach danej sekcji są często nieciągłe, ale posortowane od najniższej do najwyższej. Następne w kolejności jest skrócenie wraz z rozwinięciem w kolorze niebieskim; po nich natomiast jest umieszczony automatycznie wygenerowany opis z podsumowaniem występowania oznaczenia, a także ręcznie dopisane uwagi (jeśli takowe są) – oba segmenty zapisane są kursywą. Na samym końcu akapitu, kolorem szarym wskazana jest liczba poświadczeń oznaczenia w materiale badawczym.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.12**.

14.14. Aneks N: Dane liczbowe

Aneks N zawiera zróżnicowane dane liczbowe opracowane na podstawie sparsowanej treści SW, jak i danych względem niej pochodnych. Aneks ten podzielony jest na sekcje, z których każda dotyczy innej kategorii informacji. Wewnątrz każdej sekcji znajduje się pojedyncza tabela, która w pierwszej kolumnie zawiera opis i rodzaj danych liczbowych, a w kolumnie drugiej – obliczoną wartość liczbową.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.13**.

14.15. Aneks O: Słowniczek par derywacyjnych

Aneks O zawiera listę znalezionych przez program na podstawie odsyłaczy i informacji gramatycznej par derywacyjnych. Sparowane jednostki leksykalne zaprezentowane są w układzie kolumnowym, a pierwsza z nich jest pogrubiona.

Pary leksemów podzielone są dodatkowo na trzy sekcje:

1. Pary derywacyjne czasowników niezwrotnych i zwrotnych;
2. Pary derywacyjne czasowników i odsłowników (gerundiów);
3. Pary derywacyjne przymiotników i przysłówków.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.14**.

14.16. Aneks P: Słowotwórstwo rzeczowników żeńskich (feminatywów)

Aneks P zawiera listę rzeczowników męskich i odpowiadające im rzeczowniki żeńskie w układzie tabelarycznym. Oprócz pary leksemów w każdym wierszu umieszczona jest dodatkowo informacja o sposobie derywacji poprzez zestawienie końcówki rzeczownikowej męskiej (w tym końcówki zerowej, zaznaczonej znakiem \emptyset) z końcówką żeńską.

Metoda uzyskania danych do aneksu została opisana w **podrozdziale 10.4.15**.

15. Opis pracy

W **rozdziale 0. Wstęp** przedstawiono podwaliny teoretyczne pracy, postawiono hipotezy badawcze, określono ramy metodologiczne oraz cele praktyczne, a także nakreślono strukturę niniejszej tezy.

W **rozdziale 1. Stan badań** nakreślono sytuację relewantnych dyscyplin i umiejscowiono pracę w dotychczasowym dyskursie badawczym.

W **rozdziale 2. Cechy słownika** opisano kategorie parametrów słownika, które są istotne z perspektywy postawionych hipotez badawczych, jednocześnie demonstrując ich związek z fenomenami językowymi i światem pozajęzykowym, w tym kontekst nadawcy, odbiorcy, medium oraz funkcji.

W **rozdziale 3. Słownik warszawski – charakterystyka** zawarta jest charakterystyka *Słownika warszawskiego*, nakreślająca genezę dzieła, sylwetki autorów, proces jego powstawania, recepcję oraz wartość badawczą dzieła jako faktu społecznego.

W **rozdziale 4. Struktura SW** zaprezentowano budowę *SW*, która obejmuje mega-, makro- oraz mikrostrukturę, ze szczególnym uwzględnieniem i egzemplifikacją elementów artykułu hasłowego znaczących z punktu widzenia części inżynieryjnej niniejszej tezy doktorskiej.

W **rozdziale 5. Skrócenia** zawarty jest dokładny opis skrótów i oznaczeń użytych w słowniku, obejmujący sposoby ich analizy, wykryte nieścisłości, odtworzenie reguł ich tworzenia, klasyfikację funkcjonalną, a także proponowany sposób ustandaryzowania ich formy.

W **rozdziale 6. Metodologia retrodigitalizacyjna** przedstawiona jest teoria metodologii retrodigitalizacyjnej, a także jej konkretna realizacja na potrzeby niniejszej pracy, obejmująca wybór i analizę źródła, dobór narzędzi, utworzenie kopii cyfrowej, przygotowanie do digitalizacji, rozpoznanie tekstu, korektę rezultatów, formatowanie, analizę składniową, kontrolę jakości, wykorzystanie danych i ich prezentację, a także archiwizację i indeksowanie.

W **rozdziale 7. Narzędzia** wymienione są narzędzia oraz systemy programistyczne, zarówno autorskie jak i pochodzące od stron trzecich, które zostały wykorzystane do realizacji części inżynieryjnej niniejszej pracy doktorskiej.

W **rozdziale 8. Ekstrakcja kart hasłowych** nakreślono problematykę ekstrakcji kart hasłowych, a także opisano działanie technikę ich automatycznej ekstrakcji z wykorzystaniem dedykowanego narzędzia odautorskiego, którego rezultaty zostały następnie skonfrontowane z metodą ręczną.

W **rozdziale 9. Konwersja warstwy graficznej na tekstową** pokrótce objaśniono proces konwersji obrazu na tekst oraz napotkane problemy generyczne oraz te charakterystyczne dla opisywanego tu materiału badawczego.

W **rozdziale 10. Przetwarzanie danych SW** szczegółowo opisano proces analizy składniowej *Słownika warszawskiego*, w tym format danych wejściowych, segmentację haseł, parowanie poszczególnych segmentów (inicjalnego, centralnego i terminalnego), a także generowanie danych pochodnych i prezentacji rezultatów przetwarzania komputerowego. Dodatkowo

wymienione są zaprojektowane, ale niezaimplementowane rozwiązania programistyczne oraz potencjalne usprawnienia, które należałoby wprowadzić w kolejnej wersji programu.

W rozdziale 11. Ewolucja słowników polszczyzny przedstawiono ewolucję słowników języka polskiego, wymieniając i pokrótce charakteryzując poszczególne dzieła, a także nakreślono tendencje historyczne zachodzące zmiany metodyczne w opracowywaniu dzieł leksykograficznych, wreszcie prezentując miejsce *Słownika warszawskiego* w polskiej praktyce leksykograficznej.

W rozdziale 12. Praktyczna aplikacja wersji cyfrowej podsumowane są zalety przeprowadzonej na potrzeby niniejszej tezy digitalizacji *SW*, w tym jego zastosowania jako słownika elektronicznego, narzędzia do przetwarzania języka naturalnego, oraz wartość jako materiału do badań lingwistycznych.

W rozdziale 13. Falsyfikacja tez badawczych podjęto próbę falsyfikacji postawionych tez badawczych i przedstawiono ich uzasadnienie.

W rozdziale 14. Prezentacja danych opisano właściwą interpretację danych zawartych w aneksach, które zostały automatycznie wygenerowane na podstawie danych wyjściowych opisywanego w części inżynierskiej tej pracy analizatora składniowego.

W rozdziale 15. Opis pracy znajduje się niniejsze podsumowanie, czyniąc ten akapit rekurencyjnym.

Bibliografia

Ze względu na zróżnicowane pochodzenie użytej literatury, obok każdej pozycji została umieszczony kod języka w standardzie ISO 639-1, w którym została ona napisana, a są to:

- [PL] – język polski
- [EN] – język angielski
- [CZ] – język czeski
- [SK] – język słowacki
- [RU] – język rosyjski
- [HR] – język chorwacki
- [DE] – język niemiecki
- [FR] – język francuski
- [ES] – język hiszpański
- [IT] – język włoski
- [JP] – język japoński
- [CN] – język mandaryński

Literatura cytowana

- [autor zbiorowy] (1904): *Życie i prace Jana Karłowicza (1836-1903)*. Warszawa: Skład główny w redakcji „Wiśły”, 379 s. [PL]
- [autor zbiorowy]: „Przygotowywanie cyfrowych wersji obiektów dziedzictwa kulturowego” [na stronie:] Federacja Bibliotek Cyfrowych. URL: <http://fbc.pionier.net.pl/pro/tworzyc-repozytoria-cyfrowe/przygotowywanie-cyfrowych-wersji-obiektow-dziedzictwa-kulturowego/> (ostatnia aktualizacja: 2 listopada 2016). [PL]
- [autor zbiorowy]: „Z kanceláře Slovníku jazyka českého”, [w:] *Naše řeč* (1929): r. 13, nr 1, s. 21–22; r. 13, nr 10, s. 235–236; (1930): r. 14, nr 2, s. 40; (1931): r. 15, nr 1, s. 20–23; (1932): r. 16, nr 1, s. 21–24. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/>. [CZ]
- [autor zbiorowy]: „Zpráva z kanceláře Slovníku jazyka českého”, [w:] *Naše řeč* (1935): r. 19, nr 4, s. 123; (1937): r. 21, nr 1, s. 20–23; (1938): r. 22, nr 2, s. 56–58; (1939): r. 23, nr 2, s. 60–61; (1941): r. 25, nr 4, s. 122–124; (1942): r. 26, nr 2, s. 57–59. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/>. [CZ]
- [autor zbiorowy]: *Unicode® Standard Annex #15 – Unicode Normalization Forms*, rev. 51, rozdz. 1.1. URL: <https://unicode.org/reports/tr15/>. [EN]
- ABELSON, Harold & SUSSMAN, Gerald Jay & SUSSMAN, Julie (1996): *Structure and Interpretation of Computer Programs*, 2nd ed. Cambridge, MA – London: MIT Press. [EN]
- ADAMO, Giovanni (1998): „Terminología vs. Lessicología” [w:] *Hieronymus Complutensis*, nr 8, s. 75–86. Madrid: Instituto Universitario de Lenguas Modernas y Traductores. ISSN: 1135-304X. URL: https://cvc.cer-vantes.es/lengua/hieronymus/pdf/08/08_075.pdf. [IT]
- ALVAR, Manuel (1976): „Nota sobre CAPTENENCIA” [w:] *Revista De Filología Española*, t. 58, nr 1/4, s. 231–235. ISSN: 0210-9174. e-ISSN: 1988-8538. <https://doi.org/10.3989/rfe.1976.v58.i1/4.710>. [ES]
- ANUSIEWICZ, Janusz 1994: *Lingwistyka kulturowa. Zarys problematyki*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 180 s. ISBN: 9788322911204. [PL]
- ARGEBO, Heidi (2018): „Explaining meaning in lexicographical information tools” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 59–77. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- ASMUSSEN, Jörg (2013): „Combined products: Dictionary and corpus” [w:] Gouws, Rufus Hjalmar & Heid, Ulrich & Schweickard, Wolfgang & Wiegand, Herbert Ernst (red.): *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 1081–1090. Berlin, Boston: De Gruyter Mouton. ISBN: 9783110238129. e-ISSN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.1081>. [EN]
- AZPIAZU TORRES, Susana (1999): „Es factible un diccionario de adverbios en -mente?” [w:] *Revista de Lexicografía*, t. 5. Coruña: Universidade da Coruña, s. 7-31. ISSN: 1134-4539. e-ISSN: 2603-6673. DOI: <https://doi.org/10.17979/rlex.1999.5.0.5635>. [ES]
- BAILEY, Richard W. (1986): „Dictionaries of the next century” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): *Fulbright Papers*, t. 1, s. 111–122. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8. [EN]
- BAŃCZEROWSKI, Jerzy & POGONOWSKI, Jerzy & TADEUSZ Zgółka (1982): *Wstęp do językoznawstwa. Skrypt dla studentów studiów uniwersyteckich*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, 356 s. [PL]
- BAŃKO, Mirosław & MAJDAK, Magdalena & CZESZEWSKI, Maciej (2011): *Słowniki dawne i współczesne. Internetowy przewodnik edukacyjny (strona internetowa)*. URL: <http://www.leksykografia.uw.edu.pl>. [PL]

- BARTNICKA, Barbara (2000): „Udział słownictwa XIX-wiecznego w Słowniku języka polskiego pod redakcją Witolda Doroszewskiego” [w:] *Poradnik Językowy*, nr 1, s. 18–23. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343. [PL]
- BELYAEV, Oleg & KHOMCHENKOVA, Irina & SINITSYNA, Julia & DYACHKOV, Vadim (2021): „Digitizing print dictionaries using TEI: The Abaev Dictionary Project” [w:] *Proceedings of the Seventh International Workshop on Computational Linguistics of Uralic Languages (IWCLUL)*, s. 57–64. Syktyvkar (Russia): Association for Computational Linguistics. URL: <https://aclanthology.org/2021.iwclul-1.7>. [EN]
- BERGENHOLTZ, Henning & GOUWS, Rufus Hjalmar (2010): „A Functional Approach to the Choice between Descriptive, Prescriptive and Proscriptive Lexicography” [w:] *Lexikos*, t. 20, s. 27–51. Stellenbosch: Buro Van Die Wat, 778 s. ISBN: 978-0-9814434-2-3. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/20-0-173>. URL: <http://lexikos.journals.ac.za>. [EN]
- BERGENHOLTZ, Henning & GOUWS, Rufus Hjalmar (2012): „What is Lexicography?” [w:] du Plessis, Johannes Christiaan Marthinus Daniël (red.): *Lexikos*, t. 22, s. 92–108. Stellenbosch: Buro Van Die Wat. ISBN: 978-0-9814434-2-3. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/22-1-996>. URL: <http://lexikos.journals.ac.za>. [EN]
- BERGENHOLTZ, Henning & NIELSEN, Sandro (2013): „What Is a Lexicographical Database?” [w:] *Lexikos*, nr 23, s. 77–87. Stellenbosch: Buro Van Die Wat. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/23-1-1205>. URL: <http://lexikos.journals.ac.za>. [EN]
- BERGENHOLTZ, Henning (2003): „User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography” [w:] du Plessis, J.C.M.D. (red.): *Lexikos*, t. 13, s. 65–80. Stellenbosch: Buro Van Die Wat, 365 s. ISBN: 0-9584401-5-8. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/13-0-749>. URL: <http://lexikos.journals.ac.za>. [EN]
- BERGENHOLTZ, Henning (2005): „Falsche und richtige lexikographische Definitionen” [w:] Gottlieb, Henrik & Mogensen, Jens Erik & Zettersten, Arne (red.): *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2–4, 2002 at the University of Copenhagen*, s. 125–132. Tübingen: Max Niemeyer Verlag. ISBN: 3-484-39115-4. ISSN: 0175-9264. [DE]
- BERGENHOLTZ, Henning (2018): „Dictionary management” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 34–42. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- BERRY, David M. (2011): „The Computational Turn: Thinking About the Digital Humanities” [w:] *Culture Machine*, nr 12 (2011). London: Open Humanities Press. ISSN: 1465-4121. URL: https://sro.sussex.ac.uk/id/eprint/49813/1/BERRY_2011-THE_COMPUTATIONAL_TURN_THINKING_ABOUT_THE_DIGITAL_HUMANITIES.pdf. [EN]
- BIELSKA-BRODZIAK, Agnieszka & TOBOR, Zygmunt (2007): „Słowniki a interpretacja tekstów prawnych” [w:] Kubicki, Leszek i in. (red.): *Państwo i Prawo*, nr 5 (735). Warszawa: Komitet Nauk Prawnych PAN. ISSN: 0031-0980. [PL]
- BIEŃ, Janusz S. & BILIŃSKA, Joanna & SARNECKI, Mateusz (2014a): „An incremental approach to retrodigitization” [prezentacja na:] *European Network of e-Lexicography WG2 meeting: Retro-digitized dictionaries*. 13 s. URL: <http://bc.klf.uw.edu.pl/378/>. [EN]
- BIEŃ, Janusz S. & BILIŃSKA, Joanna & SARNECKI, Mateusz (2014b): „Przyrostowa metoda dygitalizacji słowników” [prezentacja na:] *Leksykografia polska, ukraińska, bułgarska: słowniki tradycyjne i elektroniczne Warszawa, 13.11.2014 r.* 17 s. URL: <http://bc.klf.uw.edu.pl/378/>. [EN]
- BIEŃ, Janusz S. & SZAFRAN, Krzysztof (2001): „Analiza morfologiczna języka polskiego w praktyce” [w:] *Biuletyn Polskiego Towarzystwa Językoznawczego = Bulletin de la société polonaise de linguistique*, z. LVII (2006), s. 1–17. Kraków: Universitas. ISSN: 0032–3802. [PL]
- BIEŃ, Janusz S. & SZAFRAN, Krzysztof (2010): „Słowniki elektroniczne i dygitalizacja tekstów” [prezentacja na:] *Seminarium magisterskie 2010/2011*. 28 s. URL: <http://bc.klf.uw.edu.pl/214/>. [PL]

- BIEŃ, Janusz S. (1990): „Co to jest TeX?” [w:] *Roczniki Polskiego Towarzystwa Matematycznego – Seria II: Wiadomości Matematyczne*, t. XXIX (1990). Poznań: Polskie Towarzystwo Matematyczne. ISSN. 0373-8302. [PL]
- BIEŃ, Janusz S. (1991): *Koncepcja słownikowej informacji morfologicznej i jej komputerowej weryfikacji*, [seria:] *Rozprawy Uniwersytetu Warszawskiego = Dissertationes Universitatis Varsoviensis*, t. 383. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 236 s. ISBN: 83-230-0390-4. ISSN: 0509-7177. [PL]
- BIEŃ, Janusz S. (1993): „Wybrane standardy przetwarzania tekstów” (tekst przygotowany do materiałów konferencji *Komputerowa Analiza Tekstu*, 16–18.11. 1993, Karpacz, które nigdy nie zostały wydane). [PL]
- BIEŃ, Janusz S. (1999): „Kodowanie tekstów polskich w systemach komputerowych” [w:] *Postscriptum*, nr 27–29, s. 4–27. Katowice: Wydawnictwo Uniwersytetu Śląskiego. ISSN: 1427–0501. [PL]
- BIEŃ, Janusz S. (2006a): „Aparat pojęciowy wybranych systemów przetwarzania tekstów polskich” [w:] Laskowski, Roman (red.): *Biuletyn Polskiego Towarzystwa Językoznawczego = Bulletin de la société polonaise de linguistique*, z. LXII (2006), s. 19–30. Kraków: Universitas, 283 s. ISSN: 0032–3802. [PL]
- BIEŃ, Janusz S. (2008a): „Digitalizing dictionaries of Polish” [prezentacja na:] *Metody analizy leksykalnej: założenia teoretyczne i zastosowania praktyczne Białystok*, 9–11.10.2008, 75 s. [EN]
- BIEŃ, Janusz S. (2008b): „Narzędzia do analizy tekstowej warstwy dokumentów DjVu” (Notatka do wykładu i laboratorium „Reprezentacja tekstów w systemach komputerowych”). 5 s. URL: <http://bc.klf.uw.edu.pl/105/>. [PL]
- BIEŃ, Janusz S. (2009a): „Facilitating Access To Digitalized Dictionaries in DjVu Format” [w:] Koseska-Toszewa, Violetta (red.): *Cognitive Studies / Études Cognitives*, vol. 9. Warszawa: Sławistyczny Ośrodek Wydawniczy Instytut Sławistyki PAN (SOW Publishing House), s. 161–170. ISSN: 2080-7147. [EN]
- BIEŃ, Janusz S. (2010a): „Historical Polish texts and Unicode” [prezentacja z:] 09.04.2010, 46 s. URL: <http://bc.klf.uw.edu.pl/179/>. [EN]
- BIEŃ, Janusz S. (2010b): „Dygitalizacja i komputeryzacja słowników na przykładzie *Słownika polszczyzny XVI wieku*” [w:] *Język polski – wczoraj, dziś, jutro*, s. 131–138. Kraków: Instytut Języka Polskiego PAN & Wydawnictwo LEXIS. ISBN: 978-83-89425-61-4. URL: <http://bc.klf.uw.edu.pl/165/>. [PL]
- BIEŃ, Janusz S. (2011b): „Podstawowe elementy tekstów elektronicznych” [w:] Kopcińska, Dorota & Bańko, Mirosław (red.): *Różne formy, różne treści: tom ofiarowany Profesorowi Markowi Świdzińskiemu*, s. 17–24. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 244 s. ISBN 978-83-62100-29-3. [PL]
- BIEŃ, Janusz S. (2011c): „Efficient search in hidden text of large DjVu documents” [w:] *Advanced Language Technologies for Digital Libraries*, [seria:] *Lecture Notes in Computer Science: Theoretical Computer Science and General Issues*, nr 6699, s. 1–14. Springer. ISBN: 978-3-642-23159-9. ISSN: 0302-9743. DOI: http://dx.doi.org/10.1007/978-3-642-23160-5_1. URL: <http://bc.klf.uw.edu.pl/177/>. [PL]
- BIEŃ, Janusz S. (2012a): „Od skanów do Unicode” [prezentacja na:] *Dni Technologii Językowej*, 28 września 2012 r. 65 s. URL: <http://bc.klf.uw.edu.pl/320/>. [PL]
- BIEŃ, Janusz S. (2012b): „Skanowane teksty jako korpusy” [w:] *Prace Filologiczne*, t. LXIII, s. 25–35. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. URL: <http://bc.klf.uw.edu.pl/192/>. [PL]
- BIEŃ, Janusz S. (2015): „Elektroniczny indeks do słownika Lindego” [w:] *Kwartalnik Językoznawczy*, nr 3-4/2015. Poznań: Uniwersytet im. Adama Mickiewicza, s. 1–19. ISSN: 2081-5441. DOI: <https://doi.org/10.14746/kj.2015.3-4.1>. [PL]
- BIEŃ, Janusz S. (2016a): „Repertuar znaków piśmiennych – problemy i perspektywy” [w:] *Kwartalnik Językoznawczy*, nr 4/2016. Poznań: Uniwersytet im. Adama Mickiewicza, s. 1–18. ISSN: 2081-5441. DOI: <https://doi.org/10.14746/kj.2016.4/2017.1.1>. [PL]

- BIEŃ, Janusz S. (2016b): „Problemy kodowania znaków w korpusach historycznych” [w:] Roszko, D. & Satoła-Staśkowiak, J. (red.): *Semantyka a konfrontacja językowa*, t. 5, s. 67–76. Warszawa: Slawistyczny Ośrodek Wydawniczy Instytutu Slawistyki PAN. ISBN: 978-83-64031-25-0. [PL]
- BIEŃ, Janusz S. (2016c): „Elektroniczne indeksy fiszek słownikowych” [w:] *Kwartalnik Językoznawczy*, nr 2/2016. Poznań: Uniwersytet im. Adama Mickiewicza, s. 16–27. ISSN: 2081-5441. DOI: <https://doi.org/10.14746/kj.2016.2.2>. [PL]
- BIEŃ, Janusz S. (2019): „Standard Unicode i język polski” [w:] *Acta Poligraphica*, nr 14, s. 7–28. Warszawa: Centralny Ośrodek Badawczo-Rozwojowy Przemysłu Poligraficznego. ISSN: 2299-9981. [PL]
- BIEŃ, Janusz S. (2020): „Dokumenty DjVu w Internecie. Reaktywacja” (przedruk). DOI: <https://doi.org/10.13140/RG.2.2.21561.19047>. [PL]
- BILIŃSKA, Joanna & BIEŃ, Janusz S. (2010): „Linde’s dictionary of Polish – new retro-digitisation and electronic word index” [prezentacja na:] *Retro-Digitized Dictionaries – COST ENeL WG2 meeting, Budapest, Hungary*, 24–25 lutego 2017 r. 37 s. URL: <http://bc.klf.uw.edu.pl/492/>. [EN]
- BILIŃSKA, Joanna (2006): „Dialekty i języki obce w Słowniku języka polskiego Samuela Bogumiła Lindego – zestawienie na podstawie wydania drugiego” [w:] *Prace Filologiczne*, t. LXVIII, s. 27–42. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. URL: <https://www.ceeol.com/search/article-detail?id=422683>. [PL]
- BILIŃSKA, Joanna (2011): „Describing Linde’s Dictionary of Polish for Digitalisation Purposes” [w:] Kosem, Iztok & Kosem, Karmen (red.): *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011, Bled, 10-12 November 2011*. Ljubljana: Trojina, Institute for Applied Slovene Studies, s. 43–51. ISBN: 9789619298329. URL: <http://bc.klf.uw.edu.pl/240/>. [EN]
- BILIŃSKA, Joanna (2013): *Analiza i leksykograficzny opis struktury słownika Lindego na potrzeby dygitalizacji* (praca doktorska). Warszawa, 174 s. [PL]
- BŁOCH, Agata & VASQUES FILHO, Demival & BOJANOWSKI, Michał (2020): „Networks from archives: Reconstructing networks of official correspondence in the early modern Portuguese empire” [w:] Everett, Martin & Valente, Thomas (red.): *Social Networks*. Elsevier. ISSN: 0378-8733. DOI: <https://doi.org/10.1016/j.socnet.2020.08.008>. [EN]
- BOELHOUWER, Bob & DYKSTRA, Anne & SIJENS, Hindrik (2018): „Dictionary portals” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 754–766. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- BOLINGER, Dwight (1965): „The Atomization of Meaning” [w:] *Language*, nr 41 (4), s. 555–573. Baltimore: Linguistic Society of America. ISSN: 0097-8507. e-ISSN: 1535-0665. [EN]
- BONATI, Isabella (2018): „Digital Papyrological Editions and the Experience of a Lexicographical Database” [w:] Reggiani, Nicola (red.): *Digital Papyrology II*. Berlin: De Gruyter, s. 159–173. ISBN: 9783110547450. DOI: <https://doi.org/10.1515/9783110547450-009>. [EN]
- BOUDA, Peter & CYSOUW, Michael (2012): „Treating Dictionaries as a Linked-Data Corpus” [w:] Chiarcos, Christian & Nordhoff, Sebastian & Hellmann, Sebastian (red.): *Linked Data in Linguistics: Representing and Connecting Language Data and Language Metadata*, s. 15–23. Heidelberg, Dordrecht, London, New York: Springer. ISBN: 978-3-642-28248-5. e-ISBN: 978-3-642-28249-2. DOI: <https://doi.org/10.1007/978-3-642-28249-2>. [EN]
- BREWER, Charlotte (2004): „The ‘Electronification’ of the Oxford English Dictionary” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t. 25, z. 1, s. 1–43. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.2004.0001>. [EN]
- BRÜCKNER, Aleksander (1901): „Słownik języka polskiego ułożony pod redakcją Jana Karłowicza, Adama Kryńskiego i Władysława Niedźwiedzkiego. Zeszyt I–X. Warszawa, nakładem prenumeratorów, 1898-1901).

- Słownik gwar polskich, ułożył Jan Karłowicz. 2 tomy. (Kraków, nakładem Akademii Umiejętności, 1900 i 1901)” [w:] *Przegląd Polski*, r. IV, s. 94–102. Kraków: Druk. „Czasu”. [PL]
- BRÜCKNER, Alexander (1904): „Prace słownikowe” [w:] *Życie i prace Jana Karłowicza (1836-1903)*, s. 165–175. Warszawa: Skład główny w redakcji „Wisły”, 379 s. [PL]
- BURDICK, Anne & DRUCKER, Johanna & LUNENFELD, Peter & PRESNER, Todd & JEFFREY, Schnapp (2012): *Digital_Humanities*. Cambridge: The MIT Press, 152 s. ISBN (twarda): 9780262018470. ISBN (miękka): 9780262528863. [EN]
- CAMPOS MARTIN, Natalia (2009): „Lexicographie et traduction – Les différents types de dictionnaires – Classement et typologie” [w:] *Entreculturas*, nr 1. Grupo Interuniversitario de Traducción, Comunicación y Lingüística Aplicada (HUM 767). ISSN: 1989-5097. URL: <http://entreculturasuma.comimagine.es/?p=2732>. [FR]
- CEIL, Lucas (2003): *The Role of Variation in Lexicography* [w:] *Sign Language Studies: Special Issue on Dictionaries and Lexicography, Part I: General Issues in Lexicography (Spring 2003)*, t. 3, nr 3. Washington: Gallaudet University Press, s. 322-340. ISSN: 0302-1475. e-ISSN: 1533-6263. URL: <https://www.jstor.org/stable/26204787>. [EN]
- ČERMÁK, František (2011): *Jazyk a jazykověda. Přehled a slovníky*, vyd. 4. Praha: Karolinum. ISBN: 978-80-246-1946-0. [CZ]
- ČERMÁKOVÁ, Anna (2013): „Jaký slovník uživatelé češtiny potřebují? O Slovníku současné češtiny nakladatelství Lingea” [w:] *Slovo a slovesnost*, r. 74, nr 3, s. 195–210. Praha: Ústav pro jazyk český Akademie věd České republiky. ISSN: 0037-7031. e-ISSN: 2571-0885. URL: <http://sas.ujc.cas.cz/>. [CZ]
- CHEBANNE, Andy (2010): „The Role of Dictionaries in the Documentation and Codification of African Languages: The Case of Khoisan” [w:] du Plessis, Johannes Christiaan Marthinus Daniël (red.): *Lexikos*, t. 20, s. 92–108. Stellenbosch: Buro Van Die Wat, 778 s. ISBN: 978-0-9814434-2-3. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/20-0-173>. URL: <http://lexikos.journals.ac.za>. [EN]
- CIACOMINI, Laura (2018): „Dictionaries for translation” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 284–299. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- COLEMAN, Julie & MCDERMOTT, Anne (red.) (2004): *Historical Dictionaries and Historical Dictionary Research: Papers from the International Conference on Historical Lexicography and Lexicology, at the University of Leicester, 2002*, [seria:] Allén, Sture & Corbin, Pierre & Hartmann, Reinhard Rudolf Karl & Hausmann, Franz Josef & Heid, Ulrich & Reichmann, Oskar & Zgusta, Ladislav (red.): *Lexicographica. Series Maior*, t. 123. Tübingen: Max Niemeyer Verlag. ISBN: 3-484-39123-5. ISSN: 0175-9264. [EN]
- CORBIN, Pierre (2013): „Lire les dictionnaires (3). Éléments pour une modélisation des opérations impliquées dans leur consultation : de l’indentification des articles” [w:] Gasiglia, Nathalie (red.): *Interpréter les dictionnaires : pluralité d’approches*, [seria:] Corbin, Pierre (red.): *Lexique*, t. 21, s. 32–100. Villeneuve-d’Ascq: Presses Universitaires du Septentrion, 252 s. ISBN: 978-2-7574-0437-9. ISSN: 0756-7138. [FR]
- COTTER, Colleen & DAMASO, John (2007): „Online Dictionaries as Emergent Archives of Contemporary Usage and Collaborative Codification” [w:] *Queen Mary’s Occasional Papers Advancing Linguistics*, nr 9. London: Queen Mary College, University of London. [EN]
- COWIE, A. P. (red.) (2009): *The Oxford History of English Lexicography*, t. 1 i 2: *General-Purpose Dictionaries*. Oxford: Oxford University Press, 467 + 551 s. ISBN: 978-0-19-928562-4. [EN]
- CRYSTAL, David (1986): „The ideal dictionary, lexicographer and user” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): *Fulbright Papers*, t. 1, s. 72–81. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8. [EN]

- CYGAL-KRUPA, Zofia (1979): *Szesnastowieczne edycje „Dictionarii Ioannis Murellii variarum rerum...”*; część I. *Uwagi ogólne, ortografia, fonetyka, cechy dialektyczne*. Kraków: Polskie Wydawnictwo Naukowe, 137 s. [PL]
- CYGAN, Stanisław (2020): „Dziewiętnastowieczne słowniczki gwarowe księdza Władysława Siarkowskiego z perspektywy leksykograficznej” [w:] *Gwary Dziś*, t. 12. Poznań: Wydawnictwo Poznańskiego Towarzystwa Przyjaciół Nauk, s. 229–245. ISSN: 1898-9276. DOI: <https://doi.org/10.14746/gd.2020.12.17>. [PL]
- CZELAKOWSKA, Anna & KUBICKA, Emilia & KLUBIŃSKA, Maja (2016): „Słowniki przed sądem” [w:] *LingVaria*, r. XI (2016), nr 1 (21), s. 45–64. Kraków: „Księgarnia Akademicka”. ISSN: 1896-2122. e-ISSN: 2392-1226. DOI: <https://doi.org/10.12797/LV.11.2016.21.05>. [PL]
- DAVIS, Mark (2015): „1. Corpora: An Introduction” [w:] Biber, Douglas & Reppen, Randi (red.): *The Cambridge Handbook of English Corpus Linguistics: CHECL*, s. 11–31. Cambridge: Cambridge University Press, 623 s. ISBN: 9781139764377. DOI: <https://doi.org/10.1017/CBO9781139764377.002>. [EN]
- DE SCHRYVER, Gilles-Maurice (2003): „Lexicographers’ Dreams in the Electronic-Dictionary Age?” [w:] *International Journal of Lexicography*, t. 16, z. 2, s. 143–199. ISSN: 0950-3846. e-ISSN: 1477-4577. DOI: <https://doi.org/10.1093/ijl/16.2.143>. [EN]
- DEBUS-GREGOR, Esther & HEID, Ulrich (2013): „Design criteria and ‘added value’ of electronic dictionaries for human use” [w:] Gouws, Rufus Hjalmar & Heid, Ulrich & Schweickard, Wolfgang & Wiegand, Herbert Ernst (red.): *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 1001–1013. Berlin, Boston: De Gruyter Mouton. ISBN: 9783110238129. e-ISSN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.1001>. [EN]
- DOROSZEWSKI, Witold (1952): „O pracy nad układem haseł w słowniku” [w:] Doroszewski, Witold (red.): *Poradnik Językowy*, z. 10 (105)/1952, s. 1–4. Warszawa: Państwowe Wydawnictwo „Wiedza Powszechna”.
- DOROSZEWSKI, Witold (1954): *Z zagadnień leksykografii polskiej*, wyd. I. Warszawa: Państwowy Instytut Wydawniczy, 147 s. [PL]
- DOROSZEWSKI, Witold (1966): [w:] *Wśród słów, wrażeń i myśli. Refleksje o języku polskim*, wyd. I. Warszawa: Państwowy Instytut Wydawniczy 1966, 476 s. [PL]
- DOROSZEWSKI, Witold (red.) (1968): *Słownik języka polskiego*, t. 1. Warszawa: Państwowe Wydawnictwo PWN, 1485 s. ISBN: 8301118768. [PL]
- DUBOIS, Jean & DUBOIS, Claude (1971): *Introduction à la lexicographie : le dictionnaire*, [seria:] Dubois, Jean (red.) *Langue et langage*. Paris: Librairie Larousse, 217 s. ISSN: 1766-263X. [FR]
- DUMMER, Sven & MICHAELIS, Frank & SCHLAEFER, Michael (1998): „Zur Digitalisierung historischer Wörterbücher” [w:] *Lexikos*, t. 8, s. 194–222. Stellenbosch: Buro Van Die Wat, 623 s. ISBN: 978-0-9814434-2-3. ISSN: 1684-4904. e-ISSN: 2224-0039. [DE]
- DZIAMSKA-LENART, Gabriela (2013): „O zmienności znaczeń związków frazeologicznych (na materiale *Słowniczka frazeologicznego* Antoniego Krasnowolskiego z końca XIX wieku)” [w:] Migdał, Jolanta & Piotrowska-Wojaczyk, Agnieszka (red.): *Cum reverentia, gratia, amicitia... Księga jubileuszowa dedykowana Profesorowi Bogdanowi Walczakowi*, s. 453–461. Poznań: Wydawnictwo Rys. ISBN: 978-83-63664-05-3. [PL]
- DZIEMIANKO, Anna (2018): „Electronic dictionaries” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 663–683. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- ECKER, Alexandre (2013): „*lod.lu* : dictionnaire ou dictionnaires?” [w:] Gasiglia, Nathalie (red.): *Interpréter les dictionnaires : pluralité d’approches*, [seria:] Corbin, Pierre (red.): *Lexique*, t. 21, s. 101–115. Villeneuve-d’Ascq: Presses Universitaires du Septentrion, 252 s. ISBN: 978-2-7574-0437-9. ISSN: 0756-7138. [FR]

- EFFORD, Nick (2000): *Digital Image Processing: A Practical Introduction Using Java™*. Harlow: Pearson Education. ISBN: 978-0-201-59623-6. [EN]
- FAAB, Gertrud (2018): „Lexicography and corpus linguistics” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 123–137. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- FIEDLEROVÁ, Alena (1994): „Igor Němec, nestor české historické lexikologie, sedmdesátiletý” [w:] *Naše řeč*, r. 77, nr 3, s. 113–115. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=7187>. [CZ]
- FIEDORUSZKOW, Jurij (2009): „Słowniki elektroniczne oraz środowisko hipertekstowe w leksykografii a *Słownik języka polskiego* pod red. W. Doroszewskiego” [w:] Wawrzyńczyk, Jan (red.): *Czterdzieści lat minęło... Nad „Słownikiem Doroszewskiego”*, s. 27–35. Warszawa: Wydawnictwo Takt, 254 s. ISBN: 978-83-87850-48-7. [PL]
- FILIPEC, Josef (1975): „Cesta k českému jednosvazkovému slovníku” [w:] *Naše řeč*, r. 58, nr 5, s. 225–233. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=5862>. [CZ]
- FILIPEC, Josef (1981): „Z prehistorie a historie Ústavu pro jazyk český ČSAV” [w:] *Naše řeč*, r. 64, nr 5, s. 225–232. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=6280>. [CZ]
- FILIPEC, Josef (1995): „Teorie a praxe jednoznačného slovníku výkladového” [w:] Čermák, František & Blatná, Renata (red.): *Manuál lexikografie*, s. 14–49. Jinočany: H&H, 283 s. ISBN: 80-85787-23-7. [EN]
- FUÉRTES-OLIVERA, Pedro A. & BERGENHOLTZ, Henning (2018): „Dictionaries for text production” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 267–283. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- GASIGLIA, Nathalie (2013a): „Interpréter les dictionnaires : une pluralité d’approches” [w:] Gasiglia, Nathalie (red.): *Interpréter les dictionnaires : pluralité d’approches*, [seria:] Corbin, Pierre (red.): *Lexique*, t. 21, s. 7–22. Villeneuve-d’Ascq: Presses Universitaires du Septentrion, 252 s. ISBN: 978-2-7574-0437-9. ISSN: 0756-7138. [FR]
- GASIGLIA, Nathalie (2013b): „Le {Dictionnaire Hachette benjamin} met-il à la portée des élèves du CE1 les mots des textes d’un manuel de français ?” [w:] Gasiglia, Nathalie (red.): *Interpréter les dictionnaires : pluralité d’approches*, [seria:] Corbin, Pierre (red.): *Lexique*, t. 21, s. 163–190. Villeneuve-d’Ascq: Presses Universitaires du Septentrion, 252 s. ISBN: 978-2-7574-0437-9. ISSN: 0756-7138. [FR]
- GATES, Edward (1992): „Should a Dictionary Include Only the ‘Good’ Words?” [w:] Hyldgaard-Jensen, Karl & Zettersten, Arne (red.): *Symposium on Lexicography V: Proceedings of the Fifth International Symposium on Lexicography May 3-5, 1990 at the University of Copenhagen*, s. 265–280, [seria:] Allén, Sture & Corbin, Pierre & Hartmann, Reinhard Rudolf Karl & Hausmann, Franz Josef & Heid, Ulrich & Reichmann, Oskar & Zgusta, Ladislav (red.): *Lexicographica. Series Maior*, t. 43. Tübingen: Max Niemeyer Verlag. ISBN: 3-484-30943-1. ISSN: 0175-9264. [EN]
- GLEASON, Henry Allan Jr. (1967): „The Relation of Lexicon and Grammar” [w:] Householder, Fred W. & Saporta, Sol (red.): *Problems in lexicography*, s. 85–102. Bloomington: Indiana University Press. [EN]
- GOUWS, Rufus Hjalmar (2018a): „Dictionaries and access” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 43–58. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- GOUWS, Rufus Hjalmar (2018b): „Expanding the data distribution structure” [w:] Gouws, Rufus Hjalmar & Heid, Ulrich & Herbst, Thomas et al. (red.): *Lexicographica*, t. 34 (2018), s. 225–238. Berlin: De Gruyter. ISSN: 1865-9403. e-ISSN: 22240039. DOI: <http://dx.doi.org/10.1515/lex-2018-0011>. [EN]

- GÓRSKI, Rafał L. & KRÓL, Magdalena & EDER, Maciej (2019): *Zmiana w języku. Studia kwantytatywno-korpuse*, [seria:] Bijak, Urszula & Eder, Maciej & Tyrpa, Anna & Żmigrodzki, Piotr (red.): Prace Instytutu Języka Polskiego Pan, nr 151. Kraków: Instytut Języka Polskiego PAN. ISBN: 978-83-64007-53-8. [PL]
- GRALIŃSKI, Filip & WIERZCHOŃ, Piotr (2017): „Jak powstaje słownik efemeryd leksykalnych polszczyzny XIX i XX wieku?” [w:] Wawrzyńczyk, Jan & Wierzchoń, Piotr (red.): *Wokół 300 tysięcy polskich słów. Wstęp do hasłownikologii*, s. 103–118.. Warszawa: BEL Studio Sp. z o.o., 398 s. ISBN: 978-83-7798-248-8. [PL]
- GRANGER, Sylviane (2012): „Electronic lexicography – from challenge to opportunity” [w:] Granger, Sylviane & Paquot, Magali (red.): *Electronic Lexicography*, s. 1–11. Oxford: Oxford University Press, 532 s. ISBN: 9780199654864. DOI: <https://doi.org/10.1093/acprof:oso/9780199654864.001.0001>. [EN]
- GREFENSTETTE, Gregory (1998): „The Future of Linguistics and Lexicographers: Will there be Lexicographers in the year 3000?” [w:] Fontenelle, T. & Hiligsmann, Philippe & Michiels, Archibal & Moulin, A. & Theissen, Siegfried (red.): *Actes Euralex '98 Proceedings*, s. 25–41. Liège: University of Liège. ISBN: 9782872330911. ISSN: 2521-7100. URL: <https://www.semanticscholar.org/paper/The-Future-of-Linguistics-and-Lexicographers%3A-Will-/ddd50ff1a223472d52ae4701e50fdb244474eaeed>. [EN]
- GRUSZCZYŃSKI, Włodzimierz & LAUS-MĄCZYŃSKA, Krystyna & ROGOWSKA, Marianna & SALONI, Zygmunt & SZPAKOWICZ, Stanisław & ŚWIDZIŃSKI, Marek (1981): „Stopień dokładności opisu słownikowego jako problem językoznawczy” [w:] Saloni, Zygmunt (1981): *Dwa studia z polskiej leksykografii współczesnej*, s. 7–61. Białystok: Sekcja wydawnicza Filii UW w Białymstoku, 88 s. [PL]
- GRUSZCZYŃSKI, Włodzimierz (1997): „Dwa najstarsze drukowane słowniki polskie w wydaniu faksymilowym” [w:] [Francisci Mymeri] *Dictionarium trium linguarum [i] Dictionarius Ioannis Murmellii variarum rerum [wydanie faksymilowe]*, Biblioteka Tradycji Literackich, Collegium Columbinum. Kraków: Collegium Columbinum. ISBN: 83-907059-8-2. [PL]
- GRUSZCZYŃSKI, Włodzimierz (1999): „Słowniki w warsztacie pracy dziennikarza (zarys)”, [w:] Gruszczyński, Włodzimierz & Bralczyk, Gerzy & Majkowska, Grażyna: *Polszczyzna w komunikowaniu publicznym. Prace poświęcone Profesor Halinie Satkiewicz z okazji jubileuszu Jej i Jej Zakładu*, s. 115–151. Warszawa: Oficyna Wydawnicza „Aspra”, 151 s. ISBN: 9788390893761. [PL]
- GRUSZCZYŃSKI, Włodzimierz (2011): „Nie tylko łacina, czyli o językach zestawianych z polszczyzną w słownikach z XVI, XVII i XVIII wieku” [w:] Gruszczyński, Włodzimierz & Polkowska, Laura (red.): *Problemy leksykografii. Historia – metodologia – praktyka*, s. 53–72. Kraków: Lexis, 188 s. [PL]
- GRUSZCZYŃSKI, Włodzimierz (2011): „Nie tylko łacina, czyli o językach zestawianych z polszczyzną w słownikach z XVI, XVII i XVI wieku” [w:] Gruszczyński, Włodzimierz & Polkowska, Laura (red.): *Problemy leksykografii. Historia – metodologia – praktyka*, s. 53-72. Kraków: Wydawnictwo Lexis, 20 s. ISBN: 978-83-89425-80-5. URL: <http://depot.ceon.pl/handle/123456789/1844>. [PL]
- GRZEGORCZYKOWA, Renata (2000): „Współczesne kognitywistyczne ujęcia znaczenia a koncepcja języka Witolda Doroszewskiego” [w:] *Poradnik Językowy*, nr 6/2000, s. 8–13. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343. [PL]
- HANKS, Patrick (2010a): „Compiling a Monolingual Dictionary for Native Speakers” [w:] du Plessis, J.C.M.D. (red.): *Lexikos*, t. 20, s. 580–598. Stellenbosch: Buro Van Die Wat, 778 s. ISBN: 978-0-9814434-2-3. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/20-0-173>. URL: <http://lexikos.journals.ac.za>. [EN]
- HANKS, Patrick (2012b): „Corpus Evidence and Electronic Lexicography” [w:] Granger, Sylviane & Paquot, Magali (red.): *Electronic Lexicography*. Oxford: Oxford University Press, 532 s. ISBN: 9780199654864. DOI: <https://doi.org/10.1093/acprof:oso/9780199654864.001.0001>. [Wykorzystano wersję dostępną na stronie autora:] URL: http://www.patrickhanks.com/uploads/5/1/4/9/5149363/hanks_2012f.pdf. [EN]
- HANKS, Patrick (2013): „Chapter 22: Lexicography from Earliest Times to the Present Present” (wersja w niewielkim stopniu edytowana względem oryginalnej publikacji) [w:] Allan, Keith (red.): *The Oxford Handbook of the History of Linguistics*, s. 503–536. Oxford: Oxford University Press, 924 s. DOI:

<https://doi.org/10.1093/oxfordhb/9780199585847.001.0001>. URL: http://www.patrickhanks.com/uploads/5/1/4/9/5149363/2012d-lexicography_from_earliest_times.pdf. [EN]

- HARGRAVES, Orin (2018): „Information retrieval for lexicographic purposes” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 701–714. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- HARTMANN, Reinhard Rudolf Karl & JAMES, Gregory (2002): *Dictionary of Lexicography*. London and New York: Routledge. ISBN 0-203-01768-4 (e-book). [EN]
- HARTMANN, Reinhard Rudolf Karl (2018): „Towards an international directory of lexicography” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 767–786. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- HASPELMATH, Martin (2002): *Understanding morphology*. London & New York: Arnold & Oxford University Press, 290 s. ISBN: 9780340950012. [EN]
- HAUSMANN, Franz Josef (1986): „The training and professional development of lexicographers in Germany” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): Fulbright Papers, t. 1, s. 101–110. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8. [EN]
- HAWKINS, Kevin & DALMAU, Michelle & BAUMAN, Syd (2018): *Best Practices for TEI in Libraries* (wersja 4.0). URL: <https://tei-c.org/extra/teiinlibraries/4.0.0/bptl-driver.html>. [EN]
- HEID, Ulrich (2014): „2. The impact of computational lexicography” [w:] Gouws, Rufus Hjalmar & Heid, Ulrich & Schweickard, Wolfgang & Wiegand, Herbert Ernst: *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 24–30. Berlin & Boston: De Gruyter Mouton. ISBN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.24>. [EN]
- HEMMING, T. D. (1968): „Lexicology and Old French” [w:] *The Modern Language Review*, t. 63, nr 4. Cambridge: Modern Humanities Research Association, s. 818–823. ISSN: 0026-7937. e-ISSN: 2222-4319. URL: <https://www.jstor.org/stable/3723739>. [EN]
- HORODYSKA, Halina (1986): „Rola prac nad Słownikiem języka polskiego Karłowicza-Kryńskiego-Niedźwiedzkiego w integrowaniu życia naukowego w Warszawie w końcu XIX w.” [w:] *Kwartalnik Historii Nauki i Techniki*, nr 31, z. 1, s. 149–167. Warszawa: Instytut Historii Nauki im. L. i A. Birkenmajerów PAN. ISSN: 0023-589X. e-ISSN: 2657-4020. URL: https://bazhum.muzhp.pl/media/files/Kwartalnik_Historii_Nauki_i_Techniki/Kwartalnik_Historii_Nauki_i_Techniki-r1986-t31-n1/Kwartalnik_Historii_Nauki_i_Techniki-r1986-t31-n1-s149-167/Kwartalnik_Historii_Nauki_i_Techniki-r1986-t31-n1-s149-167.pdf. [PL]
- HORODYSKA, Halina (1989): „Nieznani współpracownicy redakcyjni *Słownika języka polskiego* Jana Karłowicza, Antoniego Kryńskiego, Władysława Niedźwiedzkiego” [w:] Buttlerowa, Danuta (red.): *Poradnik Językowy*, z 9–10/1989, s. 656–661. Warszawa: Organ Towarzystwa Kultury Języka. [PL]
- HORODYSKA, Halina (2008): „Dwieście lat *Słownika języka polskiego* Samuela Bogumiła Lindego” [w:] *Kwartalnik Historii Nauki i Techniki*, nr 53, z. 1, s. 254–259. Warszawa: Instytut Historii Nauki im. L. i A. Birkenmajerów PAN. ISSN: 0023-589X. e-ISSN: 2657-4020. URL: https://bazhum.muzhp.pl/media/files/Kwartalnik_Historii_Nauki_i_Techniki/Kwartalnik_Historii_Nauki_i_Techniki-r2008-t53-n1/Kwartalnik_Historii_Nauki_i_Techniki-r2008-t53-n1-s254-259/Kwartalnik_Historii_Nauki_i_Techniki-r2008-t53-n1-s254-259.pdf. [PL]
- HRABEC, Stefan & PEPLOWSKI, Franciszek (1963): *Wiadomości o autorach i dziełach cytowanych w Słowniku Lindego*. Warszawa: Wiedza Powszechna, 271 s. [PL]
- ILSON, Robert (1986b): „British and American Lexicography” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): Fulbright Papers, t. 1, s. 51–71. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8.

- JAMES, Gregory (2000): *Colporu! : A History of Tamil Dictionaries*. Chennai: Cre-A, 908 s. ISBN: 978-8185602769. [PL]
- JANKOWIAK, Lucyna Agnieszka (2013): „O potrzebie wydzielenia kwalifikatorów etymologicznych” [w:] Migdał, Jolanta & Piotrowska-Wojaczyk, Agnieszka (red.): *Cum reverentia, gratia, amicitia... Księga jubileuszowa dedykowana Profesorowi Bogdanowi Walczakowi*, s. 609–613. Poznań: Wydawnictwo Rys. ISBN: 978-83-63664-05-3. [PL]
- JOHNSON, Samuel (1747): *The Plan of a Dictionary of the English Language: addressed to the Right Honourable Philip Dormer, Earl of Chesterfield*. London: Printed for J. and P. Knapton [etc.]. URL: <https://www.loc.gov/item/11007677/>. [EN]
- JOHNSON, Samuel (1963): „Preface to the Dictionary” [w:] McAdam, E. L. Jr. & Milne, George (red.): *Jonhson's Dictionary: a Modern Selection*. New York: Pantheon Books. [EN]
- JOST, David (2014): „Digital Change: The Benefits of Control” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t. 35, z. 1, s. 299–302. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.2014.0011>. [EN]
- JUNGMANN, Josef (1835): *Slownjk česko-německý Josefa Jungmanna*, t. I (A–J). Praha: Pomocj Českého Museum / Knížecí arcibiskupská knihtiskárna. URL: <https://vokabular.ujc.cas.cz/moduly/slovníky/digitalni-kopie-detail/JgSlov01>. [CZ]
- KAHANE, Henry & KAHANE, Renée (1992): „The Dictionary as Ideology: Sixteen Case Studies” [w:] Zgusta, Ladislav (red.): *History, Language, and Lexicographers*, s. 19–76 [seria:] Allén, Sture & Corbin, Pierre & Hartmann, Reinhard Rudolf Karl & Hausmann, Franz Josef & Heid, Ulrich & Reichmann, Oskar & Zgusta, Ladislav: (red.) *Lexicographica. Series Maior*, t. 41. Max Niemeyer Verlag: Tübingen. ISBN: 3-484-30941-5. ISSN: 0175-9264. [EN]
- KAMMERER, Matthias & WIEGAND, Herbert Ernst (1998): „Über die textuelle Rahmenstruktur von Printwörterbüchern. Präzisierung und weiterführende Überlegungen” [w:] Gouws, Rufus Hjalmar & Heid, Ulrich & Herbst, Thomas et al. (red.): *Lexicographica*, t. 14, s. 224–238. Berlin: De Gruyter. ISSN: 1865-9403. e-ISSN: 2224-0039. URL: <https://www.proquest.com/docview/1859872664>. [DE]
- KANIA, Stanisław & TOKARSKI, Jan (1984): *Zarys leksykologii i leksykografii polskiej*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne, 256 s. ISBN: 83-02-02354-X. [PL]
- KAPICOVÁ, Eva (2011): *Přístupy k řešení digitalizace dokumentů* (praca inżynierska). Praga: Vysoká škola ekonomická v Praze. [CZ]
- KARAŚ, Mieczysław (1965): „O «Słowniku gwar polskich»” [w:] *Język Polski*, r. XLV, z. 5, s. 263–278. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. [PL]
- KARŁOWICZ, Jan Aleksander & KRYŃSKI, Adam Antoni & NIEDŹWIEDZKI, Władysław Marcin (red.) (1900–1927): *Słownik języka polskiego*, t. I–VIII. Warszawa: nakładem zbiorowym. [PL]
- KARŁOWICZ, Jan Aleksander (1871): *Poradnik dla zbierających rzeczy ludowe*. Warszawa: Skład główny w Księgarni M. Olgerbranda, 13 s. URL: <http://polona.pl/item/33244>. [PL]
- KARŁOWICZ, Jan Aleksander (1876): *Przyczynki do projektu wielkiego słownika polskiego* (Osobne odbicie ze Sprawozdań Akademii Umiejętności 1876). URL: <https://fbc.pionier.net.pl/details/nnl3qhn>. [PL]
- KARŁOWICZ, Jan Aleksander (1894–1905): *Słownik wyrazów obcego a mniej jasnego pochodzenia, używanych w języku polskim*. Kraków: Nakładem autora. Skład główny w księgarni G. Gebethnera i spółki. URL: <https://fbc.pionier.net.pl/details/nnl3vm3>.
- KARŁOWICZ, Jan Aleksander (1900–1911): *Słownik gwar polskich*. Kraków, nakładem Akademii Umiejętności, 458 s. [PL]
- KARŁOWICZ, Jan Aleksander (1906): *Lud – rys ludoznawstwa polskiego* (odbitka z: Polska. Obrazy i Opisy, z. 1, s. 71–215). Lwów, 143 s. URL: <http://fbc.pionier.net.pl/id/oai:bc.wbp.lublin.pl:2951>. [PL]

- KARLOWICZ, Jan Aleksander et al. (1895): *Słownik języka polskiego. Arkusz próbny*. Warszawa: Drukarnia E. Lubowskiego i S-ki. [PL]
- KARPINSKA, Laura (2020): „English-Latvian Dictionaries in the Age of Electronic Lexicography” [w:] *Baltic Journal of English Language, Literature and Culture*, t. 10, s. 83–99. ISSN: 1691-9971. e-ISSN: 2501-0395. DOI: <https://doi.org/10.22364/BJELLC.10.2020.06>. [EN]
- KASZEWSKI, Marek (2014): *Studia z leksykografii historycznej*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych UNIVERSITAS, 134 s. ISBN: 978-83-242-2675-7. [PL]
- KASZEWSKI, Marek (2015): „Cyfrowe wersje dawnych dykcjonarzy języka polskiego jako źródło dla porównawczych badań materiałowych (M.A. Troc i J.S. Bandtkie)” [w:] *Świat Tekstów – Rocznik Słupski*, nr 13, s. 94–100. Słupsk: Wydawnictwo Naukowe Akademii Pomorskiej w Słupsku. ISSN: 2083-4721. URL: <https://swiattekstow.apsl.edu.pl/images/publikacja/13/kaszewski.pdf>. [PL]
- KAY, Christian J. (1983): „The Historical Thesaurus of English” [w:] Hartmann, Reinhard Rudolf Karl (red.): *LEXeter '83: proceedings: Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*, s. 87–100, [seria:] *Lexicographica. Series Maior*, t. 1. Tübingen: Max Niemeyer Verlag. ISBN: 9783484309012. e-ISSN: 9783111593166. ISSN: 0175-9264. [EN]
- KENSTOWICZ, Michael J. & KISSEBERTH, Charles W. (1979): *Generative Phonology: Description and Theory*. Cambridge: Academic Press, 459 s. ISBN: 978-0124051607. [EN]
- KNOWLES, Francis E. (1990): „The Computer in Lexicography” [w:] Hartmann, Reinhard Rudolf Karl (red.): *LEXeter '83: proceedings: Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*, s. 301–314, [seria:] *Lexicographica. Series Maior*, t. 1. Tübingen: Max Niemeyer Verlag. ISBN: 9783484309012. e-ISSN: 9783111593166. ISSN: 0175-9264. URL: <https://euralex.org/publications/dictionaries-and-computers/>. [EN]
- KONIUSZ, Elżbieta (2013): „Współlistnienie języków na ziemiach byłego Wielkiego Księstwa Litewskiego w świetle prac Jana Karłowicza” [w:] Pomorska, Joanna (red.): *Acta Baltico-Slavica. Archeologia, Historia, Ethnographia et Linguarum Scientia*, t. 37, s. 41–51. Warszawa: Sławistyczny Ośrodek Wydawniczy, 630 s. ISSN: 0065-1044. [PL]
- KORBUT, Gabriel (1895): „Z językoznawstwa” [w:] Kiersz, Władysław (red.): *Głos. Tygodnik literacko-społeczno-polityczny*, r. 13, 28 XII, s. 294–296. Warszawa. [PL]
- KORBUT, Gabriel (1900): „Odpowiedź na replikę” [w:] *Kurjer Warszawski*, nr 15, s. 5. Warszawa: [s.n.]. [PL]
- KRYŃSKI, Adam Antoni (1903): *Jan Karłowicz (1836–1903). Zarys życia i prac*. Warszawa: Skład główny w księgarni M. Arcta, 27 s. URL: <https://www.sbc.org.pl/dlibra/show-content/publication/edition/97192?id=97192>. [PL]
- KRYŃSKI, Adam Antoni (1904): „Jan Karłowicz jako językoznawca” [w:] *Życie i prace Jana Karłowicza (1836–1903)*, s. 120–164. Warszawa: Skład główny w redakcji „Wisły”, 379 s. [PL]
- KRYSIĄK, Patrycja (2016): „Feminytywa w polskiej tradycji leksykograficznej” [w:] Miodek, Jan & Wysoczański, Włodzimierz (red.): *Rozprawy Komisji Językowej*, t. XLII, s. 83–90. Wrocław: Wydawnictwo Wrocławskiego Towarzystwa Naukowego, 120 s. ISSN: 2451-294X. [PL]
- KUBISZYN-MĘDRALA, Zofia (2007): „Żeńskie nazwy tytułów i zawodów w słownikach współczesnego języka polskiego” [w:] *LingVaria*, r. II (2007), nr 1 (3), s. 31–40. Kraków: „Księgarnia Akademicka”. ISSN: 1896-2122. e-ISSN: 2392-1226. [PL]
- KUCAŁA, Marian (1985): „Leksykografia polska – teoria a praktyka” [w:] Nitsch, Kazimierz (red.): *Język Polski: Organ Towarzystwa Miłośników Języka Polskiego Wydawany z Zasiłkiem Polskiej Akademii Nauk*, r. LXV, nr 2–3. Kraków: Towarzystwa Miłośników Języka Polskiego, s. 181–193. [PL]

- KUPIDURA, Przemysław & MARCINIAK, Jacek & KOZA, Piotr & KOWALCZYK, Michał (2008): „Wybrane przykłady wykorzystania morfologii matematycznej w przetwarzaniu obrazów w teledetekcji”, [w:] *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, t. 18. ISBN: 978-83-61576-08-2. [PL]
- KURASZKIEWICZ, Władysław (1962–1963): *Wyrazy polskie w Słowniku łacińsko-polskim Jana Mączyńskiego, Cz. I. A–O, cz. II. P–Z*. Wrocław: Zakład Narodowy im. Ossolińskich & Wydawnictwo PAN, 321 s. [PL]
- KURZOWA, Zofia (1985): *Polszczyzna Lwowa i Kresów południowo-wschodnich do 1939 roku*, wyd. 2. Warszawa – Kraków: Państwowe Wydawnictwo Naukowe, 551 s. ISBN: 83-01-04570-1. [PL]
- KWAPIEŃ, Ewelina (2006): „Słownictwo polszczyzny XIX wieku w słowniku S.B. Lindego i w *Słowniku języka polskiego* pod redakcją W. Doroszewskiego (na przykładzie haseł rozpoczynających się na literę K)” [w:] *Prace Filologiczne*, t. LI, s. 209–230. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. [PL]
- KWAPIEŃ, Ewelina (2010): „Etymologia rzeczowników zapożyczonych w XIX wieku w SW i SJPD w świetle innych źródeł leksykograficznych” [w:] *Prace Filologiczne*, t. LVIII, s. 219–262. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. [PL]
- KWAPIEŃ, Ewelina (2014): „{Leksykon polszczyzny XIX wieku} – potrzebny, pożyteczny, realny?” [w:] *Prace Filologiczne*, t. LXXV, s. 257–268. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. [PL]
- LANDAU, Sidney I. (2001): *Dictionaries: The Art and Craft of Lexicography*, wyd. II. Cambridge: Cambridge University Press, 477 s. ISBN (twarda): 0-521-78040-3. ISBN (mięka): 0-521-78512-X. [EN]
- LARA, Luis Fernando (1990): „El diccionario entre la tradición y la realidad” [w:] *Dimensiones de la lexicografía: Dimensiones de la lexicografía*, s. 133–156, [seria:] *Jornadas*, t. 116. Ciudad de México: Colegio de México, 247 s. ISBN: 978-9681204549. ISSN: 2591-4154. DOI: <https://doi.org/10.2307/j.ctv26d9mw.9>. [ES]
- LARA, Luis Fernando (2001): „El lenguaje de la lexicografía” [w:] *Ensayos de teoría semántica: lengua natural y lenguajes científicos*, s. 149–174, [seria:] *Jornadas*, t. 135. Ciudad de México: Colegio de México, 261 s. ISBN: 978-9681210151. ISSN: 2591-4154. DOI: <https://doi.org/10.2307/j.ctv3f8p3x>. [ES]
- LATKO, Jowita (2015): „Stanisław Dubisz (red.), *Uniwersalny Słownik Języka Polskiego*, Warszawa 2003” [w:] *Poradnik Językowy*, nr 1/2015, s. 108–112. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”, 720 s. ISSN: 0551-5343. URL: http://www.poradnikjezykowy.uw.edu.pl/wydania/poradnik_jezykowy.720.2015.01.pdf. [PL]
- LE CUN, Yann & BOTTOU, Leon & EROFEEV, Andrei & HAFFNER, Patrick & RIEMERS, Bill (2001): „DjVu document browsing with on-demand loading and rendering of image components” [w:] *Internet Imaging*. San Jose: Society of Photo-Optical Instrumentation Engineers. URL: <http://leon.bottou.org/papers/lecun-2001>. [EN]
- LEMNITZER, Lothar & ROMARY, Laurent & WITT, Andreas (2013): „Representing human and machine dictionaries in markup languages (SGML, XML)” [w:] Gouws, Rufus Hjalmar et al. (red.): *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 1195–1209. Berlin: De Gruyter Mouton. ISBN: 9783110238129. e-ISBN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.1195>. [EN]
- LENARTOWICZ-ZAGRODŃA, Anna (2018): „O układzie gniazdowym artykułów hasłowych w dawnych słownikach” [w:] *Studia Językoznawcze: Synchroniczne i diachroniczne aspekty badań polszczyzny*, s. 229–241. Szczecin: Uniwersytet Szczeciński. ISSN: 1730-4180. e-ISSN: 2353-3161. DOI: <https://doi.org/10.18276/sj.2018.17-14>. URL: <https://wnus.edu.pl/sj/file/article/download/15215.pdf>. [PL]
- LEROYER, Patrick (2018): „Dictionaries for text reception” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 250–266. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]

- LEW, Robert (2007): „Linguistic semantics and lexicography: A troubled relationship” [w:] Fabiszak, Małgorzata (red.): *Language and meaning. Cognitive and functional perspectives*, [seria:] *Polish Studies in English Language and Literature*, s. 217–224. Frankfurt am Main: Peter Lang, 342 s. ISBN: 978-3631561584. [EN]
- LEWASZKIEWICZ, Tadeusz (1980): *Panslawistyczne osobliwości leksykalne S. B. Lindego i jego projekt stworzenia wspólnego języka słowiańskiego*, [seria:] Jakóbiec, Marian (red.): *Monografie slawistyczne*, t. 42. Wrocław: Zakład Narodowy im. Ossolińskich. URL: <https://kpbc.umk.pl/publication/17302>. ISSN: 0077-0531. ISBN: 83-04-00568-9. [PL]
- LEWICKI, Andrzej Maria (1993): „Leksykografia” [w:] Polański, Kazimierz: *Encyklopedia językoznawstwa ogólnego*, s. 301–304. Wrocław: Zakład Narodowy im. Ossolińskich. ISBN: 83-04-04680-6. [PL]
- LOISEAU, Sylvain (2013): „Les genres de l'exemple et de la définition dans un corpus de dictionnaires monolingues” [w:] Gasiglia, Nathalie (red.): *Interpréter les dictionnaires : pluralité d'approches*, [seria:] Corbin, Pierre (red.): *Lexique*, t. 21, s. 119–162. Villeneuve-d'Ascq: Presses Universitaires du Septentrion, 252 s. ISBN: 978-2-7574-0437-9. ISSN: 0756-7138. [FR]
- LOMBARD, F. J. (1994): „Lexicographer, Linguist and Dictionary User: An Uneasy Triangle?” [w:] *Lexikos*, t. 4, s. 204–214. Stellenbosch: Buro Van Die Wat. ISBN: 978-0-9814434-2-3. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/20-0-173>. URL: <https://lexikos.journals.ac.za/pub/article/view/1081/597>. [EN]
- LUBAŚ, Władysław (1994): „Dwa suplementy” [w:] *Język Polski*, r. LXXIV, z. 2, s. 119–123. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. URL: <http://mbc.malopolska.pl/dlibra/plain-content?id=30083>. [PL]
- ŁEMPICKA, Zofia (1958): „Prohibita w Słowniku Warszawskim” [w:] Doroszewski, Witold (red.): *Poradnik Językowy*, z. 5 (160)/1958, s. 203–208. Warszawa: Państwowe Wydawnictwo „Wiedza Powszechna”. [PL]
- ŁOPACIŃSKI, Hieronim (1901): „Gramatyka języka polskiego A. A. Kryńskiego”. Warszawa: E. Weydel, s. 738–755. (Odbitka z: *Ateneum* 1901, t. 2, z. 3) [PL]
- ŁOŚ, Jan (1900): „Из польской лексикографии” [w:] *Известия Императорской Академии наук*, r. V, t. IV, s. 1357–1374. Санкт-Петербург: в типографии Императорской Академии наук. [RU]
- ŁOŚ, Jan (1904): „Znaczenie «Słownika gwar polskich»” [w:] *Życie i prace Jana Karłowicza (1836-1903)*, s. 176–180. Warszawa: Skład główny w redakcji „Wisły”, 379 s. [PL]
- ŁUKASIK, Marek (2019): „Polska leksykografia specjalistyczna w Drugiej Rzeczypospolitej” [w:] Przybylska, Renata & Batko-Tokarz, Barbara & Pręczyk-Kisielak, Sylwia (red.): *Niepodległa wobec języka polskiego*, s. 179–189. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 375 s. ISBN: 978-83-233-4738-5. e-ISSN: 978-83-233-7021-5. [PL]
- MAJDAK, Magdalena (2006a): „Słownik warszawski w korespondencji jego twórców” [w:] *Prace Filologiczne*, t. LI, s. 231–246. Warszawa : Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. URL: <https://www.ceeol.com/search/article-detail?id=272377>. [PL]
- MAJDAK, Magdalena (2006b): „O cytatach w Słowniku warszawskim” [w:] *Studia z Filologii Polskiej i Słowiańskiej*, t. 41, s. 95–113. Warszawa: Instytut Sławistyki Polskiej Akademii Nauk. ISSN: 2392-2435. URL: <https://www.ceeol.com/search/article-detail?id=4464>. [PL]
- MAJDAK, Magdalena (2007): „Kilka uwag o definicjach w {Słowniku warszawskim}” [w:] *Prace Filologiczne*, t. LII, s. 223–240. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. URL: <https://www.ceeol.com/search/article-detail?id=94361>. [PL]
- MAJDAK, Magdalena (2008a): *Słownik warszawski: koncepcja – realizacja – recepcja*. Wydział Polonistyki Uniwersytetu Warszawskiego, 254 s. ISBN: 978-83-89663-99-3. [PL]

- MAJDAK, Magdalena (2008b): „Karta z dziejów leksykografii polskiej” [w:] Krevelj, Stela Letica (red.): *Strani jezici*, nr 37, s. 239–249. Zagreb: Croatian Philological Association. ISSN: 0351-0840. e-ISSN: 2459-671X. URL: https://stranijezici.ffzg.unizg.hr/wp-content/uploads/2020/02/SJ_37_3_4.pdf. [PL]
- MAJDAK, Magdalena (2014): „Jan Karłowicz, Adam Antoni Kryński, Władysław Niedźwiedzki (red.), *Słownik języka polskiego, Warszawa 1900–1927*” [w:] *Poradnik Językowy*, nr 6, s. 93–100. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343. [PL]
- MAJEWSKA, Małgorzata B. (2010): „Dygitalizacja bazodanowa na przykładzie *Słownika języka Jana Chryzostoma Paska*” [w:] *Prace Filologiczne*, t. LVIII, s. 263–274. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. URL: <https://www.ceeol.com/search/article-detail?id=172191>. [PL]
- MAJEWSKA, Małgorzata B. (2014): „Statystyka *Słownika wileńskiego*. Informacja etymologiczna” [w:] *Poradnik Językowy*, nr 6, s. 37–51. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343.
- MAJEWSKA, Małgorzata B. (2020): *Wileński Słownik języka polskiego... i zagadnienia jego dygitalizacji*. Warszawa: Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego, 493 s. ISBN: 978-83-8090-720-1. e-ISBN: 978-83-8090-721-8. [PL]
- MAJEWSKI, Erazm (1904): „Zarys życia” [w:] *Życie i prace Jana Karłowicza (1836-1903)*, s. 1–30. Warszawa: Skład główny w redakcji „Wisły”, 379 s. [PL]
- MALKIEL, Yakov (1968): *Essays on Linguistic Themes*. Berkeley & Los Angeles: University of California Press, 415 s. [EN]
- MAŃCZAK-WOHLFELD, Elżbieta (1988): „Anglicyzmy w *Słowniku warszawskim*” [w:] *Język Polski*, r. LXVIII, z. 1, s. 24–29. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. [PL]
- MARIAK, Leonarda (1991/1992): „Wyrazy kwalifikowane jako «mało używane» w *Słowniku warszawskim*” [w:] Kuraszkiewicz, Władysław (red.): *Slavia Occidentalis*, t. 48/49, s. 145–158. Poznań: Poznańskie Towarzystwo Przyjaciół Nauk. ISSN: 0081-0002. ISBN: 83-7063-039-1. [PL]
- MARKOWSKA, Ewa (1986): „Sesja poświęcona pamięci profesora Mieczysława Szymczaka” [w:] *Poradnik Językowy*, nr 8/1986. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”, s. 541–543. ISSN: 0551-5343. [PL]
- MARKOWSKI, Andrzej (1986): „Prace Profesora Mieczysława Szymczaka z zakresu leksykologii i leksykografii” [w:] *Poradnik Językowy*, nr 8/1986. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”, s. 530–535. ISSN: 0551-5343. [PL]
- MARKOWSKI, Andrzej (1992): *Polszczyzna końca XX wieku*. Warszawa: Wiedza Powszechna. ISBN: 83-214-0856-7. [PL]
- MARYL, Maciej (2017): „Kim są polscy humaniści cyfrowi?” [w:] *Teksty Drugie*, nr 1/2017. Warszawa: Instytut Badań Literackich Polskiej Akademii Nauk, s. 286-300. ISSN: 0867-0633. URL: <https://journals.openedition.org/td/2058#tocto1n2>. [PL]
- MATUSZCZYK, Bożena (2006): *Słownik języka polskiego S. B. Lindego. Warsztat leksykografa*. Lublin: Wydawnictwo Katolickiego Uniwersytetu Lubelskiego, 311 s. ISBN: 83-7363-371-5. [PL]
- MCCARTY, Willard (2015): „Becoming Interdisciplinary” [w:] Schreibman, Susan & Siemens, Ray & Unsworth, John (red.): *A New Companion to Digital Humanities*, s. 67–83. Hoboken: Wiley-Blackwell. ISBN: 9781118680599. e-ISBN: 9781118680605. DOI: <https://doi.org/10.1002/9781118680605.ch5>. [EN]
- MIKA, Tomasz & MASŁEJ, Dorota (2020): „Glosy jako integralny składnik średniowiecznego tekstu. W poszukiwaniu adekwatnych procedur badawczych” [w:] *LingVaria*, r. XV (2020), nr 1 (29), s. 121–133. Kraków: „Księgarnia Akademicka”, 288 s. ISSN: 1896-2122. e-ISSN: 2392-1226. DOI: <https://doi.org/10.12797/LV.15.2020.29.08>. [PL]

- MIODUNKA, Władysław (1989): *Podstawy leksykologii i leksykografii*, wyd. 1. Warszawa: Państwowe Wydawnictwo Naukowe, 269 s. ISBN: 83-01-09207-6. [PL]
- MOCHIWA, Z. S. M. (2007): „Kiswahili Verbs: A Lexicographical Challenge” [w:] *Lexikos*, t. 17, nr 1, s. 114–133. Stellenbosch: Bureau of the WAT. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.4314/lex.v17i1.51525>. [EN]
- MURRAY, James A. H. (1900): *The Romanes Lecture, 1900. The Evolution of English Lexicography*, [wykład w:] Sheldonian Theatre, Oxford, June 22, 1900. URL: <https://www.gutenberg.org/ebooks/11694>. [EN]
- NEHRING, Władysław (1901): „Pierwsze sprawozdanie z prac nadesłanych na konkurs im. Lindego, Słownik języka polskiego ułożony pod redakcją Jan Karłowicza, Adama Kryńskiego i Władysł. Niedźwiedzkiego, od r. 1898 aż dotąd (r. 1901) 8 zeszytów. Pierwsze 6 zesz. stanowi 1901) 8 zeszytów. Pierwsze 6 zesz. stanowi” [w:] *Materyały i Prace Komisji Językowej Akademii Umiejętności w Krakowie*, t. 1, z. 1, s. 155–161. *Kraków: Nakł. Akademii Umiejętności*. [PL]
- NIELSEN, Sandro (1990): „Lexicographic Macrostructures” [w:] Bergenholtz, Henning (red.): *Hermes*, nr 4, s. 49–66. Aarhus: School of Communication and Culture at Aarhus University. ISSN: 0904-1699. e-ISSN: 1903-1785. URL: <https://tidsskrift.dk/her/article/view/21422/18882>. [EN]
- NITSCH, Kazimierz (1911): „Recenzja «Słownika gwar polskich» J. Karłowicza” [w:] „Rocznik Sławistyczny”, t. IV (1911), s. 199–243. [Przedruk w:] Nitsch, Kazimierz (1958): *Wybór pism polonistycznych*, t. IV: *Pisma polonistyczne*. Wrocław & Kraków, s. 195–225. [PL]
- OBER, Krzysztof (2005): „Proces digitalizacji”, [prezentacja na konferencji:] *Biblioteki cyfrowe*, 21-23 listopada 2005, Poznań. URL: <http://lib.psnc.pl/Content/105/Proces%20digitalizacji.pdf>. [PL]
- OKONIOWA, Joanna (2013): „Słownik gwar polskich Jana Karłowicza. Dziedzictwo. Inspiracje. Wyzwania” [w:] *Prace Filologiczne*, t. LXIV, s. 245–258. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. URL: <https://www.ceeol.com/search/article-detail?id=4803>. [PL]
- OKONIOWA, Joanna (2019): „Słownik gwarowy Kazimierza Nitscha” [w:] *LingVaria*, r. XIV (2019), nr 1 (27), s. 307–315. Kraków: „Księgarnia Akademicka”, 394 s. ISSN: 1896-2122. e-ISSN: 2392-1226. DOI: <https://doi.org/10.12797/LV.14.2019.27.20>. [PL]
- PAŁKA, Patrycja & KWAŚNICKA-JANOWICZ, Agata (2017): *Przewodnik po elektronicznych zasobach językowych dla polonistów*, [seria:] Żmigrodzki, Piotr (red.): *Biblioteczka Towarzystwa Miłośników Języka Polskiego*, nr 28. Kraków: Wydawnictwo JAK. ISBN: 978-83-947837-0-9. e-ISBN: 978-83-947837-1-6. [PL]
- PIELA, Agnieszka (2018): „O frazeologii w słownikach języka polskiego z XIX i początku XX wieku” [w:] *Rozprawy Komisji Językowej Łódzkiego Towarzystwa Naukowego*, t. LXV, s. 139–153. Łódź: Łódzkie Towarzystwo Naukowe. ISSN: 0076-0390. e-ISSN: 2450-9310. DOI: <https://doi.org/10.26485/RKJ/2018/65/9>. [PL]
- PIELA, Agnieszka (2020): „Kłopotliwe kwalifikatory chronologiczne w słownikach języka polskiego” [w:] *Język Polski*, nr 3 (2020), s. 58–72. ISSN: 0021-6941. [PL]
- PIOTROWICZ, Anna (2004): *Słownictwo i frazeologia życia towarzyskiego w polskiej leksykografii XX wieku*. Poznań: Wydawnictwo Naukowe UAM, 503 s. ISBN: 83-232-1378-X. ISSN: 0554-8179. [PL]
- PIOTROWSKA-WOJACZYK, Agnieszka (2011): *Regionalizmy leksykalne w słownikach doby nowopolskiej*, [seria:] *Prace Komisji Językoznawczej*, t. 58. Poznań: Wydawnictwo Poznańskiego Towarzystwa Przyjaciół Nauk, 549 s. ISBN: 978-83-7654-066-5. ISSN: 0079-4678. [PL]
- PIOTROWSKA-WOJACZYK, Agnieszka (2011): *Regionalizmy leksykalne w słownikach doby nowopolskiej*. Poznań: Poznańskie Towarzystwo Przyjaciół Nauk, 550 s. ISBN: 9788376540665. [PL]
- PIOTROWSKI, Michael (2012). *Natural language processing for historical texts* [seria:] Hirst, Graeme (red.): *Synthesis lectures on human language technologies*, t. 17. San Rafael, CA: Morgan & Claypool, 157 s. ISBN:

9781138941601. e-ISBN: 9781315104942. DOI: <https://doi.org/10.2200/S00436ED1V01Y201207HLT017>. [EN]

- PIOTROWSKI, Tadeusz (1993): „Słowniki języka polskiego” [w:] Bartmiński, Jerzy (red.): *Współczesny język polski*. Wrocław: Wiedza o kulturze. ISBN: 83-7033-062-2. [PL]
- PIOTROWSKI, Tadeusz (1994): *Z zagadnień leksykografii*. Warszawa: Państwowe Wydawnictwo Naukowe, 221 s. ISBN: 9788301113926. [PL]
- PIOTROWSKI, Tadeusz (2001): *Zrozumieć leksykografię*, [seria:] *Kolekcja Lingwistyczna PWN*. Warszawa: Polskie Wydawnictwo Naukowe, 244 s. ISBN: 83-01-13693-6. [PL]
- PIOTROWSKI, Tadeusz (2010): „Słowniki w badaniach językowego obrazu świata” [w:] Chlebda, Wojciech (red.): *Etnolingwistyka a leksykografia: Tom poświęcony Profesorowi Jerzemu Bartmińskiemu*, s. 49–56. Opole: Wydawnictwo Uniwersytetu Opolskiego, 248 s. ISBN: 978-83-7395-386-4. [PL]
- PODHAJECKA, Mirosława (2018): „History in Lexicography and Lexicography in History: A Reappraisal” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t. 39, z. 1, s. 131–165. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.2018.0007>. [EN]
- PRESNER, Todd (2010): „Digital Humanities 2.0: A Report on Knowledge, OpenStax CNX” [w:] Bailar, Melissa (red.): *Emerging Disciplines: Shaping New Fields of Scholarly Inquiry in and beyond the Humanities*. URL: <http://cnx.org/contents/2742bb37-7c47-4bee-bb34-0f35bda760f3@6>. [EN]
- PRINKE, Rafał T. (1997): „Techniki, metody i standardy komputerowe w naukach humanistycznych” [w:] Ryszewski, Bohdan (red.): *Metody komputerowe w badaniach i nauczaniu historii: materiały II Symposium Polskiego Oddziału Association for History and Computing, Komisji Metod Komputerowych Polskiego Towarzystwa Historycznego*, Poznań 1 i 2 grudnia 1995 r. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika, 159 s. ISBN: 9788323108399. [PL]
- PRUVOST, Jean (2000): „Rapport de colloque des dictionnaires papieraux dictionnaires électroniques. VIIe Journée des dictionnaires (22 mars 2000)” [w:] *International Journal of Lexicography*, t. 13, z. 3, s. 187–193. ISSN: 0950-3846. e-ISSN: 1477-4577. DOI: <https://doi.org/10.1093/ijl/13.3.187>. [FR]
- PRZYWARA, Agnieszka (2017): „Jedenastojęzyczny słownik Calepinusa z roku 1598 przechowywany w Bibliotece Instytutu Teologiczno-Pastoralnego i Wyższego Seminarium Duchownego w Rzeszowie” [w:] *Resovia Sacra. Studia Teologiczno-Filozoficzne Diecezji Rzeszowskiej*, r. 24 (2017), s. 399–414. Rzeszów: Instytut Teologiczno-Pastoralny im. św. J. S. Pelczara w Rzeszowie. URL: <http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-4200fbeb-255b-421a-a80e-751c10d31946>. [PL]
- PUDZIKOWSKI Józef & WAWRZYŃCZYK, Jan (2017): „Hasłownikolog Linde” [w:] Wawrzyńczyk, Jan & Wierchoń, Piotr (red.): *Wokół 300 tysięcy polskich słów. Wstęp do hasłownikologii*, s. 281–283. Warszawa: BEL Studio Sp. z o.o., 398 s. ISBN: 978-83-7798-248-8. [PL]
- PUZYNINA, Jadwiga (1961): „*Thesaurus*” Grzegorza Knapiusza: *Siedemnastowieczny warsztat pracy nad językiem polskim*, [seria:] Doroszewski, Witold & Klemensiewicz, Zenon & Safarewicz, Jan & Stieber, Zdzisław (red.): *Prace Językoznawcze*, t. 29. Wrocław & Warszawa & Kraków: Zakład Narodowy im. Ossolińskich & Wydawnictwo Polskiej Akademii Nauk. [PL]
- PUZYNINA, Jadwiga (1994): „Refleksje dotyczące *Suplementu do Słownika języka polskiego* pod redakcją Mieczysława Szymczaka” [w:] Handke, Kwiryna & Dalewska-Greń, Hanna (red.): *Polszczyzna a/i Polacy u schyłku XX wieku*, s. 381–387. Warszawa: Slawistyczny Ośrodek Wydawniczy. ISBN: 83-86619-00-7. [PL]
- QUIRK, Randolph (1986): „Opening Remarks” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): *Fulbright Papers*, t. 1, s. 1–6. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8. [EN]
- RAFAJŁOWICZ, Ewaryst & RAFAJŁOWICZ, Wojciech (2010): *Wstęp do przetwarzania obrazów przemysłowych*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej. ISBN: 978-83-7493-613-2. [PL]

- RAK, Maciej (2017): „Z dziejów *Słownika gwar polskich* Jana Karłowicza. Kartoteka i działania wydawnicze” [w:] *LingVaria*, r. XII (2017), nr 2 (24), s. 263–288. Kraków: „Księgarnia Akademicka”. ISSN: 1896-2122. e-ISSN: 2392-1226. DOI: <https://doi.org/10.12797/LV.12.2017.24.17>. [PL]
- RAK, Maciej (2018): „„Czwarta piramida» Jana Karłowicza. Kartoteka pierwszego w dziejach polskiej nauki słownika etnolingwistycznego” [w:] *LingVaria*, r. XIII (2018), nr 2 (26), s. 183–200. Kraków: „Księgarnia Akademicka”. ISSN: 1896-2122. e-ISSN: 2392-1226. [PL]
- RAK, Maciej (2021a): *Jan Karłowicz w świetle materiałów archiwalnych*, [w:] Rak, Maciej (red.): *Dialektologia, etnolingwistyka i lituanistyka*, [seria:] Skarżyński, Mirosław (red.): *Biblioteka LingVariów*, t. 30. Kraków: Wydawnictwo Księgarnia Akademicka, 265 s. ISBN: 978-83-8138-419-3. e-ISBN: 978-83-8138-420-9. DOI: <https://doi.org/10.12797/9788381384209>. [PL]
- RAK, Maciej (2021b): „Początki polskiej dialektologii (na podstawie książki *Jan Karłowicz w świetle materiałów archiwalnych. Dialektologia, etnolingwistyka i lituanistyka*)”, [wystąpienie na:] Posiedzenie komisji Komitetu Językoznawstwa PAN (18.11.2021). [PL]
- REED, Joseph W. (1962): „Noah Webster’s Debt to Samuel Johnson” [w:] *American Speech*, nr 37, z. 2. Durham: Duke University Press, s. 96–105. [EN]
- REICHAN, Jerzy (red.) (1999): *Indeks alfabetyczny wyrazów z kartoteki Słownika gwar polskich*, t. 1 i 2. Kraków: Rymut, Kazimierz & Smółkowa, Teresa & Bobrowski, Ireneusz (red.): Prace Instytutu Języka Polskiego, t. 109. Kraków: Wydawnictwo Polska Akademia Nauk. ISSN: 0208-4074. [PL]
- ROSSI, Mario & ČERMÁK, Petr (2008): „Pražský lingvistický kroužek a zapomenuté příspěvky k Ottovu slovníku naučnému nové doby” [w:] *Slovo a slovesnost*, nr 1–2, s. 143–157. Praha: Ústav pro jazyk český Akademie věd České republiky. ISSN: 0037-7031. e-ISSN: 2571-0885. URL: <http://sas.ujc.cas.cz/>. [cz]
- RUDNICKA, Ewa Katarzyna (2016): „Słownik jako tekst kultury” [w:] *Język Polski*, nr 4 (2016), s. 59–69. ISSN: 0021-6941. URL: <https://www.ceeol.com/search/article-detail?id=475670>. [PL]
- RUNDELL, Michael (2012): „The road to automated lexicography: An editor’s viewpoint” [w:] Granger, Sylviane & Paquot, Magali (red.): *Electronic Lexicography*, s. 15–30. Oxford: Oxford University Press, 532 s. ISBN: 9780199654864. DOI: <https://doi.org/10.1093/acprof:oso/9780199654864.001.0001>. [EN]
- RUSTICUS, J. (1895): [*Arkusz próbny*], *Gazeta Warszawska*, nr 289, 19 X, s. 2. Warszawa: [wydawca nieznany]. [PL]
- RYKACZEWSKI, Erazm (1866): *Słownik języka polskiego podług Lindego i innych nowszych źródeł*, t. I (A-O). Berlin: B. Behr. URL: <https://polona.pl/item/slownik-jezyka-polskiego-podlug-lindego-i-innych-nowszych-zrodel-t-1,OTI4OTU2OTQ/6/#info:metadata>. [PL]
- SAARELA, Márten Söderblom (2014): „Shape and Sound: Organizing Dictionaries in Late Imperial China” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t. 35, z. 1, s. 187–208. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.2014.0012>. [EN]
- SALONI, Zygmunt (1986): „W sprawie nowego wydania Słownika języka polskiego PAN pod red. Witolda Doroszewskiego” [w:] *Język Polski*, r. LXVI, z. 5, s. 161–180; z. 5, s. 362–367. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. [PL]
- SALONI, Zygmunt (1992): „Rygorystyczny opis polskiej deklinacji przymiotnikowej” [w:] *Zeszyty Naukowe Wydziału Humanistycznego Uniwersytetu Gdańskiego. Prace Językoznawcze*, t. 16, s. 215–228. Gdańsk: Uniwersytet Gdański. ISSN: 1230-607X. [PL]
- SALONI, Zygmunt (2007): „O konwencjonalnych znakach pomocniczych w tekście polskim” [w:] *Białostockie Archiwum Językowe*, nr 7, s. 125–129. Białystok: Wydawnictwo Uniwersytetu w Białymstoku. ISSN: 1641-6961. DOI: <https://doi.org/10.15290/baj.2007.07.11>. [PL]
- SCHRÖDER, Martin (1997): „Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projektes «Deutsches Dialektwörterbuch»” [w:] Schmidt,

- Jürgen Erich: *Zeitschrift für Dialektologie und Linguistik*, nr 64, z. 1, s. 57–66. ISSN: 0044-1449. e-ISSN: 2366-2395. URL: <https://www.jstor.org/stable/40504188>. [DE]
- SECO, Manuel (1987): *Estudios de lexicografía española*. Madrid: PARANINFO, 258 s. ISBN: 9788428315432. [ES]
- SIMPSON, John (2009): „The Oxford English Dictionary: What does the Future Hold for Historical Lexicography?” [w:] Bruti, Silvia & Celia, Roberta & Foschi Albert, Marina (red.): *Perspectives on Lexicography in Italy and Europe*, s. 55–78. Cambridge: Cambridge Scholars Publishing, 341 s. ISBN: 978-1-4438-1263-4. [EN]
- SŁAWIŃSKI, Fabian Ferdynand (1873): *Obliczenie wyrazów zawartych w trzech słownikach 1) Lindego, 2) w Wi-
leńskim 3) Rykaczewskiego*. Warszawa: Skład Główny u Gebethnera i Wolffa, 38 s. [PL]
- STOKES, Peter A. & PIERAZZO, Elena (2009): „Encoding the Language of Landscape: XML and Databases at the Service of Anglo-Saxon Lexicography” [w:] Bruti, Silvia & Celia, Roberta & Foschi Albert, Marina (red.): *Perspectives on Lexicography in Italy and Europe*, s. 203–238. Cambridge: Cambridge Scholars Publishing, 341 s. ISBN: 978-1-4438-1263-4. [EN]
- STRZELCZYK, Piotr (2013). „Standard Unicode w typografii” [w:] *Acta Poligraphica*, nr 1, s. 41–50. Warszawa: Centralny Ośrodek Badawczo-Rozwojowy Przemysłu Poligraficznego. ISSN: 2299-9981. [PL]
- STRZYŻEWSKI, Mirosław (1994): Sylwetka Samuela Bogumiła Lindego z Torunia, [w:] Łapicz, Czesław (red.): *Polszczyzna dawna i współczesna. Materiały z ogólnopolskich konferencji językoznawczych*, s. 139–153. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, 174 s. ISBN 83-231-0480-8. URL: <https://kpbc.umk.pl/dlibra/publication/16504/edition/25914/content>. [PL]
- SVARTVIK, Jan (2011): „Corpus Linguistics Comes of Age” [w:] Svartvik, Jan (red.): *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*, [seria:] *Trends in Linguistics. Studies and Monographs [TiLSM]*, t. 65. Berlin: DeGruyter Mouton, 487 s. ISBN: 9783110128260. e-ISSN: 9783110867275. ISSN: 1861-4302. e-ISSN: 2199-3734. DOI: <https://doi.org/10.1515/9783110867275.7>. [EN]
- SVENSÉN, Bo (2009): *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press, 552 s. [EN]
- SVENSSON, Patrik (2010): „The Landscape of Digital Humanities” [w:] *Digital Humanities Quarterly*, nr 4 (2010), z. 1. Alliance of Digital Humanities Organizations. e-ISSN: 1938-4122. URL: <http://digitalhumanities.org:8081/dhq/vol/4/1/000080/000080.html>. [EN]
- SZAFRAN, Krzysztof (1998): „Analizator morfologiczny SAM-96” [w:] *Postscriptum Polonistyczne*, nr 24-25 (1997–1998), s. 1–35. Katowice: Uniwersytet Śląski w Katowicach, 80 s. ISSN: 1898-1593. [PL]
- SZAFRAN, Krzysztof (2004): „Ku elektronicznej wersji *Słownika polszczyzny XVI wieku*” [w:] Polański, Kazimierz (red.): *Biuletyn Polskiego Towarzystwa Językoznawczego = Bulletin de la société polonaise de linguistique*, z. LX (2004), s. 89–97. Kraków: Universitas. ISSN: 0032–3802. [PL]
- SZAFRAN, Krzysztof (2007): *Analiza i formalny opis struktury Słownika polszczyzny XVI wieku*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 218 s. ISBN: 978-83-235-0358-3. [PL]
- SZCZUREK, Grzegorz (2003): „Zastosowanie metod morfologii matematycznej do detekcji i dekompozycji obrazów”, [w:] *Telekomunikacja i Techniki informacyjne*, nr 1–2/2003. Warszawa: Instytut Łączności. ISSN: 1640-1549. e-ISSN: 1899-8933. [PL]
- SZOBER, Stanisław (1928): „Jan Karłowicz (1836-1903)” [w:] *Prace Filologiczne*, t. XIII, s. 546–557. Warszawa: Kasa im. Józefa Mianowskiego. ISSN: 0138-0567. [PL]
- SZOBER, Stanisław (1930): „Władysław Marcin Niedźwiedzki”. Warszawa: [s.n.], s. 362–364. (Odbitka z: *Prace Filologiczne*. Warszawa 1930, t. 15, s. 362–362–364). [PL]
- SZOBER, Stanisław (1934): *Adam Antoni Kryński. Życie i prace*. Warszawa: [s.n.]. (Odbitka z: *Prace Filologiczne*. Warszawa 1934, t. 16, s. 7–39). [PL]

- SZYMCZAK, Mieczysław (red.) (1978–1981): *Słownik języka polskiego*, t. I–III. Warszawa: Polskie Wydawnictwo Naukowe. ISBN: 9788301109066. [PL]
- ŚWITOŃSKI, Adam & STĄPOR, Katarzyna (2004): „Operacje morfologiczne na obrazach binarnych – zastosowanie na potrzeby wizji komputerowej”, [w:] *Studia Informatica*, t. 25, nr 2 (58). Gliwice. ISSN: 1642-0489. [PL]
- TARP, Sven (2008): *Lexicography in the Borderland between Knowledge and Non-Knowledge: General Lexicographical Theory with Particular Focus on Learner’s Lexicography*, [seria:] Corbin, Pierre & Heid, Ulrich & Herbst, Thomas & Malmgren, Sven-Göran & Reichmann, Oskar & Schweickard, Wolfgang & Wiegand, Herbert Ernst (red.): *Lexicographica. Series Maior*, t. 134. Tübingen: Max Niemeyer Verlag. ISBN: 978-3-4X4-39134-5. ISSN: 0175-9264. [EN]
- TARP, Sven (2009): „Reflections on data access in lexicographic works” [w:] Nielsen, Sandro & Tarp, Sven (red.): *Lexicography in the 21st Century: In honour of Henning Bergenholtz*, s. 43–62, [seria:] L’Homme, Marie-Claude & Kageura, Kyo (red.): *Terminology and Lexicography Research and Practice (TLRP)*. Amsterdam & Philadelphia: John Benjamins Publishing Company. ISBN (twarda): 978-90-272-2336-4. ISBN (miękka): 978-90-272-8901-8. [EN]
- TARP, Sven (2011): „Lexicographical and other e-tools for consultation purposes: Towards the individualization of needs satisfaction” [w:] Fuertes-Olivera, Pedro A. & Bergenholtz, Henning (red.): *e-Lexicography: The Internet, Digital Initiatives and Lexicography*, s. 54–70). London, New York: Continuum. ISBN: 978-1-4411-2806-5. [EN]
- TARP, Sven (2018a): „Lexicography as an independent science” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 19–33. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- TARP, Sven (2018b): „The concept of dictionary” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 237–249. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- TRAP-JENSEN, Lars (2006): „Making Dictionaries for Paper or Screen: Implications for Conceptual Design” [w:] Corino, Elisa & Marelllo, Carla & Onesti, Cristina (red.): *Proceedings of the 12th EURALEX International Congress*, s. 349-355. Torino: Edizioni dell’Orso. ISBN: 88-7694-918-6. [EN]
- UŁASZYN, Henryk (1899): „Słownik języka polskiego pod redakcją Jana Karłowicza, Adama Kryńskiego i Władysława Niedźwiedzkiego. Warszawa 1898, zeszyt 1 i 2” [w:] Morawski, Marian Ignacy (red.): *Przegląd Powszechny*, r. LXII, t. 4, s. 269–275. Kraków: Druk. W. L. Anczyca i Spółki. [PL]
- UŁASZYN, Henryk (1900): „Warszawski Słownik języka polskiego.” [w:] *Gazeta Lwowska*, nr 285. Lwów, 10 s. [PL]
- URBAŃCZYK, Stanisław (1964): *Słowniki: ich rodzaje i użyteczność*, [seria:] *Biblioteczka Towarzystwa Miłośników Języka Polskiego*, nr 16. Wrocław: Zakład Narodowy im. Ossolińskich, 47 s. URL: <https://fbc.pionier.net.pl/details/nnRns0R>. [PL]
- URBAŃCZYK, Stanisław (1966) [w:] *Polski Słownik Biograficzny*, t. 12, s. 55. Wrocław: Zakład Narodowy im. Ossolińskich. [PL]
- WALCZAK, Bogdan (1988): „Kwalifikatory w słownikach języka polskiego” [w:] Basaj, Mieczysław & Borek, Henryk & Buttler, Danuta & Kreja, Bogusław & Lubaś, Władysław & Pluta, Feliks & Skorupka, Stanisław & Skubalanka, Teresa & Śmiech, Witold & Urbańczyk, Stanisław (red.): *Wokół języka: Rozprawy i studia poświęcone pamięci profesora Mieczysława Szymczaka*, s. 413–422. Wrocław: Zakład Narodowy im. Ossolińskich, 442 s. ISBN: 978-83-04-02572-8. [PL]
- WALCZAK, Bogdan (1991): *Słownik wileński na tle dziejów polskiej leksykografii*, [seria:] *Filologia polska*, nr 44. Poznań: Uniwersytet im. Adama Mickiewicza, 174 s. ISBN: 9788323203100. [PL]

- WALCZAK, Bogdan (1995): „Norma językowa a praktyka leksykograficzna (rys historyczny)” [w:] Bugajski, Marian (red.): *Norma językowa w polszczyźnie*, s. 81–92. Zielona Góra: Wydawnictwo Wyższej Szkoły Pedagogicznej im. T. Kotarbińskiego. ISBN: 83-85693-59-9. [PL]
- WALCZAK, Bogdan (2006): „Etymologia a leksykografia (doba nowopolska)” [w:] *Prace Filologiczne*, t. LI, s. 421–437. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. [PL]
- WALCZAK, Bogdan (2011): „Co to są dawne słowa?” [w:] Seniów, Adrianna (red.): *Studia Językoznawcze*, nr 10, s. 321–329. Szczecin: Instytut Językoznawstwa Uniwersytetu Szczecińskiego. ISSN: 1730-4180. e-ISSN: 2353-3161. [PL]
- WANICOWA, Zofia (2009): „Z warsztatu badawczego leksykografa-historyka” [w:] *Język Polski*, r. LXXXIX, z. 4–5, s. 255–262. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. URL: <https://www.ceeol.com/search/article-detail?id=10692>. [PL]
- WAWRZYŃCZYK, Jan (1991): „O potrzebie opracowania wyrazów niehasłowych w {Słowniku Języka Polskiego} 1958-1960” [w:] *Zeszyty Naukowe Wyższej Szkoły Pedagogicznej w Bydgoszczy – Studia Filologiczne*, z. 30/32; *Filologia Polska* 1991 (12/13), s. 173–182. Bydgoszcz: Wydawnictwo Uczelniane WSP. ISSN: 0239-5037. [PL]
- WAWRZYŃCZYK, Jan (1995): *O wyrażeniach niehasłowych w Słowniku języka polskiego Witolda Doroszewskiego. Z wyborem przykładów*. Częstochowa: Wyższa Szkoła Pedagogiczna w Częstochowie, 34 s. [PL]
- WAWRZYŃCZYK, Jan (2009a): *Autosuplement do Słownika warszawskiego*. Poznań: SORUS SC Wydawnictwo i Drukarnia Cyfrowa. ISBN: 978-83-89949-72-1. [PL]
- WAWRZYŃCZYK, Jan (2009b): *W poszukiwaniu nowego słownictwa polskiego: Materiały z prasy lat 1932–1933*. Łódź: Instytut Rusycystyki Uniwersytetu Łódzkiego, 19 s. ISBN: 978-83-60416-65-5. [PL]
- WAWRZYŃCZYK, Jan (2010): *Inny „Doroszewski”*. Łask: Oficyna Wydawnicza LEKSEM, 40 s. ISBN 978-83-60178-86-7. [PL]
- WAWRZYŃCZYK, Jan (2011): *Słownictwo nowopolskie. Redatacje*. Warszawa: Katedra Lingwistyki Formalnej Uniwersytetu Warszawskiego. [PL]
- WAWRZYŃCZYK, Jan (2015): *300 tysięcy czy milion(y)? O stanie zasobów słownictwa polskiego w dniu 31 grudnia 2000 r.* Warszawa: Miła Hoshi, 14 s. [PL]
- WĘGRZYNEK, Katarzyna (1995): „Projekt komputerowego korpusu współczesnych tekstów polskich” [w:] *Język Polski*, r. LXXV, z. 4–5, s. 332–341. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. URL: <http://mbc.malopolska.pl/dlibra/docmetadata?id=30336&from=publication>. [PL]
- WHITCUT, Janet (1986): „Training the dictionary users” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): Fulbright Papers, t. 1, s. 111–122. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8. [EN]
- WIECZOREK, Aleksandra (2021): „Integracja *Elektronicznego słownika języka polskiego XVII i XVIII wieku* i Elektronicznego Korpusu Tekstów Polskich z XVII i XVIII Wieku okiem użytkownika i redaktora” [w:] Horyń, Ewa & Młynarczyk, Ewa & Żmigrodzki, Piotr (red.): *Język polski – między tradycją a współczesnością. Księga jubileuszowa z okazji stulecia Towarzystwa Miłośników Języka Polskiego*, s. 547–560, [seria:] *Prace Monograficzne*, t. 1030. Kraków: Wydawnictwo Naukowe Uniwersytetu Pedagogicznego, 694 s. ISSN: 2450-7865. ISBN: 978-83-8084-625-8. e-ISSN: 978-83-8084-580-0. DOI: <https://doi.org/10.24917/9788380846258>. URL: <https://rep.up.krakow.pl/xmlui/handle/11716/10824>. [PL]
- WIEGAND, Herbert Ernst & BEER, Sandra & GOUWS, Rufus Hjalmar (2013): „8. Textual structures in printed dictionaries. An overview” [w:] Gouws, Rufus Hjalmar et al. (red.): *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 31–73. Berlin: De Gruyter Mouton. ISBN: 9783110238129. e-ISSN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.253>. [EN]

- WIEGAND, Herbert Ernst (1996): „Textual condensation in printed dictionaries: a theoretical draft” [w:] *Lexikos*, t. 6, s. 133–158. Stellenbosch: Buro Van Die Wat. ISBN: 978-0-9814434-2-3. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/6-1-1029>. URL: <http://lexikos.journals.ac.za>. [EN]
- WIEGAND, Herbert Ernst (2000): „Über Suchbereiche, Suchzonen und ihre textuellen Strukturen in Printwörterbüchern Ein Beitrag zur Theorie der Wörterbuchform” [w:] Wiegand, Herbert Ernst (red.): *Wörterbücher in der Diskussion 4*, s. 233–301. Tübingen: Max Niemeyer. [DE]
- WIERZCHOŃ, PIOTR (2008): *Fotodokumentacja. Chronologizacja. Emendacja. Teoria i praktyka weryfikacji materiału leksykalnego w badaniach lingwistycznych*. Poznań. [PL]
- WIERZCHOŃ, Piotr (2009): „Fotodokumentacja 3.0” [w:] Nowak, Piotr & Nowakowski, Paweł (red.): *Język, Komunikacja, Informacja*, nr 4 (2009), s. 63–80. Poznań: Sorus, 144 s. ISBN: 978-83-63664-15-2. ISSN: 1896-9585. [PL]
- WIERZCHOŃ, Piotr (2010a): „Lingwochronografia na usługach słowotwórstwa gniazdowego” [w:] *Kwartalnik Językoznawczy*, nr 1/2010. Poznań: Uniwersytet im. Adama Mickiewicza, s. 56–71 ISSN: 2081-5441. [PL]
- WIERZCHOŃ, Piotr (2010b): *Depozytorium leksykalne języka polskiego. Nowe fotomateriały z lat 1901–2010*, t. 1. Warszawa: BEL Studio Sp. z o.o. ISBN: 978-83-61208-95-2. [PL]
- WIERZCHOŃ, Piotr (2012): *Depozytorium leksykalne języka polskiego. Nowe fotomateriały z lat 1901–2010*, t. 3. Warszawa: BEL Studio Sp. z o.o. ISBN: 978-83-7798-0631-3. [PL]
- WIERZCHOŃ, Piotr (2015): „Jak powstawał fotosuplement do {Słownika warszawskiego} i co z tego wyszło” [w:] *Scripta Neophilologica Posnaniensia*, t. XV, s. 167–178. Poznań: Wydział Neofilologii Uniwersytetu im. Adama Mickiewicza. ISSN: 1509-4146. DOI: <https://doi.org/10.7169/snp.2015.15.12>. URL: http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.ojs-doi-10_14746_snp_2015_15_12/. [PL]
- WILKOWSKI, Marcin (2013): *Wprowadzenie do historii cyfrowej*. Gdańsk: Instytut Kultury Miejskiej. ISBN: 978-83-64610-00-4. URL: <http://historiacyfrowa.ikm.gda.pl/>. [PL]
- WOLIŃSKI, Marcin & KIERAŚ, Witold (2017): „Słownik gramatyczny języka polskiego – wersja internetowa” [w:] *Język Polski*, nr 1 (2017), s. 84–93. ISSN: 0021-6941. URL: <https://www.ceeol.com/search/article-detail?id=528786>. [PL]
- WOŹNIAK, Ewa (2019): „Prestiż polszczyzny w dwudziestolecu międzywojennym” [w:] Przybylska, Renata & Batko-Tokarz, Barbara & Pręcerek-Kisielak, Sylwia (red.): *Niepodległa wobec języka polskiego*, s. 29–40. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 375 s. ISBN: 978-83-233-4738-5. e-ISSN: 978-83-233-7021-5. [PL]
- YONG Heming (雍和明) & PENG Jing (彭敬) (2008): *Chinese Lexicography: A History from 1046 BC to AD 1911*. Oxford and New York: Oxford University Press, 458 s. ISBN: 978-0199539826. [EN]
- ZARĘBSKI, Rafał (2012): „Problemy opisu czasowników z prefiksami obcymi w ujęciu historycznym” [w:] *LingVaria*, r. VII (2012), nr 2 (14), s. 111–120. Kraków: „Księgarnia Akademicka”, 286 s. ISSN: 1896-2122. e-ISSN: 2392-1226. [PL]
- ZDANIUKIEWICZ, Alojzy Adam (2006): „Atlas gwar polskich prof. K. Dejnycy i jego współpracowników a Mały atlas gwar polskich” [w:] *Język Polski*, r. LXXXV, z. 3, s. 213–219. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. [PL]
- ZDANOWICZ, Aleksander & BOHUSZ-SZYSZKO, Michał & FILIPOWICZ, January & TOMASZEWICZ, Walerjan & CZEPIELIŃSKI, Florjan & KOROTYŃSKI, Wincenty & TRENTOWSKI, Bronisław (1861): *Słownik języka polskiego, obejmujący oprócz zbioru właściwie polskich znaczną liczbę wyrazów z obcych języków polskiemu przyswojonych; nomenklatury tak dawne, jak też nowo w życie wprowadzone różnych nauk, umiejętności, sztuk i rzemiosł; nazwania monet, miar i wag główniejszych krajów i prowincji; mitologię krajów słowiańskich i innych ważniejszych, tudzież oddzielną tablicę słów polskich nieforemnych z ich odmianą*. Wilno: wydany przez Maurycego Orgelbranda. [PL]

- ZGÓŁKOWA, Halina (1994): „Praktyczność, czyli służebność słownika wobec odbiorcy” [w:] Handke, Kwiryna & Dalewska-Greń, Hanna (red.): *Polszczyzna a/i Polacy u schyłku XX wieku*, s. 363–373. Warszawa: Slawistyczny Ośrodek Wydawniczy. ISBN: 83-86619-00-7. [PL]
- ZGUSTA, Ladislav (1971): *Manual of Lexicography*. Prague: Publishing House of the Czechoslovak Academy of Sciences; The Hague, Paris: Mouton, 360 s. [EN]
- ZGUSTA, Ladislav (1986): „Summation” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): *Fulbright Papers*, t. 1, s. 138–146. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8. [EN]
- ZGUSTA, Ladislav (1989): „The Role of Dictionaries in the Genesis and Development of the Standard” [w:] Hausmann, Franz J. et al. (red.): *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography*, s. 70–79. Berlin: De Gruyter. [EN]
- ZGUSTA, Ladislav (1992): „History and Its Multiple Meaning” [w:] Zgusta, Ladislav (red.): *History, Language, and Lexicographers*, s. 1–18, [seria:] Allén, Sture & Corbin, Pierre & Hartmann, Reinhard Rudolf Karl & Hausmann, Franz Josef & Heid, Ulrich & Reichmann, Oskar & Zgusta, Ladislav: (red.) *Lexicographica. Series Maior*, t. 41. Max Niemeyer Verlag: Tübingen. ISBN: 3-484-30941-5. ISSN: 0175-9264. [EN]
- ŻMIGRODZKI, Piotr (2008): *Słowo – słownik – rzeczywistość. Z problemów leksykografii i metaleksykografii*. Kraków: Lexis, 230 s. ISBN: 978-83-89425-45-4. [PL]
- ŻMIGRODZKI, Piotr (2009): *Wprowadzenie do leksykografii polskiej*, wyd. 3 poszerzone, [seria:] Wolińska, Olga: *Językoznawstwo polonistyczne*. Katowice: Wydawnictwo Uniwersytetu Śląskiego, 310 s. ISBN: 978-83-226-1852-3. ISSN: 1644-0552. [PL]
- ŻMIGRODZKI, Piotr (2017a): „Ogólna koncepcja metainformacji w *Wielkim słowniku języka polskiego PAN*” [w:] *Annales Universitatis Paedagogicae Cracoviensis: Studia Linguistica*, nr 12 (Folia 234). Kraków: Wydawnictwo Naukowe Uniwersytetu Pedagogicznego im. Komisji Edukacji Narodowej, s. 355–365. ISSN: 2083-1765. DOI: <https://doi.org/10.24917/20831765.12.35>. [PL]
- ŻMIGRODZKI, Piotr (2017b): „Lingwistyka komputerowa – czy po prostu lingwistyka?” [w:] *Język Polski*, nr 1 (2017), s. 5–6. ISSN: 0021-6941. URL: <https://www.ceeol.com/search/article-detail?id=528871>. [PL]
- ŻMIGRODZKI, Piotr (2018a): „2. Ogólna charakterystyka słownika” [rozdział w:] Żmigrodzki, Piotr & Bańko, Mirosław & Batko-Tokarz, Barbara & Bobrowski, Jakub & Czelakowska, Anna & Grochowski, Maciej & Przybylska, Renata & Waniakowa, Jadwiga & Węgrzynek, Katarzyna (red.): *Wielki słownik języka polskiego PAN: geneza, koncepcja, zasady opracowania*, s. 17–24. Kraków: Instytut Języka Polskiego PAN, 267 s. ISBN: 978-83-64007-49-1. URL: <https://rcin.org.pl/dlibra/publication/edition/68631?id=68631&from=publication>. [PL]
- ŻMIGRODZKI, Piotr (2018b): „3. Makro- i mikrostruktura słownika.” [rozdział w:] Żmigrodzki, Piotr & Bańko, Mirosław & Batko-Tokarz, Barbara & Bobrowski, Jakub & Czelakowska, Anna & Grochowski, Maciej & Przybylska, Renata & Waniakowa, Jadwiga & Węgrzynek, Katarzyna (red.): *Wielki słownik języka polskiego PAN: geneza, koncepcja, zasady opracowania*, s. 25–38. Kraków: Instytut Języka Polskiego PAN, 267 s. ISBN: 978-83-64007-49-1. URL: <https://rcin.org.pl/dlibra/publication/edition/68631?id=68631&from=publication>. [PL]
- БАБАЕВ, Ибрагимпаша (2013): „Нормативная лексикография как этап развития гуманитарной культуры” [w:] *Žmogus ir žodis*, r. 2013, nr I, s. 15–10. [RU]
- ГАСОВ, В. М. & ЦЫГАНЕНКО, А. М. (2001): *Методы и средства подготовки электронных изданий: Учебное пособие*. Москва: МГУП. [RU]
- РЯПОЛОВА, М. А. (2018): „Толковый словарь и категория адресата”, [w:] *Гуманитарные науки*, t. 37, nr 3, s. 385–393. Белгород: Белгородский университет кооперации, экономики и права. ISSN: 2075-4574. DOI: <https://doi.org/10.18413/2075-4574-2018-37-3-385-393>. [RU]

Literatura uzupełniająca

- [autor zbiorowy] (1847): *Словарь церковнославянского и русского языка, сост. Вторым отд. имп. Акад. наук*, t. 1. [RU]
- [autor zbiorowy] (1957): „От редакции” [w:] Охманова, О. С. (red.): *Лексикографический сборник, выпуск 1*, s. 37–57. Москва: Государственное издательство иностранных и национальных словарей, 188 s. [RU]
- [autor zbiorowy] (2009): *The Digital Humanities Manifesto 2.0*. URL: http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf. [EN]
- ABHYANKAR, Richa (2014): „Evolution of Sanskrit Lexicons” [w:] *Bulletin of the Deccan College Research Institute*, t. 74 (2014), s. 267–276. URL: <https://www.jstor.org/stable/10.2307/26264707>. [EN]
- ADAMIEC, Dorota (2011): „Podhasło jako element opisu leksykograficznego w *Słowniku języka polskiego XVII i I. połowy XVIII wieku*” [w:] Dunaj, Bogusław & Rak, Maciej (red.): *Badania historycznojęzykowe: Stan, metodologia, perspektywy. Materiały konferencji naukowej Kraków 21–22 września 2010 r.*, s. 193–199, [seria:] Skarżyński, Mirosław (red.): *Biblioteka LingVariów*, t. 14. Kraków: Wydawnictwo Księgarnia Akademicka, 431 s. ISBN: 978-83-7638-175-6. [PL]
- ADAMSKA-SALACIAK, Arleta (2012): „Dictionary Definitions: Problems and Solutions” [w:] *Studia Linguistica Universitatis Jagellonicae Cracoviensis*, t. 129 (supplementum). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 323–339. ISSN: 1897-1059. e-ISSN: 2083-4624. DOI: <https://doi.org/10.4467/20834624SL.12.020.0804>. URL: <https://www.ejournals.eu/pliki/art/138/>. [EN]
- ALMIND, Richard (2005): „Dictionaries through Databases” [w:] Gottlieb, Henrik & Mogensen, Jens Erik & Zettersten, Arne (red.): *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen*, s. 101–109. Tübingen: Max Niemeyer Verlag. ISBN: 3-484-39115-4. ISSN: 0175-9264. [EN]
- ATKINS, B. T. S. (1992): „Theoretical Lexicography and Its Relation to Dictionary-making” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, nr 14 (1992/93), s. 4–43. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.1992.0011>. [EN]
- AZORÍN, Dolores & BLANCO, María Ángeles & BUENAFUENTES, Cristina & CLAVERIA, Gloria et al. (2017): „Historia interna del Diccionario de la lengua castellana de la Real Academia Española en el siglo XIX (1817-1852)” [w:] Sario López, Ignacio & Gutiérrez Cuadrado, Juan & Garriga Escribano, Cecilio (red.): *El diccionario en la encrucijada: de la sintaxis y la cultura al desafío digital*. Santander: Escuela Universitaria de Turismo Altamira. ISBN: 978-84-617-4512-8. [ES]
- BAALBAKI, Ramzi (2019): „Arabic to c. 1800” [w:] Considine, John (red.): *The Cambridge World History of Lexicography*, s. 159–181. Cambridge: Cambridge University Press, 962 s. DOI: <https://doi.org/10.1017/9781316827437.009>. ISBN: 9781107178861. [EN]
- BAHR, Joachim (1978): „Ist Lexikographie eine Wissenschaft? Bemerkungen zu einer Tagung” [w:] Ágel, Vilmos & Bubenhofer, Noah & Feilke, Helmuth & Imo, Wolfgang & Lüdeling, Anke & Tophinke, Doris (red.): *Zeitschrift für germanistische Linguistik*, t. 6, s. 97–98. De Gruyter. ISSN: 0301-3294. e-ISSN: 1613-0626. [DE]
- BAILBLÉ, Olivier (2015): „A History of the Korean Lexicon” [w:] *International Journal of Teaching, Education and Language Learning*, t. 2, z. 3, s. 34–66. ISSN: 2373-7921. e-ISSN: 2373-793X. [EN]
- BAILEY, Don Clifford (1960): „Early Japanese Lexicography” [w:] *Monumenta Nipponica*, t. 16, s. 1–52. Tokyo: Sophia University. ISSN: 0027-0741. e-ISSN: 1880-1390. [EN]
- BAŃKO, Mirosław (1995): „W sprawie Dwóch suplementów” [w:] *Język Polski*, r. LXXV, z. 1, s. 78–80. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. URL: <http://mbc.malopolska.pl/dlibra/plain-content?id=30331>. [PL]

- BAŃKO, Mirosław (2001): *Z pogranicza leksykografii i językoznawstwa. Studia o słowniku jednojęzycznym*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 336 s. ISBN: 83-87608-87-4. [PL]
- BAŃKO, Mirosław (2014): „Mieczysław Szymczak (red.), *Słownik języka polskiego*, Warszawa 1978–1981” [w:] *Poradnik Językowy*, nr 7, s. 95–101. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343. [PL]
- BAŃKO, Mirosław (2020): *Z historii antyleksykografii*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego. ISBN:978-83-235-4334-3. [PL]
- BARSANTI VIGO, María Jesús (2006): „Problemática en torno al refrán y otras categorías paremiales: definición y delimitación” [w:] Alonso Ramos, Margarita (red.): *Diccionarios y Fraseología*, [seria:] *Anexos de Revista de Lexicografía*, t. 3, s. 197–205. Coruña: Universidade da Coruña, 259 s. ISBN: 84-9749-221-8. [ES]
- BATKO-TOKARZ, Barbara (2019): *Tematyczny podział słownictwa współczesnego języka polskiego: teoria, praktyka, leksykografia*, wyd. I. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego. ISBN: 978-83-233-4554-1. e-ISBN: 978-83-233-9906-3. [PL]
- BAUER, Laurie (2002): „The Illusory Distinction between Lexical and Encyclopedic Information” [w:] Gottlieb, Henrik & Mogensen, Jens Erik & Zettersten, Arne (red.): *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2–4, 2002 at the University of Copenhagen*, s. 111–115. Tübingen: Max Niemeyer Verlag. ISBN: 3-484-39115-4. ISSN: 0175-9264. [EN]
- BEČKA, Josef (1935): „Smetánkŭv staročesky slovník” [w:] *Naše řeč*, r. 19, nr 6–7, s. 161–176. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=3103>. [CZ]
- BÉJOINT, Henri (2000): *Modern Lexicography: An Introduction*. Oxford: Oxford University Press, 288 s. ISBN: 978-0198299516. [EN]
- BÉJOINT, Henri (2010): *The Lexicography of English*. Oxford: Oxford University Press, 458 s. ISBN (miękka): 9780198299677. ISBN (twarda): 9780198299672. [EN]
- BERGENHOLTZ, Henning & AGERBO, Heidi (2014): „Extraction, selection and distribution of meaning elements for monolingual information tools” [w:] *Lexicographica*, nr 30, s. 488–510. Berlin: De Gruyter. ISSN: 0175-6206. e-ISSN: 1865-9403. DOI: <https://doi.org/10.1515/lexi-2014-0015>. [EN]
- BERGENHOLTZ, Henning & SCHAEDEER, Burkhard (1977): *Die Wortarten des Deutschen*. Stuttgart: Klett. [DE]
- BERGLUNG, Lisa (2014): „Reflecting on Digital Dictionaries” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t. 35, z. 1, s. 303–306. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.2014.0015>. [EN]
- BIELIŃSKA, Monika (2019): „Status i istota leksykografii. Stanowiska i kontrowersje” [w:] Górski, Rafał (red.): *Polonica*, t. 39. Kraków: Instytut Języka Polskiego PAN. ISSN: 0137-9712. e-ISSN: 2545-045X. DOI: <http://dx.doi.org/10.17651/POLON.39.6>. URL: <https://polonica.ijp.pan.pl/index.php/polonica/article/view/53>. [PL]
- BIEŃ, Janusz S. (2006b): „Kilka przykładów dygitalizacji słowników” [w:] *Poradnik Językowy*, nr 8 (637), s. 55–63. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551–5343. [PL]
- BIEŃ, Janusz S. (2009b): „O pojęciu wyrazu morfologicznego” [w:] *Nie bez znaczenia... Prace ofiarowane Profesorowi Zygmuntowi Saloniemu z okazji 15000 dni pracy naukowej*. Białystok: Wydawnictwo Uniwersytetu w Białymstoku, s. 67–77. ISBN 83-89031-01-9. URL: <http://bc.klf.uw.edu.pl/62/>. [PL]
- BIEŃ, Janusz S. (2011a): „Historical Polish texts and Unicode. Introduction” [prezentacja z:] 20.01.2011, 23 s. URL: <https://bc.klf.uw.edu.pl/271/>. [EN]
- BILIŃSKA, Joanna (2014a): „Język słoweński w słowniku Lindego” [prezentacja na:] *Polyslav-XVIII Budapest*, 04.09.2014. 32 s. URL: <https://bc.klf.uw.edu.pl/382/>. [PL]

- BILIŃSKA, Joanna (2014b): „Wielojęzyczność *Słownika języka polskiego* S. B. Lindego w oparciu o wydanie drugie” [prezentacja na:] *V Glosa do leksykografii*, Warszawa, 19.09.2014 r. 31 s. URL: <http://bc.klf.uw.edu.pl/384/>. [PL]
- BISCONTI, Valentina (2016): *Le sens en partage : Dictionnaires et théories du sens XIXe–XXe siècles*, [seria:] Colombat, Bernard & Van den Avenne, Cécile (red.): *Collection langages*. Lyon: ENS Éditions. ISBN: 978-2-84788-852-2. URL: <https://books.openedition.org/enseditions/8669>. [FR]
- BLANÁR, Vincent (1994): „Staročeský slovník v rámci historických slovníků slovanských jazyků” [w:] *Listy filologické / Folia philologica*, r. 117, nr 1/2 (1994), s. 66–73. Praha: Kabinet pro klasická studia Filosofického ústavu AV ČR, v. v. i. ISSN: 0024-4457. URL: <https://www.jstor.org/stable/23466815>. [SK]
- BLATNÁ, Renata (1995): „Metajazyk v lexikografii” [w:] Čermák, František & Blatná, Renata (red.): *Manuál lexicografie*, s. 72–89. Jinočany: H&H, 283 s. ISBN: 80-85787-23-7. [CZ]
- BLEVINS, James P. & ACKERMAN, Farrell & MALOUF, Robert (2019): „Word-and-Paradigm Morphology” [w:] Audring, J. & Masini, F. (red.): *Oxford Handbook of Morphological Theory*, s. 265–284. Oxford: Oxford University Press. ISBN: 9780199668984. DOI: <https://doi.org/10.1093/oxfordhb/9780199668984.001.0001>. URL: https://www.researchgate.net/publication/283712857_Word_and_Paradigm_Morphology. [EN]
- BOBROWSKI, Ireneusz (1998): „Poprzednie rozwiązania leksykograficzne a nowa propozycja słownikowa (na marginesie «Praktycznego słownika współczesnej polszczyzny»)” [w:] *Polonica*, t. 19, s. 211–217. Kraków: Instytut Języka Polskiego PAN. ISSN: 0137-9712. e-ISSN: 2545-045X. DOI: <http://dx.doi.org/10.17651/POLON.39.6>. URL: <https://polonica.ijp.pan.pl/index.php/polonica/article/view/53>. [PL]
- BOBROWSKI, Ireneusz (2018): „Norma jako składnik kompetencji językowej” [w:] Zbróg, Piotr (red.): *Wybrane aspekty badań nad normą językową*, s. 19–24. Kraków: Wydawnictwo LIBRON – Filip Lohner, 272 s. ISBN: 978-83-65705-63-1. URL: https://ilij.ujk.edu.pl/wp-content/uploads/Wybrane-aspektybadan_nad_norm%C4%85_j%C4%99zykow%C4%85_red._P.Zbr%C3%B3g.pdf. [PL]
- BOGUSŁAWSKI, Andrzej (1976): „O zasadach rejestracji jednostek języka” [w:] *Poradnik Językowy*, nr 8, s. 356–364. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343. [PL]
- BOCHNAKOWA, Anna (1991): *Le Nouveau grand dictionnaire français, latin et polonais et sa place dans la lexicographie polonaise*, [seria:] *Rozprawy Habilitacyjne*, nr 222. Kraków: Nakładem Uniwersytetu Jagiellońskiego, 161 s. ISBN: 83-233-0494-7. ISSN: 0239-782X. [FR]
- BOMBA, Radosław (2014): „Humanistyka cyfrowa między utopią a nowym paradygmatem? Wstęp do analityki kulturowej Lwa Manovicha” [w:] *Ethos*, t. 27, z. 3 (107), s. 261–279. ISSN: 0860-8024. DOI: <https://doi.org/10.12887/27-2014-3-107-15>. Lublin: Katolicki Uniwersytet Lubelski Jana Pawła II. [PL]
- BOREJSZO, Maria (2013): „Gwarowe nazwy ubiorów w Słowniku warszawskim” [w:] *Poznańskie Studia Polonistyczne*, t. 20 (40), z. 2, [seria:] *Seria Językoznawcza*. ISBN: 978-83-61573-56-2. ISSN: 1233-8672. Poznań: Wydawnictwo Poznańskiego Towarzystwa Przyjaciół Nauk, 228 s. DOI: <https://doi.org/10.14746/pspsj.2013.20.2.7>. URL: <https://pressto.amu.edu.pl/index.php/pspsl>. [PL]
- BOSQUE, Ignacio & BARRIOS RODRÍGUEZ, María Auxiliadora (2018): „Spanish lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 636–660. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- BOTHMA, Theo J. D. (2018): „Lexicography and information science” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 197–216. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- BOWKER, Lynne (2018): „Lexicography and terminology” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 138–151. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]

- BRANCA-ROSOFF, Sonia (1995): „La construction de la norme lexicographique à la fin du XVIII^e siècle : Feraud le mediateur.” [w:] *Histoire Épistémologie Langage*, r. 1995, nr 11, s. 34–46, [część numeru tematycznego:] *La genèse de la norme. Colloque de la SHESL, janvier 1994*. Paris: Société d’Histoire et d’Épistémologie des Sciences du Langage (SHESL), 10 s. e-ISSN: 1638-1580. URL: https://www.persee.fr/doc/hel_0247-8897_1995_num_11_1_3403. [FR]
- BROOKES, Ian (2007): „New Words and Corpus Frequency” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t. 28, z. 1, s. 142–145. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.2007.0017>. [EN]
- BURCHFIELD, Robert (1986): „The Oxford English Dictionary” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): Fulbright Papers, t. 1, s. 17–27. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8.
- BUTTLER, Danuta (1985): „Kodyfikacja normy” [w:] Topolewska-Jezior, Alicja (red.): *Prasa Techniczna*, z. 4/1985, s. 13–16. [PL]
- BWENGE, Charles (1989): „Lexicographical treatment of affixational morphology. A case study of four Swahili dictionaries” [w:] James, Gregory (red.): *Lexicographers and their works*, s. 5–17, [seria:] *Exeter Linguistic Studies*, t. 14. Exeter: University of Exeter. ISBN-13: 978-0859893367. [EN]
- BYNON, Theodora (1977): *Historical Linguistics*. Cambridge: Cambridge University Press, 301 s. e-ISSN (2012): 9781139165709. DOI: <https://doi.org/10.1017/CBO9781139165709>. [EN]
- CAMIÑA SALGADO, Susana & Muñiz Álvarez, Eva-María (2006): „Sobre la necesidad de marcar las colocaciones en el diccionario de uso” [w:] Alonso Ramos, Margarita (red.): *Diccionarios y Fraseología*, [seria:] *Anexos de Revista de Lexicografía*, t. 3, s. 101–110. Coruña: Universidade da Coruña, 259 s. ISBN: 84-9749-221-8. [ES]
- CARRIAZO RUIZ, José Ramón (2007): „Diccionarios históricos” [w:] *Estudios de Lingüística del Español (ELiEs)*, nr 38 (2017), s. 35–59. Granada: Laboratorio de Lingüística Experimental y Computacional de la Universidad de Granada. ISSN: 1139-8736. URL: <http://infoling.org/elies/>. [ES]
- CELIŃSKI, Piotr (2013): „Renesansowe korzenie cyfrowego zwrotu” [w:] Radomski, A., Bomba, R. (red.): *Zwrot cyfrowy w humanistyce. Internet / Nowe Media / Kultura 2.0*, s. 13–36. Lublin: E-naukowiec. URL: http://e-naukowiec.eu/wp-content/uploads/2013/05/Zwrot_cyfrowy_w_humanistyce.pdf. [PL]
- CLAVERÍA, Gloria & PLANAS, Carmen (2001): „La homonimia en la lexicografía española” [w:] *Nueva Revista de Filología Hispánica (NRFH)*, t. 49, nr 2, s. 281–306. La Ciudad de México: El Colegio de México. ISSN: 0185-0121. e-ISSN: 2448-6558. URL: <https://www.jstor.org/stable/40300016>. [ES]
- Committee on Electronic Information and Communication (2006): „Some Best Practices for Retrodigitization”. URL: https://www.mathunion.org/fileadmin/IMU/Report/retro_bestpractices.pdf. [EN]
- CONSIDINE, John (2010): *Current Projects in Historical Lexicography*. Cambridge: Cambridge Scholars Publishing. ISBN: 978-1-4438-2116-2. [EN]
- CONSIDINE, John (2019): „Early Modern Western and Central Europe” [w:] Considine, John (red.): *The Cambridge World History of Lexicography*, s. 290–314. Cambridge: Cambridge University Press, 962 s. DOI: <https://doi.org/10.1017/9781316827437.015>. ISBN: 9781107178861. [EN]
- CORBIN, Pierre & GASIGLIA, Nathalie (2017): „Un demi-siècle de conceptions du traitement de la variation dans la lexicographie d’expression française” [w:] *Revue de Sémantique et Pragmatique*, t. 41–42. e-ISSN: 2610-4377. DOI: <https://doi.org/10.4000/rsp.444>. URL: <http://journals.openedition.org/rsp/444>. [FR]
- CRAWFORTH, Hannah (2014): „Linguistics, Lexicography, and the ‘Early Modern’” [w:] *Journal for Early Modern Cultural Studies*, t. 14, nr 2 (2014), s. 94–99. Philadelphia: University of Pennsylvania Press. ISSN: 1531-0485. e-ISSN: 1553-3786. DOI: <https://doi.org/10.1353/jem.2014.0026>. URL: <https://www.jstor.org/stable/10.2307/jearlmodcultstud.14.2.94>. [EN]

- ČERMÁK, František (1995): „Paradigmatika a syntagmatika slovníku: možnosti a výhledy” [w:] Čermák, František & Blatná, Renata (red.): *Manuál lexikografie*, s. 90–115. Jinočany: H&H, 283 s. ISBN: 80-85787-23-7. [CZ]
- ČERMÁK, František (2009): „Aleš Klégr: Tezaurus jazyka českého: Slovník českých slov a frází souznačných, blízkých a příbuzných. Praha: Nakladatelství Lidové noviny, 2007. 1189 s.” [w:] *Slovo a slovesnost*, r. 70, nr 2, s. 153–157. Praha: Ústav pro jazyk český Akademie věd České republiky. ISSN: 0037-7031. e- ISSN: 2571-0885. URL: <http://sas.ujc.cas.cz/>. [CZ]
- ČERMÁK, František (2012): „Poznámky k interpretaci významu ve slovníku” [w:] Buzássyová, K. & Chocholová, B. & Janočková, N. (red.): *Slovo v slovníku: Aspekty lexikální sémantiky – gramatika – stylistika (pragmatika): Na počest Alexandry Jarošovej*, s. 76–83. Bratislava: Veda – Jazykovedný ústav Ľudovíta Štúra SAV, 300 s. ISBN 978-80-224-1268-1. [CZ]
- ČERVENÁ, Vlasta (1981): „O lexikálním archívu Ústavu pro jazyk český ČSAV” [w:] *Naše řeč*, r. 64, nr 1, s. 14–22. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=6230>. [CZ]
- D’ALEMBERT, Jean le Rond (1754) „Dictionnaire” (haslo) [w:] Diderot, Denis & d’Alembert, Jean le Rond (red.): *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers*, t. IV, s. 958–969. Paris: Briasson. [FR]
- DAL, Georgette & NAMER, Fiammetta (2012): „Faut-il brûler les dictionnaires ? Ou comment les ressources numériques ont révolutionné les recherches en morphologie” [w:] SHS Web of Conferences, t. 1 (2012), 1261–1276. e-ISSN: 2261-2424. DOI: <https://doi.org/10.1051/shsconf/20120100217>. Ulis: EDP Sciences – Web of Conferences. [FR]
- DEL BARRIO ESTÉVEZ, Laura & TORNER CASTELLS, Sergio (1994–1995): „La información diacrónica en el *Diccionario de la Lengua Española* de la Real Academia (Vigésima Primera Edición)” [w:] Gutiérrez Cuadrado, Juan & Pérez Pascual, José Ignacio (red.): *Revista de Lexicografía*, t. 1, s. 29–54. ISSN: 1134-4539. e-ISSN: 2603-6673. [ES]
- DENY Arnos Kwary & NOR Hashimah Jalaluddin (2015): „The lexicography of Indonesian/Malay” [w:] Hanks, Patrick & de Schryver, Gilles-Maurice (red.): *International Handbook of Modern Lexis and Lexicography*, s. 1–11. Berlin & Heidelberg: Springer-Verlag, 900 s. ISBN: 9783642545344. DOI: https://doi.org/10.1007/978-3-642-45369-4_83-1. [EN]
- DENY Arnos Kwary (2018): „Indonesian lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 599–607. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- DEOKAR, Lata Mahesh & CHEVILLARD, Jean-Luc (2019a): „Ancient India” [w:] Considine, John (red.): *The Cambridge World History of Lexicography*, s. 67–83. Cambridge: Cambridge University Press, 962 s. DOI: <https://doi.org/10.1017/9781316827437.005>. ISBN: 9781107178861. [EN]
- DEOKAR, Lata Mahesh & CHEVILLARD, Jean-Luc (2019b): „India and Tibet, c. 500–c. 1750” [w:] Considine, John (red.): *The Cambridge World History of Lexicography*, s. 130–158. Cambridge: Cambridge University Press, 962 s. DOI: <https://doi.org/10.1017/9781316827437.008>. ISBN: 9781107178861. [EN]
- DERKSEN, Rick (2019): „The Slavic and Baltic Languages” [w:] Considine, John (red.): *The Cambridge World History of Lexicography*, s. 445–459. Cambridge: Cambridge University Press, 962 s. DOI: <https://doi.org/10.1017/9781316827437.022>. ISBN: 9781107178861. [EN]
- DOLEŽEL, Lubomír (1955): „K stylistické a normativní charakteristice ve slovníku spisovného jazyka” [w:] *Naše řeč*, r. 38, nr 1–2, s. 12–21. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=4396>. [CZ]
- DOROSZEWSKI, Witold (1973): *Elements of Lexicology and Semiotics*. Berlin, Boston: De Gruyter Mouton, 2018. ISBN: 9783110905380. DOI: <https://doi.org/10.1515/9783110905380>. [EN]

- DORREN, Gaston (2018): *Babel: Around the World in Twenty Languages*. London: Profile Books. ISBN: 978-1781256411. e-ISBN: 978-1782832508. [EN]
- DOSTÁL, Antonín (1963): „1100 let tradice spisovných jazyků slovanských” [w:] *Naše řeč*, r. 46, nr 5, s. 225–228. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=4974>. [cz]
- DRYSDALE, Patrick (1979): „Dictionary Etymologies: What? Why? and for Whom?” [w:] Michell, Gillian (red.): *Papers of the Dictionary Society of North America*. London, Ontario: The School of Library and Information Science, University of Western Ontario, s. 39–50. [EN]
- DUBISZ, Stanisław (2011): „Kwalifikatory w słowniku – problem nie tylko leksykograficzny” [w:] *Prace Filologiczne*, t. LX, s. 99–108. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567.
- DUBISZ, Stanisław (red.) (2003): *Uniwersalny słownik języka polskiego (USJP)*, t. 1–6. Warszawa: Państwowe Wydawnictwo Naukowe. [PL]
- DUNAJ, Bogusław (2010): „Perspektywy badań historycznojęzykowych” [w:] *LingVaria*, r. V (2010), nr 2 (10), s. 53–57. Kraków: „Księgarnia Akademicka”. ISSN: 1896-2122. e-ISSN: 2392-1226. [PL]
- DUVAL, Mark (2013): „Korean Lexicography” [w:] Gouws, Rufus Hjalmar et al. (red.): *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 900–910. Berlin: De Gruyter Mouton. ISBN: 9783110238129. e-ISBN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.900>. [EN]
- FILÍPEČ, Josef (1957): „Lexikálně sémantická výstavba hesla – ústřední otázka lexikografické práce (Na materiálu jednojazyčných slovníků)” [w:] *Slovo a slovesnost*, nr 18, s. 129–150. Praha: Akademie věd České republiky & Ústav pro jazyk český Akademie věd České republiky. ISSN: 0037-7031. [cz]
- FIRTH, John Rupert (1957): *Papers in Linguistics 1934–1951*. London, New York: Oxford University Press, 233 s. ISBN: 9780197135488. [EN]
- FLAJŠHANS, Václav (1945): „Z dílny slovníkářovy” [w:] *Naše řeč*, r. 29, nr 1, s. 1–5. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=3857>. [cz]
- FRAWLEY, William (1981): „In defence of the dictionary” [w:] *Lingua*, nr 55, s. 53–61. Amsterdam: North-Holland Publishing Company. ISSN: 0024-3841. [EN]
- FUERTE-OLIVERA, Pedro A. & TARP, Sven (2014): *Theory and Practice of Specialised Online Dictionaries: Lexicography Versus Terminography*, [seria:] *Lexicographica. Series Maior*, t. 146. Berlin, Boston: de Gruyter. DOI: <https://doi.org/10.1515/9783110349023>. [EN]
- FUERTE-OLIVERA, Pedro A. (2018): „Lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- GATES, Edward (1986): „Preparation for lexicography as a career in the United States” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): *Fullbright Papers*, t. 1, s. 82–88. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8. [EN]
- GAUDIN, François & GUESPIN, Louis (2000): *Initiation à la lexicologie française. De la néologie aux dictionnaires*, wyd. 1, [seria:] Willems, Dominique & Wilmet, Marc (red.): *Champs linguistiques*. Bruxelles: Duculot, 356 s. ISSN: 1374-089X. e-ISBN: 2-8011-1264-X. [FR]
- GHATAGE, A. M. (2000): „Lexicography and Lexicology” [w:] *Annals of the Bhandarkar Oriental Research Institute (BORI)*, t. 81, nr 1/4 (2000), s. 227–236. Pune: Bhandarkar Oriental Research Institute. URL: <https://www.jstor.org/stable/41694616>. [EN]
- GLATIGNY, Michel (2015): *Les marques d'usage dans les dictionnaires français monolingues du XIXe siècle : Jugements portés sur un échantillon de mots et d'emplois par les principaux lexicographes*, [seria:] Allén, Sture & Corbin, Pierre & Hartmann, Reinhard Rudolf Karl & Hausmann, Franz Josef & Heid, Ulrich & Reichmann, Oskar & Zgusta, Ladislav (red.): *Lexicographica. Series Maior*, t. 91. Tübingen: Max

Niemeyer Verlag, 384 s. e-ISBN: 9783110920079. ISSN: 0175-9264. DOI: <https://doi.org/10.1515/9783110920079>. [FR]

- GONZÁLEZ AGUIAR, María Isabel (2006): „La definición lexicografía de las unidades fraseológicas: la aplicación de modelos formales” [w:] Alonso Ramos, Margarita (red.): *Diccionarios y Fraseología*, [seria:] *Anexas de Revista de Lexicografía*, t. 3, s. 221–233. Coruña: Universidade da Coruña, 259 s. ISBN: 84-9749-221-8. [ES]
- GOUWS, Rufus Hjalmar (2002) „Multiple Niching ” [w:] Gottlieb, Henrik & Mogensen, Jens Erik & Zettersten, Arne (red.): *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2–4, 2002 at the University of Copenhagen*, s. 261–275. Tübingen: Max Niemeyer Verlag. ISBN: 3-484-39115-4. ISSN: 0175-9264. [EN]
- GOUWS, Rufus Hjalmar (2003): „Types of articles, their structure and different types of lemmata” [w:] van Sterkenburg, Piet (red.): *A Practical Guide to Lexicography*, s. 34–43. Amsterdam & Philadelphia: John Benjamins Publishing Company. ISBN (twarda): 90-272-2329-7 (EU); 1-58811-380-9 (US). ISBN (miękka): 90-272-2330-0 (EU); 1-58811-381-7 (US). DOI: <https://doi.org/10.1075/tlrp.6.07gou>. [EN]
- GOUWS, Rufus Hjalmar (2009): „Sinous lemma files in printed dictionaries” [w:] Nielsen, Sandro & Tarp, Sven (red.): *Lexicography in the 21st Century: In honour of Henning Bergenholtz*, s. 3–21, [seria:] L’Homme, Marie-Claude & Kageura, Kyo (red.): *Terminology and Lexicography Research and Practice (TLRP)*. Amsterdam & Philadelphia: John Benjamins Publishing Company. ISBN (twarda): 978-90-272-2336-4. ISBN (miękka): 978-90-272-8901-8. [EN]
- GROCHOWSKI, Maciej (1982): *Zarys leksykologii i leksykografii: zagadnienia synchroniczne*, [seria:] *Skrypty i Teksty Pomocnicze*. Toruń: Uniwersytet Mikołaja Kopernika, 147 s. [PL]
- GRUSZCZYŃSKI, Włodzimierz (2000): *Wokabularze ryskie na tle XVI- i XVII-wiecznej leksykografii polskiej*. Warszawa: Oficyna Wydawnicza „Aspra”, 270 s. ISBN: 83-908937-7-0. [PL]
- GRZEGORCZYKOWA, Renata (1979): *Zarys słowotwórstwa polskiego. Słowotwórstwo opisowe*, wyd. 3 poprawione. Warszawa: Państwowe Wydawnictwo Naukowe, 96 s. ISBN: 83-01-00119-4. [PL]
- GRZEGORCZYKOWA, Renata (2001): *Wprowadzenie do semantyki językoznawczej*, wyd. 3 poprawione i rozszerzone. Warszawa: Wydawnictwo Naukowe PWN, 182 s. ISBN: 83-01-13335-X. [PL]
- GUERRERO SALAZAR, Susana (2002): „Los diccionarios inversos de la lengua española: Descripción del repertorio bibliográfico” [w:] Gutiérrez Cuadrado, Juan & Pérez Pascual, José Ignacio (red.): *Revista de Lexicografía*, t. 8, s. 269–295. ISSN: 1134-4539. e-ISSN: 2603-6673. DOI: <https://doi.org/10.17979/rlex.2002.8.0.5592>. [ES]
- GUOHONG Fu & LUKE, K. K. (2003): „An integrated approach for Chinese word segmentation” [w:] *Proceedings of the 17th Pacific Asia Conference on Language, Information and Computation*, s. 80–87. DOI: <http://hdl.handle.net/2065/12307>. URL: <https://aclanthology.org/Y03-1009>. [EN]
- HAIMAN, John (1980): „Dictionaries and Encyclopedias” [w:] *Lingua*, nr 50, s. 329–357. Amsterdam: North-Holland Publishing Company. ISSN: 0024-3841. [EN]
- HAIMAN, John (1982): „Dictionaries and Encyclopedias Again” [w:] *Lingua*, nr 56, s. 353–355. Amsterdam: North-Holland Publishing Company. ISSN: 0024-3841. [EN]
- HALLER, Jiří: „Poznámky k Příručnímu slovníku jazyka českého” [w:] *Naše řeč* (1935): r. 19, nr 4, s. 108–112; (1936): r. 20, nr 2–3, s. 44–49; r. 20, nr 4–5, s. 105–106; r. 20, nr 6, s. 138–142; r. 20, nr 9, s. 231–235; r. 20, nr 10, s. 256–259; (1937): r. 21, nr 8, s. 220–223; r. 21, nr 9–10, s. 250–257; (1938): r. 22, nr 2, s. 43–48; r. 22, nr 5, s. 137–142; r. 22, nr 8, s. 236–240; r. 22, nr 9, s. 266–270. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/>. [CZ]
- HANKS, Patrick (1987): „Definitions and explanations” [w:] Sinclair, John McHardy (red.): *Looking Up: An Account of the COBUILD Project in Lexical Computing*, s. 116–136. London: Collins ELT. [EN]

- HANKS, Patrick (2000): „Do Word Meanings Exist?” [w:] *Computers and the Humanities*, t. 34, nr 1/2, s. 205–215. Berlin: Springer, 243 s. e-ISSN: 1572-8412. ISSN: 0010-4817. DOI: <https://doi.org/10.1023/A:1002471322828>. [EN]
- HANKS, Patrick (2005): „Johnson and Modern Lexicography” [w:] *International Journal of Lexicography*, nr 18, z. 2, s. 243–266. ISSN: 0950-3846. e-ISSN: 1477-4577. DOI: <https://doi.org/10.1093/ijl/eci024>. [EN]
- HANKS, Patrick (2010b): „Terminology, Phraseology, and Lexicography” [w:] Dykstra, Anne & Schoonheim, Tanneke (red.): *2010 Proceedings of the XIV EURALEX International Congress. 6-10 July 2010*, s. 1299–1308. Leeuwarden/Ljouwert: Fryske Akademy – Afûk, 1564 s. ISBN: 978-90-6273-850-3. ISSN: 2521-7100. URL: https://euralex.org/elx_proceedings/Euralex2010/122_Euralex_2010_9_HANKS_Terminology,%20Phraseology,%20and%20Lexicography.pdf. [EN]
- HANKS, Patrick (2010c): „Lexicography, Printing Technology, and the Spread of Renaissance Culture” [w:] Dykstra, Anne & Schoonheim, Tanneke (red.): *Proceedings of the 14th EURALEX International Congress*, s. 988-1006. Ljouwert: Fryske Akademy. ISBN: 978-90-6273-850-3. [EN]
- HANKS, Patrick (2012a): „How people use words to make meanings: Semantic types meet valencies” [w:] Boulton, Alex & Thomas, James (red.): *Input, Process and Product: Developments in Teaching and Language Corpora*, s. 52–67. Brno: Masaryk University Press, 352 s. ISBN: 978-8021058965. e-ISSN: 9788021076365. URL: <http://clg.wlv.ac.uk/papers/hanks-2012a.pdf>. [EN]
- HARTMANN, James (2000): „Dictionary Pronunciations: Mine or Theirs or Yours?” [w:] *American Speech*, t. 75, nr 3. Durham: Duke University Press, s. 250–252. ISSN: 0003-1283. e-ISSN: 1527-2133. [EN]
- HARTMANN, Reinhard Rudolf Karl (2001): *Teaching and Researching Lexicography*, [seria:] Candlin, Christopher N. & Hall, David K. (red.): *Applied Linguistics in Action*. Edinburgh: Pearson Education Limited, 211 s. ISBN: 0-582-36977-0. [EN]
- HASPELMATH, Martin (2008): „Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability” [w:] Stolz, Thomas & Bakker, Dik & Salas Palomo, Rosa (red.): *Aspects of Language Contact: New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes*, s. 43–62, [seria:] Bossong, Georg & Comrie, Bernard & Matras, Yaron (red.): *Empirical Approaches to Language Typology*, t. 35. Berlin & New York: Mouton de Gruyter. ISBN: 978-3-11-019584-2. [EN]
- HAB, Ulrike (2019): „The Germanic Languages Other than English from c. 1700” [w:] Considine, John (red.): *The Cambridge World History of Lexicography*, s. 460–483. Cambridge: Cambridge University Press, 962 s. DOI: <https://doi.org/10.1017/9781316827437.023>. ISBN: 9781107178861. [EN]
- HAUBRICHS, Wolfgang (2013). „German I: Historical and etymological lexicography” [w:] Gouws, Rufus Hjalmar & Heid, Ulrich & Schweickard, Wolfgang & Wiegand, Herbert Ernst (red.): *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 731–741. Berlin, Boston: De Gruyter Mouton. ISBN: 9783110238129. e-ISSN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.731/>. [EN]
- HAUSMANN, Franz Josef & WIEGAND, Herbert Ernst (1989): „Component Parts and Structures of General Monolingual Dictionaries: A Survey” [w:] Hausmann, Franz Josef et al. (red.): *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography*, s. 328–360. Berlin: De Gruyter. [EN]
- HAUSMANN, Franz Josef (2005): „Isotopie, scénario, collocation et exemple lexicographique” [w:] Heinz, Michaela (red.): *L'exemple lexicographique dans les dictionnaires français contemporains. Actes des Premières Journées allemandes des dictionnaires (Klingenberg am Main, 25-27 juin 2004)*, s. 283–292, [seria:] *Lexicographica. Series Maior*, t. 128. Tübingen: Max Niemeyer Verlag. ISBN: 978-3-11-092446-6. ISSN: 0175-9264. [FR]
- HAVRÁNEK, Bohuslav (1919): „O lékařském slovníku” [w:] *Naše řeč*, r. 3, nr 3, s. 72–80. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=517>. [CZ]

- HAVRÁNEK, Bohuslav (1922): „Českoněmecký poštovní a telegrafní slovník” [w:] *Naše řeč*, r. 6, nr 7, s. 212–216. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=1803>. [CZ]
- HELLER, Louise G. (1965): „Lexicographic Etymology: Practice Versus Theory” [w:] *American Speech*, t. 40, nr 2. Durham: Duke University Press, 113–119. ISSN: 0003-1283. e-ISSN: 1527-2133. [EN]
- HERBERG, Dieter (1985): „Zur Funktion und Gestaltung von Wörterbucheinleitungen” [w:] Hyldgaard-Jensen, Karl & Zettersten, Arne (red.): *Symposium on Lexicography II: Proceedings of the Second International Symposium on Lexicography May 16-17, 1984 at the University of Copenhagen*, s. 133–154 [seria:] Allén, Sture & Corbin, Pierre & Hartmann, Reinhard Rudolf Karl & Hausmann, Franz Josef & Heid, Ulrich & Reichmann, Oskar & Zgusta, Ladislav (red.): *Lexicographica. Series Maior*, t. 5. Tübingen: Max Niemeyer Verlag. ISBN: 3-484-30905-9. ISSN: 0175-9264. [DE]
- HERBST, Thomas & KLOTZ, Michael (2003): *Lexikografie*. Paderborn: Ferdinand Schöningh, 320 s. ISBN: 3-506-97015-1. [DE]
- HIORTH, Finngeir (1955): „On the Subject Matter of Lexicography: General Remarks and Survey of Contents” [w:] *Studia Linguistica: A Journal of General Linguistics*, t. 9, z. 1–2, s. 57–65. Hoboken: Wiley-Blackwell. ISSN: 0039-3193. e-ISSN: 1467-9582. DOI: <https://doi.org/10.1111/j.1467-9582.1955.tb00518.x>. [EN]
- HLADÍKOVÁ, Jitka (1986): „Složená slova přejatá z němčiny a Jungmannův Slovník česko-německý” [w:] *Naše řeč*, r. 69, nr 1, s. 6–12. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=6590>. [CZ]
- HODURA, Kvido (1935): „Lístkové výpisky k českému slovníku” [w:] *Naše řeč*, r. 19, nr 1, s. 1–6. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=2917>. [CZ]
- HODURA, Kvido (1959): „Z dějin akademického slovníku jazyka českého” [w:] *Naše řeč*, r. 42, nr 1–2, s. 6–14. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=6711>. [CZ]
- HOOGLAND, Jan (2019): „Arabic from c. 1800” [w:] Considine, John (red.): *The Cambridge World History of Lexicography*, s. 414–430. Cambridge: Cambridge University Press, 962 s. DOI: <https://doi.org/10.1017/9781316827437.020>. ISBN: 9781107178861. [EN]
- HRBÁČEK, Josef (1974): „Lexikální ekvivalenty, dublety a varianty” [w:] *Naše řeč*, r. 57, nr 1, s. 28–33. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=5742>. [CZ]
- HUANG JinHu & POWERS, David (2003): „Chinese Word Segmentation Based on Contextual Entropy” [w:] *Proceedings of the 17th Pacific Asia Conference on Language, Information and Computation*, s. 152–158. DOI: <http://hdl.handle.net/2065/12265>. URL: <https://aclanthology.org/Y03-1017/>. [EN]
- HUDSON, Richard (1988): „The Linguistic Foundations for Lexical Research and Dictionary Design” [w:] Ilson, Robert (red.): *International Journal of Lexicography*, t. 1, z. 4. ISSN: 0950-3846. e-ISSN: 1477-4577. [EN]
- HUET, Gérard (2019): „Sanskrit lexicography, past and future” [w:] Li Wei (red.): *Research on the Language and Script in Buddhist Sutras*. Hangzhou: Hangzhou Buddhist Academy, 2019. URL: <http://gallium.inria.fr/~huet/PUBLIC/Hangzhou.pdf>. [EN]
- HUSZCZA, Romuald & PIETROW, Jarosław A. & WOJCIECHOWSKI, Bartosz T. (2011): „Leksykografia japońska i jej typologia gatunkowa – wprowadzenie” [w:] Pietrow, Jarosław A. & Wojciechowski, Bartosz T. (red.): *Lexicographica Iapono-Polonica*, t. 1, s. 23–52. Warszawa: Wydano Nakładem Wydziału Polonistyki Uniwersytetu Warszawskiego, 361 s. ISBN: 978-83-62100-97-2. [PL]
- CHAPMAN, Don (2019): „Etymology and the Doctrine of Correctness: Word Meaning in Dictionaries of English Usage” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t.

- 40, z. 2, s. 93–116. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.2019.0015>. [EN]
- IAMARTINO, Giovanni (1990): „The Lexicographer as a Biassed Witness: Social, Political, and Religious Criticism in Baretto's English-Italian Dictionary” [w:] *Aevum*, r. 64, z. 3 (settembre-dicembre 1990), s. 435-444. Milan: Vita e Pensiero – Pubblicazioni dell'Università Cattolica del Sacro Cuore. ISSN: 1827-787X. URL: <https://www.jstor.org/stable/20858561>. [EN]
- ILSON, Robert (1986a): „Introduction” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): Fulbright Papers, t. 1, s. xii–xiv. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8. [EN]
- JACKSON, Howard (2018): „English lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 540–553. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- JAKOBSON, Roman (2013): „On Linguistic Aspects of Translation” (esej z 1959 r.) [w:] Brower, Reuben Arthur (red.): *On Translation*, s. 232–239, [seria:] *Harvard Studies in Comparative Literature*, t. 23. Cambridge & London: Harvard University Press, 297 s. ISBN 9780674731615. DOI: <https://doi.org/10.4159/harvard.9780674731615>. URL: <https://web.stanford.edu/~eckert/PDF/jakobson.pdf>. [EN]
- JAKUBCZYK, Marian (2014): „O odnalezionym budapeszteńskim fragmencie {Słowarza} z 1532 roku” [w:] *Prace Filologiczne*, t. XIII, s. 546–557. Warszawa: Kasa im. Józefa Mianowskiego. ISSN: 0138-0567. [PL]
- JANKOWIAK, Lucyna Agnieszka (2018): „Odchodzenie synonimii z terminologii – na przykładzie polskiej terminologii medycznej przełomu XIX i XX wieku” [w:] Grochowski, Maciej (red.): *Linguistica Copernicana*, nr 15, s. 287–243. ISSN: 2080-1068. e-ISSN: 2391-7768. DOI: <http://doi.org/10.12775/LinCop.2018.014>. [PL]
- JANKOWIAK, Mirosław (2012): O nowym *Słowniku lotewsko-polskim* [w:] *Acta Baltico-Slavica*, t. 36. Warszawa: Instytut Sławistyki PAN. ISSN: 2392-2389. DOI: <https://doi.org/10.11649/abs.2012.020>. [PL]
- JHA, Vashishta Narayan (red.) (1997): *Proceedings of the National Seminar on the Art of Dictionary Making in Ancient India*. Pune: Center of Advanced Study in Sanskrit, University of Pune, 137 s. [EN]
- JOHNSON, Samuel (1792): *The Works of Samuel Johnson, LL. D.*, t. 2. London, 360 s. [EN]
- JONES, William Jervie (1991): „Lingua teutonum victrix ? Landmarks in German Lexicography (1500-1700)” [w:] *Histoire Épistémologie Langage*, r. 1995, nr 13, z. 2, s. 131–152, [część numeru tematycznego:] *Théories et données*. Paris: Société d'Histoire et d'Épistémologie des Sciences du Langage (SHESL), 10 s. e-ISSN: 1638-1580. DOI: <https://doi.org/10.3406/hel.1991.2336>. URL: https://www.persee.fr/doc/hel_0750-8069_1991_num_13_2_2336https://www.persee.fr/doc/hel_0247-8897_1995_num_11_1_3403. [EN]
- JURAFSKY, Dan & Martin, James H. (2021): *Speech and Language Processing*, 3. wyd. (szkic), rozdz. 8. URL: <http://web.stanford.edu/~jurafsky/slp3/8.pdf>. [EN]
- KAMINSKI, Mariusz (2013): *A History of the Chambers Dictionary*, [seria:] Corbin, Pierre & Heid, Ulrich & Herbst, Thomas & Malmgren, Sven-Göran & Reichmann, Oskar & Schweickard, Wolfgang & Wiegand, Herbert Ernst (red.): *Lexicographica. Series Maior*, t. 143. Berlin & Boston: De Gruyter, 297 s. ISBN: 9783110312508. DOI: <https://doi.org/10.1515/9783110312737>. [EN]
- KARAŚ, Mieczysław (1961): „Z historii badań nad słownictwem gwarowym” [w:] *Język Polski*, r. XLI, z. 3, s. 161–180; z. 5, s. 355–369. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. [PL]
- KARPOVA, Olga M. (2018): „Russian lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 619–635. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]

- KAY, Christian J. (2000): „Historical Semantics and Historical Lexicography: will the twain ever meet?” [w:] Coleman, Julie & Kay, Christian J. (red.): *Lexicology, Semantics and Lexicography*, s. 53–68, [seria:] Koerner, E. F. Konrad (red.): *Amsterdam Studies in the Theory and History of Linguistic Science: Series IV – Current Issues in Linguistic Theory*, t. 194. Amsterdam & Philadelphia: John Benjamins B.V., 273 s. ISBN: 90-272-3701-8 (EU); 1-55619-972-4 (US). [EN]
- KĘDELSKA, Elżbieta (1986): *Łacińsko-polskie słowniki drukowane pierwszej połowy XVI wieku i ich stosunek do źródeł czeskich*. Wrocław: Zakład Narodowy im. Ossolińskich, 119 s. ISBN: 9788304022102. [PL]
- KESSERLING, Wilhelm (1989-1991): „Das chronologische Wörterbuch” [w:] Hausmann, Franz Josef et al. (red.): *Wörterbücher. Dictionaries. Dictionnaires: ein internationales Handbuch zur Lexikographie = an international encyclopedia of lexicography = encyclopédie internationale de lexicographie*, t. 2, s. 1342-1348. Berlin & New York: de Gruyter. ISBN: 978-3-11-012420-0. e-ISBN: 978-3-11-020335-6. DOI: <https://doi.org/10.1515/9783110124200.2>. [DE]
- KIANGO, John Gongwe (2000): *Bantu Lexicography: A Critical Survey of the Principles and Process of Constructing Dictionary Entries*. Tokyo: Institute for the Study of Language and Cultures of Asia and Africa [ILCAA]. ISBN: 9784872977714. [EN]
- KILGARRIFF, Adam (1992): *Polysemy*, [seria:] *Cognitive Science Research Papers*. ISSN: 1350-3162. Brighton: University of Sussex, 146 s. [EN]
- KŁOSA, Anette (2013): „The Lexicographical Process (with special focus on online dictionaries)” [w:] Gouws, Rufus Hjalmar & Heid, Ulrich & Schweickard, Wolfgang & Wiegand, Herbert Ernst: *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 517–524. Berlin – Boston: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110238136.24>. [EN]
- KŁOSA, Annette (2014): „On Corpus Citations in Monolingual General Dictionaries” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t. 36, z. 1, s. 72–87. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.2015.0018>. [EN]
- KNAP, Eugen: „Poznámky k Příručnímu slovníku jazyka českého” [w:] *Naše řeč* (1939): r. 23, nr 6, s. 166–175; r. 23, nr 7, s. 210–215; r. 23, nr 9–10, s. 271–274; (1940): r. 24, nr 1, s. 17–21; r. 24, nr 4, s. 113–118; r. 24, nr 5–6, s. 168–172; r. 24, nr 9–10, s. 277–281; (1941): r. 25, nr 1, s. 7–11; r. 25, nr 8, s. 238–241; (1942): r. 26, nr 1, s. 13–18. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/>. [CZ]
- KNOLL, Vladislav (2017): „Sociolingvistická situace psaných slovanských jazyků v 1. polovině 18. století” [w:] Kuraskiewicz, Władysław (red.): *Slavia Occidentalis*, t. 74, z. 1, s. 7–40. Poznań: Poznańskie Towarzystwo Przyjaciół Nauk. ISSN: 0081-0002. ISBN: 83-7063-039-1. DOI: <https://doi.org/10.14746/so.2017.74.1>. [CZ]
- KOSEK, Iwona & PRZYBYSZEWSKI, Sebastian & CZEREPowicka, Monika (2017): „Realizacja pozycji podmiotu otwieranej przez zwrot frazeologiczny a jego paradygmat (na wybranych przykładach)” [w:] *Prace Językoznawcze*, t. XIX, s. 113-122. Olsztyn: Uniwersytet Warmińsko-Mazurski. ISSN: 1509-5304. URL: http://uwm.edu.pl/verbel/wp-content/uploads/2014/12/podmiot_PJUWM.pdf. [PL]
- KOSEK, Iwona & PRZYBYSZEWSKI, Sebastian (2018): „Słownik paradygmatów frazeologizmów werbalnych. Wprowadzenie”, sekcja 1.2. URL: <http://uwm.edu.pl/verbel/wp-content/uploads/2018/10/wprowadzenie-1.pdf>. [PL]
- KRIDALAKSANA, Harimurti (1979): „Lexicography in Indonesia” [w:] Pang, Alvin & Yeo, Marie Alina (red.): *RELC Journal*, t. 2 (10), s. 57-66. New York: SAGE Publications. ISSN: 0033-6882. e-ISSN: 1745-526X. DOI: <https://doi.org/10.1177/003368827901000205>. [EN]
- KRIPKE, Saul (1972): „Naming and necessity” [w:] Donald, Davidson & Gilbert, Harman (red.): *Semantics of Natural Language*, s. 253-355. Boston: D. Reidel Publishing Co. ISBN: 978-90-277-0310-1. DOI: https://doi.org/10.1007/978-94-010-2557-7_9. [EN]

- KUBIŠTA, Jan (1935): „Denní tisk jako pramen lexikografický” [w:] *Naše řeč*, r. 19, nr 6–7, s. 240–244. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=3114>. [CZ]
- LASKOWSKI, Roman (1991): „Leksykografia” [w:] Urbańczyk, Stanisław: *Encyklopedia języka polskiego*, s. 183. Wrocław: Zakład Narodowy im. Ossolińskich. ISBN: 8304045079. [PL]
- LEHMANN, Christian (2014–2021): *Christian Lehmann* (strona www). URL: <https://www.christianlehmann.eu/index.php>. [EN] [DE]
- LEMBERG, Ingrid (2001): „Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher” [w:] Lemberg, Ingrid & Schröder, Bernhard & Storrer, Angelika: *Chancen und Perspektiven computergestützter Lexikographie* (s. 71–92). Tübingen: Niemeyer (270 s.). ISBN: 3-484-39107-3. DOI [wznowienie elektroniczne z 2016 r.]: <https://doi.org/10.1515/9783110915006-005>. [DE]
- LEVINSON, Martin H. (2011): „The Oxford English Dictionary: a time-binding marvel” [w:] *ETC: A Review of General Semantics*, t. 68, z. 4 (październik 2011), s. 461–469. Fort Worth: Institute of General Semantics. ISSN: 0014-164X. URL: <https://www.jstor.org/stable/42579154>. [EN]
- LINO, Teresa (2018): „Portuguese lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 608–618. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- LITRE, Émile & DEVIC, L.-Marcel (1879): *Dictionnaire de la langue française : supplément*. Paris: Hachette. [FR]
- LIU Yèqiū 刘叶秋 (1992): 中国字典史略 [Zhōngguó zidiǎn shǐ lüè ‘Krótka historia słowników chińskich’]. Ann Arbor: University of Michigan Press, 276 s. [CN]
- ŁOZOWSKI, Przemysław (2018): „Dictionaries and culture” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 166–178. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- LUBASZEWSKI, Wiesław & BYRSKI, Wiesław (1989): „Komputerowy system prac leksykograficznych. Założenia wstępne” [w:] Lubaś, Władysław (red.): *Wokół słownika współczesnego języka polskiego*, t. 2, s. 39–58, [seria:] *Studia Leksykograficzne*, nr 3. Wrocław: Zakład Narodowy im. Ossolińskich. ISSN: 0860-6471. [PL]
- LUBASZEWSKI, Wiesław (1995): „Komputerowa kartoteka słownika fleksyjnego” [w:] Basara, Jan & Lubaś, Władysław & Rymut, Kazimierz (red.): *Studia gramatyczne XI*, [seria:] Prace Instytutu Języka Polskiego, t. 93. Kraków: Instytut Języka Polskiego PAN. ISBN: 83-85579-47-8. [PL]
- LUKSZYN, Jerzy & ZMARZER, Wanda (2006): *Teoretyczne podstawy terminologii*, wyd. 2. Katedra Języków Specjalistycznych, Warszawa. [PL]
- LURIE, David & KWON Heokseung & PHAN, John D. (2019): „Japanese, Korean, and Vietnamese from c. 1800” [w:] Considine, John (red.): *The Cambridge World History of Lexicography*, s. 340–365. Cambridge: Cambridge University Press, 962 s. DOI: <https://doi.org/10.1017/9781316827437.017>. ISBN: 9781107178861. [EN]
- LUTO-KAMIŃSKA, Anetta (2016): „Leksykografia historyczna a kompetencja językowa leksykografa” [w:] Prace Filologiczne, t. LXVIII, s. 245–260. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. URL: <https://www.ceeol.com/search/article-detail?id=423250>. [PL]
- MAC AMHLAIGH, Liam (2014): „An Examination of Irish Language Lexicographers of the Twentieth Century” [w:] *Proceedings of the Harvard Celtic Colloquium*, t. 34 (2014), s. 166–179. Newtown: Department of Celtic Languages & Literatures, Harvard University. URL: <https://www.jstor.org/stable/24640158>. [EN]
- MACHAČ, Jaroslav (1971): „Vznik a vývoj prvního českého jazykovědného pracoviště (K 60. výročí založení Kanceláře Slovníku jazyka českého)” [w:] *Naše řeč*, r. 54, nr 4, s. 190–200. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=5575>. [CZ]

- MACHAČ, Jaroslav (1979): „K nespisovným složkám frazeologie národního jazyka (Frazeologismy slangové a vulgární” [w:] *Naše řeč*, r. 62, nr 1, s. 7–17. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=6093>. [CZ]
- MAJDAK, Magdalena (2004): „Kwalifikatory jako narzędzie opisu słowa w największych powojennych słownikach języka polskiego” [w:] *Prace Filologiczne*, t. XLIX, s. 283–316. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. [PL]
- MAJEWSKA, Małgorzata B. (1998): „O potrzebie rejestrowania haseł morfemowych w słownikach języka polskiego” [w:] *Poradnik Językowy*, nr 8, s. 8–21. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343. [PL]
- MAJTCZAK, Tomasz (2011): „Charakterystyka gramatyczna w słownikach dawnej japońszczyzny (*kogo jiten*)” [w:] Pietrow, Jarosław A. & Wojciechowski, Bartosz T. (red.): *Lexicographica Iapono-Polonica*, t. 1, s. 87–96. Warszawa: Wydawnictwo Nakładem Wydziału Polonistyki Uniwersytetu Warszawskiego, 361 s. ISBN: 978-83-62100-97-2. [PL]
- MAŁEK, Eliza & WAWRZYŃCZYK, Jan (2010): „Wokół kwestii komponentów latentnych informacji leksykograficznej w *Wielkiej Encyklopedii PWN*”. Warszawa & Łódź: SEMIOSIS LEXICOGRAPHICA & Instytut Rusycystyki Uniwersytetu Łódzkiego. ISBN: 978-83-60416-50-1. [PL]
- MALMGREN, Sven-Göran (2009): „On production-oriented information in Swedish monolingual defining dictionaries” [w:] Nielsen, Sandro & Tarp, Sven (red.): *Lexicography in the 21st Century: In honour of Henning Bergenholtz*, s. 93–102, [seria:] L’Homme, Marie-Claude & Kageura, Kyo (red.): *Terminology and Lexicography Research and Practice (TLRP)*. Amsterdam & Philadelphia: John Benjamins Publishing Company. ISBN (twarda): 978-90-272-2336-4. ISBN (miękka): 978-90-272-8901-8. [EN]
- MANOVICH, Lev (2006): *Język nowych mediów*. Warszawa: Wydawnictwa Akademickie i Profesjonalne, 483 s. ISBN: 9788360501023. [PL]
- MANOVICH, Lev (2020): *Cultural Analytics*. Cambridge: MIT Press, 336 s. ISBN: 9780262037105. [EN]
- MARKOWSKI, Andrzej (2006): *Kultura języka polskiego: Teoria. Zagadnienia leksykalne*. Warszawa: Wydawnictwo Naukowe PWN, 293 s. [PL]
- MARTINEZ, Camille (2013): „La comparaison de dictionnaires comme méthode d’investigation lexicographique” [w:] Gasiglia, Nathalie (red.): *Interpréter les dictionnaires : pluralité d’approches*, [seria:] Corbin, Pierre (red.): *Lexique*, t. 21, s. 193–220. Villeneuve-d’Ascq: Presses Universitaires du Septentrion, 252 s. ISBN: 978-2-7574-0437-9. ISSN: 0756-7138. [FR]
- MAVOUNGOU, Paul Achille (2018): „French lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 554–567. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- MCARTHUR, Tom (1986): *Worlds of Reference*. Cambridge: Cambridge University Press, 224 s. ISBN: 978-0521306379. [EN]
- MEJSTRÍK, Vladimír: „Z (...) sešitu Slovníku spisovného jazyka českého” [w:] *Naše řeč* (1963): r. 46, nr 5, s. 265–268; (1966): r. 49, nr 1, s. 47–51; (1966): r. 49, nr 4, s. 237–241. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/>. [CZ]
- MEL’ČUK, Igor (2006): „Colocaciones en el diccionario” [w:] Alonso Ramos, Margarita (red.): *Diccionarios y Fraseología*, [seria:] *Anexos de Revista de Lexicografía*, t. 3, s. 11–43. Coruña: Universidade da Coruña, 259 s. ISBN: 84-9749-221-8. [ES]
- METCALF, Allan (1996): „Typography” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t. 17, z. 1, s. 17–28. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.1996.0018>. [EN]

- MICHÁLEK, Emanuel & NĚMEC, Igor (1975): „K florentskému sympoziu o velkých historických slovnících” [w:] *Listy filologické / Folia philologica*, r. 98, nr 3 (1975), s. 174–175. Praha: Kabinet pro klasická studia Filosofického ústavu AV ČR, v. v. i. ISSN: 0024-4457. URL: <https://www.jstor.org/stable/23460059>. [cz]
- MICHÁLEK, Emanuel (1979): „Slovník spisovné češtiny” [w:] *Naše řeč*, r. 62, nr 5, s. 249–253. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=6157>. [cz]
- MIKOŁAJCZAK-MATYJA, Nawoja (2005): „Hierarchiczne uporządkowanie słownictwa a tradycyjna leksykografia” [w:] Zabrocki, Władysław: *Investigationes Linguisticae*, t. XII, s. 17–32. ISSN: 1426-188X. e-ISSN: 1733-1757. URL: <https://pressto.amu.edu.pl/index.php/il/article/view/9065>. DOI: <https://doi.org/10.14746/il.2005.12.3>. [PL]
- MORTH, Karlheinz (2018): „Arabic lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 503–517. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- MÜLLER, Peter O. (2017): „Wörterbuchkritik – Historische Wörterbücher des Deutschen” [w:] Bielińska, Monika & Schierholz, Stefan J. (red.): *Wörterbuchkritik – Dictionary Criticism*, [seria:] Gouws, Rufus Hjalmar & Heid, Ulrich & Herbst, Thomas & Reichmann, Oskar & Stefan J. Schierholz & Schweickard, Wolfgang (red.): (red.): *Lexicographica. Series Maior*, t. 152, s. 391–409. Tübingen: Max Niemeyer Verlag, 416 s. ISBN: 978-3-11-054473-2. e-ISBN (PDF): 978-3-11-054710-8. e-ISBN (EPUB): 978-3-11-054487-9. ISSN: 0175-9264. DOI: <https://doi.org/10.1515/9783110547108-017>. [DE]
- MÜLLER-SPITZER, Carolin (2013): „11. Textual structures in electronic dictionaries” [w:] Gouws, Rufus Hjalmar et al. (red.): *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 367–381. Berlin: De Gruyter Mouton. ISBN: 9783110238129. e-ISBN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.367>. [EN]
- MUNI TOKE, Valelia & HABERT, Benoît (2012) „Quand les exemples embrayés rencontrent la linguistique galiléenne” [w:] Colombat, Bernard & Fournier, Jean-Marie & Raby, Valérié (red.): *Vers une histoire générale de la grammaire française. Matériaux et perspectives*, s. 275–290. Paris: Honoré Champion. [FR]
- NĚMEC, Igor (1982): „K problému slovníkové definice” [w:] *Naše řeč*, r. 65, nr 3, s. 113–118. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=6328>. [cz]
- NĚMEC, Igor (1992): „Emanuel Michálek a česká historická lexikologie a lexikografie” [w:] *Listy filologické / Folia philologica*, r. 115, nr 2 (1992), s. 12–17. Praha: Kabinet pro klasická studia Filosofického ústavu AV ČR, v. v. i. ISSN: 0024-4457. URL: <https://www.jstor.org/stable/23466735>. [cz]
- NGUYỄN Đình-Hòa (1987): „Vietnamese Lexicography” [artykuł odczytany na:] *Annual Meeting of the International Association of Applied Linguistics(8th, Sydney, New South Wales, Australia, August 16-21, 1987)*. 8 s. URL: <https://files.eric.ed.gov/fulltext/ED302082.pdf>. [EN]
- NIELSEN, Sandro (1994): *The Bilingual LSP Dictionary. Principles and Practice for Legal Language*. Tübingen: Gunter Narr Verlag, 308 s. ISBN: 978-3823345336. [EN]
- NIELSEN, Sandro (2009): „Reviewing printed and electronic dictionaries: A theoretical and practical framework” [w:] Nielsen, Sandro & Tarp, Sven (red.): *Lexicography in the 21st Century: In honour of Henning Bergholtz*, s. 23–41, [seria:] L’Homme, Marie-Claude & Kageura, Kyo (red.): *Terminology and Lexicography Research and Practice (TLRP)*. Amsterdam & Philadelphia: John Benjamins Publishing Company. ISBN (twarda): 978-90-272-2336-4. ISBN (miękka): 978-90-272-8901-8. [EN]
- NIELSEN, Sandro (2018a): „Dictionary criticism” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 78–90. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]

- NIELSEN, Sandro (2018b): „Lexicography and interdisciplinarity” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 93–104. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- NIKOLIĆ-HOYT, Anja (2002): „Izrada tezaurusa hrvatskoga jezika” [w:] *Suvremena lingvistika*, t. 53–54, nr. 1–2 Rujan. Zagreb: Hrvatsko filološko društvo, s. 73–83. ISSN: 0586-0296. e-ISSN: 1847-117X. URL: https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=24773.
- NIKOLIĆ-HOYT, Anja (2004): *Konceptualna leksikografija: prema tezaurusu hrvatskoga jezika*. Zagreb: Hrvatska sveučilišna naklada, 264 s. ISBN: 978-953-169-083-9. [HR]
- NKOMA, Dion (2018): „Dictionaries and language policy” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 152–165. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- OGRODNICZUK, Maciej (2017): „Lingwistyka komputerowa dla języka polskiego: dziś i jutro” [w:] *Język Polski*, nr 1 (2018), s. 18–28. ISSN: 0021-6941. URL: <https://www.ceeol.com/search/article-detail?id=528569>. [PL]
- OKIMORI Takuya 沖森卓也 et al. (red.) (1996): 日本辞書辞典 [*Nihon jisho jiten* ‘Encyclopedia of Dictionaries Published in Japan’]. Tokyo: ISBN 4-273-02890-5. [JP]
- OLÍMPIO DE O. SILVA, María Eugênia (2006): „Los ejemplos en el tratamiento lexicográfico de las unidades fraseológicas” [w:] Alonso Ramos, Margarita (red.): *Diccionarios y Fraseología*, [seria:] *Anexos de Revista de Lexicografía*, t. 3, s. 235–248. Coruña: Universidade da Coruña, 259 s. ISBN: 84-9749-221-8. [ES]
- OSSELTON, N. E. (1986): „Dr Johnson and the English phrasal verb” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): *Fulbright Papers*, t. 1, s. 7–16. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8.
- PATKAR, Madhukar M. (1981): *History of Sanskrit Lexicography*. New Delhi: Munshiram Manoharlal Publishers, 206 s. ISBN: 978-8121502122 . [EN]
- PENA, Jesús (1994–1995): „Formación de palabras, gramática y diccionario” [w:] Gutiérrez Cuadrado, Juan & Pérez Pascual, José Ignacio (red.): *Revista de Lexicografía*, t. 1, s. 163–181. ISSN: 1134-4539. e-ISSN: 2603-6673. DOI: <https://doi.org/10.17979/rlex.1995.1.0.5699>. [ES]
- PERSSON, Gunnar (2005): „Dictionaries as Mirrors of Social and Cultural Change” [w:] Gottlieb, Henrik & Mogensén, Jens Erik & Zettersten, Arne (red.): *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen*, s. 427–433. Tübingen: Max Niemeyer Verlag. ISBN: 3-484-39115-4. ISSN: 0175-9264. [EN]
- PETERS, Pam & FERNÁNDEZ, Trinidad (2018): „Lexicography and applied linguistics” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 105–122. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- PIETROW, Jarosław A. (2011): „Typologia definicji w słownikach języka japońskiego (kokugo jiten)” [w:] Pietrow, Jarosław A. & Wojciechowski, Bartosz T. (red.): *Lexicographica Iapono-Polonica*, t. 1, s. 213–236. Warszawa: Wydano Nakładem Wydziału Polonistyki Uniwersytetu Warszawskiego, 361 s. ISBN: 978-83-62100-97-2. [PL]
- PODHAJECKA, Mirosława & BIELIŃSKA, Monika (2008): „An English-Polish Glossary of Lexicographical Terms: A Description of the Compilation Process” [w:] Bernal, Elisenda & DeCesaris, Janet (red.): *Proceedings of the XIII EURALEX International Congress. 15-19 July 2008*, s. 1041–1049. Barcelona: Universitat Pompeu Fabra, 1591 s. ISBN: 978-84-96742-67-3. URL: http://www.euralex.org/elx_proceedings/Euralex2008/101_Euralex_2008_Mirosława%20Podhajecka_Monika%20Bielinska_An%20English-Polish%20Glossary%20of%20Lexicographical%20Terms_A%20Description.pdf. [PL]
- POWITZ, Gerhardt (1975): „Einführung und Bibliographie zu Johann Leonhard Frisch, Teutsch-Lateinisches Wörter-Buch (1741)” [w:] Henne, Helmut (red.): *Deutsche Wörterbücher des 17. und 18. Jahrhunderts*, s. 93–

108. Hildesheim & New York: Olms. ISBN: 978-3-487-05469-8. URL: <https://d-nb.info/1118508300/34>. [DE]
- PRESNER, Todd (2010): „Digital Humanities 2.0: A Report on Knowledge, OpenStax CNX” [w:] Bailar, Melissa (red.): *Emerging Disciplines: Shaping New Fields of Scholarly Inquiry in and beyond the Humanities*. URL: <http://cnx.org/contents/2742bb37-7c47-4bee-bb34-0f35bda760f3@6>. [EN]
- PRZYBYSZEWSKI, Sebastian & CZEREPOWICKA, Monika & KOSEK, Iwona (2015): „The problem of lemmatisation in the Polish inflectional dictionary of verbal MWEs” [w:] Corpas Pastor, Gloria & Buendia Castro, Miriam & Gutierrez Florido, Rut (red.): *Europhras 2015. Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives* (Book of Abstracts), s. 246–248. URL: <http://uwm.edu.pl/verbel/wp-content/uploads/2014/12/Europhras-2015-Book-of-Abstracts.pdf>. [EN]
- READ, Allen Walker (1986): „The history of lexicography” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): Fulbright Papers, t. 1, s. 28–50. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8.
- Redakce Příručního slovníku jazyka českého & Kancelář Slovníku jazyka českého (1939): „Akademický slovník” [w:] *Naše řeč*, r. 23, nr 6, s. 178–185. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=3405>. [CZ]
- RENDERS, Pascale (2019): „The Romance Languages from c. 1700” [w:] Considine, John (red.): *The Cambridge World History of Lexicography*, s. 530–552. Cambridge: Cambridge University Press, 962 s. DOI: <https://doi.org/10.1017/9781316827437.026>. ISBN: 9781107178861. [EN]
- REY, Alain (1986): „Training lexicographers: some problems” [w:] Ilson, Robert (red.): *Lexicography: An emerging international profession*, [seria:] Ilson, Robert (red.): Fulbright Papers, t. 1, s. 93–100. Manchester: Manchester University Press, 167 s. ISBN: 0-7190-1852-8. [EN]
- REY-DEBOVE, Josette (1971): *Étude linguistique et sémiotique des dictionnaires français contemporains*, [seria:] Sebeok, Thomas A. & Rey, Alain & Posner, Roland (red.): *Approaches to Semiotics [AS]*, t. 13. De Gruyter Mouton: The Hague – Paris, 329. ISSN: 0066-5576. DOI: <https://doi.org/10.1515/9783111323459>. [FR]
- RIPKA, Ivor & IMRICHOVÁ, Mária (2011): *Kapitoly z lexikológie a lexikografie*, [seria:] *Učebné texty*. Prešov: Prešovská univerzita – Filozofická fakulta, 235 s. ISBN: 978-80-555-0327-1. [SK]
- ROELCKE, Thorsten (2005): *Fachsprachen*, wyd. 2 popr. Berlin: Erich Schmidt Verlag, 253 s. ISBN: 9783503079384. [DE]
- RUDNICKA, Ewa Katarzyna (2015): „Elementy informacji gramatycznej w słowniku pod red. Tadeusza Lehra-Spławińskiego jako przykład uwarunkowań socjokulturowych w leksykografii” [w:] Rak, Maciej (red.): *LingVaria*, nr 19, s. 287–298. ISSN: 1896-2122. DOI: <https://doi.org/10.12797/LV.10.2015.19.19>. URL: <https://journals.akademicka.pl/lv/article/view/2780/2524>. [PL]
- SAGER, Juan C. (1990): *A Practical Course in Terminology Processing*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 258 s. ISBN (twarda): 9789027220769 (EU), 9781556191121 (USA). ISBN (miękka): 9789027220776 (EU), 9781556191138 (USA). e-ISSN 9789027274342. DOI: <https://doi.org/10.1075/z.44>. [EN]
- SALONI, Zygmunt & SZPAKOWICZ, Stanisław & ŚWIDZIŃSKI, Marek (1981): „Szkic koncepcji ogólnego słownika podstawowego współczesnej polszczyzny pisanej” [w:] Saloni, Zygmunt (1981): *Dwa studia z polskiej leksykografii współczesnej*, s. 63–88. Białystok: Sekcja wydawnicza Filii UW w Białymstoku, 88 s. [PL]
- SALONI, Zygmunt (2005): „Co istnieje, a co nie istnieje we fleksji polskiej?” [w:] *Język Polski*, r. LXXXV, z. 1, s. 57–58. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. [PL]
- SATKIEWICZ, Halina (2000): „Teoria kultury języka Witolda Doroszewskiego a współczesne teorie z tego zakresu” [w:] *Poradnik Językowy*, nr 1/2000, s. 14–17. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343. [PL]

- SECO, Manuel (1978): „Problemas formales de la definición lexicográfica” [w:] *Estudios ofrecidos a Emilio Alarcos Llorach*, t. 2, s. 217–239. Oviedo: Universidad de Oviedo, 537 s. ISBN: 84-400-5165-4. [ES]
- SEDLÁK, Jan V. (1928): „Deutsch-Tschechisches — Tschechisch-Deutsches Militärwörterbuch” [w:] *Naše řeč*, r. 12, nr 9, s. 208–211. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=2528>. [CZ]
- SHIVTIEL, Avihai (1993): „Root-Dictionary or Alphabetical Dictionary: A Methodological Dilemma” [w:] Dévényi, Kinga & Iványi, Tamás & Shivtiel, Avihai (red.): *Proceedings of the Colloquium on Arabic Lexicology and Lexicography (C. A. L. L.): Budapest, 1-7 September 1993. Part One*, s. 13–25, [seria:] Fodor, Alexander (red.): *The Arabist*, t. 6–7. Budapest: Eötvös Loránd University Chair for Arabic Studies: Csoma de Kőrös Society Section of Islamic Studies. ISSN: 0239-1619. [EN]
- SCHAEDER, Burkhard (1987): *Germanistische Lexikographie*. Tübingen: De Gruyter Mouton, 184 s. ISBN: 9783110993882. e-ISSN: 9783111347356. [DE]
- SCHAEDER, Burkhard (2010): „Der Abrogans und die Anfänge der deutschsprachigen Lexikographie” [w:] *Germanica Wratislaviensia*, t. 131, s. 137–157, [seria:] *Acta Universitatis Wratislaviensis*, nr 3256. Wrocław: Instytut Filologii Germańskiej Uniwersytetu Wrocławskiego. ISSN: 0435-5865. URL: <https://wuw.pl/gwr/article/view/2025>. [DE]
- SIATKOWSKI, Janusz (2000): „O leksykografii czeskiej” [w:] *Poradnik Językowy*, nr 1/2000, s. 24–37. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343. [PL]
- SIEKIERSKA, Krystyna (1992): „Rzeczywistość w słownikach Knapusza i Troca” [w:] *Prace Filologiczne*, t. XXXVII, s. 259–268. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego. ISSN: 0138-0567. [PL]
- SINCLAIR, John M. (1984): „Lexicography as an Academic Subject” [w:] Hartmann, Reinhard Rudolf Karl (red.): *LEXeter '83: proceedings: Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*, s. 3–12, [seria:] Corbin, Pierre & Heid, Ulrich & Herbst, Thomas & Malmgren, Sven-Göran & Reichmann, Oskar & Schweickard, Wolfgang & Wiegand, Herbert Ernst (red.): *Lexicographica. Series Maior*. Tübingen: Max Niemeyer Verlag. ISBN: 9783484309012. e-ISSN: 9783111593166. ISSN: 0175-9264. DOI: <https://doi.org/10.1515/9783111593166-004>. [EN]
- SINGH, Parman & TRIPATHI, Arimardan Kumar (2018): „Hindi lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 586–598. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- SINGH, Parman (2013): *Hindi Lexicography: Past, Present and Future*. Germany: Lambert Academic Publishing, 400 s. ISBN: 978-3-659-42988-0. [EN]
- SIUCIAK, Mirosława (2016): „Błąd czy innowacja? Problematyka poprawnościowa z perspektywy historyka języka” [w:] *Język Polski*, nr 1 (2016), s. 36–44. ISSN: 0021-6941. URL: <https://www.ceeol.com/search/article-detail?id=414271>. [PL]
- SKARŻYŃSKI, Mirosław & SMULKOWA, Elżbieta (2018): *Korespondencja Kazimierza Nitscha i Antoniny Obrębskiej-Jabłońskiej 1925–1958. Część 1: 1925–1945*, [seria:] Skarżyński, Mirosław (red.): *Biblioteka LingVariorów*, t. 26 – *Z Prac Wydziału Polonistyki Uniwersytetu Jagiellońskiego: Materiały do dziejów polskiego językoznawstwa*, t. III. Kraków: Księgarnia Akademicka, 617 s. ISBN: 978-83-7638-967-7. [PL]
- ŚLĄSKI, Bolesław (1926): *Polski słownik marynarski: z dołączeniem licznych nazw niemieckich, rosyjskich i francuskich*. Poznań: Odbito Czcionkami Drukarni Poradnika Gospodarskiego, 83 s. [PL]
- SMITH, Abby (2008): „37. Preservation” [w:] Schreibman, Susan & Siemens, Ray & Unsworth, John (red.): *A Companion to Digital Humanities*. Hoboken: Wiley-Blackwell. URL: <http://www.digitalhumanities.org/companion/>. [EN]
- SOCHOR, Karel (1957): „Nové odborné slovníky” [w:] *Naše řeč*, r. 40, nr 1–2, s. 40–49. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=4529>. [CZ]

- SORNIG, Karl (1981): *Lexical Innovation: A Study of Slang, Colloquialisms, and Casual Speech*. Amsterdam: John Benjamins, 117 s. ISBN: 9789027225184. e-ISBN: 9789027280800. [EN]
- SPEARS, Richard A. (1989): „Historical Principles in Dictionaries of Non-Standardized Vocabularies” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, t. 11, z. 1, s. 97–109. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.2989.0015>. [EN]
- SPLETT, Jochen (1978): „Abrogans” [hasło w:] *Die deutsche Literatur des Mittelalters. Verfasserlexikon*, wyd. 2, t. 1. Berlin & New York. ISBN: 3-11-022248-5. [DE]
- SPOCZYŃSKA, Helena (1927): „Zbieranie słownictwa ludowego. Kilka uwag o dwu wyrazach z zakresu pokrewieństwa (i części ciała)” [w:] *Język Polski*, r. XII, z. 4, s. 119–123. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. [PL]
- STORJOHANN, Petra (2018): „German lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 568–585. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISBN: 978-1-315-10494-2. [EN]
- STUMPF, Béatrice (2009): *Lexicographie et lexicologie historique du français* (praca dyplomowa). Nancy: Université Nancy. URL: <https://tel.archives-ouvertes.fr/tel-00502081>. [FR]
- SWANEPOEL, Piet (2017): „Methods in dictionary criticism” [w:] Bielińska, Monika & Schierholz, Stefan J. (red.): *Wörterbuchkritik – Dictionary Criticism*, [seria:] Gouws, Rufus Hjalmar & Heid, Ulrich & Herbst, Thomas & Reichmann, Oskar & Stefan J. Schierholz & Schweickard, Wolfgang (red.): (red.): *Lexicographica. Series Maior*, t. 152, s. 81–112. Tübingen: Max Niemeyer Verlag, 416 s. ISBN: 978-3-11-054473-2. e-ISBN (PDF): 978-3-11-054710-8. e-ISBN (EPUB): 978-3-11-054487-9. ISSN: 0175-9264. [EN]
- SZEMIŃSKA, Weronika (2014): „Struktura słownika jako narzędzia pracy lingwistycznej” [w:] Zmarzer, Wanda & Michta, Tomasz (red.): *Narzędzia pracy lingwistycznej*. Wrocław: Instytut Rusycystyki Wydział Lingwistyki Stosowanej UW, s. 240–256. ISBN: 9788361116271. [PL]
- ŠMILAUER, Vladimír (1936): „Vojenské názvosloví ručních střelných zbraní a kulometů” [w:] *Naše řeč*, r. 20, nr 2–3, s. 25–44. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=3008>. [CZ]
- TARP, Sven (1999): „Theoretical foundations of the so-called cross-reference structures” [w:] *Lexicographica*, nr 15, s. 114–137. Berlin: De Gruyter. ISSN: 0175-6206. e-ISSN: 1865-9403. [EN]
- TOKARSKI, Ryszard (1993): „Słownictwo jako interpretacja świata” [w:] Bartmiński, Jerzy (red.): *Współczesny język polski*. Wrocław: Wiedza o Kulturze. ISBN: 83-7044-062-2. [PL]
- TOOPE, Michael (1996): *Examples in the Bilingual Dictionary* (praca magisterska). Ottawa: University of Ottawa. DOI: <http://dx.doi.org/10.20381/ruor-8047>. URL: <https://ruor.uottawa.ca/handle/10393/9942>. [EN]
- TRENCH, Richard Chenevix (1857): „On Some Deficiencies In Our English Dictionaries” [w:] *Transactions of the Philological Society*, t. 4, nr 2, s. 1–70. London: Philological Society. DOI: <https://doi.org/10.1111/j.1467-968X.1857.tb00883.x>. URL: https://books.google.pl/books/about/On_Some_Deficiencies_in_Our_English_Dict.html?id=T8ozAQAAMAAJ. [EN]
- TYL, Zdeněk (1957): „Přípravy k vydání Slovníku staročeského” [w:] *Naše řeč*, r. 40, nr 7–8, s. 227–231. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=4572>. [CZ]
- TYL, Zdeněk (1983): „K historii vzniku Jungmannova Slovníku česko-německého” [w:] *Naše řeč*, r. 66, nr 3, s. 113–119. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=6387>. [CZ]

- UCIŃSKA, Joanna (2017): „Humanistyka cyfrowa - problemy definiowania pojęcia” [w:] *Zagadnienia Informatyki Naukowej – Studia Informacyjne*, t. 55, nr 1 (109). ISSN: 0324-8194. e-ISSN: 2392-2648. DOI: <https://doi.org/10.36702/zin.352>. URL: <http://ojs.sbp.pl/index.php/zin/article/view/352>. [PL]
- UHLÍŘOVÁ, Ludmila (2011): „Reflexe činnosti Kanceláře Slovníku jazyka českého v časopise Naše řeč” [w:] *Naše řeč*, r. 94, nr 1, s. 1–13. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=8158>. [CZ]
- URBAŃCZYK, Stanisław (1955): „Rozwój ortografii polskiej” [w:] *Język Polski*, r. XXXV, z. 2, s. 81–93. Kraków: Towarzystwo Miłośników Języka Polskiego. ISSN: 0021-6941. [PL]
- VAN DER SIJS, Noline (2003): „The codification of etymological information” [w:] van Sterkenburg, Piet (red.): *A Practical Guide to Lexicography*. Amsterdam & Philadelphia: John Benjamins Publishing Company, s. 314–321. ISBN (twarda): 90-272-2329-7 (EU); 1-58811-380-9 (US). ISBN (miękka): 90-272-2330-0 (EU); 1-58811-381-7 (US). DOI: <https://doi.org/10.1075/TLRP.6>. [EN]
- VETULANI, Grażyna (2011): *Kolokacje werbo-nominalne jako samodzielne jednostki języka. Syntaktyczny słownik kolokacji werbo-nominalnych języka polskiego na potrzeby zastosowań informatycznych. Część I*, [seria:] *Lingwistyka Komputerowa*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, 190 s. ISBN: 978-83-232-2369-6. ISSN: 0867-0994. [PL]
- VOGEL, Claus (2014): *Indian Lexicography – Revised edition*. München: Peter Kirchheim Verlag. ISBN: 978-3874101455. [EN]
- WALCZAK, Bogdan (1989): „Objaśnienia etymologiczne w powojennych słownikach polskich” [w:] Saloni, Zygmunt (red.): *Studia z polskiej leksykografii współczesnej*, t. 3, s. 73–98, [seria:] *Rozprawy Uniwersytetu Warszawskiego = Dissertationes Universitatis Varsoviensis*, t. 376. Białystok: Wydawnictwo Filii Uniwersytetu Warszawskiego w Białymstoku, 374 s. [PL]
- WANIAKOWA, Jadwiga (2014): „Nieistniejący słownik polskich gwarowych nazw roślin, czyli o niektórych problemach gwarowej leksykografii” [w:] *Prace Filologiczne*, t. LXV, s. 403–411. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. [PL]
- WANNER, Leo (2006): „El Corpus como un Diccionario de Colocaciones?” [w:] Alonso Ramos, Margarita (red.): *Diccionarios y Fraseología*, [seria:] *Anexos de Revista de Lexicografía*, t. 3, s. 161–173. Coruña: Universidade da Coruña, 259 s. ISBN: 84-9749-221-8. [ES]
- WAWRZYŃCZYK, Jan (1989): *Nad Słownikiem Języka Polskiego 1958–1969*. Toruń: Instytut Filologii Słowiańskiej Uniwersytetu Mikołaja Kopernika, 118 s. [PL]
- WEINGART, Miloš (1917a): „Slovník jazyka českého” [w:] *Naše řeč*, r. 1, nr 1, s. 8–11. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=4>. [CZ]
- WEINGART, Miloš (1917b): „Slovník jazyka českého” [w:] *Naše řeč*, r. 1, nr 2, s. 44–46. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=31>. [CZ]
- WEINGART, Miloš (1935): „Individuální slovník hovorové češtiny” [w:] *Naše řeč*, r. 19, nr 6–7, s. 334–351. Praha: Ústav pro jazyk český AV ČR. ISSN: 0027-8203. e-ISSN: 2571-0893. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=3129>. [CZ]
- WEINREICH, Uriel (1967): „Lexicographic definition in descriptive semantics” [w:] Householder, Fred W. & Saporta, Sol (red.): *Problems in lexicography*, s. 25–44. The Hague: Mouton. [EN]
- WELLS, Ronald A. (1973): *Dictionaries and the Authoritarian Tradition: A Study in English Usage and Lexicography*, [seria:] van Schooneveld, C. H. (red.): *Series Practica*, t. 196. The Hague & Paris: Mouton, 129 s. [EN]

- WESTENHOLZ, Aage (1985): „An Essay on the Sumerian «Lexical» Texts of the Third Millennium” [w:] *Orientalia. Nova Series*, t. 54, nr 1/2, s. 294–298. Roma: GBPress – Gregorian Biblical Press. ISSN: 0030-5367. URL: <https://www.jstor.org/stable/43075326>. [EN]
- WIEGAND, Herbert Ernst & GOUWS, Rufus Hjalmar (2013): „4. Macrostructures in printed dictionaries” [w:] Gouws, Rufus Hjalmar et al. (red.): *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Focus on Electronic and Computational Lexicography*, 73–110. Berlin: De Gruyter Mouton. ISBN: 9783110238129. e-ISBN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.73>. [EN]
- WIEGAND, Herbert Ernst (1989): „Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: alphabetische Anordnungsformen und ihre Probleme” [w:] Darquennes, Jeroen & Epps, Patience (red.): *Handbooks of Linguistics and Communication Science (HSK) 5.1*, s. 371–409. ISSN: 1861-5090. [DE]
- WIEGAND, Herbert Ernst (1998): *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband*. Berlin: De Gruyter, 1162 s. ISBN: 9783110135848. [DE]
- WIEGAND, Herbert Ernst (2002): „Wissen in der Sprachlexikografie. Ein Plädoyer für einige immer noch notwendige Differenzierungen” [w:] Ezawa, Kiyoshi & Kurschner, Wilfried & Rensch, Karl Heinz & Ringmacher, Manfred (red.): *Linguistik jenseits des Strukturalismus*, s. 265–281. Tübingen: Gunter Narr Verlag, 472 s. ISBN: 9783823358657. [DE]
- WIEGAND, Herbert Ernst (2007): „Über Zugriffspfade in Printwörterbüchern. Ein Beitrag zur Schnittstelle von Benutzungshandlungen und Wörterbuchform” [w:] *Lexikos*, t. 17, 180–211. Stellenbosch: Bureau of the WAT. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/17-0-1171>. URL: <https://www.ajol.info/index.php/lex/article/view/51532>. [DE]
- WIEGAND, Herbert Ernst (2008): „Wörterbuchregister. Grundlagen einer Theorie der Register in modernen Printwörterbüchern” [w:] *Lexikos*, t. 18, 256–302. Stellenbosch: Bureau of the WAT. ISSN: 1684-4904. e-ISSN: 2224-0039. DOI: <https://doi.org/10.5788/18-0-487>. URL: <https://www.ajol.info/index.php/lex/article/view/47256>. [DE]
- WIEGAND, Herbert Ernst & SMIT, Maria (2014): „7. Mediostructures in printed dictionaries” [w:] Gouws, Rufus Hjalmar et al. (red.): *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Focus on Electronic and Computational Lexicography*, s. 214–253. Berlin: De Gruyter Mouton. ISBN: 9783110238129. e-ISBN: 9783110238136. DOI: <https://doi.org/10.1515/9783110238136.214>. [EN]
- WIERZBICKA, Anna (1986): „Tolkovo-kombinatornyj slovar' sovremennogo russkogo jazyka By Igor' Mel'cuk and Aleksandr Žolkovskij” (recenzja) [w:] *Language*, s. 684–687. New York: Linguistic Society of America. DOI: <https://doi.org/doi:10.1353/lan.1986.0071>. [EN]
- WIERZBICKA, Anna (1992): „What Are the Uses of Theoretical Lexicography?” [w:] *Dictionaries: The Journal of the Dictionary Society of North America*, nr 14 (1), s. 44–78. DOI: <https://doi.org/10.1353/dic.1992.0014>. [EN]
- WÓJTOWICZ, Beata (2017): „Revisiting Lemma Lists in Swahili Dictionaries” [w:] Ndinga-Koumba-Binza, Hugues Steve (red.): *Lexikos*, t. 27, s. 561–577. Stellenbosch: Buro Van Die Wat, 623 s. ISBN: 978-0-9814434-2-3. ISSN: 1684-4904. e-ISSN: 2224-0039. [EN]
- WÓJTOWICZ, Marian (2008): „Dwa nowe słowniki inwersyjne: Piotr Wierzchoń, *Indeks a tergo jednostek przekładowych «Wielkiego słownika rosyjsko-polskiego z kluczem polsko-rosyjskim»*. Porządek translatowy. Wydawnictwo TAKT, Warszawa 2006, ss. XXXIX, 2019; Piotr Wierzchoń, *Indeks a tergo jednostek przekładowych «Wielkiego słownika rosyjsko-polskiego z kluczem polsko-rosyjskim»*. Porządek translandowy. Wydawnictwo TAKT, Warszawa 2007, ss. LXIV, 2011” [w:] Nowak, Piotr & Nowakowski, Paweł (red.): *Język, komunikacja, informacja*, t. 3. Poznań: Sorus, 230 s. ISBN: 978-83-89949-56-1. ISSN: 1896-9585. [PL]

- WOOLDRIDGE, Terrence Russon (1995): „Naissance et première floraison de l'exemple dans la lexicographie française : étude historique et typologique” [w:] *Langue française*, t. 106, s. 8–20. Paris: Armand Colin. ISSN: 0023-8368. e-ISSN: 1957-7982. URL: https://www.persee.fr/doc/lfr_0023-8368_1995_num_106_1_6440. [FR]
- WOTJAK, Gerd (1998): „¿Cómo tratar las unidades fraseológicas (UF) en el diccionario?” [w:] Wotjak, Gerd (red.): *Estudios de fraseología y fraseografía del español actual*. Madrid: Iberoamericana, s. 307–321. ISBN: 84-95107-18-X. [ES]
- WOŹNIAK, Ewa & ZARĘBSKI, Rafał (2018): O (nie)przydatności źródeł leksykograficznych w badaniach XIX-wiecznej terminologii prawnej” [w:] *Język Polski*, nr 4 (2018), s. 59–69. ISSN: 0021-6941. URL: <https://www.ceeol.com/search/article-detail?id=735636>. [PL]
- WOŹNIAK, Kazimierz (2006): „Propozycja ujednoczenia skrótów kwalifikatorów stosowanych w polskiej leksykografii gwarowej” [w:] Sierociuk, Jerzy (red.): *Gwary dziś*, t. 3: *Wewnętrzne zróżnicowanie języka wsi*, [seria:] *Prace Komisji Językoznawczej*, t. 45. Poznań: Poznańskie Towarzystwo Przyjaciół Nauk, s. 215–247. ISSN: 1898-9276. [PL]
- XUE Shiqi (1982): „Chinese Lexicography Past and Present” [w:] Frawley, William (red.): *Dictionaries: Journal of the Dictionary Society of North America*, nr 4 (1982), s. 151–169. Collingswood: Dictionary Society of North America. ISSN: 0197-6745. DOI: <https://doi.org/10.1353/dic.1982.0009>. URL: <https://muse.jhu.edu/article/456382/summary>. [EN]
- YONG Heming (雍和明) & PENG Jing (彭敬) (2018): „Chinese lexicography in the Internet era” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 518–539. London and New York: Routledge, 809 s. ISBN: 978-1-138-94160-1. e-ISSN: 978-1-315-10494-2. [EN]
- ZACHS, Fruma & DROR, Yehudit (2020): „The *Muḥīt al-Muḥīt* Dictionary: The Transition from Classical to Modern Arabic Lexicography” [w:] *Journal of Arabic and Islamic Studies*, t. 20, s. 15–32. Oslo: Oslo University. ISSN: 0806-198X. DOI: <https://doi.org/10.5617/jais.7853>. [EN]
- ŻMIGRODZKI, Piotr (2005): „Słownik jako korpus tekstów - korpus tekstów jako słownik. Perspektywy polskiej leksykografii naukowej” [w:] *Poradnik Językowy*, nr 6 (625), s. 3–14. Warszawa: Towarzystwo Kultury Języka – Dom Wydawniczy „Elipsa”. ISSN: 0551-5343. [PL]
- ŻMIGRODZKI, Piotr (2012): *Wielki słownik języka polskiego: zasady opracowania*. Kraków, 77 s. URL: <https://rcin.org.pl/dlibra/publication/72562/edition/53406/content>. [PL]
- ŻUROWSKI, Sebastian (2013): „Gdzie szukać w słownikach nieleksykałnych jednostek języka” [w:] *Prace Filologiczne*, t. LXIV, nr 1, s. 421–430. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego. ISSN: 0138-0567. URL: <https://www.ceeol.com/search/article-detail?id=4795>. [PL]
- АЛЕКСЕЕВ, П. М. (1975): *Статистическая лексикография (типология, составление и применение частотных словарей)*. Ленинград: Ленинградский ордена Трудового Красного Знамени Государственный педагогический институт имени А. И. Герцена, 120 s. [RU]
- БОРОВКОВ, А. К. (1957): „Из опыта составления русско-национальных словарей” [w:] Охманова, О. С. (red.): *Лексикографический сборник, выпуск 1*, s. 135–159. Москва: Государственное издательство иностранных и национальных словарей, 188 s. [RU]
- ДЕВКИН, В. Д. (2005): *Немецкая лексикография. Учебное пособие для вузов*. Москва: Высшая Школа, 670 s. ISBN: 5-06-004998-1. [RU]
- ДЕНИСОВ, П. Н. (1980): *Лексика русского языка и принципы её описания*. Москва: Русский язык, 253 s. [RU]
- ДЗЕРА, Оксана (2004): *Англійська й українська лексикографія (коротка теорія та історія)*. Львів: Видавничий центр ЛНУ імені Івана Франка, 170 s. [EN]
- КАРПОВА, О. М. (1989): *Словари языка писателей*. Москва: Издательство МПИ, 108 s. [RU]

- ПОКРОВСЬКА, О. А. & САПОЖНІКОВА, Л. Я. & РУДЕНКО, С. М. (2015): *Українська термінографія: історія, теорія, практика*. Харків: ХДУХТ, 134 s. [RU]
- ФИЛИН, Федот Петрович (1957): „От редакции” [w:] Охманова, О. С. (red.): *Лексикографический сборник, выпуск 1*, s. 3–8. Москва: Государственное издательство иностранных и национальных словарей, 188 s. [RU]
- ЧЕРНЫШЕВА, М. И. (red.) (2013): *Славянская лексикография*. Москва: Азбуковник, 886 s. ISBN: 978-5-91172-076-6. [RU]
- ЩЕРБА, Лев Владимирович (1974): „Опыт Общей Теории Лексикографии” [w:] Щерба, Лев Владимирович (red.): *Языковая система и речевая деятельность*. Москва: Едиториал УРСС, 432 s. ISBN: 55-354-00816-6. [RU]

Wykaz użytych skrótów

- DLF** – *Dictionnaire de la langue française*, t. I–IV, autorstwa Émile’a Littrégo (Paris 1863–1872).
- DWB** – *Deutsches Wörterbuch*, t. I–XVII, autorstwa Jacoba Grimma i Wilhelma Grimma (Leipzig 1854–1961).
- KSEKart** – Kartoteka słownika etnolingwistycznego (nigdy nieukończonego), autorstwa Jana Karłowicza.
- Modrz** – *Wyrazy które umarły i które umierają*, autorstwa Teofila Modrzejewskiego (Warszawa, 1936).
- MSGP** – *Mały słownik gwar polskich*, pod redakcją Renaty Kucharzyk, Anny Niezabitowskiej, Jerzego Reichana, Marii Tokarz, Wiktorii Wójcickiej i Jadwigi Wronicz (Kraków, 2010).
- OED** – *Oxford English Dictionary*.
- SAut** – *Słownik techniczny dla automobilistów polsko-francusko-niemiecki, francusko-polsko-niemiecki oraz niemiecko-polsko-francuski*, autorstwa Stanisława Szydelskiego (Lwów – Warszawa 1921).
- SBłęd** – *Słowniczek błędów językowych i najważniejszych prawideł gramatycznych*, autorstwa Artura Passendorfera (Warszawa, 1905).
- SBor** – *Słownik etymologiczny języka polskiego*, autorstwa Wiesława Borysia (Kraków, 2005).
- SBot** – *Słownik botaniczny polsko-łacińsko-hebrajski z alfabetycznymi spisami nazw łacińskich i hebrajskich*, autorstwa Lejba Pawe (Warszawa, 1933).
- SBrChem** – *Słownik polsko-francusko-niemiecko-rosyjski: Broń chemiczna*, autorstwa Marceli Wścieklicy-Pollak (Warszawa, 1932).
- SBrk** – *Słownik etymologiczny języka polskiego*, autorstwa Aleksandra Brücknera (Kraków, 1927).
- SDun** – *Słownik współczesnego języka polskiego*, autorstwa Bogusława Dunaja (Warszawa, 1996).
- SGeo** – *Słownik geograficzny*, autorstwa Józefa Haliczera (Tarnopol, 1933).
- SGJP** – *Słownik gramatyczny języka polskiego*, wyd. IV online, autorstwa Marcina Wołńskiego, Zygmunta Saloniego, Roberta Wołosza, Włodzimierza Gruszczyńskiego, Danuty Skowrońskiej i Zbigniewa Bronka (Warszawa, 2020) (<http://sgjp.pl> – stan z 16.08.2020).
- SHandl** – *Słownik handlowy Polsko-Rosyjsko-Niemiecko-Francusko-Angielski*, autorstwa Tadeusza Szulborskiego (Warszawa, 1914).

- SJPDor** – *Słownik języka polskiego*, t. 1–11, zwany „Słownikiem Doroszewskiego”, pod redakcją Witolda Doroszewskiego (Warszawa, 1955–1969).
- SJPSzym** – *Słownik języka polskiego*, t. 1–3, zwany „Słownikiem Szymczaka”, pod red. Mieczysława Szymczaka (Warszawa, 1978–1981).
- SKart** – *Słownik gwar polskich*, t. 1–6, zwany „Słownikiem Karłowicza”, autorstwa Jana Karłowicza (Kraków, 1900–1911).
- SKok** – *Słownik ortograficzny języka polskiego*, autorstwa Władysława Kokowskiego (1903).
- SKol** – *Słowniczek kolejowy*, autorstwa Karola Stadtmüllera (Kraków, 1919).
- SKop** – *Słownik wyrazów obcych i zwrotów obcojęzycznych z almanachem*, autorstwa Władysława Kopalińskiego (Warszawa, 2007).
- SL** – *Słownik języka polskiego*, zwany „Słownikiem Lindego”, autorstwa Samuela Bogumiła Lindego (Warszawa, 1807–1814).
- SLeśn** – *Słownik leśniczy polsko-niemiecko-francusko-angielski*, autorstwa Marjana Małaczyńskiego (Lwów, 1923).
- SLot** – *Niemiecko-polski słownik lotniczy*, autorstwa Karola Stadtmüllera (Lwów – Warszawa, 1920).
- Słow** – *Słownik łowiecki*, autorstwa Jerzego Orensteina-Oreńskiego (Warszawa, 1913).
- SMor** – *Słownik morski polsko-angielsko-francusko-niemiecko-rosyjski*, pod redakcją Bogusława Bagniewskiego (Warszawa, 1929-1936).
- SMyśl** – *Polsko-niemiecki i niemiecko-polski najczęściej używanych wyrazów i zwrotów gwary myśliwskiej*, autorstwa X. Ludwika Niedbała (Poznań, 1917).
- SNazwGeo** – *Atlas nazw geograficznych słowiańszczyzny zachodniej*, autorstwa Stanisława Kozierowski (Poznań, 1934).
- SOkr** – *Niemiecko-polski słownik okrętowy*, autorstwa Karola Stadtmüllera (Lwów – Warszawa, 1921).
- SOrt** – *Słownik ortograficzny* (Szkocja, 1946).
- SPedag** – *Podręczny leksykon pedagogiczny*, autorstwa Kazimierza Królińskiego (Poznań, 1935).
- SPrzędz** – *Słowniczek przędzalniczy*, autorstwa Adama Trojanowskiego (Warszawa, 1910).
- SRyk** – *Słownik języka polskiego*, autorstwa Erazma Rykaczewskiego (Berlin – Poznań, 1866).
- SRzem** – *Słownictwo rzemieślnicze*, autorstwa Karola Stadtmüllera (Kraków, 1921).
- SSJČ** – *Slovník spisovného jazyka českého*, t. I–IV, autorstwa B. Havránka, J. Běliče, M. Helcla i A. Jedličky (Praž 1960–1971).
- SStużGeo** – *Słownictwo służby geograficznej angielsko-polskie* (Edinburgh, 1946).

- SSok** – *Słownik francusko-polski i polsko-francuski zawierający wykazy, terminy i zwroty z dziedziny bankowości, handlu, skarbowości, finansów, giełdy, księgowości i innych*, autorstwa Józefa Sokołowskiego (Warszawa, 1933).
- SStarop** – *Słownik staropolski*, t. I-XI, pod redakcją Stanisława Urbańczyka (Warszawa, 1953-1955).
- SSzob** – *Słownik poprawnej polszczyzny*, wyd. II, autorstwa Stanisława Szobera (Warszawa, 1948).
- SŚlus** – *Słownik niemiecko-polski dla ślusarzy, tokarzy, kowali, kotlarzy i blachnierzy* (Poznań, 1913).
- STakt** – *Słownik taktyczny polsko-francusko-niemiecko-rosyjski*, autorstwa Tadeusza Bałabana, Wacława Berki i Tadeusza Frasunkiewicza (Warszawa, 1924).
- STech** – *Słownik techniczny*, t. 1-2, autorstwa Karola Stadtmüllera i Karola Stadtmüllera (Warszawa, 1923).
- STey** – *Słownik elektrotechniczny francusko-polski polsko-francuski: telefonja, telegrafja, radjotelegrafja*, autorstwa Teyssiera, Georges-a-Maurice'a (Lwów – Warszawa, 1925).
- STkac** – *Słownik tkacko-wykończalniczy w 5-u językach*, autorstwa Adama Trojanowskiego (Warszawa, 1927).
- STow** – *Słownik towaroznawczo-przemysłowy polsko-rosyjsko-francusko-niemiecko-angielski a częściowo i łaciński (chemikalja i rośliny)*, autorstwa Artura Drewnowskiego (Wilno, 1939).
- SW** – *Słownik języka polskiego*, t. 1–8, zwany „Słownikiem warszawskim”, pod. red. Jana Karłowicza, Adama Antoniego Kryńskiego i Władysława Niedźwiedzkiego (Warszawa, 1900–1927).
- SWil** – *Słownik języka polskiego, obejmujący oprócz zbioru właściwie polskich, znaczną liczbę wyrazów z obcych języków polskiemu przyswojonych*, zwany „Słownikiem wileńskim”, autorstwa Aleksandra Zdanowicza, Michała Bohusza-Szyszki, Januarego Filipowicza, Waleriana Tomaszewicza, Floriana Czeplińskiego i Wincentego Korotyńskiego, z udziałem Bronisława Trentowskiego, a wydany przez Maurycego Orgelbranda (Wilno, 1861).
- SWojsk** – *Słownik wojskowy angielsko-polski = English-Polish Military Dictionary* (Tel-Aviv, 1942).
- SWOKart** – *Słownik wyrazów obcych*, autorstwa Jana Karłowicza (Kraków, 1895).
- SWyrTech** – *Słownik polskich wyrazów technicznych* (Warszawa, 1929-1939).
- SXVI** – *Słownik polszczyzny XVI wieku*, autorstwa Marii Renaty Mayenowej (<https://spxvi.edu.pl/> – stan z 11.08.2020).
- SXVII** – *Elektroniczny słownik języka polskiego XVII i XVIII wieku* (<https://sxvii.pl/> – stan na 21.02.2017).

- SZeg** – *Słownik techniczny dla rękodziela / zegarmistrzowskiego / złotniczego / jubilerskiego / ratowniczego / optycznego i pokrewnych branż*, autorstwa Tadeusza Pawlickiego (Poznań, 1926).
- SZiel** – *Słownik zielarski*, autorstwa Kazimierza Gierynga (Wilno, 1938).
- SŻegl** – *Współczesna żegluga morska oraz słownik żeglarski z licznymi rysunkami w tekście*, autorstwa Marjusza Zaruskiego (Poznań, 1920).
- SŻer** – *Słownik elektrotechniczny do praktycznego użytku w biurach, składach i fabrykach elektrotechn.*, autorstwa Tadeusza Żerańskiego (Lwów, 1921).
- ŚląsM** – *Polski Słownik Marynarski*, autorstwa Bolesława Śląskiego (Poznań, 1926).
- ŚląsMR** – *Słownik morsko-rybolówczy*, autorstwa Bolesława Śląskiego (Poznań, 1922).
- TEM** – *Polski słownik ortograficzny*, autorstwa Władysława Trzaskę, Ludwika Józefa Everta i Jana Michalskiego (Warszawa, 1936).
- USJP** – *Uniwersalny słownik języka polskiego*, pod redakcją Stanisława Dubisza (Warszawa, 2003).
- WSJP** – *Wielki słownik języka polskiego PAN* pod red. Piotra Źmigrodzkiego (<https://wsjp.pl/> – stan na 01.06.2021).
- WSWOiT** – *Wielki słownik wyrazów obcych i trudnych*, autorstwa Andrzeja Markowskiego & Radosława Pawelca (2017).

Wykaz tabel

Tabela 1. Elementy słownika a jego funkcje pierwotne.....	46
Tabela 2. Elementy słownika a jego funkcje wtórne.	47
Tabela 3. Porównanie treści zewnętrznych tomów SW.....	78
Tabela 4. Typy danych wejściowych.....	236
Tabela 5. Tabelaryczna reprezentacja wzorów gramatycznych.....	240
Tabela 6. Struktura informacji o walencji.....	251
Tabela 7. Przykłady normalizacji haseł.....	263
Tabela 8. Fragment przetworzonego pliku podmian.	265
Tabela 9. Porównanie cech słowników tradycyjnych i elektronicznych.....	316
Tabela 10. Fragment aneksu I.....	325

Wykaz listingów

Listing 1. Pierwsze 10 wierszy z pliku dla litery F.	236
Listing 2. Pierwsze dziesięć wierszy indeksu lokalizacji.....	239
Listing 3. Pary derywacyjne końcówek męskich i żeńskich	241
Listing 4. Wyrażenie regularne <code>Rx_header1</code> (poziom 1).....	244
Listing 5. Wyrażenie regularne <code>Rx_lemma_comp</code> (poziom 2).....	244
Listing 6. Wyrażenie regularne <code>Rx_gram</code> (poziom 2).....	244
Listing 7. Wyrażenie regularne <code>Rx_pronun</code> (poziom 2).	244
Listing 8. Wyrażenie regularne <code>Rx_syn</code> (poziom 2).....	244
Listing 9. Wyrażenie regularne <code>Rx_abbr_combo_ch</code> (poziom 2).	244
Listing 10. Wyrażenie regularne <code>Rx_pos</code> (poziom 2).	245
Listing 11. Wyrażenie regularne <code>Rx_ending</code> (poziom 3).....	245
Listing 12. Wyrażenie regularne <code>Rx_gen</code> (poziom 3).	245
Listing 13. Element składowy wyrażenia regularnego <code>Ltr_p1_s</code> (poziom 4).	245
Listing 14. Wyrażenie regularne <code>Rx_abbr_combo</code>	245
Listing 15. Wyrażenie regularne <code>Rx_abbr_combo</code> (po interpretacji).	245
Listing 16. Wyrażenie regularne <code>Rx_abbr_combo_ch</code> (po interpretacji).....	246
Listing 17. Wyrażenie regularne do przechwytywania informacji temporalnej.....	248
Listing 18. Hasło „Ferycyjan” w foncie Cambria Math.....	267
Listing 19. Wybrane znaki w foncie Cambria Math.	268

Wykaz ilustracji

Ilustracja 1. Leksykografia praktyczna i teoretyczna (metaleksykografia).....	25
Ilustracja 2. Tom II, strona 119. – sekcja z literą J rozpoczęta na środku strony.	80
Ilustracja 3. Tom VII, strona 1160. – sekcje z literami X oraz Y.	81
Ilustracja 4. Hasło „Flejtuch” – wyróżnione segmenty najwyższego poziomu.....	84
Ilustracja 5. Hasło „Flejtuch” – wyróżnione segmenty wielokrotne.	84
Ilustracja 6. Hasło „Furmański” oraz podhasło „Po furmańsku”.	85
Ilustracja 7. Hasło „Furmański” oraz podhasło „Po furmańsku” – wyróżnione segmenty...85	
Ilustracja 8. Hasło „Fałdować” oraz podhasło „F. ś.” („Fałdować się”).	86
Ilustracja 9. Początek sekcji z literą „F”.	93
Ilustracja 10. Sekcja z literą „V”.	93
Ilustracja 11. Początek sekcji z literą „I”.	94
Ilustracja 12. Początek sekcji z literą „Ó”.	94
Ilustracja 13. Sekcja z literą „X”.	95
Ilustracja 14. Początek sekcji z literą „Y”.	95
Ilustracja 15. Fragment hasła Figurować” z podhasłem „Figurowany”.	127
Ilustracja 16. „Dalsze skrócenia” w tomie V.....	129
Ilustracja 17. Układ klawiatury Polyglot – bez klawiszy modyfikujących.	189
Ilustracja 18. Układ klawiatury Polyglot – z klawiszem modyfikującym SHIFT.	189
Ilustracja 19. Układ klawiatury Polyglot z klawiszem modyfikującym ALT.....	189
Ilustracja 20. Układ klawiatury Polyglot z klawiszami modyfikującymi SHIFT oraz ALT. ..	190
Ilustracja 21. Zrzut z programu WinMerge ze strony domowej programu.....	191
Ilustracja 22. Hasło „Ichtjotomia” – przed obróbką graficzną.	197
Ilustracja 23. Hasło „Ichtjotomia” – po obróbce graficznej.....	197
Ilustracja 24. Strona 72. – obraz przed binaryzacją.	206
Ilustracja 25. Strona 72. – obraz po binaryzacji metodą Otsu.....	207
Ilustracja 26. Strona 110. – obraz przed binaryzacją.....	208
Ilustracja 27. Strona 110. – obraz po zwykłej binaryzacji z niewłaściwymi parametrami.	209
Ilustracja 28. Strona 110. – po zastosowaniu progowania metodą Otsu.	210
Ilustracja 29. Strona 112. – nieprawidłowo wykryty obrys (z artefaktem).....	211
Ilustracja 30. Strona 112. – prawidłowo wykryty obrys.	212
Ilustracja 31. Strona 112. – po obroceniu.....	214
Ilustracja 32. Strona 112. – po prawidłowym przycięciu.	215
Ilustracja 33. Strona 112. – kolumna lewa: po podziale.	216
Ilustracja 34. Strona 112. – kolumna prawa: po podziale.....	216

Ilustracja 35. Strona 112 – kolumna lewa: regiony nałożone na tekst.	218
Ilustracja 36. Strona 112 – kolumna prawa: regiony nałożone na tekst.....	218
Ilustracja 37. Fiszka nr 1 („Istoczenie”) ze strony 112.	219
Ilustracja 38. Fiszka nr 3 („Istorja”) ze strony 112.	219
Ilustracja 39. Fiszka nr 6 („Istość”) ze strony 112.	219
Ilustracja 40. Fiszka nr 8 („Istotimie”) ze strony 112.	219
Ilustracja 41. Fiszka nr 12 („Iszczeć”) ze strony 112.	219
Ilustracja 42. Fiszka z hasłem „Ichnografja”.....	220
Ilustracja 43. Fiszka z hasłem „Ichnografja” – usunięte kontury.....	220
Ilustracja 44. Fiszka z hasłem „Ichnografja” – kontury usunięte Metodą II.....	221
Ilustracja 45. Fiszka z hasłem „Ichnografja” – kontury usunięte Metodą III.....	221
Ilustracja 46. Hasło „Ibungsmarś” – przed korektą.....	221
Ilustracja 47. Hasło „Ibungsmarś” – po korekcie.....	221
Ilustracja 48. Hasło „Ichni” – przed korektą.....	222
Ilustracja 49. Hasło „Ichni” – po korekcie.....	222
Ilustracja 50. Hasło „Ichnograficzny” – przed korektą.....	222
Ilustracja 51. Hasło „Ichnograficzny” – po korekcie.....	222
Ilustracja 52. Hasło „Ichtjograficzny” – przed korektą.....	222
Ilustracja 53. Hasło „Ichtjograficzny” – po korekcie.....	222
Ilustracja 54. Hasło „Icek” – przed korektą.	222
Ilustracja 55. Hasło „Icek” – po korekcie.....	223
Ilustracja 56. Fragment obrazu strony.	230
Ilustracja 57. Fragment strony po OCR.	231
Ilustracja 58. Hasło „Ferycyjan” w foncie Courier New (Notepad++).	267
Ilustracja 59. Wybrane znaki w foncie Courier New.....	267
Ilustracja 60. Wybrane znaki złożone wyświetlane w konsoli.....	268
Ilustracja 61. Wygląd programu w konsoli.	269
Ilustracja 62. Etapy przetwarzania danych.	270
Ilustracja 63. Możliwe relacje między leksykografią a historią.	278
Ilustracja 64. Modyfikacja hasła „Fulminacja”.....	327

ANEKSY