



**Phonotactics and morphonotactics  
of Polish and English**

**Theory, description, tools and applications**



ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ  
SERIA FILOLOGIA ANGIELSKA NR 49

PAULINA ZYDOROWICZ, PAULA ORZECHOWSKA, MICHAŁ JANKOWSKI,  
KATARZYNA DZIUBALSKA-KOŁACZYK, PIOTR WIERZCHOŃ,  
DAWID PIETRALA

# **Phonotactics and morphonotactics of Polish and English**

**Theory, description, tools and applications**



POZNAŃ 2016

ABSTRACT: Paulina Zydorowicz, Paula Orzechowska, Michał Jankowski, Katarzyna Dziubalska-Kołączyk, Piotr Wierchoń, Dawid Pietrala, *Phonotactics and morphonotactics of Polish and English. Theory, description, tools and applications* [Fonotaktyka i morfonotaktyka języka polskiego i angielskiego. Teoria, opis, narzędzia i zastosowania]. Adam Mickiewicz University Press. Poznań 2016. Pp. 342, 40 Figs., 79 Tabs. ISBN 978-83-232-3009-0. ISSN 0554-8144. Text in English with a summary in Polish.

*Phonotactics* determines phonological conditions and constraints on the occurrence or co-occurrence of sounds (vowels and consonants) in a given language. *Morphonotactics* refers to the interaction between phonotactics and morphotactics. The main approach to consonant clusters is based on the division into phonotactic and morphonotactic clusters. Phonotactic clusters occur within morphemes. Morphonotactic clusters arise as a result of morphological operations such as concatenation or apophony. Morphonotactic clusters are often phonotactically marked and often fail to surface as phonotactic clusters. The research focuses both on morphologically simple and complex clusters. The leitmotif of the present book is the verification of the hypothesis concerning the markedness of clusters in relation to their phonotactic or morphonotactic character. The dictionary and corpus data was examined in terms of markedness, based on universal phonotactic preferences. In order to evaluate cluster preferability, a phonotactic calculator was devised, which enables making measurements on a large scale.

KEY WORDS: phonotactics, morphonotactics, corpus linguistics, markedness.

Paulina Zydorowicz, Paula Orzechowska, Michał Jankowski, Katarzyna Dziubalska-Kołączyk, Piotr Wierchoń, Dawid Pietrala, Faculty of English, Adam Mickiewicz University, al. Niepodległości 4, 61-678 Poznań, Poland; email: dkasia@wa.amu.edu.pl

Reviewer/Recenzent: Prof. Mag. Dr. Nikolaus Ritt

Publikacja finansowana przez  
Wydział Anglistyki UAM

© Paulina Zydorowicz, Paula Orzechowska, Michał Jankowski,  
Katarzyna Dziubalska-Kołączyk, Piotr Wierchoń, Dawid Pietrala 2016

This edition © Uniwersytet im. Adama Mickiewicza w Poznaniu,  
Wydawnictwo Naukowe UAM, Poznań 2016

Cover design/Projekt okładki: Pracownia Wydawnicza WA UAM

Cover image generated from the authors' data courtesy of <http://www.wordle.net/>  
Typsetting and formatting/Skład i formatowanie: Pracownia Wydawnicza WA UAM

ISBN 978-83-232-3109-7

ISSN 0554-8144

# Contents

|   |    |
|---|----|
| List of tables .....  | 7  |
| List of figures .....   | 11 |
| The alphabetical list of abbreviations .....  | 13 |
| <br>  |    |
| Preamble .....  | 15 |
| <br>  |    |
| <b>Chapter 1</b>  |    |
| <b>Theoretical preliminaries</b> .....  | 17 |
| 1.1. Phonotactics and morphonotactics .....   | 17 |
| 1.2. Beats-and-Binding Phonology and NAD Principle .....                                      | 22 |
| 1.2.1. Beats-and-Binding Phonology: An overview .....   | 22 |
| 1.2.2. Beats-and-Binding (B&B) phonotactics and NAD .....                                     | 25 |
| 1.2.3. Beats-and-Binding (B&B) phonotactics in<br>Natural Linguistics explanatory model ..... | 30 |
| 1.3. Hypotheses: Preview .....  | 31 |
| <br>  |    |
| <b>Chapter 2</b>  |    |
| <b>The Phonotactic Calculator</b> .....   | 33 |
| 2.1. Introduction .....   | 33 |
| 2.2. Technology .....   | 34 |
| 2.3. Usage and future development .....   | 45 |
| <br>  |    |
| <b>Chapter 3</b>  |    |
| <b>The resources</b> .....  | 47 |
| 3.1. The data for Polish .....  | 47 |
| 3.1.1. Corpus data .....  | 48 |
| 3.1.2. Dictionary data .....  | 50 |
| 3.1.3. Paradigmatic data .....  | 52 |
| 3.1.4. Phonemic transcription of Polish .....   | 57 |
| 3.2. The data for English .....   | 66 |
| 3.2.1. The list of inflectional forms (paradigm) .....  | 66 |
| 3.2.2. The dictionary list .....  | 69 |
| 3.2.3. The corpus data .....  | 69 |

**Chapter 4**

|  |    |
|--|----|
| <b>Descriptive phonotactics of Polish</b> .....  | 71 |
| 4.1. Resources .....                             | 71 |
| 4.2. Quantitative phonotactics .....             | 73 |
| 4.3. Cluster overlap in the three sources .....  | 75 |
| 4.4. Segment distribution in clusters .....      | 82 |
| 4.5. Segmental composition of Polish words ..... | 88 |
| 4.6. Summary .....                               | 92 |

**Chapter 5**

|   |     |
|---|-----|
| <b>Polish and English (mor)phonotactics: An empirical study</b> ..... | 93  |
| 5.1. Hypotheses .....   | 93  |
| 5.2. Polish .....   | 94  |
| 5.2.1. Principles of morphological parsing .....                      | 94  |
| 5.2.2. Results: Hypothesis 1 (dictionary) .....                       | 101 |
| 5.2.3. Results: Hypothesis 2 .....                                    | 105 |
| 5.2.4. Results: Hypothesis 3 .....                                    | 119 |
| 5.3. English .....  | 121 |
| 5.3.1. Principles of morphological parsing .....                      | 121 |
| 5.3.2. Results: Hypothesis 1 .....                                    | 122 |
| 5.3.3. Results: Hypothesis 2 .....                                    | 126 |
| 5.3.4. Results: Hypothesis 3 .....                                    | 128 |

|                          |     |
|--------------------------|-----|
| <b>Conclusions</b> ..... | 131 |
|--------------------------|-----|

|   |     |
|---|-----|
| Appendix A. Cluster distribution in Polish .....                      | 133 |
| Appendix B. Polish cluster statistics .....                           | 137 |
| Appendix C. Polish cluster examples.....                              | 205 |
| Appendix D. Distribution of English consonant clusters: raw data .... | 273 |
| Appendix E. Cluster distribution in English .....                     | 275 |
| Appendix F. English cluster statistics .....                          | 279 |
| Appendix G. English cluster examples .....                            | 305 |

|                         |     |
|-------------------------|-----|
| <b>References</b> ..... | 331 |
|-------------------------|-----|

|  |     |
|--|-----|
| <b>Fonotaktyka i morfonotaktyka języka polskiego i angielskiego.<br/>Teoria, opis, narzędzia i zastosowania (Streszczenie)</b> ..... | 339 |
|--|-----|

# List of tables

|           |   |    |
|-----------|---|----|
| Table 1.  | Distances in MOA and POA .....  | 26 |
| Table 2.  | Distances in MOA and POA: Polish .....  | 27 |
| Table 3.  | Distances in MOA and POA: English .....   | 27 |
| Table 4.  | The explanatory model of Natural Linguistics<br>(Dziubalska-Kołodziej 2009, 2014) .....                 | 30 |
| Table 5.  | Default POA and MOA values .....  | 39 |
| Table 6.  | POA, MOA and voicing values for input sounds .....  | 42 |
| Table 7.  | Preferability calculations for double-initials including<br>all parameters (MOA, POA and voicing) ..... | 42 |
| Table 8.  | Preferability calculations for double-initials<br>excluding MOA .....                                   | 43 |
| Table 9.  | Preferability calculations for double-initials<br>excluding POA .....                                   | 43 |
| Table 10. | Preferability calculations for double-initials<br>excluding voicing .....                               | 43 |
| Table 11. | Preferability calculations for double-medials including<br>all parameters (MOA, POA and voicing) .....  | 43 |
| Table 12. | Preferability calculations for double-medials<br>excluding MOA .....                                    | 44 |
| Table 13. | Preferability calculations for double-medials<br>excluding POA .....                                    | 44 |
| Table 14. | Preferability calculations for double-medials<br>excluding voicing .....                                | 44 |
| Table 15. | Preferability calculations for double-finals including<br>all parameters (MOA, POA and voicing) .....   | 44 |
| Table 16. | Preferability calculations for double-finals<br>excluding MOA .....                                     | 44 |
| Table 17. | Preferability calculations for double-finals<br>excluding POA .....                                     | 44 |
| Table 18. | Preferability calculations for double-finals<br>excluding voicing .....                                 | 45 |
| Table 19. | A sample data structure for clusters .....  | 50 |
| Table 20. | The inventory of consonants adopted in the transcription .....  | 61 |
| Table 21. | The treatment of the graphemes <ą> and <ę> .....  | 65 |
| Table 22. | Terminology .....   | 72 |
| Table 23. | Cluster types in the dictionary .....   | 73 |
| Table 24. | Word types in the dictionary .....  | 73 |

|           |  |     |
|-----------|--|-----|
| Table 25. | Cluster types in the paradigm .....  | 74  |
| Table 26. | Word types in the paradigm .....   | 74  |
| Table 27. | Cluster types in the corpus .....  | 74  |
| Table 28. | Word types in the corpus .....   | 74  |
| Table 29. | Word tokens in the corpus .....  | 75  |
| Table 30. | The participation of consonants<br>in cluster formation in Polish .....                          | 83  |
| Table 31. | Positional order of consonants in Polish clusters .....  | 85  |
| Table 32. | Positional order of consonants in Polish clusters:<br>Examples of initial clusters .....         | 86  |
| Table 33. | Positional order of consonants in Polish clusters:<br>Examples of medial clusters .....          | 87  |
| Table 34. | Positional order of consonants in Polish clusters:<br>Examples of final clusters .....           | 87  |
| Table 35. | Skeletal structure of Polish words .....   | 89  |
| Table 36. | Patterns of Polish words .....   | 89  |
| Table 37. | Examples of words with the clear morphological<br>parsing rule .....                             | 95  |
| Table 38. | Examples of words parsed according to the active<br>stem rule .....                              | 97  |
| Table 39. | Examples of words parsed with<br>the part-of-speech-transfer rule .....                          | 98  |
| Table 40. | Examples of words parsed according to the<br>prepositions rule .....                             | 99  |
| Table 41. | The list Polish affixes and morphological operations<br>generating morphonotactic clusters ..... | 100 |
| Table 42. | Hypothesis 1: Size vs. morphological composition<br>(initial cluster types) .....                | 101 |
| Table 43. | The total number of word types with phonotactic<br>and morphonotactic initial clusters .....     | 102 |
| Table 44. | The number of word types with phonotactic<br>and morphonotactic initial CC clusters .....        | 102 |
| Table 45. | The number of word types with phonotactic<br>and morphonotactic initial CCC clusters .....       | 102 |
| Table 46. | The number of word types with phonotactic<br>and morphonotactic initial CCCC clusters .....      | 102 |
| Table 47. | Hypothesis 1: Size vs. morphological composition<br>(final cluster types) .....                  | 104 |
| Table 48. | The total number of word types with phonotactic<br>and morphonotactic final clusters .....       | 104 |

|           |  |     |
|-----------|--|-----|
| Table 49. | The number of word types with phonotactic and morphonotactic final CC clusters .....   | 105 |
| Table 50. | The number of word types with phonotactic and morphonotactic final CCC clusters .....  | 105 |
| Table 51. | Hypothesis 2: Morphological composition vs. preferability (word types with initial CC clusters) .....                        | 106 |
| Table 52. | 20 top-frequency CC lexical clusters (P=preferred, D=dispreferred) .....   | 107 |
| Table 53. | Low-frequency CC lexical clusters represented by 1, 2 or 3 dictionary words (P=preferred, D=dispreferred) .....              | 107 |
| Table 54. | Hypothesis 2: Morphological composition vs. preferability (word types with initial CCC clusters) .....                       | 109 |
| Table 55. | Hypothesis 2: Morphological composition vs. preferability (initial CC cluster types) .....                                   | 109 |
| Table 56. | Hypothesis 2: Morphological composition vs. preferability (initial CCC cluster types) .....                                  | 110 |
| Table 57. | Hypothesis 2: Morphological composition vs. preferability (word types with final CC clusters) .....                          | 111 |
| Table 58. | Hypothesis 2: Morphological composition vs. preferability (word types with final CCC clusters) .....                         | 112 |
| Table 59. | Hypothesis 2: Morphological complexity vs. preferability (final CC cluster types) .....                                      | 113 |
| Table 60. | Hypothesis 2: Morphological composition vs. preferability (final CCC cluster types) .....                                    | 114 |
| Table 61. | An alphabetical list of initial double clusters with morphological divisions and the NAD status .....                        | 114 |
| Table 62. | An alphabetical list of initial triple clusters with morphological divisions and the NAD status .....                        | 116 |
| Table 63. | An alphabetical list of initial quadruple clusters with morphological divisions .....  | 117 |
| Table 64. | The alphabetical list of word-final double and triple clusters ordered with morphological divisions and the NAD status ..... | 118 |
| Table 65. | A list of English suffixes generating morphonotactic clusters .....  | 121 |
| Table 66. | Hypothesis 1: Size vs. morphological composition (cluster types in the word list) .....                                      | 122 |
| Table 67. | The total number of word types with phonotactic and morphonotactic final clusters (in the word list) .....                   | 123 |
| Table 68. | The number of word types with phonotactic and morphonotactic final CC clusters (in the word list) .....                      | 123 |

10 List of tables

|           |   |     |
|-----------|---|-----|
| Table 69. | The number of word types with phonotactic and morphonotactic final CCC clusters (in the word list) .....  | 123 |
| Table 70. | The number of word types with phonotactic and morphonotactic final CCCC clusters (in the word list) ..... | 124 |
| Table 71. | The total number of word tokens with phonotactic and morphonotactic final clusters (in the corpus) .....  | 124 |
| Table 72. | The number of word tokens with phonotactic and morphonotactic final CC clusters (in the corpus) .....     | 124 |
| Table 73. | The number of word tokens with phonotactic and morphonotactic final CCC clusters (in the corpus) .....    | 125 |
| Table 74. | The number of word tokens with phonotactic and morphonotactic final CCCC clusters (in the corpus) .....   | 125 |
| Table 75. | Hypothesis 2: Morphological composition vs. preferability (cluster types in the word list) .....          | 126 |
| Table 76. | Hypothesis 2: Morphological composition vs. preferability (word types in the word list) .....             | 126 |
| Table 77. | Hypothesis 2: Morphological composition vs. preferability (in the corpus) .....                           | 127 |
| Table 78. | Cluster raking based on type (word-list) frequency .....  | 129 |
| Table 79. | Cluster raking based on token (corpus) frequency (*copied from Table 78) .....                            | 129 |

# List of figures

|            |  |     |
|------------|--|-----|
| Figure 1.  | The universals-to-performance quintuple<br>(Dressler 1985: 292) .....                        | 21  |
| Figure 2.  | NAD-based predictions for initial doubles .....  | 29  |
| Figure 3.  | NAD-based predictions for final doubles .....  | 29  |
| Figure 4.  | NAD-based predictions for medial doubles .....   | 30  |
| Figure 5.  | The Phonotactic Calculator – main window .....   | 35  |
| Figure 6.  | The Phonotactic Calculator – Other settings window .....                                     | 38  |
| Figure 7.  | The Phonotactic Calculator – bar chart window .....  | 40  |
| Figure 8.  | Number of “new” word types over a 24-month period .....                                      | 48  |
| Figure 9.  | Number of word tokens over a 24-month period .....   | 49  |
| Figure 10. | Number of word tokens and types plotted together<br>over a 24-month period .....             | 49  |
| Figure 11. | Close-up plot of the number of types over<br>a 24-month period .....                         | 49  |
| Figure 12. | Initial CC clusters found in the dictionary, paradigm<br>and corpus .....                    | 76  |
| Figure 13. | Initial CCC clusters found in the dictionary, paradigm<br>and corpus .....                   | 77  |
| Figure 14. | Initial CCCC clusters found in the dictionary, paradigm<br>and corpus .....                  | 78  |
| Figure 15. | Final CC clusters found in the dictionary, paradigm<br>and corpus .....                      | 79  |
| Figure 16. | Final CCC clusters found in the dictionary, paradigm<br>and corpus .....                     | 80  |
| Figure 17. | Final CCCC clusters found in the dictionary, paradigm<br>and corpus .....                    | 81  |
| Figure 18. | Frequency (type) distribution of Polish word patterns .....                                  | 91  |
| Figure 19. | Frequency of CV patterns in Polish .....   | 91  |
| Figure 20. | Hypothesis 1: Size vs. morphological composition<br>(initial cluster types) .....            | 101 |
| Figure 21. | Hypothesis 1: Size vs. morphological composition<br>(word types with initial clusters) ..... | 103 |
| Figure 22. | Hypothesis 1: Size vs. morphological composition<br>(final cluster types) .....              | 104 |
| Figure 23. | Hypothesis 1: Size vs. morphological composition<br>(word types with final clusters) .....   | 105 |

|            |  |     |
|------------|--|-----|
| Figure 24. | Hypothesis 2: Morphological composition vs. preferability (word types with initial CC clusters) .....  | 106 |
| Figure 25. | Hypothesis 2: Morphological composition vs. preferability (word types with initial CCC clusters) ..... | 109 |
| Figure 26. | Hypothesis 2: Morphological composition vs. preferability (initial CC cluster types) .....             | 110 |
| Figure 27. | Hypothesis 2: Morphological composition vs. preferability (initial CCC cluster types) .....            | 110 |
| Figure 28. | Hypothesis 2: Morphological composition vs. preferability (word types with final CC clusters) .....    | 111 |
| Figure 29. | Hypothesis 2: Morphological composition vs. preferability (word types with final CCC clusters) .....   | 112 |
| Figure 30. | Hypothesis 2: Morphological complexity vs. preferability (final CC cluster types) .....                | 113 |
| Figure 31. | Hypothesis 2: Morphological composition vs. preferability (final CCC cluster types) .....              | 114 |
| Figure 32. | Hypothesis 3: Preferability vs. frequency (initial CC in the paradigm and corpus) .....                | 119 |
| Figure 33. | Hypothesis 3: Preferability vs. frequency (final CC in the paradigm and corpus) .....                  | 120 |
| Figure 34. | Hypothesis 1: Size vs. morphological composition (cluster types in the word list) .....                | 123 |
| Figure 35. | Hypothesis 1: Size vs. morphological composition (word types in the word list) .....                   | 124 |
| Figure 36. | Hypothesis 1: Size vs. morphological composition (in the corpus) .....                                 | 125 |
| Figure 37. | Hypothesis 2: Morphological composition vs. preferability (cluster types in the word list) .....       | 126 |
| Figure 38. | Hypothesis 2: Morphological composition vs. preferability (word types in the word list) .....          | 127 |
| Figure 39. | Hypothesis 2: Morphological composition vs. preferability (in the corpus) .....                        | 127 |
| Figure 40. | Hypothesis 3: Preferability vs. frequency .....  | 129 |

# The alphabetical list of abbreviations

|         |   |
|---------|---|
| ADJ     | adjective   |
| AS      | active stem   |
| B&B     | Beats-and-Binding   |
| COCA    | Corpus of Contemporary American English                             |
| CP      | clear parsing   |
| CUVOALD | Computer Usable Version of the Oxford Advanced Learner's Dictionary |
| DIM     | diminutive  |
| GEN     | genitive  |
| IMP     | imperative  |
| IMPERF  | imperfective  |
| INSTR   | instrumental  |
| MOA     | manner of articulation  |
| N       | noun  |
| NAD     | Net Auditory Distance   |
| NOM     | nominative  |
| PERF    | perfective  |
| PDI     | phonetic difficulty index   |
| PL      | plural  |
| POA     | place of articulation   |
| PoS     | part of speech boundary transfer                                    |
| SG      | singular  |
| *       | non-existent form   |



## Preamble

The title of the present volume says it all. We provide a thorough description of the consonantal phonotactics and morphonotactics of Polish in a comparative perspective with English. The description is guided by the theoretical assumptions of the models we adopted, i.e. the Beats-and-Binding phonology and natural morphonotactics. The tools we have used consist in the statistical analyses of language corpora, which supplied the other guideline to our research, alongside the theory. Thus, methodologically, we have combined the deductive with the inductive approach, or, in modern sociolinguistic terms, a corpus-based with a corpus-driven one. Eventually, we also arrive at some actual and potential applications of both the results of the research as well as of the data corpus itself.

The aim of the research was twofold. The first aim was to verify the applicability of the Beats-and-Binding model to the study of phonotactics and of the natural morphonotactics model to the study of morphonotactics. The second aim was to provide extensive descriptive observations concerning the consonantal phonotactics and morphonotactics of Polish and English, with emphasis on the former.

Chapter 1 provides theoretical background for the research. It presents phonotactics and morphonotactics within the framework of Beats-and-Binding Phonology and the natural linguistic explanatory model. The principal tool of analysis – the Net Auditory Distance (NAD) – is introduced. The hypotheses guiding the research are previewed. Chapter 2 describes the phonotactic calculator which was created in order to automatize the calculation of NAD values for consonantal clusters. The phonotactic calculator is available online for the study of phonotactics of further languages. In Chapter 3, the resources for Polish and English used in the analyses are described. Chapter 4 constitutes the most substantial contribution of the present volume: an extensive and novel view on the descriptive phonotactics of Polish. Chapter 5 deals with a comparative analysis of the selected aspects of Polish and English (mor)phonotactics, with a view to indicating some typological and language-specific differences and similarities between the two languages. The volume is equipped with ample appendices containing a realm of data on which the analyses were based. The data constitutes a unique resource for further investigations.

The research presented in the volume has been conducted within the National Science Foundation (NCN) Project no. N N104 382540. The book constitutes an extensive final report of the project research and the obtained results. As much as the project has been a joint effort of the project participants, this volume is also a collaborative endeavour of writing up the outcomes of the common work. The project participants are jointly the authors of the present volume, although it must be acknowledged that it was Michał Jankowski, Paulina Zydorowicz and Paula Orzechowska, who played a crucial role in the editing process.

Katarzyna Dziubalska-Kołaczyk, PI

# Chapter 1

## Theoretical preliminaries

### 1.1. Phonotactics and morphonotactics

The term *phonotactics*<sup>1</sup> is uncontroversial. It describes the co-occurrence restrictions among segments in a string of sounds. The term comes from Ancient Greek *phōnḗ* ‘voice, sound’ and *taktikós* ‘having to do with arranging’ and was coined by Robert Stockwell in 1954 (cf. Hill 1958: 68-88). The interest in the combinatorial possibilities of speech segments, however, is as old as the interest in the syllable, i.e. it dates back to the ancient times.

Phonotactics is an essential topic in phonology. Every phonological theory needs to be able to provide an account of permissible combinations of phonemes within a given language. It needs to be able to explain why particular phonemes “stick together” while others do not. For example, initial /tr-/ is possible in English, but /rt-/ is not. In Polish both sequences are allowed. A phonological theory looks for the criteria or rules which govern the possible phoneme sequences in a language, i.e. it looks for language-specific phonotactic constraints. More interestingly, however, it also looks for explanations of such constraints on a more general, universal level. One very well known explanation is the Sonority Sequencing Principle (Hooper 1976, Kiparsky 1979, Steriade 1982, Selkirk 1984, Clements 1990, Zec 1995, among others), which organizes sequences within syllables according to the universally preferred pattern of rising sonority towards the peak and falling sonority away from it.

Explanatory models of phonotactics very often rely on the syllable as the domain of phonotactic constraints. The co-occurrence restrictions refer to the shape of onsets and codas as well as syllable rhymes. They define the combinatorial possibilities among consonants as well as between consonants and vowels (filling in syllable nuclei). Such models need to

---

<sup>1</sup> “The sequential arrangements of phonological units that are possible in a language. In English, for example, initial /spr-/ is a possible phonotactic sequence, whereas /spm-/ is not.” Crystal (1992: 301)

observe a particular definition of the syllable which dictates the required phonological shape of the syllable (cf. the Sonority Sequencing Principle mentioned above) and the criteria of syllabification (e.g. the Contact Law by Vennemann 1988). Consonants which violate the required (unmarked) shape of the syllable receive a special treatment, since they are relegated outside of the syllable. Thus, they become extra-syllabic. Still, they remain part of words in a language, and need to be accounted for. Therefore, the word is a better candidate for a phonotactic domain. Descriptive models of phonotactics most often describe word-internal combinatorics. However, they tend to ignore word-internal morphological complexity. It is clear that monomorphemic phonotactics differs from the phonotactics of complex words derived by morphological operations. The present approach is to consider both the purely phonotactic constraints governing the structure of *intramorphemic* sequences<sup>2</sup> and the morphotactic regularities which give rise to partially or entirely morphologically motivated sequences. In this way credit is given to phonology and morphology in their roles of shaping sound sequences in a language without pretending that phonology itself does it.

In this book we provide an account of the language-specific consonantal phonotactics of Polish and English from the perspective of a phonological preference theory (Beats-and-Binding Phonology, see below for the presentation). The theory governs the choice of the domain for the phonotactic generalizations – the word, and of the measures of markedness – manner and place of articulation and some other phonetic cues. The theory offers a clear interface to morphology, allowing us to account for morphonotactics alongside phonotactics. This differentiates our approach from those models of phonotactics which rely on the syllable and ignore morphological structure.

*Morphonotactics* is part of morphonology in the sense of Dressler (1985, 1996) and has been defined as “the area of interaction between morphotactics and phonotactics” (Dressler and Dziubalska-Kołodziej 2006: 250). The interactions may be cooperative or conflictual. Morphotactics and phonotactics converge when morphological operations produce phonotactic sequences which already exist intramorphemically, e.g.

---

<sup>2</sup> By ‘*intramorphemic*’ sequences we mean the non-derived lexical sequences. In this sense, the cluster /ln/ in Pol. *lnu* (derived from *len* ‘flax’) is not *intramorphemic*. In the following chapters, the terms ‘phonotactic’, ‘lexical’ and ‘intramorphemic’ are used interchangeably to signify the same entity – the non-derived phonological sequences.

the final /-nd/ in *screen*+ed (morphonotactic) and *find* (intramorphemic). Conflictual interactions occur when morphological operations (of both concatenative and non-concatenative type) produce clusters which either overlap with the already existing intramorphemic clusters or are totally 'new' in the sense that they only appear in morphologically complex words. Thus created morphonotactic clusters may be systematically distinguished along the following scale (cf. Dressler and Dziubalska-Kołodziejczyk 2006: 253ff):

- the exclusively morphologically motivated clusters (e.g. the word-final clusters /-fs, -vz/ as in *laughs*, *wife's*, *loves*, *wives*);
- morphologically motivated as a strong default (e.g. the word-final clusters /-ts, -dz/, as in *cats*, *kids*, which occur in just a few monomorphemic words, i.e. *waltz*, *grits*, *adz(e)*, and the loan-words *quartz*, *kibbutz*);
- morphologically motivated as a weak default (e.g. the word-final cluster /-ks/, as in *docks*, *lacks*, which occurs in monomorphemic Latinate words such as *tax*, *sex*, *box*, *flux*, *fix*, *six*);
- morphologically motivated in the majority (e.g. in Italian, the word-initial clusters /zl-, zm-, zn-, skw-/ contain a morpheme boundary in the majority of cases, as in *s+leale* 'disloyal' vs. *slang*, *s+membrare* 'dismember' vs. *smog*, *s+naturare* 'denaturate' vs. *snack*, *s+qualificare* 'disqualify' vs. *squadra* 'team');
- morphologically motivated in the minority (e.g. in Polish, the word-initial clusters /sp-, /sk-, /st- and /vj-/ contain a morpheme boundary in the minority of words they head, as in *s+palić* 'burn' vs. *spać* 'sleep', *s+kończyć* 'finish' vs. *skoczyć* 'jump', *s+toczyć* 'tumble' vs. *stać* 'stand', *w+jechać* 'drive in' vs. *wiem* 'I know').

From the perspective of universal phonotactic preferences (elaborated on below), morphonotactic clusters are likely to be marked due to their complexity. They are either complex in size, i.e. the number of consonants in a sequence exceeds the number allowed within a morpheme, or complex in quality, i.e. they violate universal phonotactic preferences. Of course, language-specific phonotactic (intramorphemic) clusters may also be marked. However, they are legal in the language system (this is where language-specific phonotactics diverges from universal phonotactics).

A theory of universal phonotactics predicts that universal phonotactic preferences govern the general layout of the phonotactic systems of specific languages. The deviations are, among others, due to morphological operations which strive to guarantee the salience of morphological markers. Therefore, one can expect morphological operations to produce phonologically marked clusters.

Other deviations from optimal phonotactics are historical-evolutionary and are largely due to a wide array of performance and statistical factors. For instance, phonostylistic processes of casual speech may give rise to ‘improved’, less marked sequences by modification of the marked ones, e.g. *text(s)* /-kst(s)/ → /-ks/, *wszystko* ‘everything’ /fʂ-/ → /ʂ-/. However, they may also give rise to new, marked clusters, e.g. *przynajmniej* ‘at least’ /pʂɨn-/ → /pʂn-/ (cf. Dressler et al. 2001 for the discussion).

In this book, the data is restricted to written sources, so we do not consider the influence of the processes of connected speech on the phonotactic systems of the two languages. Similarly, other performance data, such as child language or aphasic speech, are not examined here. The analysis is synchronic, so we do not study the history of the clusters under scrutiny<sup>3</sup>. We do, however, use both dictionary and corpus data of sufficient size to be able to verify our results and look for explanations which escape the phonotactic and morphonotactic generalizations. For example, initial cluster /kt-/ is legal in Polish. It is both phonotactically marked and monomorphemic, so morphonotactics is of no help in explaining its occurrence. Its corpus frequency is high, but in a dictionary it figures in a few items only. The high corpus frequency is due to the 2-3 words *kto* ‘who’, *który* ‘which’, *ktoś* ‘somebody’, whose function explains their frequency.

Our approach is holistic. Clusters of sounds are seen as phonological representations in the speaker’s/listener’s mind which can be influenced by morphological motivations as well as performance-related factors. Phonology and morphology are governed by principles of adequacy on the universal, typological and system-specific level. Phonological and morphological structures also undergo modifications due to a variety of performance factors related to sociolinguistic and psycholinguistic vari-

---

<sup>3</sup> But see Dziubalska-Kołaczyk 2005, Ritt (2010), Ritt and Kaźmierski (in press), as well as the Austrian-Polish project in progress on *Final consonant clusters in Middle English: phonotactics, morphology and the effects of schwa loss*. Grant number: P 27592.

ability. In the latter, statistical considerations are highly significant. The holistic approach remains in accord with the model of explanation in natural linguistics, a bedrock framework for Beats-and-Binding Phonology and its phonotactics (Dziubalska-Kołodziejczyk 2002, 2009, 2014) and Natural Morpho(no)logy and its morphonotactics (Dressler 1985, 1996, Dressler and Dziubalska-Kołodziejczyk 2006).

The explanatory cycle of natural linguistics is very adequately visualized by Dressler's (1985) quintuple.

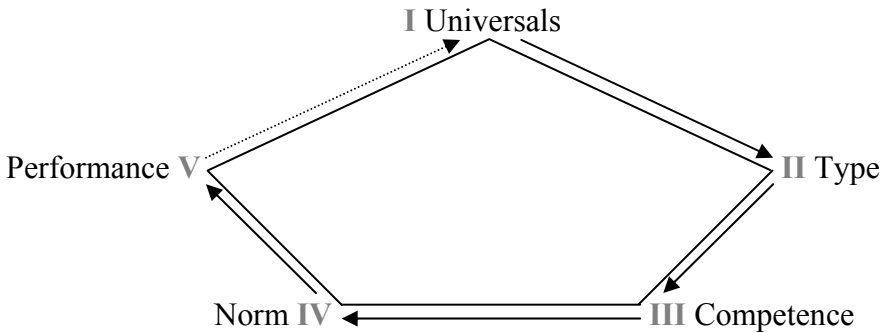


Figure 1. The universals-to-performance quintuple (Dressler 1985: 292).

The quintuple (originally established by Hjelmslev and Coseriu, cf. Dressler 1985: 292) has been adapted by Dressler to replace Chomskyan triple (I, III and V) and Saussurean quadruple (I, III, IV and V), since it shows the path from universal properties of language to individual performance in steps compatible with the naturalist framework. Each of the five elements is simultaneously the basis of and is filtered by the next one. One needs to consider the whole quintuple when attempting to account for performance. Let me interpret the diagram: one can distinguish three subtheories of naturalness, i.e. (I) the theory of universals, (II) the theory of type adequacy and (III) the theory of language-specific system adequacy. (IV) normative, i.e. sociolinguistic factors and (V) psycholinguistic factors further contribute to the final shape of performance. Performance, in turn, has the potential to modify universals.

The analysis in this book refers primarily to (I) and (III), i.e. universal phonotactics and language-specific phonotactics and morphonotactics. It makes reference to (II) for the sake of morphonotactics, since Polish and English represent different morphological types. Lexical and corpus frequencies reflect language-specific use of lemmas in the written texts. Spo-

ken language frequencies and variation, influenced by sociolinguistic, pragmatic and individual factors, are beyond the scope of the present study.

## 1.2. Beats-and-Binding Phonology and NAD Principle

In this section, the relevant characteristics of the Beats-and-Binding Phonology (B&B henceforth) will be presented and followed by an outline of the B&B model of phonotactics and its Net Auditory Distance (NAD) Principle.

### 1.2.1. Beats-and-Binding Phonology: An overview

Beats-and-Binding (B&B) Phonology (Dziubalska-Kołączyk 2002, 2009) is a syllable-less theory of phonology, embedded in Natural Phonology. Its basic units are beats and non-beats, basic relations are bindings, and preferences for structure consist in specifying the patterning, strength and realization of beats and non-beats in a sequence.

In B&B Phonology it is proposed that the notions of a *beat* and *non-beat* make it possible for the functions of the syllable to be taken over without maintaining it as a unit. A basic rhythmic speech skeleton consists of regularly recurring beats. Beats are primary rhythmic entities realized preferably by vocalic figures against the consonantal ground (non-beats). They do not possess any inherent articulatory characteristics since they are functions rather than units, i.e. they are intentional (in the sense of Baudouin de Courtenay 1895) and perceptual rather than articulatorily actual in nature. Their realization is preferably vocalic due to the saliency potential inherent to vowels, although consonants may take over a beat function in a number of circumstances. Interrelationships between beats and non-beats consist in bindings which are binary and whose strength relates to sonority.

A beat is a regularly recurring skeletal prosodic unit of phonological representation, of a size corresponding to that of a segment. The most basic organizational principle of a sequence is the alternation of beats (relatively more prominent) and non-beats (relatively less prominent). This can be described as the preference for the alternation between beats and non-beats. The other two preferences for structure are the preference for a trochee and the preference for the vocalic beat.

Beats and non-beats have direct phonetic correlates both in production and in perception. Most preferably, a beat is realized by a vowel (short, long, or a diphthong, the difference resulting in a difference in the weight count), but occasionally also by a consonant. Two types of beats can be distinguished. The first one is the timing beat, which is a unit of the rhythmic or timing tier; recurring timing beats are what speakers are able to count. The other type is the quantity beat, which is a unit of weight activated in quantity-sensitive languages; it is comparable in its function to the mora.

A non-beat is always a consonant; it cannot stand alone without support of a beat, either directly by a respective binding, or indirectly, when it appears in a cluster with other consonants, regulated by NAD (Net Auditory Distance, see below).

The beats/non-beats alternation also has a direct phonetic grounding: “To construct a useful signalling system out of sound, there must be some differentiation between different parts of the signal in time. It appears that a basic organization of this differentiation of sound in all (spoken) languages consists of an alternation between louder and quieter levels of sound, with a period not too far from 150-200 ms” (Maddieson 1999: 2525). This amounts to:

“[a] fairly regular wave-like alternation of amplitude peaks and valleys. The occurrence and timing of this pattern have been suggested to be related to a natural frequency of the jaw, which can be approximately equated with a comfortable mastication rate” (Maddieson 1999: 2525).

The alternating sequence is cut into pieces called words and morphemes on its way from the prelexical to lexical level. Clusters of non-beats are “accidents” due to some distortions of the alternating sequence (e.g. dropping or breaking of some beats). A long or a diphthongal beat counts as double in quantity-sensitive languages, and a  $B \leftarrow n$  binding (see below) might also contribute to weight; this resembles mora counting, but does not require the introduction of a mora.

Bindings connect beats (B) and non-beats (n) in a sequence. They are binary:  $n \rightarrow B$  and  $B \leftarrow n$ . They are motivated by the principle of perceptual contrast (more generally known as the figure and ground principle). From the phonetic perspective, larger modulations have more survival value than lesser ones and therefore will persist in languages (Ohala 1990a). Of

the two bindings, the **n**→**B** binding is stronger than the **B**←**n** one due to the CV preference. The latter receives phonetic grounding in the following observations:

- “it is generally the case that the most salient acoustic modulations in a syllable occur near the CV interface”, and “auditory cues present in CV’s are more robust than those in VC’s” (Ohala and Kawasaki 1984: 117f);
- “since there is a richer, more reliable set of place cues in the CV transition than the VC transition, listeners weight the former more heavily than the latter in deciding what they’ve heard” (Ohala 1990b: 265);
- “the speaker actively tries to create temporally more well defined, more precise, articulations near the CV as opposed to VC interface” (Ohala and Kawasaki 1984: 119).

B&B representations vary depending on the rhythmic type of a language. In English, the **B**←**n** binding is always present, which is typical of the so-called stress-timed languages. In the prototypical syllable-timed languages, this binding is absent. In B&B phonology, a word, rather than a syllable, is the domain for bindings as well as phonotactic constraints. It is supported by the fact that a speaker is able to supply the citation form of a lexical item with clear-cut boundaries, which is not possible in the case of a syllable. The creation of words can be treated as a sequence of events by which one understands the occurrences of {**nB**}, {**Bn**} and {**B**}. The most probable sequences are: {**nBnB**} /CVCV/, {**nBn**} /CVC/, and {**BnB**} /VCV/.

Sequences of non-beats, realised by consonant clusters, result from reorganisation of the basic beat/non-beat structure of sequences. The resulting clusters are subject to phonotactic preferences which are responsible for different degrees of intersegmental cohesion and thus determine the behaviour of segments within words, in word-initial, medial and final position. The different degrees of cluster cohesion create the impression of syllable structure: in this sense the latter is emergent or epiphenomenal.

Bindings relate to sonority, and operate between beats and non-beats on the prosodic level, before melody is filled into the prosodic structure. Phonotactic preferences, which correspond to constraints or syntagmatic well-formedness conditions in other models, rely on the auditory distance

between segments as a measure (cf. Net Auditory Distance below) and operate on the actual segments, when beats and non-beats receive actual phonetic realization.

### 1.2.2. Beats-and-Binding (B&B) phonotactics and NAD

The model of phonotactics operating with NAD (cf. Dziubalska-Kołodziej 2009, 2014) is part of the Beats-and-Binding Phonology. The rationale behind this model of phonotactics is to counteract the preference for CV (but cf. Hyman 2008) which overwhelmingly forces clusters to wither either via reduction or epenthesis/prothesis or at least substitution (CCV→CV, CCV→CVCV, CCV→VCCV). Clusters of consonants persist thanks to auditory contrast and its proper distribution across the word. The principle of perceptual contrast is grounded in the semiotic figure-and-ground principle, on the one hand, and in phonetics, on the other.

The Net Auditory Distance (NAD) Principle defines cluster preferability in relation to the position in the word (initial, medial and final). It reads:

A cluster is preferred if it satisfies a pattern of distances specified by the universal phonotactic preference relevant for its position in the word.

For example, for the word initial cluster C1C2V-, the NAD preference reads:

**NAD (C1,C2) ≥ NAD (C2,V)**

It is believed that auditory (perceptual) distance can be expressed by respective combinations of articulatory features which eventually bring about the auditory effect. An indefinite number of articulatory features as well as detailed acoustic cues would have to be investigated in terms of the degree of their contribution to the overall auditory effect obtained in a cluster. As a reliable starting point, manner (MOA) and place of articulation (POA) are selected, as well as the voiced/voiceless contrast (Lx)<sup>4</sup>.

---

<sup>4</sup> In the research following the project reported on in this book, the Lx parameter will be replaced with the S/O (sonorant/obstruent) distinction.

Manners refer to the most generally acknowledged version of the so-called sonority scale, while places are taken from Ladefoged (2006: 258). The tables below illustrate the way in which number values are being assigned to particular POAs and MOAs. Table 1 is general, while Table 2 and 3 are adapted to Polish and English, respectively. The values for Lx are 1 for VOICED and 0 for VOICELESS.

The choice of consonants and their characteristics in the table for Polish (Table 2) has been dictated by a number of phonological (phonemic and phonetic) considerations which will be discussed in detail in Chapter 3. For instance, nasalized glides / $\tilde{w}$ ,  $\tilde{j}$ / have been distinguished alongside the oral ones, and placed closer on the manner-of-articulation axis to vowels than the latter. Of the two liquids, /r/ has been placed closer to vowels than /l/. Affricates have been situated between stops and fricatives. On the place-of-articulation axis, labials received a two-way distinction, coronals – a three-way distinction and dorsals – a two-way one.

Table 1. Distances in MOA and POA.

| 4                     |  | 3                               |  | 2             |  | 1           |  | 0                           |  |
|-----------------------|--|---------------------------------|--|---------------|--|-------------|--|-----------------------------|--|
| OBSTRUENT             |  |                                 |  | SONORANT      |  |             |  |                             |  |
| STOP                  |  | FRICATIVE                       |  | SONORANT STOP |  | APPROXIMANT |  | VOWEL                       |  |
|                       |  | AFFRICATE                       |  |               |  |             |  | semiV                       |  |
| p b                   |  | ɸ β<br>f v                      |  | m<br>ŋ        |  | w           |  | LABIAL<br>1                 |  |
| t̪ d̪<br>t d<br>t̬ d̬ |  | θ ð<br>ʃ ʒ<br>s z<br>ʂ ʐ<br>ʧ ʑ |  | p̪<br>n       |  | r l         |  | CORONAL<br>2                |  |
| k g<br>c ɟ            |  | ç ʒ<br>x ɣ                      |  | ɲ<br>ŋ        |  | j           |  | DORSAL<br>3                 |  |
|                       |  |                                 |  |               |  |             |  | RADICAL<br>4                |  |
| ʔ                     |  | h                               |  |               |  |             |  | LARYNGEAL<br>(GLOTTAL)<br>5 |  |

Table 2. Distances in MOA and POA: Polish.

| OBSTRUENT |           |      | SONORANT |         |        |       | VOWEL |                  |         |
|-----------|-----------|------|----------|---------|--------|-------|-------|------------------|---------|
| STOP      | FRICATIVE |      | NASAL    | LIQUID  |        | GLIDE |       |                  |         |
|           | AFFRICATE |      |          | lateral | rhotic |       |       |                  |         |
| 5.0       | 4.5       | 4.0  | 3.0      | 2.5     | 2.0    | 1.0   | 0     |                  |         |
| p b       |           |      | m        |         |        | w w̃  | 1.0   | bilabial         | LABIAL  |
|           |           | f v  |          |         |        |       | 1.5   | labio-dental     |         |
| t d       | ts dẓ    | s z  | n        | l       |        |       | 2.0   | (post-)dental    | CORONAL |
|           | tʂ dẓʂ   | ʂ ẓ |          |         | r      |       | 2.3   | alveolar         |         |
|           | tɕ dẓɕ   | ɕ z  | ɲ        |         |        |       | 2.6   | alveolar-palatal |         |
|           |           |      |          |         |        | j j̣  | 3.0   | palatal          | DORSAL  |
| k g       |           | x    | ŋ        |         |        | w w̃  | 3.5   | velar            |         |
|           |           |      |          |         |        |       | 4.0   | palatal          | RADICAL |
|           |           |      |          |         |        |       | 5.0   | glottal          | GLOTTAL |

Table 3. Distances in MOA and POA: English.

| OBSTRUENT |           |     | SONORANT |         |        |       | VOWEL |               |         |
|-----------|-----------|-----|----------|---------|--------|-------|-------|---------------|---------|
| STOP      | FRICATIVE |     | NASAL    | LIQUID  |        | GLIDE |       |               |         |
|           | AFFRICATE |     |          | lateral | rhotic |       |       |               |         |
| 5.0       | 4.5       | 4.0 | 3.0      | 2.5     | 2.0    | 1.0   | 0     |               |         |
| p b       |           |     | m        |         |        | w     | 1.0   | bilabial      | LABIAL  |
|           |           | f v |          |         |        |       | 1.5   | labio-dental  |         |
|           |           | θ ð |          |         |        |       | 2.0   | inter-dental  | CORONAL |
| t d       |           | s z | n        | l       |        |       | 2.3   | alveolar      |         |
|           | tʃ dʒ     | ʃ ʒ |          |         | r      |       | 2.6   | post-alveolar |         |
|           |           |     |          |         |        | j     | 3.0   | palatal       | DORSAL  |
| k g       |           |     | ŋ        |         |        | w     | 3.5   | velar         |         |
|           |           |     |          |         |        |       | 4.0   | palatal       | RADICAL |
| ʔ         |           | h   |          |         |        |       | 5.0   | glottal       | GLOTTAL |

For example, NAD preference value for an initial Polish cluster  $prV$  will be calculated as illustrated below:

$$\text{NAD CC} = |\text{MOA1} - \text{MOA2}| + |\text{POA1} - \text{POA2}| + |\text{Lx1} - \text{Lx2}|$$

$$\text{NAD CV} = |\text{MOA1} - \text{MOA2}| + |\text{Lx1} - \text{Lx2}|$$

$$\text{NAD CC} = |5-2| + |1-2.3| + 1 = 3+1.3+1 = 5.3$$

$$\text{NAD CV} = |2-0| + 1 = 2+1 = 3$$

$$5.3 > 3$$

So, the preference  $\text{NAD}(C1, C2) \geq \text{NAD}(C2, V)$  is observed since  $5.3 > 3$ .

The NAD preferences, describing two- and three-consonant clusters in word-initial, medial and final position, are listed below.

C1C2V-

$$\text{NAD}(C1, C2) \geq \text{NAD}(C2, V)$$

-VC1C2

$$\text{NAD}(V, C1) \leq \text{NAD}(C1, C2)$$

-V1C1C2V2-

$$\text{NAD}(V1, C1) \geq \text{NAD}(C1, C2) < \text{NAD}(C2, V2)^5$$

C1C2C3V-

$$\text{NAD}(C1, C2) < \text{NAD}(C2, C3) \geq \text{NAD}(C3, V)$$

-VC1C2C3

$$\text{NAD}(V, C1) \leq \text{NAD}(C1, C2) > \text{NAD}(C2, C3)$$

-V1C1C2C3V2-

$$\text{NAD}(V, C1) \geq \text{NAD}(C1, C2) \& \text{NAD}(C2, C3) < (C3, V2)$$

NAD allows for formulating two types of predictions. On the one hand, it distinguishes between two sets of clusters, preferred and dispreferred ones, which constitutes a binary qualification into unmarked = preferred = natural vs. marked = dispreferred. On the other hand, it allows building a hierarchy of preference of clusters ranging from the most preferred to the most dispreferred (least preferred), which constitutes a scalar qualification of relative markedness from the most natural = unmarked = preferred to the heavily marked. This clearly is a strong attribute of this measure, since it reflects the human ability to both categorize and build continua. The matrices below (Figures 2, 3 and 4) provide an adequate visualisation of the nature of NAD predictions.

---

<sup>5</sup> With the reservation that  $\text{NAD}(C1, C2) > 0$ .

preferred status of NAD for initial doubles

|    | p  | b | t  | d  | k  | g | ts | dz | tʃ | dʒ | tʃ | dʒ | f | v | s | z  | ʃ  | ʒ  | ʃ  | x  | n  | ɲ   | m   | ŋ   | l   | r   | w   | j   | ɰ   | j   |  |
|----|----|---|----|----|----|---|----|----|----|----|----|----|---|---|---|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| p  |    |   | NO |    |    |   |    |    |    |    |    |    |   |   |   | NO | NO | NO | NO | NO | NO | YES |     |     | YES | YES | YES | YES |     |     |  |
| b  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    | NO | NO | NO | NO | NO | NO  | YES | YES |     | YES | YES | YES | YES |     |  |
| t  |    |   |    |    | NO |   |    |    |    |    |    |    |   |   |   |    |    | NO | NO | NO | NO | YES | YES |     | YES | YES | YES | YES |     |     |  |
| d  |    |   | NO |    |    |   |    |    |    |    |    |    |   |   |   |    | NO | NO | NO | NO | NO | NO  | YES |     | YES | YES | YES | YES |     |     |  |
| k  |    |   |    | NO |    |   |    |    |    |    |    |    |   |   |   |    |    | NO | NO | NO | NO | NO  | YES |     | YES | YES | YES | YES |     |     |  |
| g  |    |   |    |    | NO |   |    |    |    |    |    |    |   |   |   |    |    | NO | NO | NO | NO | NO  | YES |     | YES | YES | YES | YES |     |     |  |
| ts |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    | NO  | YES |     | YES | YES | YES | YES |     |     |  |
| dz |    |   | NO |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     | YES |     | YES | YES | YES | YES |     |     |  |
| tʃ |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     | YES |  |
| dʒ |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| tʃ |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| dʒ |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| f  | NO |   | NO |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     | YES | YES | YES | YES |     |  |
| v  |    |   |    | NO |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     | NO  | YES | YES | YES | YES |  |
| s  | NO |   | NO |    | NO |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     | YES | YES | YES | YES |     |  |
| z  |    |   | NO |    | NO |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     | NO  | YES | YES | YES | YES |  |
| ʃ  | NO |   | NO |    | NO |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     | YES | YES | YES | YES |     |  |
| ʒ  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     | YES | YES | YES | YES |     |  |
| x  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     | NO  | YES | YES | YES |     |  |
| n  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| ɲ  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| m  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| ŋ  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| l  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| r  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| w  |    |   | NO |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| j  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| ɰ  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |
| j  |    |   |    |    |    |   |    |    |    |    |    |    |   |   |   |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |  |

Figure 2. NAD-based predictions for initial doubles.

preferred status of NAD for final doubles

|    | p   | b | t   | d   | k   | g   | ts  | dz  | tʃ  | dʒ  | tʃ | dʒ | f | v | s | z  | ʃ  | ʒ   | ʃ | x | n | ɲ | m | ŋ | l | r   | w   | j   | ɰ   | j |  |
|----|-----|---|-----|-----|-----|-----|-----|-----|-----|-----|----|----|---|---|---|----|----|-----|---|---|---|---|---|---|---|-----|-----|-----|-----|---|--|
| p  |     |   | NO  |     |     |     |     |     |     |     |    |    |   |   |   | NO | NO |     |   |   |   |   |   |   |   |     | NO  |     |     |   |  |
| b  |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| t  |     |   |     |     |     |     |     |     |     | NO  |    |    |   |   |   |    | NO |     |   |   |   |   |   |   |   | NO  | NO  | NO  | NO  |   |  |
| d  |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| k  |     |   |     | NO  |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     | NO  | NO  | NO  |   |  |
| g  |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| ts |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| dz |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| tʃ | NO  |   | NO  |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| dʒ |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| tʃ |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| dʒ |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| f  |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| v  |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| s  | NO  |   | NO  |     | NO  |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     | NO  |     | NO  |   |  |
| z  |     |   | NO  |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| ʃ  | NO  |   | NO  |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| ʒ  | NO  |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| x  |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| n  |     |   |     | NO  |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| ɲ  |     |   |     | YES |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| m  | YES |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| ŋ  |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| l  |     |   |     | YES |     |     | YES | YES | YES | YES |    |    |   |   |   |    |    | YES |   |   |   |   |   |   |   | NO  | NO  | NO  |     |   |  |
| r  | YES |   | YES | YES | YES | YES | YES | YES | YES | YES |    |    |   |   |   |    |    | YES |   |   |   |   |   |   |   | NO  | NO  | NO  |     |   |  |
| w  | YES |   | YES | YES | YES | YES | YES | YES | YES | YES |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   | YES | YES | YES | YES |   |  |
| j  | YES |   | YES | YES | YES | YES | YES | YES | YES | YES |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   | YES | YES | YES | YES |   |  |
| ɰ  |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |
| j  |     |   |     |     |     |     |     |     |     |     |    |    |   |   |   |    |    |     |   |   |   |   |   |   |   |     |     |     |     |   |  |

Figure 3. NAD-based predictions for final doubles.

preferred status of NAD for medial doubles

|    | p   | b   | t   | d   | k   | g   | ts  | dz  | tʃ  | dʒ  | tʂ  | dʑ  | f   | v   | s   | z   | ʃ   | ʒ   | ʧ   | ʤ   | x   | ɲ   | ɳ   | m   | ŋ   | l   | r   | w   | j   | ɰ   | ɱ   |     |  |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| p  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| b  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| t  | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| d  | YES | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| k  | YES | YES | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| g  |     |     | YES | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| ts |     |     | YES | YES | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| dz |     |     | YES | YES | YES | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| tʃ | YES | YES |     |     | YES | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| dʒ | YES | YES |     |     | YES | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| tʂ | YES |     |     |     |     |     | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| dʑ | YES |     |     |     |     |     | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| f  |     |     | YES | YES | YES | YES | YES | YES |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| v  |     |     | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| s  | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| z  | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| ʃ  | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| ʒ  | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| ʧ  | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| ʤ  | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| x  |     |     | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| ɲ  |     |     | YES | YES | NO  | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| ɳ  |     |     | YES | YES | NO  | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| m  | YES | YES | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  |  |
| ŋ  |     |     | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |  |
| l  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  |  |
| r  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  |  |
| w  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  |  |
| j  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  |  |
| ɰ  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  |  |
| ɱ  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  | NO  |  |

Figure 4. NAD-based predictions for medial doubles.

### 1.2.3. Beats-and-Binding (B&B) phonotactics in Natural Linguistics explanatory model

Table 4 below shows the path of explanation in Natural Linguistics. Let us see how useful it becomes for Beats-and-Binding phonotactics.

Table 4. The explanatory model of Natural Linguistics (Dziubalska-Kolaczyk 2009, 2014).

|  |   |
|--|---|
| <b>higher principles</b><br>(e.g. the principle of the least effort, of cognitive economy) | <i>non-linguistic</i><br>(cognitive, phonetic, psychological, sociological etc) |
| <b>preferences</b><br>(e.g. a preference for simple phonotactics, for a CV structure)      | <i>linguistic</i>   |
| <b>preference parameters</b><br>(pronounceability, perceptibility)                         | <i>functional and semiotic</i>  |
| <b>consequences of preferences</b><br>(absence of clusters in a language)                  | <i>linguistic</i>   |

One of the higher principles, the cognitive principle of least effort dictates that it is less effortful to produce a single consonant than a cluster or a less rather than more complex cluster. The effort is better managed when a produced cluster is well perceived. The semiotic principle of

figure and ground guarantees that the contrast between a single consonant and a vowel is a better figure-against-ground structure than a cluster. The phonetic principle of alternation favours sequences of interspaced consonants and vowels. The above principles give rise to the linguistic CV-preference and the phonotactic preferences defined with reference to it (i.e. necessarily counteracting it). Once the preferences are defined, they can be parametrised using functional and semiotic measures as well as by means of morphonotactic and statistical parameters. Eventually, one arrives at the consequences of phonotactic preferences among the languages of the world. Typologically, only 30% of the languages have complex structures (C)(C)(C)V(C)(C)(C)(C) (Maddieson 2008). Additionally, preferred clusters are typologically attested. Finally, dispreferred clusters tend to be reduced by universal and language-specific processes in diachrony, acquisition, phonostylistics, speech pathology, i.e. in all manifestations of language performance.

### 1.3. Hypotheses: Preview

A number of hypotheses concerning consonantal clusters have been derived on the basis of the theoretical apparatus demonstrated in this chapter. They will be previewed here and elaborated on in the empirical chapters of the book.

The first hypothesis refers to the size of a cluster:

Cluster size corresponds to morphological complexity. The longer a cluster, the more likely it is to be morphonotactic.

The second hypothesis refers to the difference between phonotactic and morphonotactic clusters:

The degree of phonological preferability is inversely proportional to morphological complexity. Morphonotactic clusters are expected to have a lower degree of preferability than phonotactic ones.

The third hypothesis refers to frequency of a cluster:

The degree of cluster preferability is directly proportional to frequency. Preferred clusters are expected to be more frequent than dispreferred clusters.

In this chapter the theoretical ground for the investigations of consonant clusters has been laid out. The next chapter will present a tool worked out to support the analysis of data – the NAD Phonotactic Calculator.

## Chapter 2

# The Phonotactic Calculator

### 2.1. Introduction

The Phonotactic Calculator is a tool functionally designed by Katarzyna Dziubalska-Kołaczyk (Dziubalska-Kołaczyk and Krynicki 2007) in her research on Beats-and-Binding phonotactics. The computation of the Net Auditory Distance includes a number of parameters. It is a complicated process and calculation by hand is a laborious task. Thus, in order to improve the computation process experience, a special tool must be used to ensure the high speed of the process as well as correctness of the results.

The main functionality of the Calculator was to perform the NAD calculations on the basis of the parameters provided by the user. Thus, the program operates in the following way:

- (1) The user provides input, i.e. a set of clusters. The clusters can be copied into the program or the user may click on the symbols provided in the form of the symbol chart to place the input in the input field.
- (2) The user sets the parameters for the calculation, e.g. whether to include POA, MOA, voicing etc. to determine NAD, the location of the cluster for which the computation is being done, etc.
- (3) The program provides the outcome in the form of a table containing the results of the calculations as well as their parameters for future reference.
- (4) The results may be copied to clipboard or exported to a TXT file for further processing.

The tool described in this chapter is Version 2 of the application. Version 1, developed by Grzegorz Krynicki in 2007, was an online tool which performed computations on the basis of hardcoded parameters. Version 2, developed in years 2008-2013 by Dawid Pietrala, is a desktop program developed in Java that allows for the customization of the parameters provided in a readable external data source that can be easily modified in

the course of research. The selection of Java as the main framework for the application stemmed from the necessity to maintain interoperability of the program as well as flexibility in terms of development and user experience (ability to modify settings and phone values). The design implemented a model-view-controller approach which assured scalability of the application as well as its portability to other interface designs.

Version 2.0 was largely based on version 1.0+ in terms of functionality and interface design. The most important difference was that the 2.0+ version has been changed into a desktop application. Moreover, the tool has a number of advantages over the one by Krynicki:

- (1) Version 2.0+ does not require Internet access.
- (2) It gives additional options which allow the user to experiment with various phone values and revert to the original ones.
- (3) Changes to phone values can be saved to become permanent or they can be made for one session only.

## 2.2. Technology

The choice of Java for the development of the application was due to the necessity to make the application work both in Linux and Windows environments. Java is a multiplatform, general-purpose, concurrent, class-based, object-oriented programming language and computing platform first released by Sun Microsystems (now subsidiary of Oracle Corporation) in 1995 as a core component of Sun Microsystems' Java platform. Its multiplatform capabilities are possible thanks to compilation of the source code to bytecode (class file) that can run on any Java Virtual Machine (JVM) regardless of computer architecture provided it is equipped with the Java Runtime Environment (hence, JRE). Since 2007, most of Java technologies are available under the GNU General Public License (more information can be found at <http://java.com>). Each update of the application was tested against the Fedora and openSUSE Linux distributions as well as Windows XP and 7 and Mac OS 10.6+. The application was distributed as a JAR file which is usually immediately recognized by an operating system as a file which should be run by the JRE.

An additional precondition for an operating system to run the program (apart from having JRE installed) was to provide a font able to display all IPA symbols. The recommended font used in the process of designing the

GUI was the Doulos SIL font<sup>1</sup>. Doulos SIL is a single Unicode font which contains a comprehensive inventory of symbols and glyphs needed for any Roman- or Cyrillic-based writing system. The font was released under the Open Font License and can be used for any purposes free of charge, however the usage of the font in commercial applications is restricted in a number of cases. The font is designed to work on systems and with applications that provide support for TrueType fonts and for Unicode character encoding. Thus, it includes the most popular operating systems such as Microsoft Windows and OS X and the most popular Linux distributions. Upon loading, the application checks if the font is installed on the system. If the result is *true*, the application continues loading the GUI. Otherwise, it displays a pop-up message asking the user to install the font and warning of potential malfunction until/unless the font is installed; the label symbols, user input textfield and results table may fail to display the content correctly. This does not affect standard Latin symbols and should not affect the calculation process, however if the user wants to calculate an IPA sequence without having the Doulos SIL font installed, the application may not provide correct results or function in an unexpected way.

**Main view.** The main view of the application consists of one screen as well as additional subviews which can be switched between using tab switching. Figure 5 presents a screenshot of the main view of the calculator.

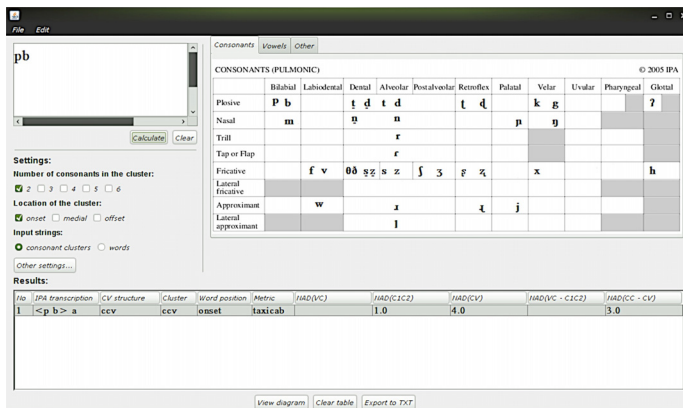


Figure 5. The Phonotactic Calculator – main window.

<sup>1</sup> More information on the font can be found here: [http://scripts.sil.org/cms/scripts/page.php?site\\_id=nrsl&id=doulossilfont](http://scripts.sil.org/cms/scripts/page.php?site_id=nrsl&id=doulossilfont)

The screen consists of a number of components:

- (1) A textfield which allows the user to enter clusters for which NAD should be calculated – clusters should be delimited by the symbols: ‘.’, ‘;’, and ‘,’. The preferred one is ‘,’. If the user enters a sequence of letters which are not delimited by separation symbols, the Calculator will automatically generate all possible combinations of clusters of the length indicated in the section below.
- (2) Buttons *Calculate* and *Clear* – the former button starts the calculation process, while the latter button clears the textfield.
- (3) *Settings* – in this area, the user can set the parameters for the calculation. The parameters are the following:
  - The number of consonants in the cluster (one or more settings are permissible).
  - The location of the cluster (one or more settings are permissible) – this setting allows the user to decide whether calculation should be performed for onset, medial and/or offset clusters.
  - Input strings (one setting is permissible) – this setting allows the user to decide whether the calculator should search for clusters as they are entered by the user or whether it should scan words in search for clusters.
- (4) *Other settings* button – opens a new window which allows the user to set other settings including phoneme parameters (see below).
- (5) *Results* area – an area containing a table with calculation results. The table consists of a number of columns and results are put into the table in rows. The columns are:
  - *Number* – indicates the calculation sequence number performed by the calculator.
  - *IPA transcription* – shows the notation of a cluster. The vowel /a/ is added automatically in the process of result generation.
  - *CV structure* – shows the structure of the cluster. Consonants are represented by the letter C while vowels are represented by V.
  - *Cluster* – shows the structure of the cluster. Consonants are represented by the letter C while vowels are represented by V.
  - *Word position* – indicates one of the three parameters which should be set before calculation (onset/medial/offset). Calculation results for each location are presented in separate rows.
  - *Metric* – indicates whether the calculation metric was taxicab or euclidean.

- *NAD (VC)* – shows the NAD value for the VC sequence. This column is used for calculations including medial and offset clusters. Results are presented in a floating point value. If the location of cluster is set to onset, it displays no value.
- *NAD (C1C2)* – shows the NAD value for the first and second consonants in the cluster. Used for all locations of clusters. The result is presented in the form of a floating point value.
- *NAD (C2C3)* – this column is not shown in the screenshot. It shows the NAD value for the second and third consonants in the cluster. Used for all locations of clusters provided the number of consonants in the cluster is set to 3 and the application has found a 3-symbol cluster in the input. The result is presented in the form of a floating point value.
- *NAD (CV)* – shows the NAD value for the CV sequence. This column is used for calculations including medial and onset clusters. The results are presented in a float value. If the location of the cluster is set to onset, it displays no value.
- *NAD (VC – C1C2)* – shows the result (in the form of a floating point value) of subtraction of NAD values for VC and C1C2 which indicates the preferability of the cluster in medial and offset location.
- *NAD (C1C2 – C2C3)* – this column is not shown in the screenshot. It shows the result of subtraction of NAD values for C1C2 and C2C3 which indicates the preferability of the cluster in all locations provided the number of consonants in the cluster is set to 3 and the application has found a 3-symbol cluster in the input. The result is presented in the form of a floating point value.
- *NAD (CC – CV)* – shows the result (in the form of a floating point value) of subtraction of NAD values for CC and VC which indicates the preferability of the cluster in medial and onset locations.
- The buttons *View diagram*, *Clear table*, and *Export to TXT* – these buttons allow the user to perform additional actions which may follow the calculation process.
- *View diagram* – opens a new window displaying bar charts for the results which enhances the readability of the results (see below).
- *Clear table* – removes the results from the Results table. The same action is performed when a new calculation is performed.

- *Export to TXT* – exports the results to a plain text file in which all results are presented on separate lines and the values are delimited by the ‘;’ sign. This file can be easily imported to Excel spreadsheets for further processing of the results.
- (6) IPA symbol area – this area shows a collection of IPA symbols which allows the user to enter the symbols into the cluster textfield by clicking on the symbols. This is provided for user convenience as the user does not have to generate IPA symbols in a separate application. The labels representing IPA symbols are put as a layer over the IPA chart image which also allows the IPA-aware user to easily find symbols representing various sounds. Due to the fact that not all symbols could be included in the main view of the IPA symbol area, they are divided into 3 tabs:
- *Consonants* – default view showing pulmonic consonants.
  - *Vowels* – a view showing cardinal vowels.
  - *Other* – a view showing affricates.

The design described above employed the most standard usability principles as the process of using the calculator started from the upper left corner and finished at the bottom of the screen.

**Subviews.** As has been stated above, the calculator includes two subviews which are opened in separate windows. The first subview presents a settings screen which allows the user to set additional parameters used in calculation. The screen is shown in Figure 6.

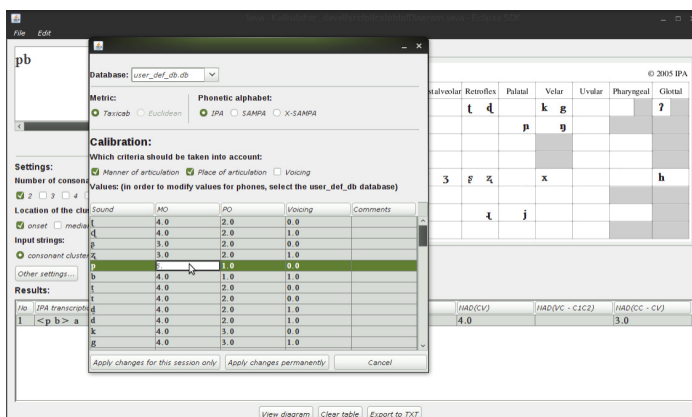


Figure 6. The Phonotactic Calculator – Other settings window.

This window allows the user to:

- (1) Select a database of phone values – the default one is the Original Database. The phone values are shown in Table 5.

Table 5. Default POA and MOA values.

| OBSTRUENT |           |     | SONORANT |         |        |       | VOWEL |               |                |
|-----------|-----------|-----|----------|---------|--------|-------|-------|---------------|----------------|
| STOP      | FRICATIVE |     | NASAL    | LIQUID  |        | GLIDE |       |               |                |
|           | AFFRICATE |     |          |         |        |       |       |               |                |
|           |           |     |          | lateral | rhotic |       |       |               |                |
| 5.0       | 4.5       | 4.0 | 3.0      | 2.5     | 2.0    | 1.0   | 0     |               |                |
| p b       |           |     | m        |         |        | w     | 1.0   | bilabial      | <b>LABIAL</b>  |
|           |           | f v |          |         |        |       | 1.5   | labio-dental  |                |
|           |           | θ ð |          |         |        |       | 2.0   | inter-dental  |                |
| t d       |           | s z | n        | l       |        |       | 2.3   | alveolar      | <b>CORONAL</b> |
|           | tʃ dʒ     | ʃ ʒ |          | r       |        |       | 2.6   | post-alveolar |                |
|           |           |     |          |         |        | j     | 3.0   | palatal       | <b>DORSAL</b>  |
| k g       |           |     | ŋ        |         |        | w     | 3.5   | velar         |                |
|           |           |     |          |         |        |       | 4.0   | palatal       | <b>RADICAL</b> |
| ʔ         |           | h   |          |         |        |       | 5.0   | glottal       | <b>GLOTTAL</b> |

In this database, the user cannot modify values (fields containing values are non-editable). If the user wants to test his/her values, s/he needs to switch the database to *user\_def\_db*, fields which contain phone values become editable, and the user may change the values to their liking. If the user saves the new values permanently (to hard drive), this database will become his/her default database.

- (2) Change the metrics from taxicab to euclidean.
- (3) Switch between phonetic alphabets – the user may choose IPA (default), SAMPA, and X-SAMPA.
- (4) Enable or disable NAD criteria. Originally, all three criteria (i.e. place of articulation, manner of articulation and voicing) were enabled by default. After further research, the voicing criterion has been abandoned yet not removed from the interface in case further analyses with the criterion included were to be performed.

- (5) The rest of the screen comprises a table containing a list of sounds as well as their values for manner of articulation, place of articulation, and voicing. An additional column allows the user to enter comments next to each sound for future reference. As it has been stated above, if the user changes the database to User Defined Database, s/he will be able to modify the values, add comments and save their modifications either to memory (for testing and/or quick reference; the changes will disappear after the user has closed the calculator) or make the changes permanent by saving them to file. This feature considerably enhances the calibration process for the user as s/he may freely experiment with data in order to calibrate the values.
- (6) Another subview contains a bar chart which presents the results of NAD calculation in a coherent way. An example of a bar chart for the /pb/ cluster is presented in Figure 7.

The data presented using the bar chart is exactly the same as presented in the table.

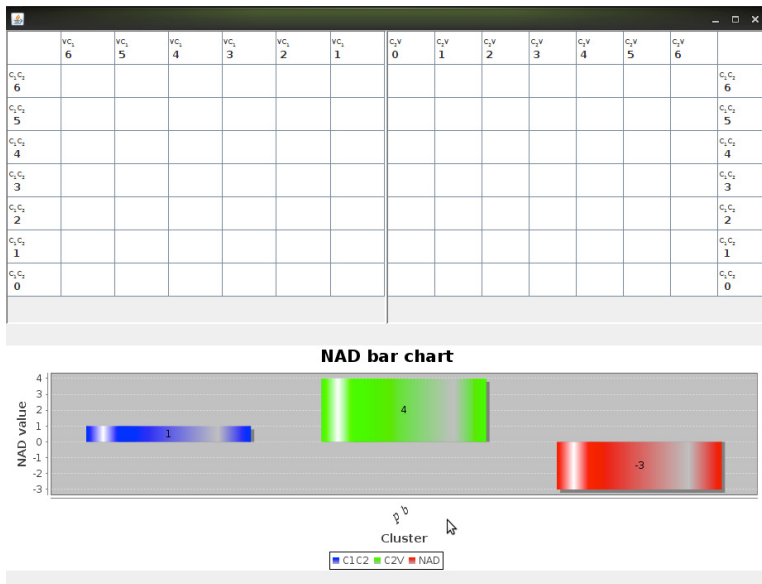


Figure 7. The Phonotactic Calculator – bar chart window.

**Other files.** The application makes use of 4 additional files which are stored in the application folder. These files are:

- 2 database files (*orig\_db.db* and *user\_def\_db.db*) – these files contain phone values as originally set by Dziubalska-Kołodziejczyk (*orig\_db.db*) or modified by the user (*user\_def\_db.db*). If the application does not find the files in the folder, it makes an attempt to create them. The phone values for both databases have been hardcoded in the program, and in the process of creating the files, the calculator will set the values to default.

The files are simple plain text files with a .db extension. They contain entries analyzed by the application line by line. Values in each line are delimited by tabs. Each line should consist of the following values:

- <phoneme symbol in X-SAMPA><tab><manner of articulation value><tab><place of articulation value><tab><voicing value – binary>.
- Upon launch, the application starts reading the file line by line and validating the lines in a simple way (by checking if the number of entries in each line is sufficient; no additional validation methods were applied), delimiting each line according to tabulation signs and assigning proper values as read from the files.
- 1 log file – contains a log of the application since its launch. If the file does not exist, the programme attempts to create it. The values put into the file are hardcoded in the application and they are produced by a standard *out.println()* Java method.
- 1 config.conf file – contains default user values as set in the *Settings* subview.

**Testing.** As it has been stated at the beginning of the chapter, the Calculator allows for fast generation of data with different calculation parameters. The results can then be used for further analysis. In order to present the capabilities of the tool, a short study was performed. Generating the data using the tool took less than a minute. Such speed allows for efficient generation of data which is extremely useful when forming and testing various hypotheses related to Beats-and-Binding Phonotactics.

In the study, 3 most popular double-initial clusters in English were chosen, namely /pr, st, fr/<sup>2</sup>. For these 3 clusters, the Calculator was used to calculate the preferability of the clusters in their initial positions. The calculations were done with different parameters, namely:

- (1) POA, MOA and voicing,
- (2) POA and MOA (without voicing),
- (3) POA and voicing (without MOA),
- (4) MOA and voicing (without POA).

Table 6 presents the values for POA, MOA and voicing for all components of the calculations.

Table 6. POA, MOA and voicing values for input sounds.

| Sound | POA | MOA | Voicing |
|-------|-----|-----|---------|
| /p/   | 1   | 5   | 0       |
| /r/   | 2.6 | 2   | 1       |
| /s/   | 2.3 | 5   | 0       |
| /t/   | 2.3 | 5   | 0       |
| /f/   | 1.5 | 4   | 0       |
| vowel | n/a | 0   | 1       |

The results of the calculations are presented in Tables 7-10.

Table 7. Preferability calculations for double-initials including all parameters (MOA, POA and voicing).

| Cluster | NAD CC | NAD CV | NAD CC-CV |
|---------|--------|--------|-----------|
| /prV/   | 5.6    | 2      | 3.6       |
| /stV/   | 0      | 6      | -6        |
| /frV/   | 4.1    | 2      | 2.1       |

<sup>2</sup> The popularity of the clusters was determined on the basis of the analysis of the 9 638 English words with double initial clusters selected from among the 500 000 most frequent words in the Corpus of Contemporary American English (Davies 2011, Pietrala 2014).

Table 8. Preferability calculations for double-initials excluding MOA.

| Cluster | NAD CC | NAD CV | NAD CC-CV |
|---------|--------|--------|-----------|
| /prV/   | 2.6    | 0      | 2.6       |
| /stV/   | 0      | 1      | -1        |
| /frV/   | 2.1    | 0      | 2.1       |

Table 9. Preferability calculations for double-initials excluding POA.

| Cluster | NAD CC | NAD CV | NAD CC-CV |
|---------|--------|--------|-----------|
| /prV/   | 4      | 2      | 2         |
| /stV/   | 0      | 6      | -6        |
| /frV/   | 3      | 2      | 1         |

Table 10. Preferability calculations for double-initials excluding voicing.

| Cluster | NAD CC | NAD CV | NAD CC-CV |
|---------|--------|--------|-----------|
| /prV/   | 4.6    | 2      | 2.6       |
| /stV/   | 0      | 5      | -5        |
| /frV/   | 3.1    | 2      | 1.1       |

The analysis of the data shows that for these particular clusters the selection of parameters does not have an impact on their (dis)preferred status as in all four computations as /pr/ and /fr/ are preferred and /st/ is dispreferred. However, manipulating the parameters largely influences the level of (dis)preferability. /st/'s dispreferability is high in all cases apart from the one in which MOA was excluded. It is the most striking case as in other clusters differences are not that large. Nevertheless, manipulating the parameters may have a significant impact on less frequent clusters and may be important when high precision of the preferability scores plays an important role.

Additionally, a similar study was done for these clusters in their medial and final positions. The results are presented in Tables 11-18.

Table 11. Preferability calculations for double-medials including all parameters (MOA, POA and voicing).

| Cluster | NAD VC | NAD CC | NAD CV | NAD VC-CC | NAD CC-CV |
|---------|--------|--------|--------|-----------|-----------|
| /VprV/  | 6      | 5.6    | 2      | 0.4       | 3.6       |
| /VstV/  | 6      | 0      | 6      | 6         | -6        |
| /VfrV/  | 5      | 4.1    | 2      | 0.9       | 2.1       |

Table 12. Preferability calculations for double-medials excluding MOA.

| Cluster | NAD VC | NAD CC | NAD CV | NAD VC-CC | NAD CC-CV |
|---------|--------|--------|--------|-----------|-----------|
| /VprV/  | 1      | 2.6    | 0      | -1.6      | 2.6       |
| /VstV/  | 1      | 0      | 1      | 1         | -1        |
| /VfrV/  | 1      | 2.1    | 0      | -1.1      | 2.1       |

Table 13. Preferability calculations for double-medials excluding POA.

| Cluster | NAD VC | NAD CC | NAD CV | NAD VC-CC | NAD CC-CV |
|---------|--------|--------|--------|-----------|-----------|
| /VprV/  | 6      | 4      | 2      | 2         | 2         |
| /VstV/  | 6      | 0      | 6      | 6         | -6        |
| /VfrV/  | 5      | 3      | 2      | 2         | 1         |

Table 14. Preferability calculations for double-medials excluding voicing.

| Cluster | NAD VC | NAD CC | NAD CV | NAD VC-CC | NAD CC-CV |
|---------|--------|--------|--------|-----------|-----------|
| /VprV/  | 5      | 4.6    | 2      | 0.4       | 2.6       |
| /VstV/  | 5      | 0      | 5      | 5         | -5        |
| /VfrV/  | 4      | 3.1    | 2      | 0.9       | 1.1       |

Table 15. Preferability calculations for double-finals including all parameters (MOA, POA and voicing).

| Cluster | NAD VC | NAD CC | NAD VC-CC |
|---------|--------|--------|-----------|
| /Vpr/   | 6      | 5.6    | 0.4       |
| /Vst/   | 6      | 0      | 6         |
| /Vfr/   | 5      | 4.1    | 0.9       |

Table 16. Preferability calculations for double-finals excluding MOA.

| Cluster | NAD VC | NAD CC | NAD VC-CC |
|---------|--------|--------|-----------|
| /Vpr/   | 1      | 2.6    | -1.6      |
| /Vst/   | 1      | 0      | 1         |
| /Vfr/   | 1      | 2.1    | -1.1      |

Table 17. Preferability calculations for double-finals excluding POA.

| Cluster | NAD VC | NAD CC | NAD VC-CC |
|---------|--------|--------|-----------|
| /Vpr/   | 6      | 4      | 2         |
| /Vst/   | 6      | 0      | 6         |
| /Vfr/   | 5      | 3      | 2         |

Table 18. Preferability calculations for double-finals excluding voicing.

| Cluster | NAD VC | NAD CC | NAD VC-CC |
|---------|--------|--------|-----------|
| /Vpr/   | 5      | 4.6    | 0.4       |
| /Vst/   | 5      | 0      | 5         |
| /Vfr/   | 4      | 3.1    | 0.9       |

The analysis of double-medials and finals shows that /pr, fr/, despite being preferred in initial positions and one exception in which we remove MOA in final positions, are dispreferred in their medial and final positions which corroborates the results presented in the previous chapter.

As for /st/, the cluster is dispreferred in all positions. It is noteworthy that in medial position the NAD formula is satisfied but the cluster violates the condition that NAD CC must be greater than 0, which makes it dispreferred as well.

All this data was calculated in seconds which clearly shows that the tool can be effectively used for performing analyses of cluster preferability with the B&B framework.

### 2.3. Usage and future development

The version of the tool described above was used in a number of studies, for instance Dziubalska-Kołaczyk et al. (2012, 2013), Dziubalska-Kołaczyk and Zielińska (2010, 2011), Zydorowicz (2010) and Pietrala (2014).

For the sake of future studies and popularizing the tool, the Phonotactic Calculator 3 is currently being developed. This version is available online at <http://wa.amu.edu.pl/nadcalc/>. It incorporates modern web technologies which allow the user to make use of all the options included in version 2 and sets the tool in a widely-accessible environment (Dziubalska-Kołaczyk et al. 2014).



## Chapter 3

### The resources

The analyses that were crucial to the project could not have been performed without access to the appropriate data. In this section we describe the textual resources used in the project.

#### 3.1. The data for Polish

Since the objective of the analysis was a quantitative and qualitative study of the distribution of consonant clusters in Polish and English, it was necessary to provide access to the appropriate linguistic data. Through the study of three different collections of data, we wanted to show how the character of the data studied determines the outcome. The main focus of the study was the frequency of consonants and clusters of consonants in three different language resources:

- raw text (e.g. nation-wide newspaper text designed for the general public),
- basic word forms (lemmas), representative of the core lexical inventory of a language (eg. list of entries in an “essential” type dictionary),
- inflectional forms of the core vocabulary, generated on the basis of full paradigms of the words.

Based on this directive, a decision was made to acquire the following resources:

- the full text of the *Rzeczpospolita* newspaper from the 2000-2001 period,
- the entire list of approx. 8 000 entries from the *Słownik Podstawowego Języka Polskiego dla Cudzoziemców* (Essential Polish Dictionary for Foreign Learners) by Bartnicka-Dąbkowska and Sinielnikoff (1999),
- inflectional forms generated on the basis of full paradigms for the 8 000 word list

### 3.1.1. Corpus data

The purpose of using raw newspaper text is to take advantage of the word frequency information that can be obtained for a word list extracted from such a corpus. Newspaper text is rich in topics covered and, consequently, rich in the vocabulary used. Another reason for the choice of this type of resource was its accessibility and controllable structure, where the content and meta information (such as author, title, date, domain etc) are easy to identify and process. The newspaper corpus studied in the project covered two years, which in technical terms meant: approximately 190 article texts per day (4.5 thousand texts in a single one-month folder), that is approximately 108 thousand files, 48.6 million words (tokens), which produced a list of approximately 630 thousand unique word forms (types). The graphs below show the growth of the vocabulary resource with respect to the two dimensions (tokens, types) in one-month intervals. As can be seen, in spite of the fact that the total number of words in the corpus studied was relatively large (40 million words), the number of “new” vocabulary in the last month was still relatively large (approx. 12 thousand), although much lower than in the first months (approx. 30-50 thousand).

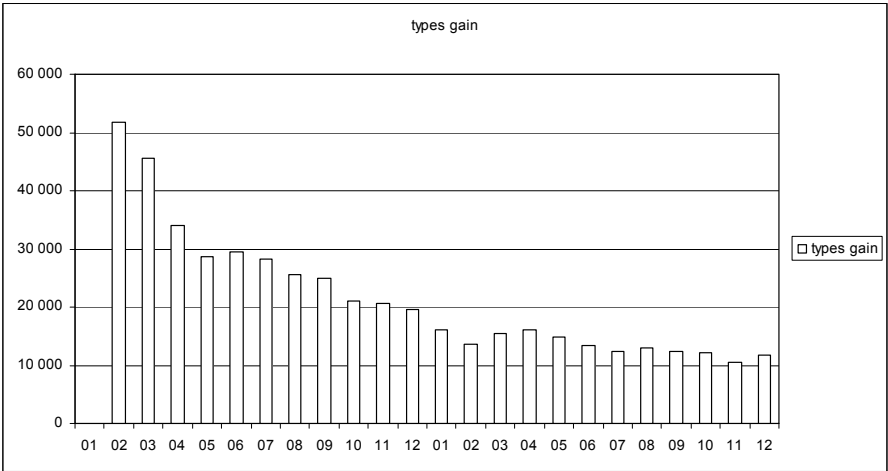


Figure 8. Number of “new” word types over a 24-month period.

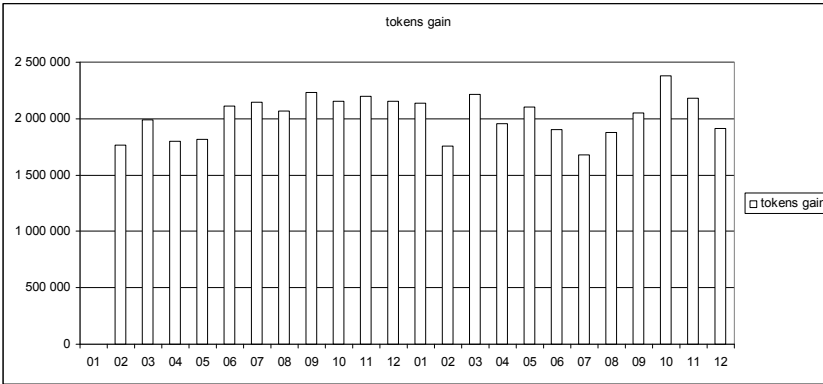


Figure 9. Number of word tokens over a 24-month period.

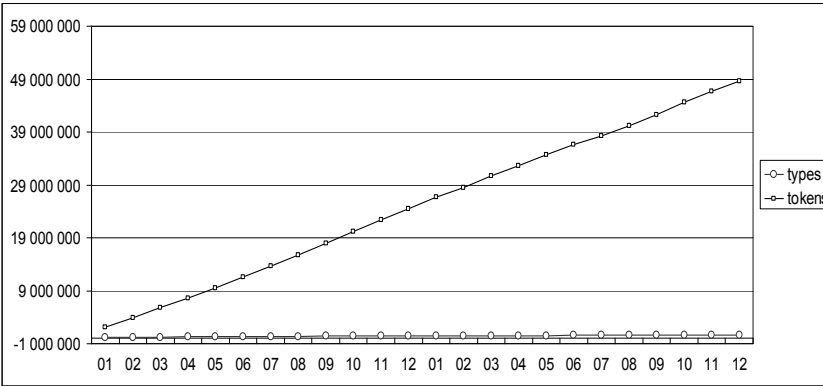


Figure 10. Number of word tokens and types plotted together over a 24-month period.

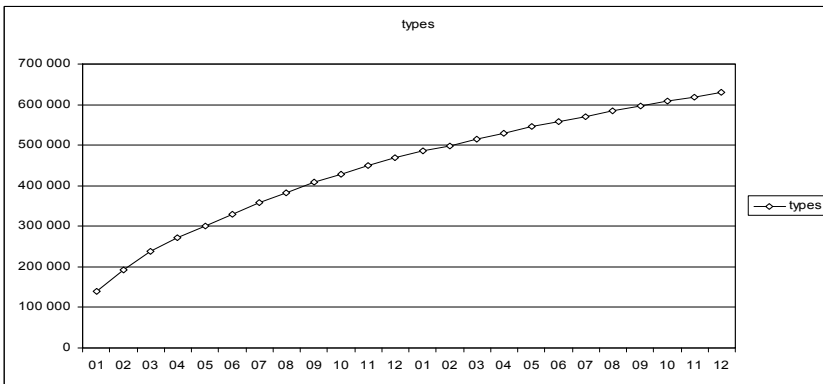


Figure 11. Close-up plot of the number of types over a 24-month period.

The corpus data was used to acquire frequency data for consonant clusters.

The table below features a sample data structure for the clusters in a number of most frequent words.

Table 19. A sample data structure for clusters.

| cluster | transcription | spelling | position | frequency | length |
|---------|---------------|----------|----------|-----------|--------|
| st      | jest          | jest     | f        | 317453    | 2      |
| pś      | pśes          | przez    | i        | 177921    | 2      |
| dl      | dla           | dla      | i        | 146173    | 2      |
| ...     |               |          |          |           |        |

### 3.1.2. Dictionary data

A list of lemmas was another resource to be studied. The assumptions about its characteristics were the following: it features basic dictionary forms (lemmas), all the items are in current usage (ie. it could not come from any historical type dictionary). As far as the domain is concerned, it should be as varied as possible (Markowski 1991), i.e. common to a large number of stylistic varieties of Polish. In a sense, a word frequency list would have met this criterion, which will be referred to further below.

The Polish lexicographic research has produced several resources which could provide a list of basic forms for the project. It has to be emphasized that such dictionaries would have to cover the vocabulary of the 20<sup>th</sup> century, with no limitations of geography, style, or domain. That is why we initially considered such collections as e.g. *Słownik języka polskiego* (Szymczak 1981), *Uniwersalny Słownik Języka Polskiego* (Dubisz 2003) and *Inny Słownik Języka Polskiego* (Bańko 2000). All of these turned out to be too large (approximately 70 000 – 100 000 entries) for the study to be feasible. After considering several other possibilities, a decision was made to use a “core” dictionary of essential Polish (called *słownik minimum*).

The reason for making this choice was the desire to obtain a representative collection (words to be encountered relatively frequently in Polish texts) of a manageable size. From among several possible “core” monolingual and bilingual dictionaries, the *Słownik Podstawowy Języka Polskiego dla Cudzoziemców* (Bartnicka and Sinielnikoff 1999) was selected, based on the information given by the publisher about its size being five thousand items. Another dictionary that was considered was the *Słownik Frekwencyjny Polszczyzny Współczesnej* (Kurcz et al. 1990), featuring approximately 10 000 entries, which was viewed as too large to be manageable. Although,

as it later turned out, the Bartnicka and Sinielnikoff dictionary word list was larger (8 000), the discrepancy resulting from a different understanding of the structure of the entry, this dictionary remained the dictionary of choice for this part of the project.

Below is a fragment of the dictionary entry list, where the items in the second and third column are either alternatives, perfective forms of the verbs, or derivatives (adjectives, adverbs etc.) of the items in the first column.

|                |                |           |
|----------------|----------------|-----------|
| dopływ         |                |           |
| dopływać       | dopłynąć       |           |
| dopóki         |                |           |
| doprowadzać    | doprowadzić    |           |
| dopuszczać     | dopuszczać     |           |
| dopuszczać się | dopuszczać się |           |
| doradca        | doradczy       |           |
| dorobek        |                |           |
| dorosły        |                |           |
| dorsz          |                |           |
| do rzeczy      |                |           |
| doskonali      | doskonale      |           |
| dostarczać     | dostarczyć     |           |
| dostateczny    | dostatecznie   |           |
| dostawać       | dostać         |           |
| dostawać się   | dostać się     |           |
| dostęp         |                |           |
| dostępny       |                |           |
| dostosowywać   | dostosować     |           |
| dosyć          | dość           |           |
| do syta        |                |           |
| dość           |                |           |
| doświadczenie  | doświadczalny  |           |
| doświadczony   |                |           |
| dotąd          |                |           |
| dotrzymywać    | dotrzymać      |           |
| dotychczas     |                |           |
| dotychczasowy  |                |           |
| dotyczyć       |                |           |
| dotykać        | dotknąć        |           |
| dowcip         | dowcipny       | dowcipnie |
| dowiadzać się  | dowiedzieć się |           |
| do widzenia    |                |           |
| do woli        |                |           |
| dowolny        |                |           |
| dowód          |                |           |
| dowódca        | dowódczy       |           |
| doznawać       | doznać         |           |
| dozorca        |                |           |
| dożynki        | dożynkowy      |           |
| dół            |                |           |
| dramat         | dramatyczny    |           |
| drapać         | drapać         |           |
| drapać się     | podrapać (się) |           |

Here is a fragment of the resulting word list.

|             |               |               |             |
|-------------|---------------|---------------|-------------|
| dopływać    | dostarczać    | dotąd         | dowód       |
| dopóki      | dostarczyć    | dotknąć       | dowódca     |
| doprowadzać | dostatecznie  | dotrzymać     | dowódczy    |
| doprowadzić | dostateczny   | dotrzymywać   | doznać      |
| dopuszczać  | dostawać      | dotychczas    | doznawać    |
| dopuszczyć  | dostęp        | dotychczasowy | dozorca     |
| doradca     | dostępny      | dotyczyć      | dożynki     |
| doradczy    | dostosować    | dotykać       | dół         |
| dorobek     | dostosowywać  | dowcip        | dramat      |
| dorosły     | dosyć         | dowcipnie     | dramatyczny |
| dorsz       | dość          | dowcipny      | drapać      |
| doskonale   | doświadczalny | dowiadzać     | drapieźny   |
| doskonały   | doświadczenie | dowiedzieć    | drapnąć     |
| dostać      | doświadczony  | dowolny       |             |

In a fashion similar to the other resources, initial, medial, and final clusters were extracted from the words and prepared for further study. It is important to add at this point that, as expected, the items on the “core” list were in fact “basic” forms, in the sense of “uninflected”. This characteristic had its consequences in the results obtained in our study.

### 3.1.3. Paradigmatic data

A separate part of the study was the analysis of clusters in inflected forms which were generated from the “core” dictionary word list with the use of the morfologik.org project rules. The original 8 000 list grew to approximately 190 000 forms, which can be interpreted as an average of approximately 25 inflectional forms for each word. This resource, apart from being representative of the inflectional variety of Polish, contained detailed grammatical labels, which proved helpful in identifying lexical categories for many selection operations. Below a sample of the data is shown.

A paradigm for the noun *mapa* ‘map’ – 10 forms:

|        |                                       |
|--------|---------------------------------------|
| map    | subst:pl:gen:f                        |
| mapa   | subst:sg:nom:f                        |
| mapach | subst:pl:loc:f                        |
| mapami | subst:pl:inst:f                       |
| mapie  | subst:sg:dat.loc:f                    |
| mapo   | subst:sg:voc:f                        |
| mapom  | subst:pl:dat:f                        |
| mapy   | subst:pl:acc.nom.voc:f+subst:sg:gen:f |
| mapą   | subst:sg:inst:f                       |
| mapę   | subst:sg:acc:f                        |

A paradigm for the verb *marnować* ‘to waste’ – 69 forms (not all shown):

|                 |   |
|-----------------|---|
| marnowanej      | ppas:sg:dat.gen.loc:f                                   |
| marnowaliby     | verb:pot:praet:pl:ter:m1:imperf                         |
| marnowalibyście | verb:pot:praet:pl:sec:m1:imperf                         |
| marnowalibyśmy  | verb:pot:praet:pl:pri:m1:imperf                         |
| marnowaliście   | verb:praet:pl:sec:m1:imperf                             |
| marnowaliśmy    | verb:praet:pl:pri:m1:imperf                             |
| marnowana       | ppas:sg:nom.voc:f                                       |
| marnowane       | ppas:pl:acc.nom.voc:f.m2.m3.n+ppas:sg:<br>acc.nom.voc:n |
| marnowanego     | ppas:sg:acc:m1.m2+ppas:sg:gen:m.n                       |
| marnowanemu     | ppas:sg:dat:m.n   |
| ...             |   |

Paradigms for the adjectives *jednopokojowy* ‘one-room’ and *jednopololeniowy* ‘single-generation’:

|                  |   |
|------------------|---|
| jednopokojowy    | adj:sg:nom.voc:m1.m2.m3:pos                   |
| jednopokojowych  | adj:pl:acc:m1.p1:pos                          |
| jednopokojowych  | adj:pl:gen:m1.m2.m3.f.n1.n2.<br>p1.p2.p3:pos  |
| jednopokojowych  | adj:pl:loc:m1.m2.m3.f.n1.n2.<br>p1.p2.p3:pos  |
| jednopokojowym   | adj:pl:dat:m1.m2.m3.f.n1.n2.<br>p1.p2.p3:pos  |
| jednopokojowym   | adj:sg:inst:m1.m2.m3.n1.n2:pos                |
| jednopokojowym   | adj:sg:loc:m1.m2.m3.n1.n2:pos                 |
| jednopokojowymi  | adj:pl:inst:m1.m2.m3.f.n1.n2.<br>p1.p2.p3:pos |
| jednopololeniowa | adj:sg:nom.voc:f:pos                          |
| jednopololeniową | adj:sg:acc:f:pos                              |
| jednopololeniową | adj:sg:inst:f:pos                             |
| jednopololeniowe | adj:pl:acc:m2.m3.f.n1.n2.p2.p3:pos            |
| jednopololeniowe | adj:pl:nom.voc:m2.m3.f.n1.n2.p2.p3:pos        |
| jednopololeniowe | adj:sg:acc:n1.n2:pos                          |
| jednopololeniowe | adj:sg:nom.voc:n1.n2:pos                      |

*The methods of morphological analysis.* The data, regardless of the resource, were subject to a number of essentially identical operations. The transcribed form of the word was used to extract clusters and record their position in the word. Where possible, the token frequency was recorded for the items present in the newspaper corpus. At this stage the format of the data was similar to those shown below:

| #  | cluster                | sym | trans       | ortho        | p | freq |
|----|------------------------|-----|-------------|--------------|---|------|
| 1. | $\widehat{\text{ts}}$  | Cc  | ućeCce      | ucieczce     | m | 263  |
| 2. | $\widehat{\text{ts}}$  | Cc  | zwolenńiCce | zwolenniczce | m | 0    |
| 3. | $\widehat{\text{tsf}}$ | Cf  | Cfartek     | czwartek     | i | 6658 |
| 4. | $\widehat{\text{tsf}}$ | Cf  | Cfartkom    | czwartkom    | i | 0    |
| 5. | $\widehat{\text{tsk}}$ | Ck  | pożyCki     | pożyczki     | m | 2320 |
| 6. | $\widehat{\text{tsk}}$ | Ck  | zebraCko    | zebraczko    | m | 0    |

An important part of the study was the identification of the status of a cluster with respect to morphology, that is classifying a cluster as morphonotactic (i.e. containing a morphological boundary) or lexical (*intramorphemic*, i.e. without a morphological boundary)<sup>1</sup>. This analysis was performed initially “by hand” for the “core” dictionary resource and did not bring satisfactory results due to the differences in interpretation of the morphological structure of a significant number of Polish words among the members of the research team. This phenomenon can be seen as resulting from variations in sensitivity to morphology reflecting the natural blurring of morphological boundaries due to lexicalization and other diachronic processes, which as simultaneous processes influence the linguistic awareness of the native speaker.

At one stage in the project the research team considered adopting the information on the morphological structure of clusters from two extensive resources on the Polish morphology. One was the five-volume dictionary of Polish derivation *Słownik Gniazd Słowotwórczych Współczesnego Języka Ogólnopolskiego* (Jadacka 2001a) and the other was the structured *a tergo* index of Polish inflectional forms *Schematyczny Indeks A Tergo Polskich Form Wyrazowych* (Tokarski 1993). Neither of the resources, however, proved usable in the project for the following reasons:

- poor theoretical motivation, i.e. no explicit division directives,
- no inflectional forms,
- mechanical, entirely asemantic approach to morphological structure.

Based on the analyses of the existing resources, including those of lexicographic, derivational, inflectional and grammatical nature, another

<sup>1</sup> Cf. Chapter 1, section 1.1., footnote 2.

method emerged involving the observation of a set of prefixes and suffixes which concatenate with consonants in nouns and adjectives, e.g. {-nica, -nik, -jada, s-, przed-} (i.e. morphemes with a consonantal onset for suffixes and with a consonantal offset for prefixes. The sources considered here were: (a) *Gramatyka Współczesnego Języka Polskiego* (Grzegorzczkowska et al. 1998), (b) *System Słowotwórczy Współczesnej Polszczyzny (1945–2000)* (Jadacka 2001b). Other sources on Polish morphology were also consulted on an informal basis. In the end, the morphemes studied in (a) were not considered due to their often archaic nature and low frequency. Among the morphemes used in this method are: prefixes: {przeciw-, kontr-, nad-, super-, hiper-, pod-, eks-, śród-, przed-, bez-} and suffixes: {-nica, -ka, -nia, -nik} etc. for the noun category and {-niczy, -ski} etc. for the adjective category. The suffixes were expanded to their inflectional forms, e.g. in *bruzdownica* ‘wall chaser’:

|                  |             |                        |
|------------------|-------------|------------------------|
| bruzdow /nic/    | bruzdownica | subst:pl:gen:f         |
| bruzdow /nica/   | bruzdownica | subst:sg:nom:f         |
| bruzdow /nicach/ | bruzdownica | subst:pl:loc:f         |
| bruzdow /nicami/ | bruzdownica | subst:pl:inst:f        |
| bruzdow /nice/   | bruzdownica | subst:pl:acc.nom.voc:f |

In other words, once it was established that the cluster /vn/ in *bruzdownica*, resulting from clipping {-nica} to *bruzda*, was morphological, the cluster /vn/ in all the other forms of *bruzdownica* was to be marked as morphological automatically. This way, the number of words where the assignment of a cluster to the morphonotactic or lexical category was to be performed manually, was significantly reduced. Each nominal suffix was expanded to all its possible inflectional forms. In addition to that, for the category of noun a division to prefix and stem was introduced to obtain the following data structure: *absolut+na*, *absolut+ne*, *absolut+nego*, *absolut+nej*, *absolut+nemu*, *absolut+ni*, *absolut+ny*, *absolut+nych*, *absolut+nym*, *absolut+nymi* (the cluster being /t+n/) ‘absolute’ and: *aktor+ce*, *aktor+ek*, *aktor+ka*, *aktor+kach*, *aktor+kami*, *aktor+ka*, *aktor+kę*, *aktor+ki*, *aktor+ko* (the cluster being /r+ts/ and /r+k/) ‘actress’. This approach, however, did not reduce the set of clusters to be analyzed manually to a manageable number. From the approximately 190 000 inflectional forms generated from the “core” vocabulary list, a list of approximately 170 000 words remained to be analyzed manually

Eventually, yet another method was devised to approach the issue of morphological boundaries. This method takes advantage of sorting the data according to specific criteria based on the cluster and its context (including *a tergo*), which allows for large groups of clusters to be assigned to one or the other category in an optimized semi-automatic fashion. This speeded up the process of boundary assignment considerably. Examples of data formats from various stages of this process are shown below.

| orth            | tr-rev          | orth-rev        | pos | freq | type | LEFT    | M | RIGHT   |
|-----------------|-----------------|-----------------|-----|------|------|---------|---|---------|
| administracja   | ajcartsińimda   | ajcartsinimda   | m   | 988  |      | adminis | t | racja   |
| administracjach | hajcartsińimda  | hajcartsinimda  | m   | 3    |      | adminis | t | racjach |
| administracjami | imajcartsińimda | imajcartsinimda | m   | 4    |      | adminis | t | racjami |
| administracje   | ejcartsińimda   | ejcartsinimda   | m   | 37   |      | adminis | t | racje   |
| administracje   | ejcartsińimda   | ejcartsinimda   | m   | 454  |      | adminis | t | rację   |
| administracji   | ijcartsińimda   | ijcartsinimda   | m   | 7294 |      | adminis | t | racji   |
| administracja   | ojcartsińimda   | ojcartsinimda   | m   | 0    |      | adminis | t | racjo   |
| administracja   | Xojcartsińimda  | ajcartsinimda   | m   | 184  |      | adminis | t | racja   |
| administracjom  | mojcartsińimda  | mojcartsinimda  | m   | 1    |      | adminis | t | racjom  |
| administracyj   | jycartsińimda   | jycartsinimda   | m   | 0    |      | adminis | t | racyj   |

| orth-rev       | pos | freq | type | LEFT     | M  | RIGHT |
|----------------|-----|------|------|----------|----|-------|
| awtsrezc       | m   | 0    |      | cz       | st | wa    |
| awtsrezc       | m   | 1    |      | cz       | st | wą    |
| ewtsrezc       | m   | 2    |      | cz       | st | we    |
| ogewtsrezc     | m   | 1    |      | cz       | st | wego  |
| jewtsrezc      | m   | 1    |      | cz       | st | wej   |
| umewtsrezc     | m   | 0    |      | cz       | st | wemu  |
| iwtsrezc       | m   | 0    |      | cz       | st | wi    |
| ywtsrezc       | m   | 3    |      | cz       | st | wy    |
| hcywtsrezc     | m   | 1    |      | cz       | st | wych  |
| mywtsrezc      | m   | 0    |      | cz       | st | wym   |
| imywtsrezc     | m   | 0    |      | cz       | st | wymi  |
| wtsradopsog    | f   | 1742 |      | gospodar | st | w     |
| awtsradopsog   | m   | 3496 |      | gospodar | st | wa    |
| hcawtsradopsog | m   | 335  |      | gospodar | st | wach  |
| imawtsradopsog | m   | 26   |      | gospodar | st | wami  |
| mewtsradopsog  | m   | 90   |      | gospodar | st | wem   |
| owtsradopsog   | m   | 813  |      | gospodar | st | wo    |
| mowtsradopsog  | m   | 42   |      | gospodar | st | wom   |
| uwtsradopsog   | m   | 9    |      | gospodar | st | wu    |

|              |   |    |  |        |    |      |
|--------------|---|----|--|--------|----|------|
| wtsrecrah    | f | 0  |  | harcer | st | w    |
| awtsrecrah   | m | 84 |  | harcer | st | wa   |
| hcawtsrecrah | m | 0  |  | harcer | st | wach |
| imawtsrecrah | m | 0  |  | harcer | st | wami |
| mewtsrecrah  | m | 0  |  | harcer | st | wem  |
| owtsrecrah   | m | 22 |  | harcer | st | wo   |

### 3.1.4. Phonemic transcription of Polish

As mentioned in Chapter One, the following set of 31 segments was used for the analysis of consonant clusters in Polish:

**p, b, t, d, k, g,  $\widehat{ts}$ ,  $\widehat{dz}$ ,  $\widehat{tʃ}$ ,  $\widehat{dʒ}$ ,  $\widehat{tɕ}$ ,  $\widehat{dʑ}$ , f, v, s, z, ʂ, ʐ, ɕ, ʐ, x, m, n, ɲ, ɳ, l, r, w, j,  $\widetilde{w}$ ,  $\widetilde{j}$**

**Automatic phonemic transcription.** All the resources used for the analysis of consonant clusters in the project were transcribed with the use of software and rules designed by a member of the research team (originally described in Jankowski 1994). Thanks to the flexibility of the software, it was possible to modify the transcription rules according the changing research needs.

**Transcription rules.** A rule in the automatic transcription system is a sequence of symbols of the following simple form:

**X, Y, Z**

where *X* denotes a grapheme with a possible context, *Y* denotes the output text (the intermediate or final form of the transcription), and *Z* denotes a *shift* argument, which is a number instructing the algorithm's internal pointer to move to the right by the specified number of characters. For example, a simplest possible rule:

**t, t, 1**

says: *if the character 't' is encountered in the text processed, output the text 't' and move the pointer one position to the right, i.e. to the next character following 't'*. As a result the character 't' is copied to the output unchanged, as often happens in Polish, as e.g. in the word *tak* 'yes'. Another rule:

**cz,  $\widehat{ts}$ , 2**

reads: *if the characters ‘c’ and ‘z’ are encountered one after the other in the text processed, output the text ‘ $\widehat{ts}$ ’ and move the pointer two positions to the right, i.e. skipping over two characters.* The rules for the letter ‘b’ in various contexts illustrate more complex cases:

**bi[aeouaęó],bj,2**

**bi,b,1**

**b\$,p,1**

**b[ptkfstch],p,1**

**b,b,1**

The first rule covers words such as *biały* ‘white’, *biel* ‘whiteness’, *biorą* ‘they take’, *biuro* ‘office’ etc., where the phonemes /b/, /j/ plus a vowel are expected in the transcription (*biały* → /bjawi/). If there is no vowel following <i>, as in *bić* ‘to fight’, *bitwa* ‘battle’ etc., the phonemes /b/, /i/ plus a vowel are expected in the transcription (*bić* → /bić/). The next rule covers the word final <b>, which is rendered as /p/, as in the word *ryb* ‘fish’ (GEN.PL). <b> is rendered as <p> when it precedes graphemes representing voiceless consonants <p>, <t>, <k>, <f>, <s>, <c>, <ć>, <h> (rule 4), as in *subpolarny* ‘subpolar’, *subtelny* ‘subtle’, *babka* ‘cake’, *subfolder*, *babski* ‘feminine’, *babcia* ‘granny’, *babć* ‘granny’ (GEN.PL) etc. The last rule covers cases not taken care of by the previous rules. The difference in the value of the Z argument results from the fact that in rule 1 the algorithm moves by two characters to the right, that is skips over the character ‘i’ and points to the vowel as the character processed next, whereas in the in rule 2 the pointer moves by one character and points to ‘i’ as the character processed next. In a similar fashion, in rule 4 the algorithm does not alter any of the consonants which are the context for ‘b’. All it does is changes ‘b’ to ‘p’ and moves the pointer by one position to the character to be processes next. The system for the transcription of Polish consists of two sets of approximately 360-380 rules.

**Transcription stages.** Where the final form of the transcriptions depends on longer contexts, the transcription algorithm is designed to perform several more complex operations in stages (as was established in tests on corpus data, a maximum of three). The first stage is where complex graphemes

such as “sz”, “cz”, “ch” etc. are translated to their intermediate form ( $\mathfrak{s}$ ,  $\mathfrak{t}\mathfrak{s}$ ,  $\mathfrak{h}$  respectively), whereas the subsequent stages are where the algorithm appears to follow the natural mechanism of sequentially cascading assimilations present in the Polish language. For example, in the word *krzywǫd* ‘harm’ (GEN.PL), the consequence of the fact that the “assimilation” rules cover pairs of symbols is that the final cluster is processed in two steps: first “wd” is changed to  $\mathfrak{vt}$ , in accordance with the rule changing a final voiced consonant to its voiceless counterpart ( $\mathfrak{d}\mathfrak{s}$ ,  $\mathfrak{t}$ ,  $\mathbf{1}$ ), and then  $\mathfrak{vt}$  is transformed into  $/\mathfrak{ft}/$  in accordance with the rule changing a voiced consonant to its voiceless counterpart if the following consonant is voiceless ( $\mathfrak{vt}$ ,  $\mathfrak{f}$ ,  $\mathbf{1}$ ). Similarly, the initial cluster in *krzywǫd* is processed in two steps: first, “rz” changes to an intermediate symbol  $\mathfrak{z}$ , in accordance with the rule simplifying the digraph  $\langle rz \rangle$  to  $\mathfrak{z}$  ( $\mathfrak{rz}$ ,  $\mathfrak{z}$ ,  $\mathbf{2}$ ), and then  $\mathfrak{kz}$  changes to  $/\mathfrak{k}\mathfrak{s}/$  in accordance with the rule changing a voiced consonant to its voiceless counterpart if the preceding symbol is a voiceless consonant ( $\mathfrak{kz}$ ,  $\mathfrak{k}\mathfrak{s}$ ,  $\mathbf{2}$ ).

As can be seen, this system does not have a separate rule for the text “wd” to be transformed to  $/\mathfrak{ft}/$  in word-final position in one step (this rule would have the form  $\mathfrak{wd}\mathfrak{s}$ ,  $\mathfrak{ft}$ ,  $\mathbf{2}$ ). Instead, this model uses a smaller number of simple rules and allows more passes over the processed text. The text “*zdzk*”, in words such as *rzdzka* ‘wand’, *mzdzku* ‘cerebellum’ (GEN.SG.) and *przejazdzka* ‘ride’, is processed in three stages:  $\mathfrak{z}$   $\mathfrak{d}$   $\mathfrak{z}$   $\mathfrak{k}$  (original sequence)  $\rightarrow$   $\mathfrak{z}$   $\mathfrak{d}\mathfrak{z}$   $\mathfrak{k}$  (intermediate  $\mathfrak{z}$  (from  $\langle \mathfrak{z} \rangle$ ), intermediate  $\mathfrak{d}\mathfrak{z}$  (from  $\langle \mathfrak{dz} \rangle$ ) and  $\mathfrak{k}$ )  $\rightarrow$   $\mathfrak{z}$   $\mathfrak{t}\mathfrak{s}$  (from  $\mathfrak{d}\mathfrak{z}$  in the context of  $\mathfrak{k}$ )  $\mathfrak{k} \rightarrow$   $\mathfrak{s}$  (from  $\mathfrak{z}$  in the context of  $\mathfrak{t}\mathfrak{s}$ )  $\mathfrak{t}\mathfrak{s}$   $\mathfrak{k}$  (final form). Similarly, the text “*zdzci*” in the word *gwizdzcie* ‘whistle’ (IMP.PL.) undergoes a three-stage transformation:  $\mathfrak{z}$   $\mathfrak{d}$   $\mathfrak{z}$   $\mathfrak{c}$   $\mathfrak{i}$   $\rightarrow$   $\mathfrak{z}$   $\mathfrak{d}\mathfrak{z}$   $\mathfrak{t}\mathfrak{c}$   $\rightarrow$   $\mathfrak{z}$   $\mathfrak{t}\mathfrak{s}$   $\mathfrak{t}\mathfrak{c}$   $\rightarrow$   $\mathfrak{s}$   $\mathfrak{t}\mathfrak{s}$   $\mathfrak{t}\mathfrak{c}$ . This approach is different from that used in the Steffen-Batogowa (1975) one-pass system, where single rules in the form of regular expressions are applied to longer orthographic contexts.

***The application of contextual rules in exceptional cases.*** Since the transcription algorithm takes advantage of contextual rules, the compilation of the phonematic transcription rules followed a thorough analysis of the graphemics of the samples of the Polish language designed for the analysis of clusters. In an effort for the rules to cover all possible graphemic and intermediate patterns, all possible contexts present in the data had to be taken into account to make sure all cases would receive the correct transcription. To illustrate the problem, let us consider the sequence of letters *dzi*, which is ambiguous and can be interpreted as the grapheme

<dzi> and transcribed as a single phoneme /**d̥z̥**/ (as in *podzielić* ‘to divide’, or as two separate graphemes <d> and <zi> and transcribed as two single phonemes /d/ and /z/ (as in *podziemny* ‘underground’ (ADJ)). A graphemic analysis of the resources revealed two large groups of words which feature the sequence *d z i*:

- words with the prefix {*pod-*} and a stem related to the word *ziemia* ‘ground’, such as (the numbers in parentheses are raw token frequencies):

podziemnych (396), podziemia (320), podziemnej (291),  
podziemne (281), podziemnego (241), podziemnym (171),  
podziemny (138), podziemiach (135), podziemiu (127),  
podziemną (74), podziemna (74), podziemie (68), pod-  
ziemi (61), podziemiem (37), podziemnymi (33), podzie-  
miami (4), podziemni (2), podziemnie (1)

- words with the prefix {*po-*}, such as *podzielić* ‘to divide’, *podziewać* ‘to be’, *podziękować* ‘to thank’

This sort of analysis allows for such formulation of rules that words from either group are treated in a different way. One rule has to take into account the context ‘*ziem-*’ (‘ground’) for the letter ‘d’:

**dziem, d, 1**

which says: *copy ‘d’ to output and move the pointer by one position (to the beginning of ‘dziem’)*, and the other covers all other contexts, including those where the grapheme <dzi> precedes a vowel symbol:

**dzi [aeouaęó] , d̥z̥, 3**

which says: *write **d̥z̥** to output and skip over to the vowel*. This notion, besides the mere correctness of the final transcription, has tangible consequences for the project: the words of the {*pod-*} + *ziem* group contain the cluster *d+z* (the plus sign indicates that the cluster is morphotactic), whereas the remaining words contain a single phoneme /**d̥z̥**/.

**Exceptions.** The rules in the automatic transcription system can take into account longer contexts and exceptions, such as the case of the word *pod-*

*zielonogórski* ‘of the area around Zielona Góra in western Poland’ ADJ., which is an example of a word with a prefix {*pod-*} (not {*po-*}), but cannot be processed with the rules for this group. If it was applied the rules given above for the {*pod-*} and {*po-*} contexts, the transcription would be wrong. Since it does not contain the component “*ziem*”, it matches the second rule, which would not give the expected transcription /*pod|zɛ1.../*. The system allows for such contexts as *podzielonogórski* to have exception rules, which in this case would have to specify the string “*podzielonog*” as the context, because only such a long context can correctly exclude words such as *podzielony* ‘divided’, *podzielono* ‘division was performed’ etc. The same approach is used to take into account other better or lesser known problematic cases of the Polish orthography, such as the words of the *marznąć* ‘to freeze’, *marzły* ‘they NVIR were freezing’, *zmarzłak* ‘a chilly person’ group (all featuring a four-grapheme <m><a><r><z> ‘to freeze’ stem), the words of the *sinus*, *silos* type, and various loanwords such as *anchois*, *atelier*, *foyer*, *gros* etc.

***The inventory of consonants adopted in the process of transcription.***

Since the establishment of the repertoire of consonants requires making some phonological decisions concerning the treatment of labial consonants, palatal stops, the realization of nasalized vowels, hard and soft resonance before fricatives, as well as obstruent final devoicing, below we present the relevant sources that influenced our choices.

Table 20. The inventory of consonants adopted in the transcription.

| ipa | examples  | references to the literature   | comments  |
|-----|---|--|---|
| p   | puk ‘knock’, pik ‘spades’, pies /pjes/ ‘dog’, zupa ‘soup’, drób ‘poultry’ | Wierzchowska 1967, 1971, 1980; Jassem 1958; Dukiewicz 1958; Koneczna 1934, 1965; Biedrzycki 1963; Benni 1926; Klemensiewicz 1930; Doroszewski and Wieczorkiewicz 1947; Bartnicka-Dąbkowska 1968; Dukiewicz, and Sawicka 1995; Steffen-Batogowa 1975; Steffen-Batóg, Nowakowski 1993; Nowakowski 1997; Jassem 1973; Osowicka-Kondratowicz 2005, Osowicka-Kondratowicz and Serowik 2007; Sawicka 1988; Sawicka, Grzybowski 1999; Madejowa 1989, 1990, 1992, 1993; Dunaj 1991, 2006 | There are no soft labial phonemes, only hard ones, as a result of asynchronous pronunciation of labial consonants |

|                |   |   |   |
|----------------|---|---|---|
| b              | buk 'beech', bij 'to beat' (IMP.SG), bierz /bjeʃ/ 'to take' (IMP.SG), żaba 'frog' | Same as above   | Same as above   |
| t              | tak 'yes', tik 'tick', sympatia /...tja/ 'liking'                                 | Same as above and Stieber 1966, 1972  | Palatalized sounds have always been positional variants, which results from the fact that foreign borrowings feature in Polish prominently  |
| d              | dal 'far', diler 'dealer', sty-pendium /...djum/ 'scholarship'                    | Same as above   | Same as above   |
| k              | kat 'executioner', kita 'tail', okiem /okjem/ 'eye' (INSTR.SG)                    | Koneczna 1934, 1965; Wierzchowska 1967, 1971, 1980; Dukiewicz, Sawicka 1995; Roślowski 1986; Nowakowski 1997; Madelska and Witaszek-Samborska 2015; Osowicka-Kondratowicz 2005, Osowicka-Kondratowicz and Serowik 2007; Madejowa 1989, 1990; Dunaj 1994, 2006 | The consequence of the increasing asynchronous pronunciation of postpalatal consonants is the elimination of postpalatal phonemes from phonological systems. The relevant sound have become hard. |
| g              | gol 'goal', gitara 'guitar', czoł-giem /...gjem/ 'tank' (INSTR.SG)                | Same as above   | Same as above   |
| $\widehat{ts}$ | cal 'inch', porcja 'portion', koc 'blanket'                                       | Standard in all sources   |   |
| $\widehat{dz}$ | sadza 'soot'  | Standard in all sources   |   |
| $\widehat{tʃ}$ | czas 'time'   | Standard in all sources   |   |
| $\widehat{dʒ}$ | Azerbejdżan 'Azarbayjan', dojeżdżać 'to commute'                                  | Standard in all sources   |   |
| $\widehat{tʃ}$ | ćwiczyć 'practise', ciarki /tearki/ 'chill (PL), jedź 'to go' (IMP.SG)            | Standard in all sources   |   |

|    |  |   |   |
|----|--|---|---|
| ɖz | niedźwiedzi<br>'bear' (ADJ),<br>dziad /ɖzat/<br>'grandfather'      | Standard in all sources   |   |
| f  | futro 'furcoat',<br>fiat /fjat/, paw<br>'peacock'                  | Standard in all sources   | Labial, analogical to /p/<br>and /b/  |
| v  | wór 'sack', wiór<br>/vjur/ 'wood<br>shaving'                       | Standard in all sources   | Same as above   |
| s  | sok 'juice', wóz<br>'cart'   | Standard in all sources   |   |
| z  | poza 'pose'  | Standard in all sources   |   |
| ɣ  | szał 'fury', beż<br>'beige'  | Standard in all sources   |   |
| ẓ | żał 'sorrow'   | Standard in all sources   |   |
| ɕ  | śląd 'trail', sito<br>/ɕito/ 'sieve',<br>weż 'to take'<br>(IMP.SG) | Standard in all sources   |   |
| z̥ | źle 'badly',<br>zima /zima/<br>'winter'                            | Standard in all sources   |   |
| x  | hala 'hall',<br>hiena /xjena/<br>'hyena'                           | Standard in all sources   | Voiced and postpalatal<br>allophones are<br>combinatorial variants<br>since they are bound by<br>optional variation – they<br>do not form minimal pairs |
| m  | mama 'mom',<br>miska 'bowl',<br>mieć /mjete/ 'to<br>have' (INF)    | Standard in all sources   | Labial, analogical to /p/<br>and /b/  |
| n  | noc 'night',<br>ponczo<br>'poncho'                                 | Standard in all sources, Dukiewicz<br>1967; Dunaj 2006; Doroszewski and<br>Wieczorkiewicz 1947; Klemensiewicz<br>1930, Madejowa 1989, 1990, 1992,<br>1993; Steffen-Batóg 1996 | The alveolar and possibly<br>voiceless allophones are<br>positional   |
| ɲ  | niania 'nanny',<br>woń 'scent',<br>nieco 'a tad', nic<br>'nothing' | Same as above   | Same as above   |
| ŋ  | męka 'ordeal',<br>łągi 'water<br>meadow', bank,<br>tango           | Same as above   | Nasalized vowels do not<br>occur before stops   |

|   |   |  |   |
|---|---|--|---|
| l | lok 'hair lock',<br>linia 'line', liana<br>/ljana/              | Same as above  | The palatalized and possibly voiceless allophones are positional  |
| r | rola 'role',<br>riposta<br>'comeback'                           | Same as above  | Same as above   |
| w | łuk 'arch',<br>połowa 'half',<br>auto 'car'                     | Zwoliński 1949 and the above.  | The non-syllabic /u/ has substituted the dental /w/ in the Polish standard. It also occurs in loan words, where it is represented graphically as <u>  |
| j | jest 'to be'<br>3P.SG., łajba<br>'boat', daj<br>'give' (IMP.SG) | Standard in all sources  |   |
| ẃ | wąs /woṽs/<br>'moustache'                                       | Jassem 1973; Dukiewicz 1967; Sawicka 1988; Dukiewicz and Sawicka 1995, Madejowa 1989, 1990, 1992, 1993.; Madelska and Witaszek-Samborska 2015; Dunaj 1991; Biedrzycki 1963, 1978 | The nasal resonance occurs as the second segment in a detailed notation of nasalized vowels, featuring asynchronous pronunciation and polisegmental structure   |
| ĩ | gdański<br>/gdaĩski/<br>'Gdańsk' (ADJ)                          | Same as above and Benni 1924, 1959   | The soft nasal resonance occurs as the second segment in a detailed notation of nasalized vowels with a raised articulation and as a representation of the <ń> grapheme after a vowel and preceding a fricative |

We also provide the treatment of the graphemes <ą> and <ę>, since they contribute to the formation of clusters. The table below shows representative examples of words with the so-called “nasal” <ą> and <ę> graphemes in various contexts and their transcriptions in the model followed in the project. The relevant literature consulted before this approach was adopted is given below with additional comments.

Table 21. The treatment of the graphemes &lt;ą&gt; and &lt;ę&gt;.

| example                    | transcription | references to the literature   | comments  |
|----------------------------|---------------|--|---|
| ząbek 'tooth'<br>(DIM)     | zɔmbɛk        | Standard in all systems  | Nasalized vowels occur either in word final or in initial and medial position before fricatives. In other cases we can only see them as the representation (in the form of various sounds) of the <ą> and <ę> letters before sounds other than fricatives.                    |
| ząbki 'tooth'<br>(DIM.PL)  | zɔmpki        |  |   |
| sąd 'court'                | sɔnt          |  |   |
| sądy 'courts'              | sɔndɨ         |  |   |
| mąka 'flour'               | mɔŋka         |  |   |
| żądza 'lust'               | ʒɔndza        |  |   |
| dać 'to blow'              | ɔɲtɛ          |  |   |
| kąsek 'bite'               | kɔŋsɛk        |  |   |
| łączyć 'to join'           | wɔntʂite      |  |   |
| gorąco 'hot'<br>(ADV)      | gɔrɔntso      |  |   |
| kaślivość 'biting'         | kɔʂɛlivoʂɛ    |  |   |
| wąs 'moustache'            | vɔŋs          |  |   |
| kęs 'bite'                 | kɛŋs          | Dukiewicz 1967; Dukiewicz and Sawicka 1995, Steffen-Batogowa 1975, Madejowa 1989, 1990, 1992, 1993 | This bi-phonemic interpretation is not supported within the classical Polish theory. Phoneticians (on the other hand) believe that the synchronous nasalized vowels no longer exist in Polish and emphasize a diphthongal nature: first oral resonance, then nasal resonance. |
| męka 'ordeal'              | mɛŋka         |  |   |
| wkręt 'screw'              | fkrent        |  |   |
| błędy 'errors'             | bwɛndɨ        |  |   |
| ręce 'hands'               | rɛntɛ         |  |   |
| tęcz 'rainbow'<br>(GEN.PL) | tɛntʂ         |  |   |

## 3.2. The data for English

In this section we present the sources we used to examine the morphonotactics of English, namely, the list of inflectional forms, the dictionary list and the corpus data.

### 3.2.1. The list of inflectional forms (paradigm)

The basic resource used in the project for the study of English was a list of phonemic transcriptions of basic and inflectional forms for a large number of core vocabulary, based on a well-established general purpose dictionary. This full lexicon may thus be seen as equivalent to the ‘paradigm’ resource used for Polish.

***The Lexicon for English.*** The lexicon used in the project is CUVOALD (Computer Usable Version of the *Oxford Advanced Learner’s Dictionary*) produced in 1985-86 by Roger Mitton. As the author states in the README file which is distributed along with the resource, CUVOALD is “a partial dictionary of English in computer-usable form – “partial” because each entry contained only some of the information from the original dictionary, and “computer-usable” (rather than merely “computer-readable”) because it was in a form that made it easy for programs to access it” (Mitton 1992: 2).

The lexicon was released in two dictionary files: CUVOALD and CUV2. Both were developed at the same time. These files were derived originally from the *Oxford Advanced Learner’s Dictionary of Current English*, third edition, published by the Oxford University Press the machine-readable version of which is available to researchers from the Oxford Text Archive (Mitton 1992: 2). The files were derived from the machine-readable OALDCE as part of a research project funded by the Leverhulme Trust, into spelling correction. There were additions performed by Roger Mitton in his own research as a lecturer in Computer Science at Birkbeck College, however no further information is provided on that process (Mitton 1992: 2).

***CUVOALD vs CUV2.*** As has been stated above, the dictionary is stored in two files: CUVOALD and CUV2. CUV2 is derived from CUVOALD and it also contains all inflected form provided explicitly. This feature

made it more robust to use the CUV2 files for the research purposes as it avoided the necessity to provide morphological processing of the input.

In Mitton's work, CUV2 contained 70 646 entries. Each entry occupied one line and consisted of a spelling, a pronunciation, one or more syntactic tags (part-of-speech) with rarity flags, a syllable count, and a set of verb patterns for verbs (Mitton 1992: 3).

The first file derived from the OALDCE (CUVOALD) contained all the headwords and subentries from the original dictionary – subentries are words like “abandonment” which comes under the headword “abandon” – except for a handful that contained funny characters (such as “Lsd” where the “L” was a pound sign). Subentries were not included if they consisted of two or three separate words that occurred individually elsewhere in the dictionary, such as “division bell” which comes under the headword “division”, except when the combination formed a syntactic unit not immediately predictable from its constituents, e.g. “above board”, which is listed as an adverb. To this list of about 35,000 entries, I added about 2,500 proper names – common forenames, British towns with a population of over 5,000, countries, nationalities, states, counties and major cities of the world. I would like to have added many more proper names, but I didn't have the time (Mitton 1992: 3).

In the years 1986-1992, Mitton made a number of corrections, added the rarity flags and the syllable counts and inserted about 2 000 new entries. The new entries, nearly all of which were derived forms of words already in the dictionary, were selected from a list of several thousand words that occurred in the LOB Corpus but were not included previously in CUVOALD. Mitton also admits having added about 300 reasonably common abbreviations and making changes to the existing entries where these were implied by the new entries; for example, when adding a plural form of a word whose existing tag was “uncountable”, it was necessary to change the tag of the singular form (Mitton 1992: 3).

Moreover, Mitton admits that

in the CUV2 file, each word has only one entry unless it has two different pronunciations, e.g. “abuse” (noun and verb). I have departed from this rule in the case of compound adjectives, such as “hard-working”, which have a slightly different stress pattern depending on whether they are used attributively (“she's a hard-working girl”) or predicatively (“she's very hard-working”). These are entered only once; they generally have the attributive stress pattern except when the

predicative one seemed the more natural. [...] I have also given only one entry to those words that have strong and weak forms of pronunciation, such as “am” (which can be pronounced &m, @m or m). Generally, it is the strong form that is entered (Mitton 1992: 3).

Unfortunately, the file in which the lexicon was described does not contain any detailed changelog which would particularly indicate the changes and additions performed by the author of the CUV2 file.

**The spellings and pronunciations.** In CUV2, the spelling contains the characters “A to Z”, “a to z”, hyphen, apostrophe, space, umlaut or dieresis (HEX 22), cedilla (3C), circumflex (5E), acute (5F), grave (60) and tilde (7E). These diacritic characters precede the letter that they mark, e.g. “se~nor”. Additionally, the author of the lexicon added two numbers – 5 and 6, which were used in the entries for “MI5” and “MI6” (Mitton 1992: 4).

**Additional changes to CUV2.** As has been stated above, Mitton’s lexicon, before it was used for the purpose of research, had undergone a number of additional changes. The version of CUV2 used for the present study was not a straight derivative of CUV2. It was extracted from the Phonetic Difficulty Index application designed by Sobkowiak (Sobkowiak 2006; Sobkowiak and Ferlacka 2011). In the course of preparation of the lexicon for their purposes, Sobkowiak (2006) adapted the dictionary by:

- adding a number of new words with their transcriptions (the version of the lexicon used for the study contains 84 501 entries 13 861 of which were added by Sobkowiak),
- providing General American transcriptions of words – this process was done semi-automatically by Sobkowiak,
- adding various additional tags that are used by the PDI application (Sobkowiak, personal communication).

In further steps, taken in the process of the dictionary’s design, the lexicon was adapted so that each entry consisted of four strings:

- orthographical representation of the word,
- part of speech marking,

- Received Pronunciation transcription in SAMPA – the lexicon was based on the original SAMPA except for the symbol for the TRAP vowel ('{') which was changed into 'ae' in order to enhance the readability of the lexicon,
- GenAm pronunciation in SAMPA with the same change concerning the TRAP vowel.

Each of these strings was enclosed in brackets and separated with commas. Thus, the entry for the word *dictator* looked as follows:

**(dictator) , (n) , (dIk'teIt@) , (dIk'teIt@r)**

Moreover, the current authors supplemented the entries with additional symbols for strong and weak forms which were concatenated with the words to which the dichotomy applied in order to distinguish the two pronunciation variants. Additionally, symbols for parts of speech were also changed for the sake of readability of the lexicon entries.

The changes applied in both steps were necessary to design a version of CUV2 which could be processed for the sake of the project.

### **3.2.2. The dictionary list**

The lexicon list limited to basic dictionary forms of words was used in the English part of the project as an equivalent of the 'core' dictionary list used for Polish. The number of words in this resource was approximately 60 000.

### **3.2.3. The corpus data**

Frequency data for the English 'corpus' part of the project were extracted from a frequency list based on the 410 million word Corpus of Contemporary American English (COCA) (Davies 2011). In other words, the corpus was used solely as a source of word frequency information for the lexicon resource. This list contains approximately 500,000 word forms, along with their grammar codes, number of occurrences, and number of sources in which they appear.



## Chapter 4

# Descriptive phonotactics of Polish

This section is devoted to the quantitative description of Polish phonotactics. Previous research on Polish phonotactics has primarily focused on the classification, frequency and ranking of clusters in written (Dobrogowska 1984, 1990, 1992, Dukiewicz 1980, 1985, Śledziński 2005) and spoken texts (Bargielówna 1950, Dunaj 1985, 1986, Madejowa 1990). Phonotactics has also supplied material for the study of phonostylistics (Durand and Gubrynowicz 1999) and language acquisition (Łukaszewicz 2007, Milewski 2005). The early work of Bargielówna (1950) and Kuryłowicz (1952) and the accounts which followed were structuralist descriptions. Theory-oriented accounts of phonotactics involved (e.g. Cyran and Gussmann 1999, Dressler and Dziubalska-Kołaczyk 2006, Dziubalska-Kołaczyk 2001, 2002, Gussmann and Cyran 1998, Rochoń 2000, Scheer 2007, 2008).

We supplement the existing descriptions with the presentation of Polish phonotactic potential realised in the three sources: the dictionary of core Polish, the paradigm as well as the corpus. Subsequently, we will discuss the importance of using the three sources by demonstrating types of clusters which are found in each dataset. As will be shown, a vast majority of clusters are present all the resources, but a certain set of clusters is exclusive to one source only. Finally, we discuss the participation of individual segments in clusters as well as the CV patterns of Polish words.

### 4.1. Resources

For the purposes of our study, we used data extracted from the dictionary, paradigm and corpus (for more details cf. Chapter 3). To ensure an exhaustive and multi-dimensional analysis of Polish and English phonotactics, we exploited these datasets in various ways. As far as Polish is concerned, the dictionary of 8 000 items *Słownik Podstawowy Języka Polskiego dla Cudzoziemców* (Bartnicka and Sinielnikoff 1999) was used for morphological parsing. The paradigm resource was created on the basis of the dictionary, providing us with an exhaustive list of 190 000 inflectional

forms. The corpus of newspaper texts containing 500 000 inflected word forms served as a reference for the frequency of use.

To compare with the English data, we used different sources due to the impoverished nature of English morphology, in particular non-existence of word-initial morphotactics. Therefore, we considered a list of inflectional forms based on a well-established dictionary and a word-frequency list from a large well-balanced corpus to be sufficiently reliable resources for the purpose of our analysis. The list of English inflected words was generated on the basis of the *Oxford Advanced Learner's Dictionary of Current English* (Hornby 1974), while the word-frequency data was extracted from the 500 million word *Corpus of Contemporary American English COCA* (Davis 2011). The corpus list contained approximately 500 000 word forms, and included their grammar codes, as well as a number of occurrences.

**Terminology.** Since we discuss phonotactics on the basis of three levels of analysis, we adopted the following terminology, which is presented in Table 22 with Polish and English examples.

Table 22. Terminology.

| language | cluster types | word types  | word tokens |
|----------|---------------|---|-------------|
| Polish   | kr-           | kraj 'country'<br>kreda 'chalk'<br>kraść 'to steal' ...                       | 221 134     |
|          | str-          | strajk 'strike'<br>s+trata 'loss'<br>s+trawić 'to digest' ...                 | 107 047     |
|          | fstr-         | ws+tręć 'disgust',<br>ws+trętny 'disgusting'<br>ws+trętnie 'disgustingly' ... | 138         |
| English  | -st           | fast, last, mist, miss+ed ...   | 5 446 745   |
|          | -ld           | old, cold, build, bill+ed ...   | 2 143 394   |
|          | -mp           | lamp, jump, hemp ...  | 115 767     |

The term *cluster type* refers to an individual cluster with a unique composition, e.g. the final English CC /st/. This specific cluster can occur in different *word types*, such as *fast*, *last*, *must*, *mist*, and *dentist*. Therefore the term *word types* refers to all words featuring a particular cluster.

Each inflected form is treated as a separate word type, as in *ask*, *asks* and *asked*. Repetitions of these words in the corpus are referred to as *word tokens*.

#### 4.2. Quantitative phonotactics

All the sources varied significantly in size, with the dictionary containing 8 000 lexical entries, the paradigm and corpus containing 190 000 and 500 000 inflected word forms, respectively. All of them generated different sets of cluster types. Tables 23-29 show the quantitative data for Polish in terms of cluster types, word types and word tokens.

Cluster types and word types were extracted from the dictionary (Table 23 and 24) and the paradigm (Tables 25 and 26). The corpus provided us with information not only on cluster and word types but also on the number of occurrences of each word, i.e. tokens (Tables 27-29). The presented are cumulative results obtained for clusters of various lengths and in different word positions. The visual representation of the data shown in Tables 23-29. can be found in Appendix A. Cluster-specific frequencies are given in Appendix B. The parallel data for English is presented in Appendix D and E (distribution of English consonant clusters in terms of size, word position and source), F (cluster-specific frequencies depending on the source, size and word position), and G (cluster overlap in the three sources).

Table 23. Cluster types in the dictionary.

| size  | initial | medial | final | total |
|-------|---------|--------|-------|-------|
| 2     | 147     | 295    | 54    | 496   |
| 3     | 85      | 216    | 10    | 311   |
| 4     | 9       | 28     | 0     | 37    |
| 5     | 0       | 2      | 0     | 2     |
| total | 241     | 541    | 64    | 846   |

Table 24. Word types in the dictionary.

| size  | initial | medial | final | total |
|-------|---------|--------|-------|-------|
| 2     | 2 144   | 5 429  | 565   | 8 138 |
| 3     | 385     | 870    | 34    | 1 289 |
| 4     | 15      | 68     | 0     | 83    |
| 5     | 0       | 4      | 0     | 4     |
| total | 2 544   | 6 371  | 599   | 9 514 |

Table 25. Cluster types in the paradigm.

| size  | initial | medial | final | total |
|-------|---------|--------|-------|-------|
| 2     | 162     | 357    | 100   | 619   |
| 3     | 97      | 457    | 27    | 581   |
| 4     | 14      | 95     | 8     | 117   |
| 5     | 0       | 12     | 1     | 13    |
| 6     | 0       | 1      | 0     | 1     |
| total | 273     | 922    | 136   | 1 331 |

Table 26. Word types in the paradigm.

| size  | initial | medial  | final | total   |
|-------|---------|---------|-------|---------|
| 2     | 47 607  | 165 459 | 2 334 | 215 400 |
| 3     | 10 272  | 20 071  | 82    | 30 425  |
| 4     | 349     | 1 842   | 28    | 2 219   |
| 5     | 0       | 73      | 2     | 75      |
| 6     | 0       | 2       | 0     | 2       |
| total | 58 228  | 187 447 | 2 446 | 248 121 |

Table 27. Cluster types in the corpus.

| size  | initial | medial | final | total |
|-------|---------|--------|-------|-------|
| 2     | 217     | 455    | 142   | 814   |
| 3     | 207     | 985    | 52    | 1 244 |
| 4     | 30      | 300    | 9     | 339   |
| 5     | 0       | 50     | 1     | 51    |
| 6     | 0       | 3      | 0     | 3     |
| total | 454     | 1 793  | 204   | 2 451 |

Table 28. Word types in the corpus.

| size  | initial | medial  | final | total   |
|-------|---------|---------|-------|---------|
| 2     | 89 118  | 357 708 | 8 874 | 455 700 |
| 3     | 14 589  | 62 720  | 244   | 77 553  |
| 4     | 730     | 12 196  | 70    | 12 996  |
| 5     | 0       | 464     | 5     | 469     |
| 6     | 0       | 13      | 0     | 13      |
| total | 104 437 | 433 101 | 9 193 | 546 731 |

Table 29. Word tokens in the corpus.

| size  | initial    | medial     | final     | total      |
|-------|------------|------------|-----------|------------|
| 2     | 8 653 268  | 20 313 563 | 1 524 889 | 30 491 720 |
| 3     | 1 301 445  | 3 347 894  | 66 238    | 4 715 577  |
| 4     | 59 689     | 557 881    | 18 233    | 635 803    |
| 5     | 0          | 32 433     | 2 706     | 35 139     |
| 6     | 0          | 782        | 0         | 782        |
| total | 10 014 402 | 24 252 553 | 1 612 066 | 35 879 021 |

### 4.3. Cluster overlap in the three sources

Below, we present the repertoire of hundreds of consonantal combinations, classifying them according to their position within a word (initial, medial, final) and size (CC, CCC, CCCC, CCCCC, CCCCCC). Our aim is to illustrate cluster overlap, i.e. presence or absence of initial and final clusters in particular sources. Since the data obtained for the word-medial position are sizeable, a report on the overlap of medial clusters is included in Appendix C.

**Initial and final clusters.** Figures 12-17 list the inventory of Polish clusters found in the dictionary, paradigm and corpus. Each higher-level source encompasses holistically the data found in the lower-level one. The lowest-level set contains clusters found in the dictionary. The mid-level set, together with its constituent set, represents the data extracted from the paradigm. The data found in the paradigm encompass all the dictionary clusters and an additional range of clusters that the dictionary is not able to generate, i.e. all inflected forms. As regards the highest-level set, it is an extension of the two smaller sets and lists extra cluster contributions from the corpus to the smaller-sized sources.

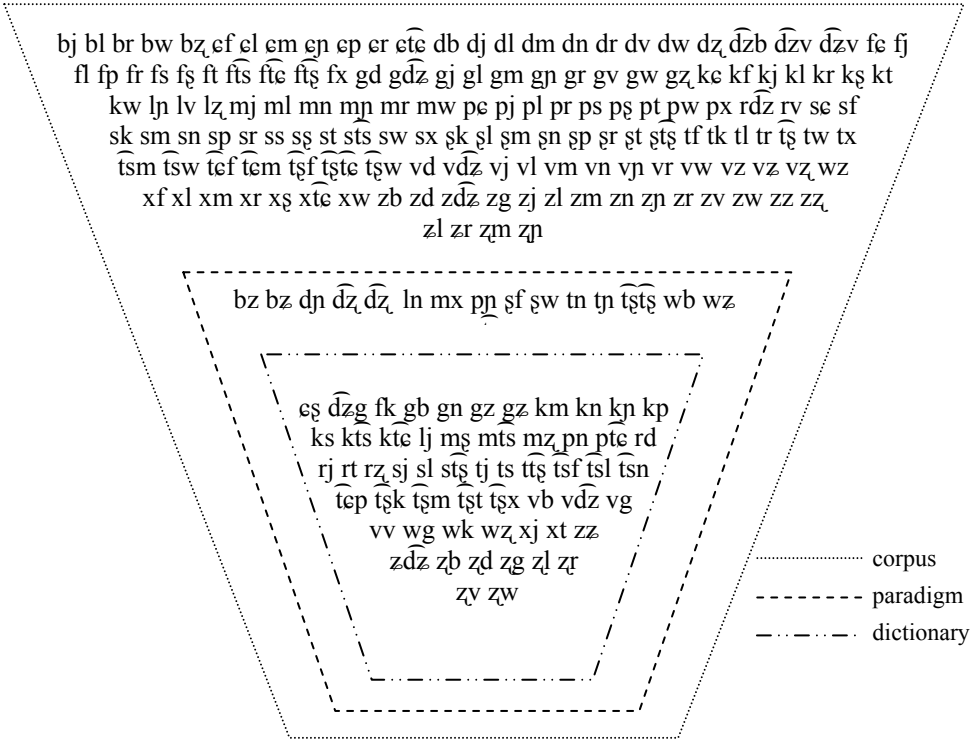


Figure 12. Initial CC clusters found in the dictionary, paradigm and corpus.

Among double word-initial clusters, all clusters types which appear in the dictionary are found in the paradigm and the corpus. /wz/ is the only initial double cluster present in the paradigm and absent from the corpus and the dictionary.

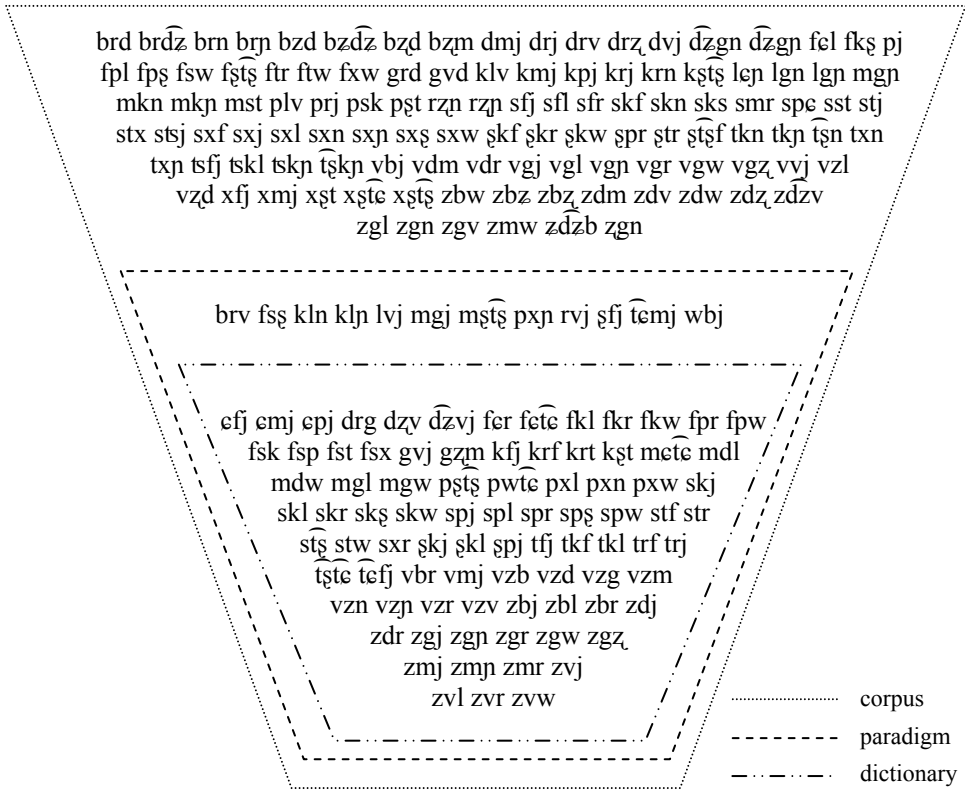


Figure 13. Initial CCC clusters found in the dictionary, paradigm and corpus.

Among triple word-initial clusters, all clusters types which appear in the dictionary are found in the paradigms and the corpus. /f̃s̃s̃/ is the only initial triple cluster present in the paradigms and absent from the corpus and the dictionary.

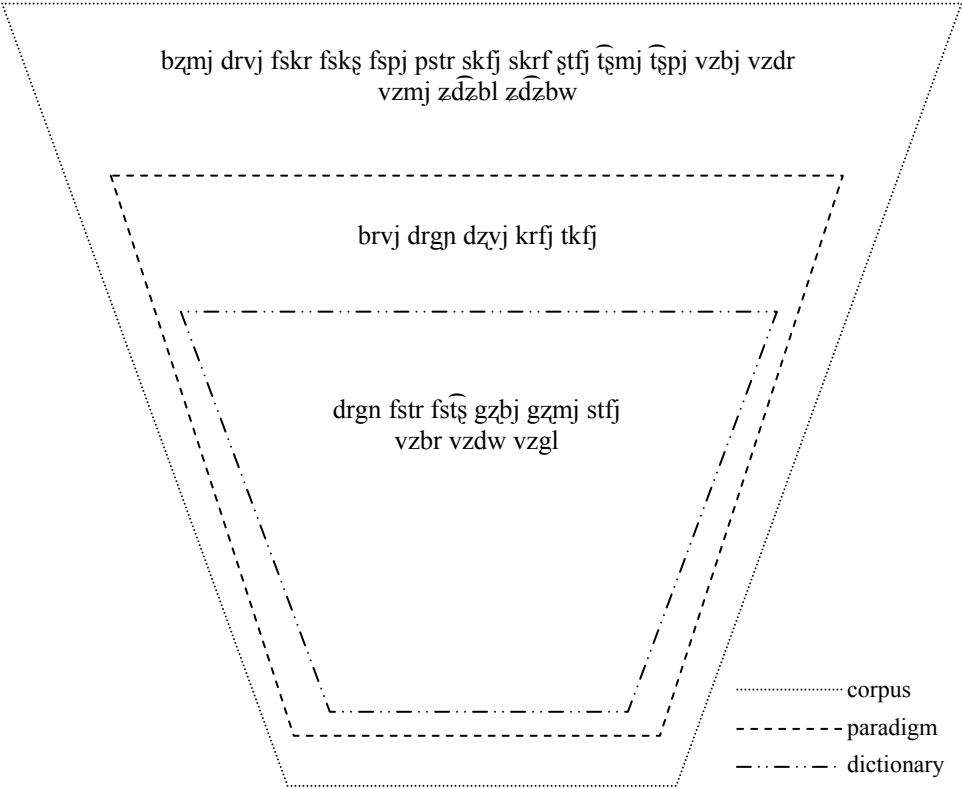


Figure 14. Initial CCCC clusters found in the dictionary, paradigm and corpus.

Among quadruple word-initial clusters, all cluster types which appear in the dictionary occur in the paradigm and the corpus. All the clusters found in the paradigm are present in the corpus as well.

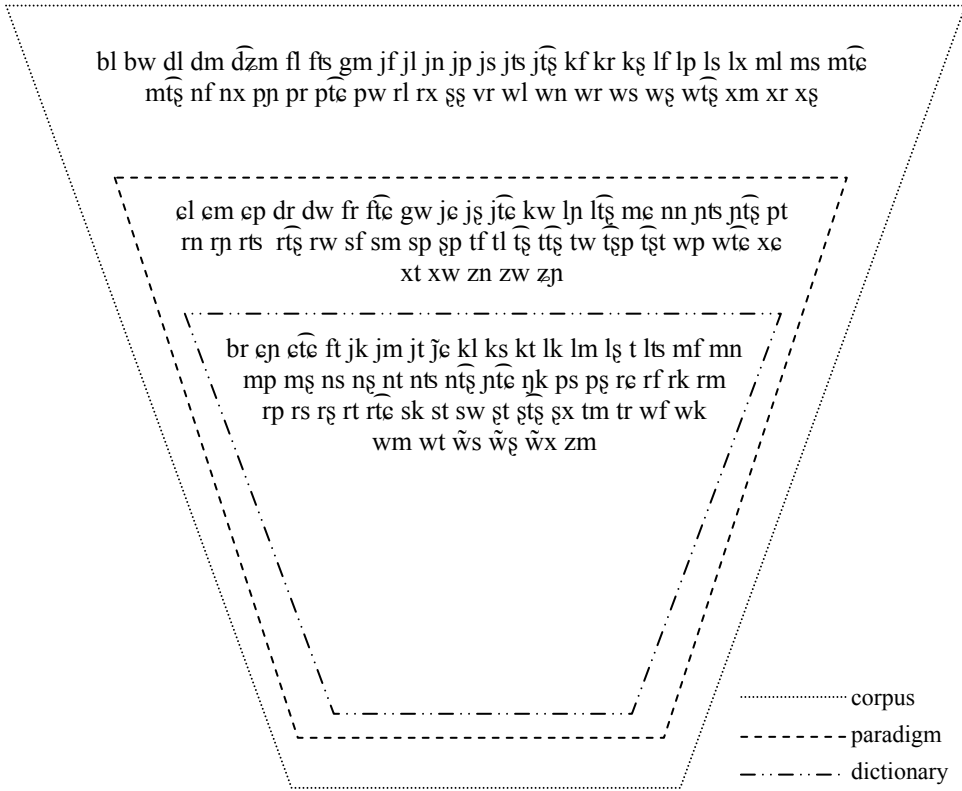


Figure 15. Final CC clusters found in the dictionary, paradigm and corpus.

Among double word-final clusters, all clusters types which appear in the dictionary are found in the paradigm and the corpus. Three double word-final clusters, / $\widehat{pts}$ , şf, wj/, were present in the paradigm and absent from the corpus and the dictionary.

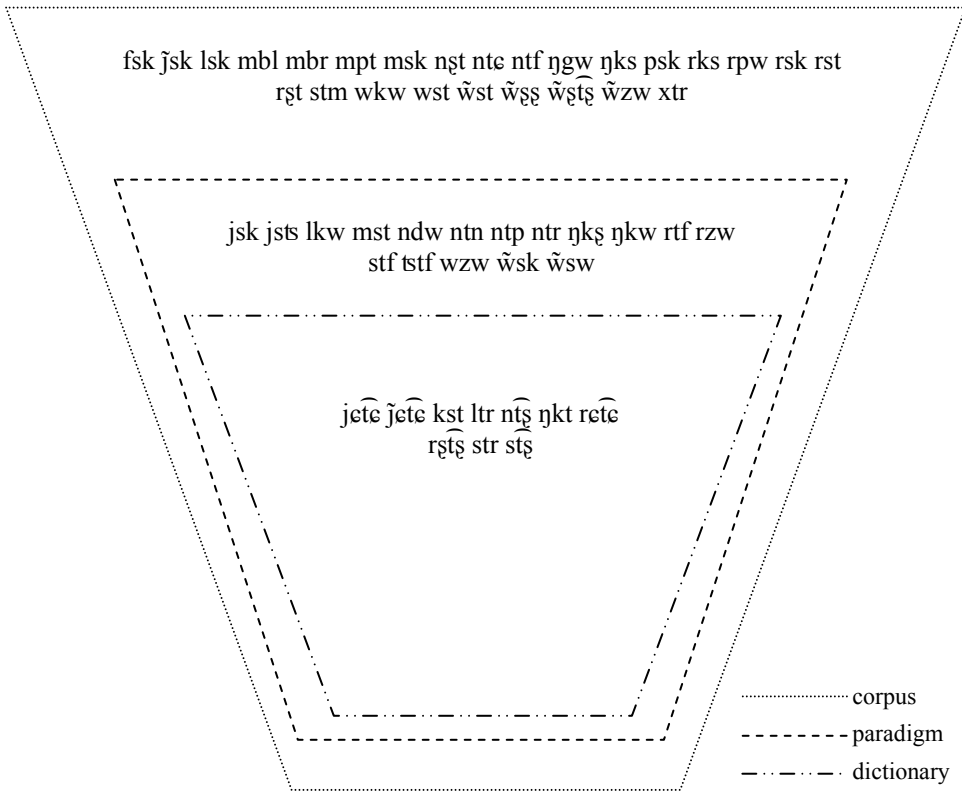


Figure 16. Final CCC clusters found in the dictionary, paradigm and corpus.

Among triple word-final clusters, all clusters types which appear in the dictionary are found in the paradigm and the corpus. The only triple word-final cluster present in the paradigm but absent from the corpus is /ntn/, e.g. in *t̄t̄n* ‘pulse’ (GEN.PL).

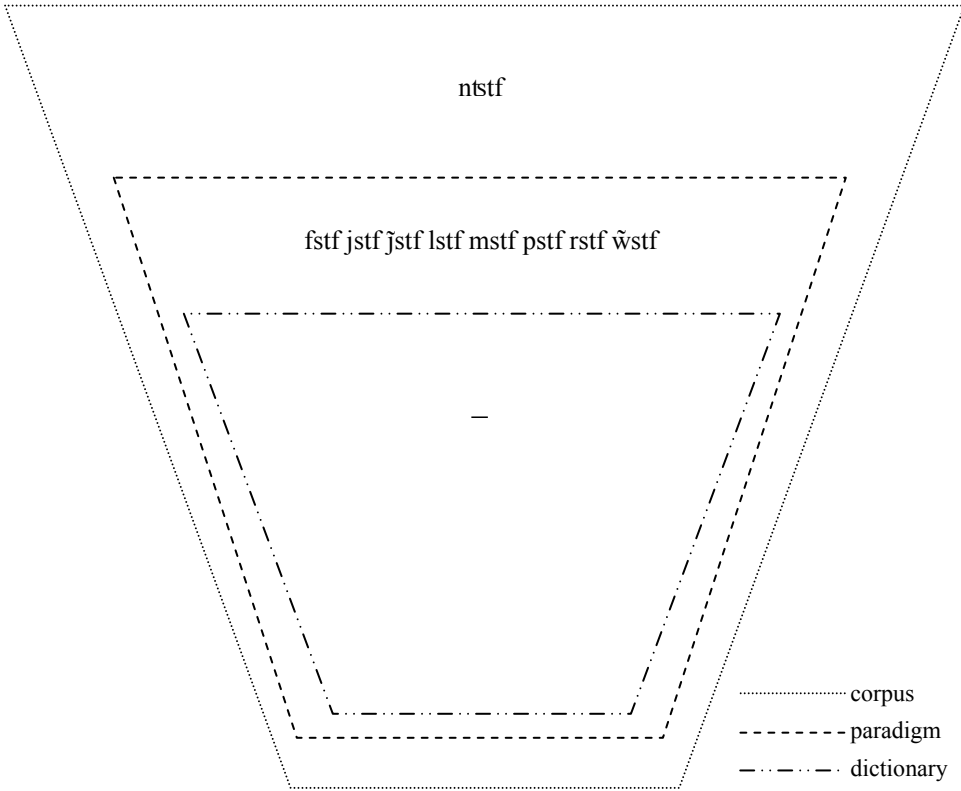


Figure 17. Final CCCC clusters found in the dictionary, paradigm and corpus.

Our data also includes one example of a five-member cluster, ie. /mpstf/ in *przestęstw* ‘crime’ (GEN.PL), which is present in the paradigm and the corpus. The dictionary does not feature any word-final clusters, as these appear in inflected forms, whereas the dictionary features only basic forms.

In Figures 12-17, we show an increase in the number of cluster types as we move from the dictionary, through the paradigm, to the corpus. The lowest-level set is always present in the larger set, which means that each level provides a contribution in terms of cluster types to the other levels. The exceptions to this observation include:

- a) /wz/ is the only initial double cluster present in the paradigm and absent from the corpus

- b) /fʂʂ/ is the only initial triple cluster present in the paradigm and absent from the corpus
- c) Three double word-final clusters, /ptʂ̩, ʂf, wɲ/, were present in the paradigm and absent from the corpus
- d) The only triple word-final cluster present in the paradigm but absent from the corpus is /ntn/, e.g. in *tętn*.

Additionally, as stated earlier, Polish has one 5-element cluster /mpstf/ in *przestępstw* ‘crime’ (GEN.PL), found in the word-final position only in the paradigm and corpus. The dictionary does not feature any 5-member word-final clusters, as these appear in inflected forms, whereas the dictionary lists only basic forms.

#### 4.4. Segment distribution in clusters

Let us now turn our attention to the structure and content of Polish clusters, i.e. the productivity of single consonants in cluster formation, and position occupancy by consonants.

Below we present the frequency of occurrence of 31 Polish consonants in clusters in the dictionary, paradigm and corpus. Table 30 contains the inventory of Polish consonants ordered by their cumulative frequency of occurrence (total) in clusters in each source separately. More detailed information about the occurrence of a cluster in a given word position is provided in columns ‘ini’ (initial), ‘med’ (medial), and ‘fin’ (final). The numbers in the table refer to the occurrence of a given consonant in words featuring this consonant in a cluster. To illustrate with an example, /n/ occurs in an initial cluster in 36 words in the dictionary, 898 in the paradigm and 1 561 in the corpus. It occupies the first position in all the sources in terms of its cumulative frequency in clusters. However, its participation in cluster formation varies across position, with the medial position prevailing.

Table 30. The participation of consonants in cluster formation in Polish.

| dictionary |       |     |      |     | paradigm |       |       |       |      | corpus |        |       |        |      |
|------------|-------|-----|------|-----|----------|-------|-------|-------|------|--------|--------|-------|--------|------|
| C          | total | ini | med  | fin | C        | total | ini   | med   | fin  | C      | total  | ini   | med    | fin  |
| n          | 1752  | 36  | 1630 | 86  | n        | 39601 | 898   | 37475 | 1228 | n      | 109344 | 1561  | 104034 | 3749 |
| t          | 1747  | 332 | 1277 | 138 | j        | 36713 | 5940  | 30734 | 39   | r      | 98593  | 23661 | 74258  | 674  |
| r          | 1674  | 587 | 1034 | 53  | p        | 36351 | 21634 | 14651 | 66   | t      | 97421  | 12256 | 83756  | 1409 |
| s          | 1582  | 551 | 972  | 59  | r        | 34612 | 11242 | 23216 | 154  | s      | 97364  | 21334 | 75235  | 795  |
| k          | 1518  | 292 | 1164 | 62  | s        | 33868 | 12960 | 20747 | 161  | k      | 93603  | 12073 | 80753  | 777  |
| j          | 1324  | 346 | 962  | 16  | t        | 33437 | 6610  | 26561 | 266  | j      | 92226  | 13363 | 78782  | 81   |
| p          | 1320  | 828 | 476  | 16  | m        | 26586 | 3708  | 22795 | 83   | p      | 67960  | 35312 | 32464  | 184  |
| ɲ          | 942   | 35  | 735  | 172 | ɛ        | 26455 | 2281  | 23967 | 207  | ts̄    | 46759  | 454   | 43272  | 3033 |
| ʂ          | 830   | 509 | 298  | 23  | k        | 25902 | 6040  | 19757 | 105  | ʂ      | 45338  | 23320 | 21858  | 160  |
| l          | 717   | 159 | 548  | 10  | ʂ        | 25582 | 14939 | 10575 | 68   | l      | 43382  | 6389  | 36798  | 195  |
| v          | 702   | 231 | 471  | 0   | w        | 21788 | 4962  | 16662 | 164  | m      | 37436  | 5582  | 31200  | 654  |
| d          | 602   | 94  | 508  | 0   | ts̄      | 19684 | 1005  | 18078 | 601  | v      | 37225  | 10121 | 27102  | 2    |
| w          | 598   | 249 | 340  | 9   | v        | 17980 | 5501  | 12479 | 0    | d      | 36215  | 6031  | 30090  | 94   |
| m          | 588   | 150 | 401  | 37  | ɲ        | 17461 | 1363  | 15678 | 420  | f      | 34810  | 9584  | 25022  | 204  |
| ts̄        | 572   | 46  | 518  | 8   | z        | 17300 | 9168  | 8093  | 39   | ɲ      | 34809  | 1540  | 32685  | 584  |
| f          | 530   | 213 | 312  | 5   | ts̄      | 16965 | 151   | 15708 | 1106 | w      | 31425  | 7410  | 23561  | 454  |
| ɛ          | 502   | 129 | 220  | 153 | b        | 16632 | 2711  | 13919 | 2    | z      | 28071  | 12213 | 15327  | 531  |
| z          | 500   | 254 | 226  | 20  | d        | 14188 | 2093  | 12052 | 43   | ts̄    | 27515  | 1984  | 25432  | 99   |
| ts̄        | 493   | 33  | 140  | 320 | l        | 13358 | 2957  | 10362 | 39   | b      | 25822  | 5801  | 20005  | 16   |
| g          | 419   | 155 | 264  | 0   | f        | 11956 | 4603  | 7282  | 71   | g      | 24485  | 5649  | 18781  | 55   |
| b          | 412   | 131 | 280  | 1   | g        | 10464 | 3238  | 7209  | 17   | ɛ      | 23286  | 3343  | 17897  | 2046 |
| ts̄        | 316   | 13  | 296  | 7   | ts̄      | 9940  | 662   | 9217  | 61   | ts̄    | 16727  | 1134  | 13015  | 2578 |
| ẓ         | 215   | 45  | 170  | 0   | ẓ       | 7935  | 807   | 7128  | 0    | ɲ      | 11095  | 0     | 10824  | 271  |
| ɲ          | 175   | 0   | 158  | 17  | x        | 3998  | 1451  | 2536  | 11   | ẓ     | 10290  | 1742  | 8548   | 0    |
| x          | 134   | 54  | 77   | 3   | ɲ        | 3626  | 0     | 3599  | 27   | x      | 8071   | 2347  | 5678   | 46   |
| ŵ          | 96    | 0   | 90   | 6   | ŵ        | 2319  | 0     | 2294  | 25   | ʃ      | 6642   | 0     | 6587   | 55   |
| dz̄        | 75    | 15  | 60   | 0   | dz̄      | 2270  | 254   | 2016  | 0    | ŵ      | 4341   | 0     | 4289   | 52   |
| ʃ          | 59    | 0   | 48   | 11  | dz̄      | 1889  | 120   | 1769  | 0    | dz̄    | 3236   | 398   | 2837   | 1    |
| z          | 55    | 8   | 47   | 0   | z        | 1202  | 110   | 1091  | 1    | dz̄    | 2451   | 153   | 2298   | 0    |
| dz̄        | 35    | 8   | 27   | 0   | ʃ        | 791   | 0     | 759   | 32   | z      | 2108   | 146   | 1958   | 4    |
| dẓ        | 11    | 0   | 11   | 0   | dẓ      | 485   | 18    | 467   | 0    | dẓ    | 588    | 28    | 560    | 0    |

The data presented in the table are ordered according to frequency of occurrence of a consonant in clusters of all positions. Within the first top 10 consonants making up clusters, 8 appear in all the three sources /n, t, r, s, k, j, p, ʃ/, whereas the remaining three positions are occupied by different consonants. Actually, the dictionary and the corpus are compatible in terms of 9 consonants (/n, t, r, s, k, j, p, ʃ, l/), differing only in the presence of /ɲ/ in the dictionary and /tʂ/ in the corpus. The consonant distribution in the paradigm is slightly different: the segments /m/ and /ɕ/, which did not occur among the most productive consonants in the dictionary and the corpus, occupy the 7th and 8th position, respectively.

Let us move on to the bottom of the frequency list, the 9 consonants participating in cluster formation the least in all the three sources are /d͡z, d͡z̄, z, ʃ, d͡z̄, w̃, x, ɲ, z̄/. The only consonant differs is /tʂ/ present in the dictionary, /tʂ̣/ present in the paradigm, and /tʂ̣/ present in the corpus.

If one considers the properties of the least frequent consonants present in all the sources, one may conclude that they are not involved in cluster formation for two reasons, namely, their limited distribution and inherent properties. Limited distribution concerns /d͡z, d͡z̄, z, ʃ, d͡z̄, w̃, ɲ, z̄/ in the following way:

- a) voiced obstruents do not appear as the last element of a word-final cluster due to the process of obstruent final devoicing,
- b) the emergence of the velar nasal /ɲ/ is conditioned by the direct right-hand neighbourhood of /k/ and /g/ in the medial position, and /k/ in the final position, while word-initially /ɲ/ does not appear even as a singleton, let alone a cluster,
- c) the soft resonance /ʃ/ occurs as a realisation of the so called nasal vowels <a, ɛ> before alveolo-palatal fricatives /ɕ, z/, e.g. *wąsik* ‘moustache’ (DIM), and of <ń> before fricatives, e.g. *koński* ‘horse’ (ADJ). As a result, /ʃ/ can neither participate in the formation of word-initial clusters nor serve as a final element of a word-final cluster. Similar observations hold true for the hard resonance /w̃/, which occurs before hard fricatives, e.g. *wąs* ‘moustache’,
- d) /d͡z, d͡z̄, d͡z̄/ are affricates and their status in phonological theories is dubious. Moreover, there is a dispreference for the participation of affricates in clusters.

Another important observation on segment distribution in clusters regards positional occupancy of segments in a cluster. Below, we show an implicational hierarchy for consonants' participation in a cluster.

Table 31. Positional order of consonants in Polish clusters.

| C   | i1 | i2 | i3 | i4 | m1 | m2 | m3 | m4 | m5 | m6 | f1 | f2 | f3 | f4 | f5 |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| j   |    | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  |    |    |    |    |
| p   | *  | *  | *  |    | *  | *  | *  | *  | *  | *  | *  | *  | *  |    |    |
| r   | *  | *  | *  | *  | *  | *  | *  | *  | *  |    | *  | *  | *  |    |    |
| s   | *  | *  | *  |    | *  | *  | *  | *  | *  |    | *  | *  | *  |    |    |
| t   | *  | *  | *  |    | *  | *  | *  | *  | *  |    | *  | *  | *  | *  |    |
| k   | *  | *  | *  |    | *  | *  | *  | *  | *  |    | *  | *  | *  |    |    |
| m   | *  | *  | *  |    | *  | *  | *  | *  |    |    | *  | *  | *  |    |    |
| ʃ   | *  | *  | *  | *  | *  | *  | *  | *  | *  |    | *  | *  | *  |    |    |
| n   |    | *  | *  | *  | *  | *  | *  | *  |    |    | *  | *  |    |    |    |
| e   | *  | *  | *  |    | *  | *  | *  | *  | *  |    | *  | *  | *  |    |    |
| w   | *  | *  | *  | *  | *  | *  | *  | *  | *  |    | *  | *  | *  |    |    |
| Ń   |    |    |    |    | *  |    |    |    |    |    | *  |    |    |    |    |
| tc  | *  | *  | *  |    | *  | *  | *  | *  |    |    |    | *  | *  |    |    |
| v   | *  | *  | *  |    | *  | *  | *  | *  | *  |    | *  |    |    |    |    |
| ɲ   |    | *  | *  | *  | *  | *  | *  | *  | *  |    | *  | *  |    |    |    |
| z   | *  | *  |    |    | *  | *  | *  | *  |    |    | *  | *  |    |    |    |
| b   | *  | *  | *  |    | *  | *  | *  | *  |    |    | *  | *  |    |    |    |
| ts  | *  | *  |    |    | *  | *  | *  | *  |    |    | *  | *  | *  |    |    |
| l   | *  | *  | *  | *  | *  | *  | *  | *  | *  |    | *  | *  | *  |    |    |
| d   | *  | *  | *  |    | *  | *  | *  | *  |    |    | *  | *  |    |    |    |
| f   | *  | *  | *  | *  | *  | *  | *  | *  | *  |    | *  | *  | *  | *  | *  |
| g   | *  | *  | *  |    | *  | *  | *  | *  |    |    | *  | *  |    |    |    |
| tʃ  | *  | *  | *  |    | *  | *  | *  | *  |    |    | *  | *  | *  |    |    |
| ʒ   | *  | *  | *  |    | *  | *  | *  | *  | *  |    |    |    |    |    |    |
| x   | *  | *  | *  |    | *  | *  | *  | *  |    |    | *  | *  |    |    |    |
| dz  | *  | *  | *  |    | *  | *  | *  | *  |    |    | *  |    |    |    |    |
| dz̄ | *  | *  |    |    | *  | *  | *  |    |    |    |    |    |    |    |    |
| z   | *  | *  | *  |    | *  | *  | *  |    |    |    | *  |    |    |    |    |
| ŋ   |    |    |    |    | *  | *  |    |    |    |    | *  |    |    |    |    |
| ʃ̄  |    |    |    |    | *  | *  |    |    |    |    | *  |    |    |    |    |
| dz̄ | *  | *  |    |    | *  | *  | *  |    |    |    |    |    |    |    |    |

Table 31 presents the inventory of Polish consonants ordered alphabetically for all word positions. For each word position, we distinguish several positions within a cluster, 1-4 for initials, 1-6 for medials and 1-5 for finals. An asterisk marks the presence of a consonant in a given position in a cluster. Except for three segments marked in dark gray in the initial position (cluster position 1), no consonant is recorded in position x if it fails to appear in position x-1. In other words, if a segment is found in po-

sition 3, we can be sure that (with the exception of /j, n, ɲ/ in position 1 in initial clusters) it also appears in positions 2 and 1 in some other clusters. If a segment is not recorded in position 2, it is banned from positions 3 and 4. There are, however, some segments which do not participate in cluster formation in a given word position. For instance, /ŋ, ʃ, ɰ/ do not constitute an element in an initial cluster, while /d̥z, d̥z, d̥z, z/ do not form final clusters. Let us illustrate the positional occupancy with some examples.

Table 32. Positional order of consonants in Polish clusters: Examples of initial clusters.

| C    | i1          | i2         | i3           | i4          |
|------|-------------|------------|--------------|-------------|
| j    | –           | więc       | świata       | stwierdził  |
| p    | przez       | spółki     | współpracy   | –           |
| r    | rdzenia     | proc       | sprawie      | wskroś      |
| s    | spółki      | współpracy | skserowane   | –           |
| t    | trzeba      | które      | kształcenia  | –           |
| k    | które       | skarbu     | wskazuje     | –           |
| m    | między      | zmiany     | brzmi        | –           |
| ɕ    | szczególnie | przez      | sprzedaży    | wstrzymanie |
| n    | –           | znacznie   | wznowienie   | drgnęło     |
| ɛ    | świata      | wśród      | spsiało      | –           |
| w    | łzy         | właśnie    | zwłaszcza    | wzdłuż      |
| ɰ    | –           | –          | –            | –           |
| t̥e  | ćwierci     | wciąż      | plci         | –           |
| v    | więc        | dwa        | drzwi        | –           |
| ɲ    | –           | dni        | zmniejszenie | drgnie      |
| z    | zdaniem     | wzrost     | –            | –           |
| b    | brak        | zbyt       | wzbudza      | –           |
| t̥s  | cia         | chce       | –            | –           |
| l    | lwowie      | dla        | sklepów      | względu     |
| d    | dla         | gdy        | wzdłuż       | –           |
| f    | wczoraj     | świata     | stwierdził   | skrwawione  |
| g    | gdy         | zgodnie    | względu      | –           |
| t̥ɕ  | członków    | wczoraj    | wszczęcia    | –           |
| z̥   | żniwa       | września   | zgrzewalna   | –           |
| x    | chce        | wchodzi    | wschodniej   | –           |
| d̥z  | dźwięku     | gdzie      | brdzie       | –           |
| d̥z  | dzwoni      | rdzenia    | –            | –           |
| z    | źródło      | wziąć      | zbzikowani   | –           |
| ŋ    | –           | –          | –            | –           |
| ʃ    | –           | –          | –            | –           |
| d̥z̥ | dźdzownic   | dźdzownic  | –            | –           |

Table 33. Positional order of consonants in Polish clusters: Examples of medial clusters.

| C   | m1        | m2           | m3                       | m4                            | m5                           | m6                         |
|-----|-----------|--------------|--------------------------|-------------------------------|------------------------------|----------------------------|
| j   | miejsce   | powiedział   | akcji                    | wszystkie                     | austria                      | przestępstwie              |
| p   | dopiero   | gospodarki   | wątpliwości              | najwspanialszych              | wewnątrz-<br>spółdzielczego  | wewnątrz-<br>wspólnotowego |
| r   | bardzo    | ochrony      | centrum                  | austria                       | armstrong                    | –                          |
| s   | ustawy    | polsce       | przestępstwa             | wewnątrz-<br>spółdzielczego   | wewnątrz-<br>wspólnotowego   | –                          |
| t   | podczas   | ustawy       | państwa                  | przestępstwa                  | wewnątrz-<br>stanowym        | –                          |
| k   | akcji     | tylko        | polski                   | dębski                        | wewnątrz-<br>szkolnym        | –                          |
| m   | natomiast | firmy        | transmisji               | rozbrzmiewa                   | –                            | –                          |
| ʃ   | jeszcze   | warszawie    | pierwszy                 | rozstrzygnięcia               | kontrprzykład                | –                          |
| n   | będą      | jednak       | miesięczne               | wewnętrznych                  | –                            | –                          |
| ɕ   | jeśli     | części       | pierwsi                  | eksksięży                     | –                            | –                          |
| w   | spółki    | według       | obsługi                  | bezwzględnie                  | wewnątrz-<br>właścicielskich | –                          |
| ɨ̃  | związku   | –            | –                        | –                             | –                            | –                          |
| ʧe  | dziećmi   | wartości     | części                   | rozwścieczony                 | –                            | –                          |
| v   | również   | rozwoju      | nadzwyczajne             | kontrwywiadu                  | wewnątrz-<br>związkowe       | –                          |
| ɲ   | będzie    | również      | najmniej                 | wewnętrznie                   | kontr-<br>wniosków           | –                          |
| z   | decyzji   | związane     | bezwzględnie             | kontrze                       | –                            | –                          |
| b   | sobie     | albo         | mógłby                   | augsburskiego                 | –                            | –                          |
| ʦ   | obecnie   | akcji        | polsce                   | mędrców                       | –                            | –                          |
| l   | tylko     | jeśli        | handlu                   | wątpliwości                   | bezwzględnie                 | –                          |
| d   | jednak    | będą         | najzdolniejszych         | folksdojcz                    | –                            | –                          |
| f   | zawsze    | pierwszy     | województwa              | państwa                       | przestępstwa                 | –                          |
| g   | także     | ciągu        | uwzględnieniem           | bezwzględnie                  | –                            | –                          |
| ʦ̣  | stycznia  | jeszcze      | mężczyzn                 | chełmszczyźnie                | –                            | –                          |
| ʐ   | można     | także        | andrzej                  | wewnątrz-<br>żydowską         | wewnątrz-<br>brzusznym       | –                          |
| x   | ochrony   | powierzchni  | środkowo-<br>wschodniej  | wewnątrzchrześ-<br>cijańskich | –                            | –                          |
| ɕz  | choćby    | będzie       | dziewięć-<br>dziesiątych | wewnątrzdziel-<br>nicowe      | –                            | –                          |
| ɕẓ | powiedzmy | bardzo       | odrdzewiania             | –                             | –                            | –                          |
| z   | później   | więzienia    | przedsięwzięcia          | –                             | –                            | –                          |
| ɲ   | rynku     | garnki       | –                        | –                             | –                            | –                          |
| ʃ   | państwa   | bahrajńskiej | –                        | –                             | –                            | –                          |
| ɕẓ | liczba    | przyjeżdża   | eksdzokeja               | –                             | –                            | –                          |

Table 34. Positional order of consonants in Polish clusters: Examples of final clusters.

| ph | f1      | f2    | f3              | f4          | f5 |
|----|---------|-------|-----------------|-------------|----|
| j  | miejsce | –     | –               | –           | –  |
| p  | wieprz  | skarb | wątp            | –           | –  |
| r  | firm    | piotr | rejestr         | –           | –  |
| s  | jest    | kurs  | przestępstw     | –           | –  |
| t  | piotr   | jest  | przedsiębiorstw | przestępstw | –  |
| k  | projekt | bank  | gdańsk          | –           | –  |

|     |            |             |            |                 |             |
|-----|------------|-------------|------------|-----------------|-------------|
| m   | dostęp     | firm        | astm       | –               | –           |
| ʃ   | koszt      | wciąż       | mistrz     | –               | –           |
| n   | więc       | mężczyzn    | –          | –               | –           |
| ɛ   | wartość    | część       | skądś      | –               | –           |
| w   | aut        | mógł        | pękl       | –               | –           |
| ŵ   | wciąż      | –           | –          | –               | –           |
| te  | –          | wartość     | część      | –               | –           |
| v   | manewr     | –           | –          | –               | –           |
| ɲ   | pięć       | kopalń      | –          | –               | –           |
| z   | mężczyzn   | zamarzł     | –          | –               | –           |
| b   | dóbr       | ansambl     | –          | –               | –           |
| ts  | województw | więc        | miejsc     | –               | –           |
| l   | film       | myśl        | ansambl    | –               | –           |
| d   | spadł      | zwiądl      | –          | –               | –           |
| f   | luft       | najpierw    | mistrzostw | przedsiębiorstw | przestępstw |
| g   | mógł       | poprzysiągl | –          | –               | –           |
| ts̄ | liczb      | wręcz       | barszcz    | –               | –           |
| z̄  | –          | –           | –          | –               | –           |
| x   | jakichś    | zmierzch    | –          | –               | –           |
| dz̄ | wiedźm     | –           | –          | –               | –           |
| dz̄ | –          | –           | –          | –               | –           |
| z   | przyjaźń   | –           | –          | –               | –           |
| ɲ   | bank       | –           | –          | –               | –           |
| ʃ   | część      | –           | –          | –               | –           |
| dz̄ | –          | –           | –          | –               | –           |

#### 4.5. Segmental composition of Polish words

The following subsection aims at discussing phonotactic aspects of the word structure in Polish. For this purpose, all the words from the paradigm were reanalysed in terms of their constituent segments. Vowels and consonants were marked as V and C respectively. On the basis of 194 014 words, we found 1 572 word patterns of various degrees of complexity such as CVC *lot* ‘flight’ and CCVCCCVCV in *błędny* ‘erroneous’ (ADJ.PL).

We are interested in examining the following:

- the proportion of consonants to vowels in words,
- the presence of consonant sequences of various lengths in words,
- the proportion of patterns devoid of consonant clusters to patterns containing a cluster (i.e. at least one CC).

One of the intriguing questions which emerges in the course of our considerations is the following: What is the skeletal structure of Polish? By the skeletal structure we mean the alternation of consonants and vowels in

a word. Phonological accounts point to the consonantal character of the Polish segmental inventory. With 31 consonants, and only 6 vowels (Jassem 2003), the consonant/vowel ratio is very high (Maddieson 2008). According to our estimate the ratio amounts to 1.46 in favour of consonants. This fact suggests cumulation of consonants, and for this reason, emergence of clusters in a word. The results of this investigation are tabulated in 35 below, where patterns stand for the sequences of vowels and consonants in all words, and examples refer to the number of lexical items with these patterns. For instance, the CCVCC pattern is represented by *skarb* /skarp/ ‘treasure’, *kraść* /kract̩/ ‘to steal’, *trucht* /truxt/ ‘jog trot’.

Table 35. Skeletal structure of Polish words.

|            | word patterns |       | word examples |       |
|------------|---------------|-------|---------------|-------|
|            | number        | %     | number        | %     |
| no cluster | 69            | 4.39  | 29 548        | 15.23 |
| cluster    | 1 503         | 95.61 | 164 466       | 84.77 |

As shown in Table 35, over 95% of all the examined word patterns feature at least one cluster of two or more consonants. Almost 85% of the inflectional forms contain at least one cluster. Patterns vary in length from one vowel in, e.g. *o* ‘about’, to more than 20 vowels and consonants (see Table 36 below).

Table 36. Patterns of Polish words.

| pattern length | number of word patterns | number of words |
|----------------|-------------------------|-----------------|
| 1              | 2                       | 9               |
| 2              | 2                       | 70              |
| 3              | 4                       | 733             |
| 4              | 10                      | 2 923           |
| 5              | 18                      | 8 135           |
| 6              | 31                      | 15 064          |
| 7              | 43                      | 22 518          |
| 8              | 70                      | 28 614          |
| 9              | 102                     | 31 065          |
| 10             | 132                     | 27 686          |
| 11             | 160                     | 21 725          |
| 12             | 198                     | 14 978          |

|       |       |         |
|-------|-------|---------|
| 13    | 209   | 9 793   |
| 14    | 210   | 6 000   |
| 15    | 166   | 2 807   |
| 16    | 109   | 1 220   |
| 17    | 65    | 465     |
| 18    | 28    | 156     |
| 19    | 8     | 32      |
| 20    | 3     | 12      |
| 21    | 2     | 8       |
| 22    | 1     | 3       |
| total | 1 573 | 194 016 |

The longest word found in the corpus was composed of 22 segments, VCCVCCVCCVCCVCCVCCVCCV in *internacjonalistycznymi* ‘internationalistic’ (INSTR.PL). However, a number of long words also had a simple CV structure, among which the longest pattern was composed of 15 alternating consonants and vowels (CVCVCCVCCVCCVCCV) and was represented by 10 word forms:

- *realizowałabym* ‘I would realise’, *realizowałabyś* ‘you would realise’
- *wylądowywałabym* ‘I would disembark’, *wylądowywałabyś* ‘you would disembark’,
- *zadecydowałabym* ‘I would decide’, *zadecydowałabyś* ‘you would decide’,
- *załadowywałabym* ‘I would embark’, *załadowywałabyś* ‘you would embark’,
- *zatarasowałabym* ‘I would encumber’, *zatarasowałabyś* ‘you would encumber’.

The length of patterns which occur most frequently in the paradigm resource is medium, with the average of 12.3 segments, which is shown in Figure 18. The *x* axis shows the number of segments in a pattern (length), whereas the *y* axis refers to the number of patterns of a particular length. The average length of all the inflectional forms in the paradigm is 9.23.

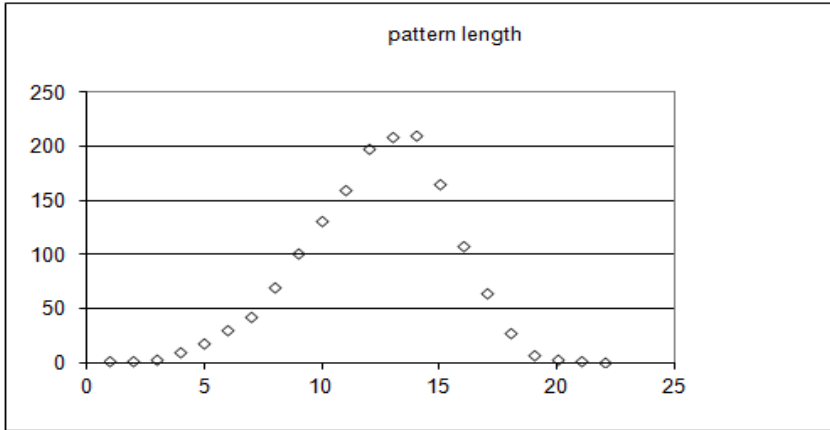


Figure 18. Frequency (type) distribution of Polish word patterns.

**Mean length of clusters.** This section is devoted to the analysis of the 190 000 inflected forms of a 8 000 dictionary list, transcribed in IPA and converted to approximately 1 600 CV patterns. Some patterns are very common, e.g. 4 700 words have this cluster-free pattern CVCVCVCV as in *bałaganu* ‘mess’ (GEN.SG), *cebulam* ‘onion’ (GEN.SG), *czokolada* ‘chocolate’, *dziekanatu* ‘dean’s office’ (GEN.SG), *karuzela* ‘merry-go-round’, *lakomego* ‘greedy’ (GEN.SG), *rzucalaby* ‘she would throw’, *wypasiona* ‘fancy’, *zobaczymy* ‘we will see’. The least common pattern is CVVCCC in *naostrz* ‘sharpen’ (IMP.SG). When the patterns are sorted according to frequency (descending) and their lengths are displayed as dots on a 1-25 scale, we obtain the following image.

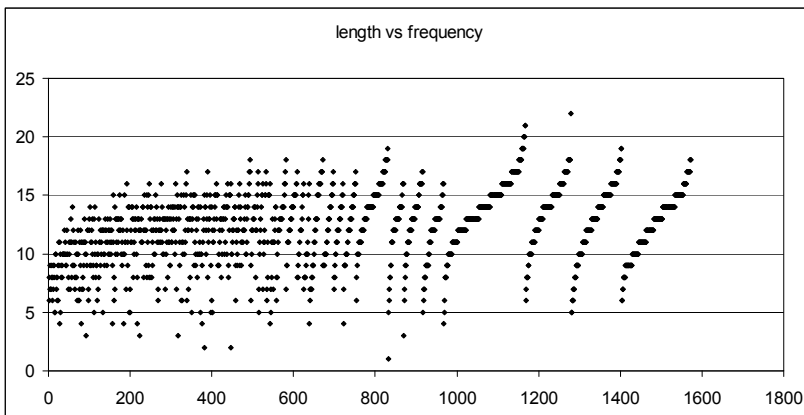


Figure 19. Frequency of CV patterns in Polish.

The 0–25  $y$  axis represents the length of the pattern and the 0–1800  $x$  axis displays the frequency-based rank of a pattern. The pattern with length 1, just right of the 800 mark, is a V, e.g. the conjunction *i* ‘and’, the preposition *o* ‘about’ etc., and the one at 1 300 is a C, e.g. the prepositions *w* ‘in’ and *z* ‘with’. As can be seen, short patterns (fewer than 5 segments) are virtually absent right of the 900 mark, i.e. they are present solely among the more frequent ones. The first 200 most frequent patterns are characteristically short as they do not exceed 15 segments in length. There are only two patterns that are longer than 20 phonemes, which are CCVCVCCVCVCVCVCVCVCCV (21) in *scharakteryzowatybyśmy* ‘we would characterise’ and VCCVCCVCCVCVCVCVCVCCV (22) in *internacjonalistycznymi* ‘internationalist’ (PL.INST), found in position 1 168 and 1 280 respectively. The mean pattern length is 11.44 segments in the upper half and 13.17 in the lower half.

#### 4.6. Summary

The goal of this chapter was to provide a quantitative account of Polish phonotactics. We presented consonant clusters in all word positions (initial, medial, final) in the three sources (the dictionary, the paradigm and the corpus). We performed calculations over cluster and word types in the dictionary and paradigm, as well as cluster and word types as and word tokens in the corpus. The importance of the varied sources was illustrated in Figures 12-17, which show that although the majority of clusters is shared by the three resources, a larger database guarantees a richer output. Additionally, we analysed the participation of individual segments in the formation of clusters and segment distribution within clusters. Finally, the analysis of Polish word patterns revealed that almost all word patterns and a vast majority of words representing them contain at least one cluster. This exhaustive description provides a representative database for testing three hypotheses which constitute the core of our research.

## Chapter 5

# Polish and English (mor)phonotactics: An empirical study

The objective of this chapter is to analyse the phonotactic inventory of Polish and English according to three variables; morphological composition of clusters, their preferability and frequency. The cluster set extracted from the three sources (dictionary, paradigm, corpus) described in detail in Chapter 3 will serve as the input for testing our hypotheses. In measuring cluster preferability, we follow the NAD model presented in Chapter 1. With a view to providing an exhaustive morpho-phonological description of Polish and English consonant clusters, three hypotheses have been put forward.

### 5.1. Hypotheses

The first hypothesis concerns the relation between cluster size and morphological complexity. Our supposition regarding the interdependence of the two factors is based on a phonological universal stating that CV is a preferred syllable template in the languages of the world (Greenberg 1978, Maddieson 2008). In consequence, the optimal word structure consists of alternating consonants and vowels. In languages such as Polish and English, which permit CC(C...), it is assumed that there must be a force sustaining such sequences. One of the reasons why complex clusters emerge is attributed to the influence of morphology. Therefore, we expect longer clusters to imply the presence of a morphological boundary, and shorter clusters to be morphologically simple.

Hypothesis 2 assumes that the degree of phonological preferability is inversely proportional to morphological complexity. Morphotactic clusters are expected to have a lower degree of preferability than phonotactic ones. This assumption stems from the semiotic precedence/superiority of morphology over phonology: the morphological function may override the phonological one. In phonotactics, signalling a morphological boundary may result in obeying phonological preferability of a cluster.

Hypothesis 3 states that the degree of cluster preferability is directly proportional to frequency. Preferred clusters are expected to be more frequent than dispreferred clusters in both languages.

## 5.2. Polish

As demonstrated in Chapter 3, establishing morphological boundaries turned out to be one of the most challenging tasks. Since the methods of morphological parsing undertaken by the research team did not allow for a consistent, unanimous, individual treatment of words of varying morphological transparency, we arrived at a set of rules which were applied consistently to word-initial and word-final contexts. The subsequent section (5.2.1.) presents the methodology of morphological parsing.

### 5.2.1. Principles of morphological parsing

**Word-initial context.** Determining the morphological composition of clusters necessitated making decisions on (1) morphemic boundaries, and (2) the treatment of morphophonological phenomena emerging at morpheme juncture. Polish has non-syllabic consonantal prefixes, {s-}, {z-}, {w-}, {ws-}, {wz-} (a subset of which is bi-consonantal), which when attached to a consonant-initial word stem, generate the whole range of morphologically complex initial clusters. Examples of such clusters are numerous: *s+chodźić* ‘to go down’, *z+robić* ‘to do’ (PERF), *w+robić* ‘to set sb up’, *ws+chodźić* ‘to rise’, *wz+braniać* (*się*) ‘to refrain’. In the context before  $\widehat{t\epsilon}$ , the prefix {s-} is realized as an alveolo-palatal fricative /ç/, as in *ś+ciać* ‘to cut down’, *ś+ciągnąć* ‘to take down’, *ś+ciskać* ‘to squeeze’. However, one should bear in mind that /s/ can also function as a non-morphemic segment and may lead to the coincidence in semantically and etymologically unrelated words as in *kok – skok* ‘hair bun – jump’, *lew – zlew* ‘lion – sink’, *wędzić – swędzić* ‘to smoke – to itch’. Additionally, the suffix {-nąć}, added to non-syllabic stems, forms morphonotactic clusters in words such as *pch+nąć* ‘to push’ (PERF) ← *pch+ać* ‘to push’ (IMPERF) and *drg+nąć* ‘to vibrate’ (PERF) ← *drg+ać* ‘to vibrate’ (IMPERF). Below we present examples of words for all clusters with **clear parsing** (henceforth CP).

Table 37. Examples of words with the clear morphological parsing rule.

|    | cluster | example                       |    | cluster | example                          |
|----|---------|-------------------------------|----|---------|----------------------------------|
| 1  | fe      | w+siadać 'to get in'          | 37 | spw     | s+plukać 'to wash away'          |
| 2  | fp      | w+padać 'to fall into'        | 38 | ss      | z+sunać 'to slide down'          |
| 3  | fpr     | w+prowadzić 'to move in'      | 39 | sş      | z+szywać 'to sew'                |
| 4  | fpw     | w+płacić 'to pay in'          | 40 | st      | s+tajać 'to melt'                |
| 5  | fs      | w+sypać 'to pour'             | 41 | stf     | s+tworzyć 'to create'            |
| 6  | fsk     | w+skoczyć 'to jump into'      | 42 | stfj    | s+twierdzić 'to state'           |
| 7  | fsp     | ws+pomnieć 'to mention'       | 43 | str     | s+tracić 'to lose'               |
| 8  | fst     | w+stać 'to stand up'          | 44 | stw     | s+tłuc 'to break'                |
| 9  | fstş    | ws+trząs 'shock'              | 45 | sx      | s+chodźić 'to go down'           |
| 10 | fşx     | ws+chodźić 'to rise'          | 46 | sxr     | s+chronić 'to shelter'           |
| 11 | fte     | w+ciągnąć 'to drag in'        | 47 | zb      | z+badać 'to examine'             |
| 12 | fx      | w+chodzić 'to enter'          | 48 | zbj     | z+bieg 'fugitive' (N)            |
| 13 | vj      | w+jechać 'to drive into'      | 49 | zbl     | z+blokować 'to combine'          |
| 14 | vl      | w+lać 'to pour into'          | 50 | zd      | z+denerwować 'to upset'          |
| 15 | vm      | w+mawiać 'to persuade'        | 51 | zdż     | z+dziwić 'to surprise'           |
| 16 | vmj     | w+mieszać 'to stir into'      | 52 | zg      | z+gasić 'to extinguish'          |
| 17 | vr      | w+rażenie 'to impression'     | 53 | zgj     | z+giąć 'to bend'                 |
| 18 | vw      | w+lączyć 'to turn on'         | 54 | zgn     | z+gnić 'to decay'                |
| 19 | vzb     | wz+budzić 'to arouse'         | 55 | zgr     | z+gromadzić 'to gather together' |
| 20 | vzd     | wz+dychać 'to sigh'           | 56 | zgw     | z+głosić 'to propose'            |
| 21 | vzn     | wz+nosić 'raise' (IMPF)       | 57 | zj      | z+jawić 'to appear'              |
| 22 | wzpn    | wz+nieść 'raise' (PERF)       | 58 | zl      | z+lać 'to pour off'              |
| 23 | vzr     | wz+rosnąć 'to grow'           | 59 | zm      | z+marnować 'to waste'            |
| 24 | vz      | w+rzucić 'to throw'           | 60 | zmj     | z+mierzyć 'to measure'           |
| 25 | ete     | ś+ciąć 'to cut down'          | 61 | zmr     | z+mrok 'dusk'                    |
| 26 | sf      | s+falszować 'to falsify'      | 62 | zn      | z+nosić 'to carry off' (IMPF)    |
| 27 | sk      | s+komplicować 'to complicate' | 63 | zpn     | z+nieść 'to carry off' (PERF)    |
| 28 | skj     | s+kierować 'to direct'        | 64 | zr      | z+robić 'to do'                  |
| 29 | skl     | s+kleić 'to glue together'    | 65 | zv      | z+ważyć 'to weigh'               |
| 30 | skr     | s+kraść 'to steal'            | 66 | zvj     | z+wiązać 'to tie'                |
| 31 | skş     | s+krzywić 'to hurt'           | 67 | zvr     | z+wracać 'to return'             |
| 32 | skw     | s+kłamać 'to lie'             | 68 | zw      | z+lamać 'to break'               |
| 33 | sp      | s+pakować 'to pack'           | 69 | zż      | z+rzucić 'to knock down'         |
| 34 | spj     | s+piąć 'to clip together'     | 70 | drgn    | drg+nać 'to vibrate'             |
| 35 | spl     | s+pleść 'to weave'            | 71 | pxn     | pch+nać 'to push'                |
| 36 | spr     | s+prostować 'to put right'    |    |         |                                  |

Some clusters can have their origin in non-concatenative morphology. Thus, consonantal strings in the derivatives *krwawy* ← *krew* ‘bloody’ ← ‘blood’, *krwotok* ← *krew* ‘hemorrhage’ ← ‘blood’, *lniany* ← *len* ‘linen’ (ADJ) ← (N), *lwi* ← *lew* ‘lion’ (ADJ) ← (N), *plciowy* ← *pleć* ‘sexual’ ← ‘sex’ and *psi* ← *pies* ‘dog’ (ADJ) ← (N) were also classified as morphotactic.

The prefixes can also come in contact with stems whose decomposition is possible only after a several-step analysis. Let us now outline the problematic areas and justify the choice of principles used in morphological parsing. Unlike the examples given in Table 37, we found cases such as *s+cierać* ‘to wipe away’, whose prefixless form \**cierać* does not exist in isolation. However, Polish features numerous other words with the stem \**cierać*, preceded by various prefixes, e.g. *wy+cierać* ‘to wipe off’, *na+cierać* ‘to rub’, *prze+cierać* ‘to wipe’, *za+cierać* ‘to erase’ and many others. To account for such examples, we introduced the principle referred to as **active stem** (henceforth AS). The clusters which emerge in this context are morphotactic, as in *s+przątać* ‘to clean’, *wy+przątać* ‘to clean up’, *za+przątać* ‘to absorb, to preoccupy’. Synchronic semantic proximity between the items does not have to be observed. Much as *s+przątać* and *wy+przątać* are related to ‘freeing from dirt or pollution’, *za+przątać* is used in an ‘intellectual’ sense. Similarly, the items such as *s+prawa* ‘issue’ and *w+prawa* ‘skill’ are classified according to the AS principle, as *na+prawa* ‘repair’, *u+prawa* ‘cultivation’, *wy+prawa* ‘expedition’, *za+prawa* ‘mortar’ exist but are devoid of semantic proximity. Should the vocalic alternations occur in the compared words, the morphological analysis of the words is still conducted according to the active stem rule. The word *z+wrot* ‘return’ represents the case of an active stem because of the existence of items such as *po+wrót* ‘return’, *na+wrót* ‘recurrence’ and *za+wrót* ‘vertigo’, in which case the vocalic alternation /o/ – /u/ does not prevent us from identifying the common stem. For more examples of words parsed according to the AS rule, see Table 38.

Table 38. Examples of words parsed according to the active stem rule.

|    | cluster | example                    |    | cluster | example                     |
|----|---------|----------------------------|----|---------|-----------------------------|
| 1  | fkw     | w+kładać 'to put into'     | 23 | spr     | s+pragniony 'thirsty'       |
| 2  | fpr     | w+prawiać 'to insert'      | 24 | spś     | s+przątać 'to clean'        |
| 3  | fpw     | w+płacać 'to pay in'       | 25 | spw     | s+płukiwać 'to flush'       |
| 4  | fsk     | w+skakiwać 'to hop in'     | 26 | stf     | s+tworzyć 'to create'       |
| 5  | fsp     | ws+pominać 'to recall'     | 27 | stfj    | s+twierdzać 'to state'      |
| 6  | fst     | w+stępny 'preliminary'     | 28 | str     | s+trata 'loss'              |
| 7  | fstr    | ws+tręt 'disgust'          | 29 | zb      | z+baczać 'to get off track' |
| 8  | fsx     | ws+chodni 'eastern'        | 30 | zbl     | z+bliżać 'to approach'      |
| 9  | vj      | w+jeźdźać 'to drive in'    | 31 | zd      | z+darzać 'to happen'        |
| 10 | vl      | w+lewać 'to pour in'       | 32 | zg      | z+gadnąć 'to guess'         |
| 11 | vw      | w+łączać 'to switch on'    | 33 | zgw     | z+głaszać 'to report'       |
| 12 | vz      | w+zzywać 'to call'         | 34 | zj      | z+jawiać 'to appear'        |
| 13 | vzb     | wz+budzać 'to engage'      | 35 | zl      | z+lecień 'order'            |
| 14 | vzg     | wz+górze 'hill'            | 36 | zmn     | z+mniejszać 'to reduce'     |
| 15 | vzgl    | wz+gląd 'regard'           | 37 | zn      | z+nikać 'to disappear'      |
| 16 | vzm     | wz+macniać 'to strengthen' | 38 | zr      | z+rywać 'to pick'           |
| 17 | vzr     | wz+rastać 'to grow'        | 39 | zv      | z+walniać 'to slow down'    |
| 18 | ete     | ś+cierać 'to wipe off'     | 40 | zvj     | z+wiedzać 'to visit'        |
| 19 | se      | z+siadły 'sour'            | 41 | zvl     | z+wlekać 'to linger'        |
| 20 | skr     | s+kracać 'to shorten'      | 42 | zvr     | z+wrot 'return'             |
| 21 | skw     | s+kładać 'to fold'         | 43 | zvw     | z+włoka 'delay'             |
| 22 | sp      | s+pełniać 'to fulfil'      |    |         |                             |

Another principle of morphological analysis handles cases where a morphological division in one part of speech was maintained in another. This principle, here referred to as **part-of-speech boundary transfer** (henceforth PoS), explains the existence of a morphological boundary across all parts of speech. We shall illustrate this principle with two examples. The noun *spadochron* 'parachute' (= *preventing from falling*) can be derived from the prefixed verb *spadać* (*s+padać*) 'to fall down'. Therefore, the /sp/ cluster in the first element of the compound, i.e. *spad* 'fall', was analysed as morphonotactic (*s+pad*). Similarly, the adjective *spozstrzegawczy* comes from the verb *spozstrzegać* (*s+postrzegać*) 'to perceive', as a result of which the cluster /sp/ was considered morphologically parsable. More complex examples require a several-step analysis, as shown below:

- a. *s+przedawca* ‘seller’ (PoS) ← *s+przedawać* ‘sell’ (AS) ← *\*przedawać* but see: *wy+przedawać* ‘to sell off’, *za+przedawać* ‘to betray’ (synchronic semantic proximity maintained),
- b. *sprężyna* ‘mainspring’ (PoS) ← *s+prężyć* ‘to compress’ (CP) ← *prężyć* ‘to flex’,
- c. *z+niechęcić* ‘to discourage’ (PoS) ← *\*niechęcić* ← *nie+chęć* ‘reluctance’.

Examples of words for each cluster parsed according to the PoS rule are tabulated below.

Table 39. Examples of words parsed with the part-of-speech-transfer rule.

|    | cluster | example                 |    | cluster | example                   |
|----|---------|-------------------------|----|---------|---------------------------|
| 1  | fsp     | ws+pólnota ‘community’  | 12 | zd      | z+dechnąć ‘die’           |
| 2  | skr     | s+krót ‘abbreviation’   | 13 | zj      | z+jawisko ‘phenomenon’    |
| 3  | skw     | s+kłonny ‘prone’        | 14 | zm      | z+marszczka ‘wrinkle’     |
| 4  | sp      | s+padochron ‘parachute’ | 15 | zn      | z+nów ‘again’             |
| 5  | spr     | s+prężyna ‘spring’      | 16 | zɲ      | z+niechęcać ‘discourage’  |
| 6  | spɕ     | s+przedawca ‘seller’    | 17 | zr      | z+rozpaczony ‘desperate’  |
| 7  | str     | s+treszczać ‘summarise’ | 18 | zv      | z+wolenik ‘supporter’     |
| 8  | sxr     | s+chronisko ‘shelter’   | 19 | zvj     | z+wierzchnik ‘superior’   |
| 9  | zb      | z+będny ‘unnecessary’   | 20 | zvr     | z+wrotnik ‘tropic’        |
| 10 | zj      | z+biornik ‘container’   | 21 | zw      | z+ludzenie ‘illusion’     |
| 11 | zbr     | z+brojny ‘armed’        | 22 | zɹ      | z+rzeczenie ‘association’ |

As in Polish consonantal prepositions coincide with prefixes ({z-} phonetically realized as /s/ or /z/ and {w-} realized as /f/ or /v/), they were also found in fossilized prepositional phrases and complex prepositions (Karpowicz 2009: 99f). These words also underwent morphological parsing and exemplify the principle of **prepositions**. Fossilized *prepositional phrases* containing a morphonotactic cluster and found in the dataset include: *s+kąd* ‘where from’, *s+tamtąd* ‘from there’, *s+tąd* ‘from here’, *w+cale* ‘at all’, *w+ciaż* ‘still’, *w+krótce* ‘soon’, *w+prost* ‘directly’, *w+reszcie* ‘finally’, *w+szerz* ‘across’, *w+śród* ‘among’, *wz+dłuż* ‘along’, *wz+wyż* ‘upwards’, *z+nowu* ‘again’, *z+resztą* ‘besides’. The group of *complex prepositions* comprises *s+pod* ‘from under’, *s+pode* ‘from under’, *s+po+śród* ‘from among’.

Table 40. Examples of words parsed according to the prepositions rule.

|   | cluster         | example            |    | cluster | example               |
|---|-----------------|--------------------|----|---------|-----------------------|
| 1 | f <sup>er</sup> | w+śród ‘among’     | 8  | sp      | s+pośród ‘from among’ |
| 2 | f <sup>kr</sup> | w+krótce ‘soon’    | 9  | st      | s+ąd ‘from here’      |
| 3 | f <sup>pr</sup> | w+prost ‘straight’ | 10 | vr      | w+reszcie ‘at last’   |
| 4 | f <sup>sz</sup> | w+szerz ‘across’   | 11 | vzdw    | wz+dłuż ‘along’       |
| 5 | f <sup>te</sup> | w+ciąż ‘still’     | 12 | vzv     | wz+wyż ‘upwards’      |
| 6 | f <sup>ts</sup> | w+cał ‘at all’     | 13 | zn      | z+nowu ‘again’        |
| 7 | sk              | s+ąd ‘where from’  | 14 | zr      | z+resztą ‘besides’    |

All the remaining words containing clusters which did not fall into any of the aforementioned categories were considered to be lexical. Additionally, clusters affected by morpho-phonological alternations were classified as lexical. This is true for clusters containing the /w – l/ alternation in *zły – źle* /zw – zl/ ‘bad’ – ‘badly’, *mdły – mdlić* /mdw – mdl/ ‘bland’ – ‘to nauseate’, *pchła – pchli* /pxw – pxl/ ‘flee’ (N) – (ADJ), *mgła – mglisty* /mgw – mgl/ ‘fog’ – ‘foggy’ as well as alternations between soft and hard segments in *mścić – mszczyć* /m<sup>te</sup> – m<sup>st</sup>/ ‘to take revenge’ – ‘I take revenge’, *chcieć – chcę* /x<sup>te</sup> – x<sup>t</sup>/ ‘to want’ – ‘I want’, *czcić – czczyć* /t<sup>st</sup> – t<sup>st</sup>/ ‘to worship’ – ‘I worship’, and the reverse *spać – śpię* /sp – ep/ ‘to sleep’ – ‘I sleep’. In our dataset, we encountered examples of suppletion. Words with initial clusters such as /lz/ in *lżej* ← *lekko* ‘more easily’ – ‘easily’, /sf/ in *swój* ‘one’s’, /tf/ in *twój* ‘yours’, and in /vz/ in *wziąć* ← *brać* ‘to take’ (PERF) ← (IMPERF) were considered to have a lexical status. It must be noted that clusters appearing within bound morphemes were regarded as lexical, e.g. /pʂ/ in {przed-} ‘pre-’.

Since the dictionary contains only uninflected forms, additional decisions had to be made for the study of the paradigm and the corpus, which contain non-concanenative morphonotactic clusters. Clusters arising as a result of vowel – zero alternation were considered to be morphonotactic, e.g. *pies – psy* ‘dog’ (NOM.SG) – (NOM.PL), *krew – krwi* ‘blood’ (NOM.SG) – (GEN.SG).

**Word-final context.** As regards the morphological division of clusters in the word-final position, the following decisions were taken. Polish possesses two non-syllabic suffixes, namely the preterit {-ł} and the infinitival {-ć}, which yield morphonotactic clusters. The former generates clusters such as /tw, rw, sw/ in *poszed+ł* ‘he went’, *umar+ł* ‘he died’, *niós+ł* ‘he was carrying’, while the latter triggers consonantal combina-

tions / $\widehat{c\acute{t}e}$ ,  $\widehat{nt\acute{e}}$ ,  $\widehat{j\acute{e}t\acute{e}}$ / as in *nies*+ $\acute{c}$  ‘to carry’, *zaczq*+ $\acute{c}$  ‘to begin’, *przyjs*+ $\acute{c}$  ‘to come’. Non-concatenative morphology in the paradigm and the corpus gives rise to morphonotactic clusters in imperative or genitive plural forms, e.g. / $\widehat{c\acute{t}e}$ / in *puś* $\acute{c}$  ← *puś**ci* $\acute{c}$  ‘to let go’ (IMP) ← ‘to let go’ (INF), /*nt*/ in *kont* ← *konto* ‘account’ (GEN.PL – NOM.SG). The infinitival ending was found in the dictionary, paradigm and corpus, while examples of the remaining inflected forms were found only in the paradigm and the corpus. A comprehensive summary of the aforementioned context is presented in Table 5.

Similarly to the word initial context, clusters appearing in bound morphemes were regarded as lexical, e.g. / $\widehat{c\acute{t}e}$ / in {-oś $\acute{c}$ } [suffix for abstract nouns], and /*nts*/ in {-*q*} [suffix for the present adverbial participle].

Table 41. The list Polish affixes and morphological operations generating morphonotactic clusters.

| pronunciation          | function                 | example                                      |
|------------------------|--------------------------|--|
| s / $\epsilon$         | e.g. aspect<br>semantics | s+kończyć ‘to end’<br>ś+cierać ‘to wipe off’ |
| z                      | e.g. aspect              | z+robić ‘to do’                              |
| f                      | semantics                | w+chodzić ‘to enter’                         |
| v                      | semantics                | w+jechać ‘to drive into’                     |
| $\widehat{t\acute{e}}$ | infinitive               | iś+ $\acute{c}$ ‘to go’                      |
| $\emptyset$            | vowel ~ zero alternation | pies ‘dog’ vs. psy ‘dogs’                    |
| $\emptyset$            | imperative               | puś $\acute{c}$ ‘to let go’ (IMP)            |
| $\emptyset$            | genitive plural          | miejs $\acute{c}$ ‘places’ (GEN.PL)          |

Let us now proceed to the presentation of the data extracted from the 3 different sources: the dictionary (hypotheses 1 and 2), the paradigm and the corpus (hypothesis 3). The initial statistical testing assumed performing Chi-square analyses of the relationship between 1) cluster size and its morphological complexity, 2) the morphological status of a cluster and cluster preferability. Due to insufficient data (numbers below 5), some of the calculations could not be performed. Therefore, in the following discussion, we include the *p*-values only when possible.

### 5.2.2. Results: Hypothesis 1 (dictionary)

Below we present the data coming from the dictionary, with the subdivision into types and words as well and positions within a word.

**Initial cluster types.** In the word-initial position, we found the total of 238 cluster types (see Chapter 4). We grouped the types according to their size (CC, CCC, CCCC) and their morphological composition (characterized by simple or complex morphological structure). The three-way division reflects the fact that we can distinguish clusters which are purely lexical, e.g. /pt/ in *ptak* ‘bird’, /kr/ in *kreda* ‘chalk’, /pʂtʂ/ in *pszczoła* ‘bee’; purely morphonotactic, e.g. /zg/ in *z+gubić* ‘to lose’, /skw/ in *s+kłamać* ‘to lie’, /fsk/ in *w+skoczyć* ‘to jump in’; and a group of types which may take both realizations, e.g. /sk/ in *skok* ‘jump’ vs. *s+kończyć* ‘to finish’, /ɛtɛ/ in *ściana* ‘wall’ vs. *ś+cierać* ‘to wipe away’, /fst/ in *wstążka* ‘ribbon’ vs. *w+stać* ‘to get up’. Table 42 and Figure 20 summarize the results. The values obtained were calculated for the total of clusters occurring in each size group.

Table 42. Hypothesis 1: Size vs. morphological composition (initial cluster types).

|      | lexical |    | morphonotactic |    | both |    | total<br>no. |
|------|---------|----|----------------|----|------|----|--------------|
|      | no.     | %  | no.            | %  | no.  | %  |              |
| CC   | 106     | 73 | 21             | 14 | 18   | 12 | 145          |
| CCC  | 36      | 42 | 35             | 41 | 14   | 16 | 85           |
| CCCC | 2       | 25 | 6              | 75 | 0    | 0  | 8            |

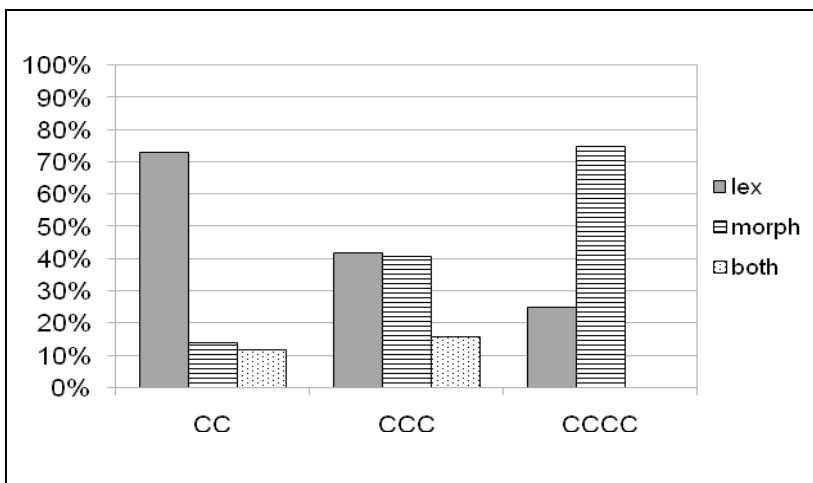


Figure 20. Hypothesis 1: Size vs. morphological composition (initial cluster types).

**Word types with initial clusters.** As regards word types with initial clusters, we analysed 2532 words which represented the 238 types discussed earlier. Table 43 shows the total number of words found in each category. This data does not contain the group referred to as ‘both’ as one word can be analysed either as lexical or morphological.

Table 43. The total number of word types with phonotactic and morphonotactic initial clusters.

|             | lexical | mophonotactic | total |
|-------------|---------|---------------|-------|
| words (no.) | 2 076   | 456           | 2 532 |
| words (%)   | 82%     | 18%           | 100%  |

In total, 82% of words with word-initial clusters feature morphologically simplex clusters, whereas 18% of words contain morphologically complex clusters. Tables 44, 45, 46 summarize the data for each cluster length separately: CC, CCC, and CCCC.

Table 44. The number of word types with phonotactic and morphonotactic initial CC clusters.

|             | lexical | mophonotactic | total |
|-------------|---------|---------------|-------|
| words (no.) | 1 884   | 253           | 2 137 |
| words (%)   | 88%     | 12%           | 100%  |

Table 45. The number of word types with phonotactic and morphonotactic initial CCC clusters.

|             | lexical | mophonotactic | total |
|-------------|---------|---------------|-------|
| words (no.) | 189     | 192           | 381   |
| words (%)   | 50%     | 50%           | 100%  |

Table 46. The number of word types with phonotactic and morphonotactic initial CCCC clusters.

|             | lexical | mophonotactic | total |
|-------------|---------|---------------|-------|
| words (no.) | 3       | 11            | 14    |
| words (%)   | 21%     | 79%           | 100%  |

Figure 21 provides an overview of Tables 44, 45, 46 with the percentage values calculated for the total of each size group separately, i.e. CC, CCC and CCCC.

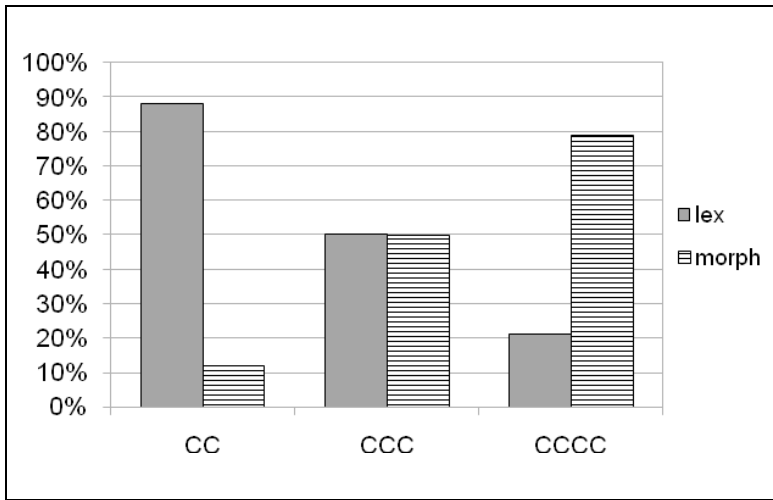


Figure 21. Hypothesis 1: Size vs. morphological composition (word types with initial clusters).

For the word-initial context, we conclude that hypothesis 1 is corroborated. The number of morphonotactic clusters increases alongside cluster length. This tendency is observed for both cluster types ( $p = 0$ ) and word types ( $p = 0$ ).

**Final cluster types.** Similarly in final cluster types, the probability of the emergence of a morphological boundary in a cluster increases along with the number of consonants in a cluster. However, one has to be cautious when formulating generalizations on the results for the final cluster types. As can be seen in Table 47 and Figure 22, the total number of CCC types in question is limited to 10 items only. There are 8 lexical cluster types, 1 morphonotactic type and 1 cluster subsumed under the category ‘both’. The lexical clusters include /kst, ltr, ntʂ, ŋkt, rɛtɛ, rʂtʂ, str, stʂ/ and are found in words such as *tekst* ‘text’, *filtr* ‘filter’, *wewnątrz* ‘inside’, *instynkt* ‘instinct’, *garść* ‘handful’, *barszcz* ‘borsch’, *semestr* ‘semester’, *mistrz* ‘master’. The remaining categories are represented by only 1 cluster type each; the morphonotactic /jɛtɛ/ in *wyjść*+ć ‘to leave’, and /jɛtɛ/ in *pięść* ‘fist’ and *siąść*+ć ‘to sit down’. Therefore, the increase from CC and CCC by 10% visible in Figure 22 is not representative and meaningful as it stands for the increase from 0 to 1.

Table 47. Hypothesis 1: Size vs. morphological composition (final cluster types).

|     | lexical |    | morphonotactic |    | both |    |
|-----|---------|----|----------------|----|------|----|
|     | no.     | %  | no.            | %  | no.  | %  |
| CC  | 52      | 96 | 0              | 0  | 2    | 4  |
| CCC | 8       | 80 | 1              | 10 | 1    | 10 |

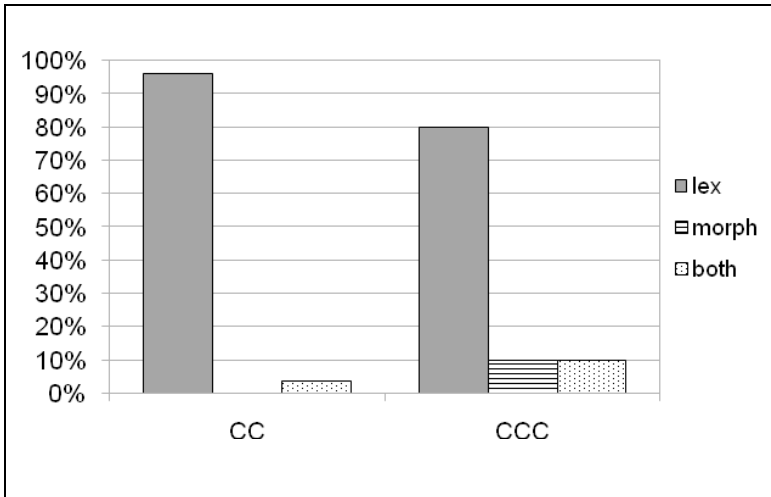


Figure 22. Hypothesis 1: Size vs. morphological composition (final cluster types).

**Word types with final clusters.** The total number of word types with final lexical and morphonotactic clusters is given in Table 48 and Figure 23. Tables 49 and 50 show the results for each size group separately. Please note that no words with sequences of 4 consonants are found in the word-final position. It is due to the fact that CCCC clusters, such as /pstf/ in *glupstw* ‘foolishness’, /rstf/ in *warstw* ‘layer’, and /jstf/ in *dobrodziejstw* ‘good deed’, occur in inflected genitive plural forms, which are not included as lexical entries in a dictionary. As can be seen, we can observe a correlation between cluster size and morphological composition ( $p = 0.0168$ ).

Table 48. The total number of word types with phonotactic and morphonotactic final clusters.

|             | lexical | morphonotactic | total |
|-------------|---------|----------------|-------|
| words (no.) | 373     | 229            | 602   |
| words (%)   | 62%     | 38%            | 100%  |

Table 49. The number of word types with phonotactic and morphonotactic final CC clusters.

|             | lexical | mophonotactic | total |
|-------------|---------|---------------|-------|
| words (no.) | 359     | 209           | 568   |
| words (%)   | 63%     | 37%           | 100%  |

Table 50. The number of word types with phonotactic and morphonotactic final CCC clusters.

|             | lexical | mophonotactic | total |
|-------------|---------|---------------|-------|
| words (no.) | 14      | 20            | 34    |
| words (%)   | 41%     | 59%           | 100%  |

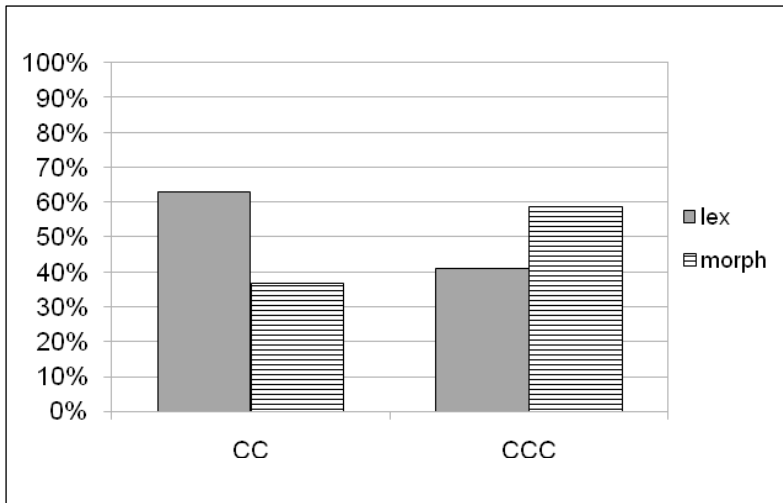


Figure 23. Hypothesis 1: Size vs. morphological composition (word types with final clusters).

### 5.2.3. Results: Hypothesis 2

Let us recall the assumption of hypothesis 2, namely the degree of cluster preferability is inversely proportional to morphological complexity. It is predicted that morphonotactic clusters will have a lower degree of preferability than phonotactic ones. We start our analysis with the description of initial clusters and then proceed to the description of final ones.

**Word types with initial clusters.** The summary of the results obtained for initial CC and CCC clusters occurring in the word types is provided in Tables 51 and 54, respectively. The corresponding graphic visualisations are given below in 24 and 25. Percentages were calculated for the lexical and morphonotactic category separately.

Table 51. Hypothesis 2: Morphological composition vs. preferability (word types with initial CC clusters).

|       | preferred |    | dispreferred |    |
|-------|-----------|----|--------------|----|
|       | no.       | %  | no.          | %  |
| lex   | 1 020     | 54 | 864          | 46 |
| morph | 35        | 14 | 218          | 86 |

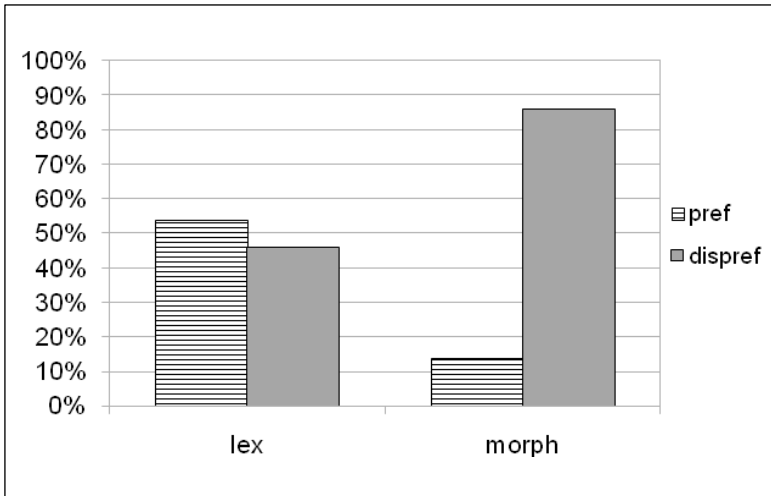


Figure 24. Hypothesis 2: Morphological composition vs. preferability (word types with initial CC clusters).

As shown above, the vast majority (86%) of morphonotactic clusters have a dispreferred status. Within the lexical category, the proportion of preferred and dispreferred clusters is relatively balanced, with a slight preference towards good clusters (54%). Following the hypothesis, we observe a correlation between the morphological status and preferability of clusters ( $p = 0$ ).

The most and the least frequent clusters provide some interesting insights into cluster preferability. If we consider, say, the first 20 most frequent clusters (see Table 52), we observe that only 5 out of 20 are dispreferred. The remaining 75% of the clusters are preferred. The dispreferred

clusters are represented by /pʂ/, /s/ + stop sequences /st, sp, sk/, and /ʂtʂ/. The most frequent cluster in Polish in the dictionary is a dispreferred /pʂ/ (353 out of 1884 word types with lexical clusters), as it occurs in highly productive prefixes {przed-}, {przy-} and {prze-}, which generate numerous derivatives. As regards /s/ + stop clusters, they are problematic for in any theoretical model.

Table 52. 20 top-frequency CC lexical clusters (P=preferred, D=dispreferred).

|    | lexical cluster | number of words | NAD |    | lexical cluster | number of words | NAD |
|----|-----------------|-----------------|-----|----|-----------------|-----------------|-----|
| 1  | pʂ              | 353             | D   | 11 | kl              | 36              | P   |
| 2  | pr              | 134             | P   | 12 | mj              | 36              | P   |
| 3  | st              | 95              | D   | 13 | sp              | 35              | D   |
| 4  | pj              | 69              | P   | 14 | ʂtʂ             | 33              | D   |
| 5  | tr              | 68              | P   | 15 | sk              | 30              | D   |
| 6  | vj              | 64              | P   | 16 | br              | 30              | P   |
| 7  | kr              | 62              | P   | 17 | gw              | 29              | P   |
| 8  | sw              | 54              | P   | 18 | pl              | 29              | P   |
| 9  | gr              | 50              | P   | 19 | dr              | 27              | P   |
| 10 | pw              | 37              | P   | 20 | bj              | 26              | P   |

Let us now turn our attention to the bottom of the frequency list. We arbitrarily isolated cluster types represented by 1, 2 or 3 word types, and found that 67% of these types are dispreferred. We also noticed that in the case of clusters represented by 2 or 3 entries, most belong to the same derivational nest. For example, the cluster /vd/ occurs 3 times in the dictionary in *wdowa* ‘widow’, *wdowi* ‘widow’ (ADJ), *wdowieć* ‘to widow’. The exhaustive list of items under consideration is given below.

Table 53. Low-frequency CC lexical clusters represented by 1, 2 or 3 dictionary words (P=preferred, D=dispreferred).

| lexical cluster | no of words | words                 | NAD |
|-----------------|-------------|-----------------------|-----|
| ep              | 1           | śpiwór ‘sleeping bag’ | D   |
| db              | 1           | dbać ‘to care’        | D   |
| dn              | 1           | dno ‘bottom’          | D   |
| px              | 1           | pchać ‘to push’       | D   |
| rv              | 1           | rwać ‘to tear’        | D   |
| tʂte            | 1           | czcić ‘to worship’    | D   |
| vɲ              | 1           | wniosek ‘conclusion’  | D   |
| vz              | 1           | wziąć ‘to take’       | D   |
| wz              | 1           | łza ‘tear’            | D   |
| zl              | 1           | źle ‘badly’           | D   |

|     |   |  |   |
|-----|---|--|---|
| zm  | 1 | żmija ‘viper’  | D |
| gj  | 1 | gielda ‘(stock) market’  | P |
| sr  | 1 | szron ‘frost’  | P |
| tem | 1 | ćma ‘moth’   | P |
| tsw | 1 | clo ‘customs’  | P |
| xl  | 1 | chleb ‘bread’  | P |
| zr  | 1 | zraz ‘beef collop’   | P |
| dzb | 2 | dzban ‘jug’, dzbanek ‘jug’ (DIM)   | D |
| lz  | 2 | lżej ‘light’ (ADV), lżejszy ‘lighter’ (ADJ)  | D |
| ml  | 2 | mleć ‘to grind’, mleczny ‘milky’   | D |
| mn  | 2 | mnożyć ‘to multiply’, mnóstwo ‘plenty’   | D |
| tef | 2 | ćwiczyć ‘to exercise’, ćwiczenie ‘exercise’  | D |
| tk  | 2 | tkać ‘to weave’, tkanina ‘fabric’  | D |
| tsf | 2 | czwartek ‘Thursday’, czwartkowy ‘Thursday’ (ADJ)   | D |
| xte | 2 | chcieć ‘to want’, chciwy ‘greedy’  | D |
| zb  | 2 | zboże ‘cereal’, zbożowy ‘cereal’ (ADJ)   | D |
| zd  | 2 | zdanie ‘sentence’, zdaniowy ‘sentence’ (ADJ)   | D |
| zj  | 2 | żniwo ‘harvest’, żniwny ‘harvest’ (ADJ)  | D |
| em  | 2 | śmigło ‘airscrew’, śmigłowy ‘airscrew’ (ADJ)   | P |
| dm  | 2 | dmuchać ‘to blow’ (IMPERF), dmuchnąć ‘to blow’ (PERF)  | P |
| tl  | 2 | tlen ‘oxygen’, tlenowy ‘oxygen’ (ADJ)  | P |
| tsm | 2 | cmentarz ‘cemetery’, cmentarny ‘cemetery’ (ADJ)  | P |
| tsw | 2 | człowiek ‘human being’, członek ‘member’   | P |
| dzv | 3 | dzwon ‘bell’, dzwonek ‘bell’ (DIM), dzwonić ‘to ring’  | D |
| dzv | 3 | dzwigo ‘crane’, dźwigać ‘to carry’ (IMPERF), dźwignąć ‘to carry’ (PERF)                                | D |
| dż  | 3 | drzewo ‘tree’, drzewny ‘tree’ (ADJ), drzeć ‘to tear’   | D |
| ft  | 3 | wtedy ‘then’, wtorek ‘Tuesday’, wtorkowy ‘Tuesday’ (ADJ)   | D |
| gd  | 3 | gdy ‘when’, gdyby ‘if’, gdyż ‘because’   | D |
| pt  | 3 | ptak ‘bird’, ptaszek ‘birdy’, ptasi ‘bird’ (ADJ)   | D |
| rdz | 3 | rdza ‘rust’, rdzawy ‘rust-coloured’, rdzawo ‘rust-coloured’ (ADV)                                      | D |
| sn  | 3 | sznur ‘rope’, sznurek ‘string’, sznurowadło ‘shoe lace’  | D |
| tx  | 3 | tchórz ‘coward’, tchórzliwy ‘cowardly’, tchórzliwie ‘in a cowardly fashion’                            | D |
| vd  | 3 | wdowieć ‘to become a widow’, wdowa ‘widow’, wdowi ‘widow’ (ADJ)  | D |
| vdz | 3 | wdzięczność ‘gratitude’, wdzięczny ‘grateful’, wdzięcznie ‘gratefully’                                 | D |
| xş  | 3 | chrześcijanin ‘Christian’ (male), chrześcijanka ‘Christian’ (female), chrześcijański ‘Christian’ (ADJ) | D |
| dl  | 3 | dla ‘for’, dlaczego ‘why’, dlatego ‘therefore’   | P |
| gl  | 3 | gleba ‘soil’, glina ‘clay’, gliniany ‘clay’ (ADJ)  | P |
| gm  | 3 | gmach ‘building’, gmina ‘district’, gminny ‘district’ (ADJ)  | P |
| sr  | 3 | srebro ‘silver’, srebrny ‘silver’ (ADJ), srebrno ‘silver’ (ADV)  | P |
| xm  | 3 | chmura ‘cloud’, chmurka ‘cloudlet’, chmurny ‘sulky’  | P |
| xr  | 3 | chrapać ‘to snore’, chronić ‘to protect’, chryпка ‘hoarseness’   | P |

As regards initial triples, as much as 69% of lexical CCC clusters are preferred. Interestingly, the proportion of preferred to dispreferred clusters in the morphological group is exactly the same as in the lexical group: 69% to 31% ( $p = 1$ ). The data for 3-element clusters is found in Table 54.

Table 54. Hypothesis 2: Morphological composition vs. preferability (word types with initial CCC clusters).

|       | preferred |    | dispreferred |    |
|-------|-----------|----|--------------|----|
|       | no.       | %  | no.          | %  |
| lex   | 130       | 69 | 59           | 31 |
| morph | 133       | 69 | 59           | 31 |

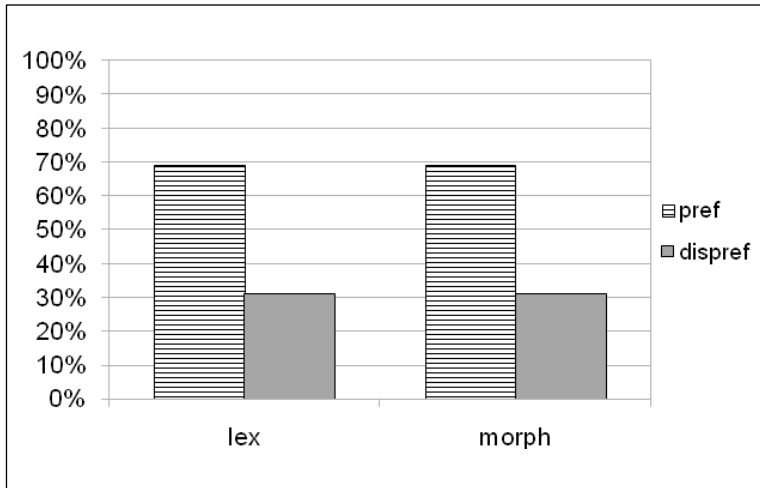


Figure 25. Hypothesis 2: Morphological composition vs. preferability (word types with initial CCC clusters).

**Initial cluster types.** Another level of our analysis is based on cluster types. As can be seen in Tables 55 and 56 as well as Figures 26 and 27, 100% is assumed for the subcategories referred to as lexical, morphonotactic and ‘both’. This division stands for exclusively phonotactic clusters, exclusively morphonotactic clusters, and clusters which function both as phonotactic and morphonotactic types.

Table 55. Hypothesis 2: Morphological composition vs. preferability (initial CC cluster types).

|       | preferred |    | dispreferred |    |
|-------|-----------|----|--------------|----|
|       | no.       | %  | no.          | %  |
| lex   | 49        | 46 | 57           | 54 |
| morph | 1         | 5  | 20           | 95 |
| both  | 5         | 28 | 13           | 72 |

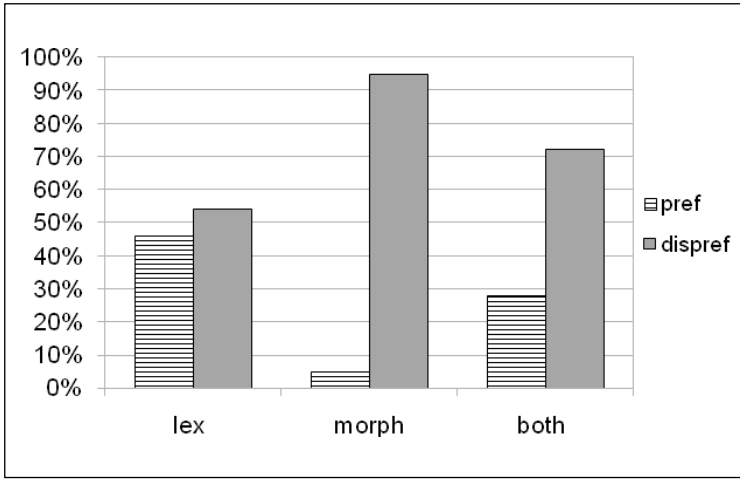


Figure 26. Hypothesis 2: Morphological composition vs. preferability (initial CC cluster types).

Table 56. Hypothesis 2: Morphological composition vs. preferability (initial CCC cluster types)

|       | preferred |    | dispreferred |    |
|-------|-----------|----|--------------|----|
|       | no.       | %  | no.          | %  |
| lex   | 20        | 56 | 16           | 44 |
| morph | 19        | 54 | 16           | 46 |
| both  | 10        | 71 | 4            | 29 |

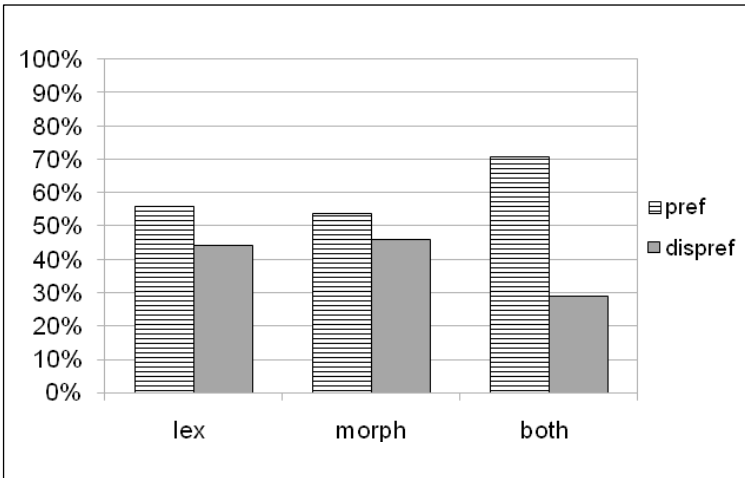


Figure 27. Hypothesis 2: Morphological composition vs. preferability (initial CCC cluster types).

In the case of CC cluster types, we can observe the same tendency as for word types, i.e. within the exclusively lexical category, the proportion of preferred and dispreferred clusters is comparable (46% vs. 54%), while exclusively morphonotactic clusters are largely dispreferred (95%). Types whose realisation may be phonotactic or morphonotactic ('both') tend to be dispreferred (72%). As for CCC sequences, the percentages obtained for the group of preferred and dispreferred types are comparable, while the 'both' types are dominated by preferred clusters.

**Word types with final clusters.** Let us now move on to the analysis of word-final sequences. The data for words with 2-element clusters found in Table 57 shows that morphonotactic clusters are dispreferred with no exception, which corroborates our hypothesis. The second observation is that, again, the lexical category is represented by both preferred and dispreferred clusters, with dispreferred clusters prevailing (58%).

Table 57. Hypothesis 2: Morphological composition vs. preferability (word types with final CC clusters).

|       | preferred |    | dispreferred |     |
|-------|-----------|----|--------------|-----|
|       | no.       | %  | no.          | %   |
| lex   | 151       | 42 | 208          | 58  |
| morph | 0         | 0  | 209          | 100 |

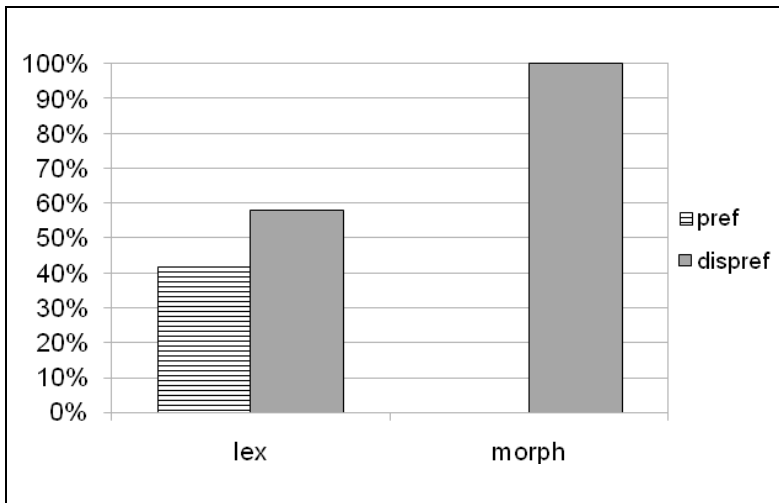


Figure 28. Hypothesis 2: Morphological composition vs. preferability (word types with final CC clusters).

It must be noted that 54 final lexical CC types are represented by 359 words, out of which 82 feature the dispreferred /*ete*/. This cluster appears in a productive suffix {-ość}, which generated 69 abstract nouns, e.g. *miłość* ‘love’ and *radość* ‘joy’. The remaining 13 words ending in /*ete*/ include *cześć* ‘hi’, *dość* ‘enough’, *gość* ‘guest’, *gwóźdź* ‘nail’, *korzyść* ‘advantage’, *kość* ‘bone’, *liść* ‘leaf’, *maść* ‘ointment’, *napaść* ‘assault’, *ość* ‘fishbone’, *powieść* ‘novel’, *przepaść* ‘abyss’, *treść* ‘contents’. We also noticed that 37 cluster types are represented by 1, 2, or 3 word types. This means that 69% of all final lexical doubles occur in very few words, which may suggest that clusters in the word-final position are disfavoured.

Interestingly, when it comes to the preferability of triples, opposite results are found (see Table 58). Firstly, all morphonotactic clusters are preferred (for more details see section *Final cluster types*). Secondly, the lexical group demonstrates a slight prevalence of preferred clusters (58%). These observations illustrate reversed tendencies holding for final CC and CCC clusters.

Table 58. Hypothesis 2: Morphological composition vs. preferability (word types with final CCC clusters).

|       | preferred |     | dispreferred |    |
|-------|-----------|-----|--------------|----|
|       | no.       | %   | no.          | %  |
| lex   | 7         | 58  | 5            | 42 |
| morph | 20        | 100 | 0            | 0  |

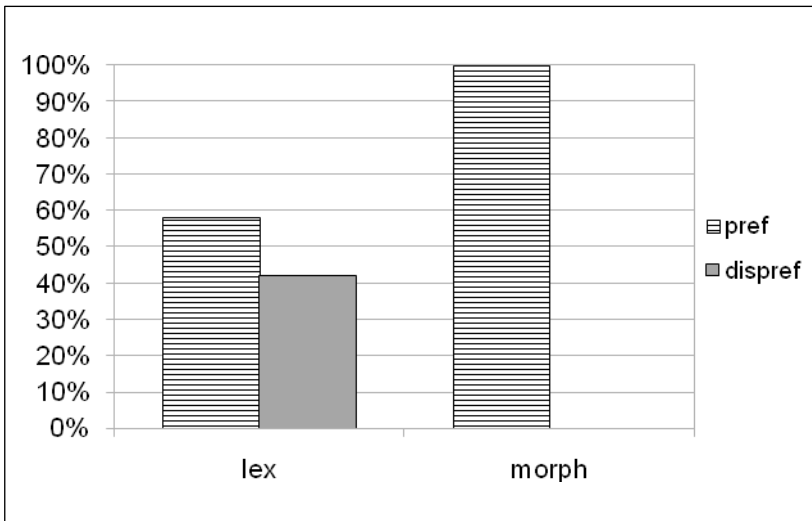


Figure 29. Hypothesis 2: Morphological composition vs. preferability (word types with final CCC clusters).

**Final cluster types.** As regards CC cluster types, exclusively lexical clusters display are preferred and dispreferred to a similar degree (52% vs. 48%). Exclusively morphonotactic clusters do not exist in the analysed sample. Clusters which may have dual motivation (phonological and morphological) are dispreferred. No generalisation in terms of percentages can be made as only two types represent the group: /jɛtɛ/ and /ɛtɛ/. The data for 2-element cluster types is found in Table 59.

Table 59. Hypothesis 2: Morphological complexity vs. preferability (final CC cluster types).

|       | preferred |    | dispreferred |     |
|-------|-----------|----|--------------|-----|
|       | no.       | %  | no.          | %   |
| lex   | 28        | 52 | 26           | 48  |
| morph | 0         | 0  | 0            | 0   |
| both  | 0         | 0  | 2            | 100 |

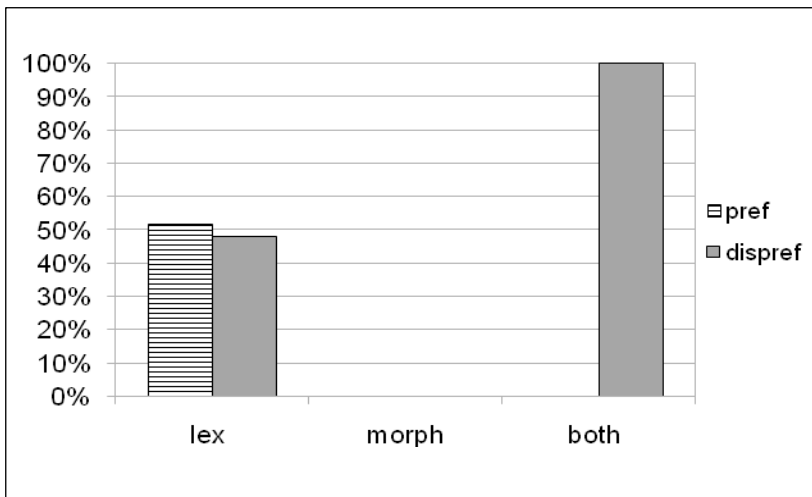


Figure 30. Hypothesis 2: Morphological complexity vs. preferability (final CC cluster types).

As far as CCC types are concerned, lexical clusters are balanced in terms of their preferability (50% vs. 50%), whereas clusters referred to as ‘morph’ and ‘both’ are preferred. However, these results are by no means generalisations, as each group is represented by either 1 or 4 clusters: exclusively morphonotactic /jɛtɛ/, ‘both’ /jɛtɛ/, exclusively phonotactic preferred /ntɕ, ŋkt, rɛtɛ, rɕtɕ/, and exclusively phonotactic dispreferred /kst, ltr, str, stɕ/. The data for 3-element cluster types is tabulated below.

Table 60. Hypothesis 2: Morphological composition vs. preferability (final CCC cluster types).

|       | preferred |     | dispreferred |    |
|-------|-----------|-----|--------------|----|
|       | no.       | %   | no.          | %  |
| lex   | 4         | 50  | 4            | 50 |
| morph | 1         | 100 | 0            | 0  |
| both  | 1         | 100 | 0            | 0  |

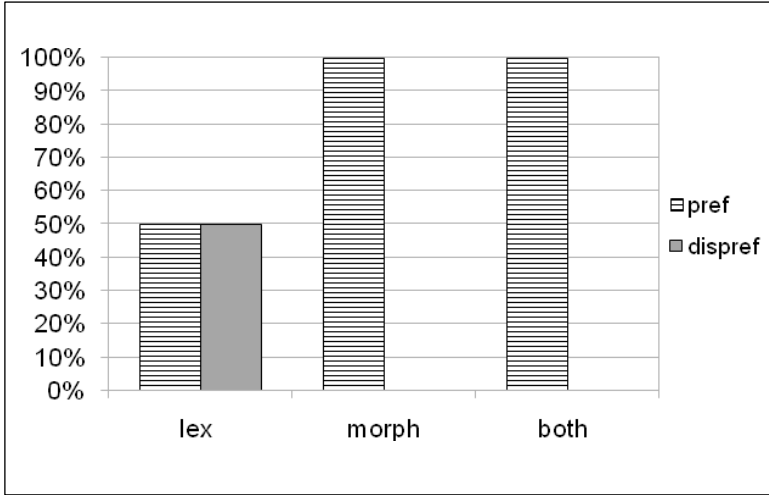


Figure 31. Hypothesis 2: Morphological composition vs. preferability (final CCC cluster types).

The raw data, on the basis of which the calculations for hypothesis 2 were performed, is presented in Tables 61-64 below. The tables show cluster preferability (shaded cells = preferred; (quadruples are not evaluated by NAD) and information about their morphological status.

Table 61. An alphabetical list of initial double clusters with morphological divisions and the NAD status.

|   | CC | lex | mor | total |    | CC  | lex | mor | total |
|---|----|-----|-----|-------|----|-----|-----|-----|-------|
| 1 | bj | 26  |     | 26    | 9  | en  | 6   |     | 6     |
| 2 | bl | 15  |     | 15    | 10 | ep  | 1   |     | 1     |
| 3 | br | 30  |     | 30    | 11 | er  | 13  |     | 13    |
| 4 | bw | 13  |     | 13    | 12 | etc | 8   | 8   | 16    |
| 5 | bz | 12  |     | 12    | 13 | db  | 1   |     | 1     |
| 6 | ef | 4   |     | 4     | 14 | dj  | 9   |     | 9     |
| 7 | el | 25  |     | 25    | 15 | dl  | 3   |     | 3     |
| 8 | em | 2   |     | 2     | 16 | dm  | 2   |     | 2     |

|    | CC   | lex | mor | total |     | CC  | lex | mor | total |
|----|------|-----|-----|-------|-----|-----|-----|-----|-------|
| 17 | dn   | 1   |     | 1     | 60  | mn  | 2   |     | 2     |
| 18 | dr   | 27  |     | 27    | 61  | mr  | 10  |     | 10    |
| 19 | dv   | 4   |     | 4     | 62  | mw  | 8   |     | 8     |
| 20 | dw   | 8   |     | 8     | 63  | pe  |     | 1   | 1     |
| 21 | dz   | 3   |     | 3     | 64  | pj  | 69  |     | 69    |
| 22 | dz̄b | 2   |     | 2     | 65  | pl  | 29  |     | 29    |
| 23 | dz̄v | 3   |     | 3     | 66  | pr  | 134 |     | 134   |
| 24 | dz̄v | 3   |     | 3     | 67  | ps  | 6   |     | 6     |
| 25 | fe   |     | 2   | 2     | 68  | pş  | 353 |     | 353   |
| 26 | fj   | 4   |     | 4     | 69  | pt  | 3   |     | 3     |
| 27 | fl   | 5   |     | 5     | 70  | pw  | 37  |     | 37    |
| 28 | fp   |     | 6   | 6     | 71  | px  | 1   |     | 1     |
| 29 | fr   | 7   |     | 7     | 72  | rdz | 3   |     | 3     |
| 30 | fs   |     | 2   | 2     | 73  | rv  | 1   |     | 1     |
| 31 | fş   | 6   | 1   | 7     | 74  | se  |     | 1   | 1     |
| 32 | ft   | 3   |     | 3     | 75  | sf  | 9   | 2   | 11    |
| 33 | ftş  | 6   |     | 6     | 76  | sk  | 30  | 13  | 43    |
| 34 | ft̄e |     | 3   | 3     | 77  | şk  | 10  |     | 10    |
| 35 | fts  |     | 1   | 1     | 78  | sl  | 7   |     | 7     |
| 36 | fx   |     | 1   | 1     | 79  | sm  | 14  |     | 14    |
| 37 | gd   | 3   |     | 3     | 80  | şm  | 5   |     | 5     |
| 38 | gdz  | 4   |     | 4     | 81  | sn  | 4   |     | 4     |
| 39 | gj   | 1   |     | 1     | 82  | şn  | 3   |     | 3     |
| 40 | gl   | 3   |     | 3     | 83  | sp  | 35  | 55  | 90    |
| 41 | gm   | 3   |     | 3     | 84  | şp  | 8   |     | 8     |
| 42 | gn   | 7   |     | 7     | 85  | sr  | 3   |     | 3     |
| 43 | gr   | 50  |     | 50    | 86  | şr  | 1   |     | 1     |
| 44 | gv   | 8   |     | 8     | 87  | ss  |     | 2   | 2     |
| 45 | gw   | 29  |     | 29    | 88  | sş  |     | 1   | 1     |
| 46 | gz   | 9   |     | 9     | 89  | st  | 95  | 10  | 105   |
| 47 | ke   | 8   |     | 8     | 90  | st̄ | 9   |     | 9     |
| 48 | kf   | 17  |     | 17    | 91  | stş | 33  |     | 33    |
| 49 | kj   | 16  |     | 16    | 92  | sts | 9   |     | 9     |
| 50 | kl   | 36  |     | 36    | 93  | sw  | 54  |     | 54    |
| 51 | kr   | 62  |     | 62    | 94  | sx  | 6   | 8   | 14    |
| 52 | kş   | 10  |     | 10    | 95  | tf  | 12  |     | 12    |
| 53 | kt   | 7   |     | 7     | 96  | tşf | 2   |     | 2     |
| 54 | kw   | 9   |     | 9     | 97  | tef | 2   |     | 2     |
| 55 | l̄n  |     | 1   | 1     | 98  | tk  | 2   |     | 2     |
| 56 | lv   |     | 1   | 1     | 99  | tl  | 2   |     | 2     |
| 57 | lz   | 2   |     | 2     | 100 | tem | 1   |     | 1     |
| 58 | mj   | 36  |     | 36    | 101 | tr  | 68  |     | 68    |
| 59 | ml   | 2   |     | 2     | 102 | tş  | 14  |     | 14    |

|     | CC    | lex | mor | total |     | CC           | lex   | mor | total |
|-----|-------|-----|-----|-------|-----|--------------|-------|-----|-------|
| 103 | t̄ste | 1   |     | 1     | 125 | xr           | 3     |     | 3     |
| 104 | tw    | 14  |     | 14    | 126 | xš           | 3     |     | 3     |
| 105 | t̄sw  | 2   |     | 2     | 127 | xte          | 2     |     | 2     |
| 106 | tx    | 3   |     | 3     | 128 | xw           | 9     |     | 9     |
| 107 | t̄sm  | 2   |     | 2     | 129 | zb           | 2     | 11  | 13    |
| 108 | t̄sw  | 1   |     | 1     | 130 | zd           | 2     | 12  | 14    |
| 109 | vd    | 3   |     | 3     | 131 | zdž          |       | 1   | 1     |
| 110 | vdž   | 3   |     | 3     | 132 | zg           |       | 17  | 17    |
| 111 | vj    | 64  | 4   | 68    | 133 | zj           |       | 5   | 5     |
| 112 | vl    |     | 2   | 2     | 134 | zl           |       | 4   | 4     |
| 113 | vm    |     | 2   | 2     | 135 | zl           | 1     |     | 1     |
| 114 | vn    | 4   |     | 4     | 136 | zm           |       | 15  | 15    |
| 115 | vñ    | 1   |     | 1     | 137 | zn           | 13    | 6   | 19    |
| 116 | vr    | 12  | 3   | 15    | 138 | zñ           |       | 12  | 12    |
| 117 | vw    | 16  | 4   | 20    | 139 | zr           | 1     | 13  | 14    |
| 118 | vz    | 7   | 1   | 8     | 140 | zr           | 6     |     | 6     |
| 119 | vž    | 1   |     | 1     | 141 | zv           | 9     | 12  | 21    |
| 120 | vž    | 6   | 2   | 8     | 142 | zw           | 10    | 6   | 16    |
| 121 | wz    | 1   |     | 1     | 143 | zž           |       | 2   | 2     |
| 122 | xf    | 7   |     | 7     | 144 | zm           | 1     |     | 1     |
| 123 | xl    | 1   |     | 1     | 145 | zñ           | 2     |     | 2     |
| 124 | xm    | 3   |     | 3     |     | <b>total</b> | 1 884 | 253 | 2 137 |

Table 62. An alphabetical list of initial triple clusters with morphological divisions and the NAD status.

|    | CCC    | lex | mor | total |    | CCC    | lex | mor | total |
|----|--------|-----|-----|-------|----|--------|-----|-----|-------|
| 1  | efj    | 28  |     | 28    | 18 | gvj    | 3   |     | 3     |
| 2  | emj    | 14  |     | 14    | 19 | gzñ    | 1   |     | 1     |
| 3  | epj    | 6   |     | 6     | 20 | kfj    | 6   |     | 6     |
| 4  | dr̄g   | 1   |     | 1     | 21 | krf    |     | 2   | 2     |
| 5  | dzv    | 1   |     | 1     | 22 | krt    | 2   |     | 2     |
| 6  | d̄zv̄j | 4   |     | 4     | 23 | kšt    | 2   |     | 2     |
| 7  | fer    |     | 1   | 1     | 24 | met̄e  | 1   |     | 1     |
| 8  | fete   | 1   |     | 1     | 25 | mdl    | 1   |     | 1     |
| 9  | fl̄    | 1   |     | 1     | 26 | mdw    | 2   |     | 2     |
| 10 | fkr    |     | 1   | 1     | 27 | mgl    | 1   |     | 1     |
| 11 | fk̄w   |     | 2   | 2     | 28 | mgw    | 1   |     | 1     |
| 12 | fpr    |     | 6   | 6     | 29 | p̄st̄š | 2   |     | 2     |
| 13 | fp̄w   |     | 7   | 7     | 30 | pwte   |     | 2   | 2     |
| 14 | fsk    |     | 5   | 5     | 31 | pxl    | 1   |     | 1     |
| 15 | fsp    | 1   | 10  | 11    | 32 | pxñ    |     | 1   | 1     |
| 16 | fst    | 7   | 7   | 14    | 33 | pxw    | 1   |     | 1     |
| 17 | fsx    |     | 3   | 3     | 34 | skj    |     | 1   | 1     |

|    | CC   | lex | mor | total |    | CC           | lex | mor | total |
|----|------|-----|-----|-------|----|--------------|-----|-----|-------|
| 35 | ʃkj  | 1   |     | 1     | 61 | vzb          |     | 2   | 2     |
| 36 | skl  | 3   | 1   | 4     | 62 | vzd          |     | 1   | 1     |
| 37 | ʃkl  | 2   |     | 2     | 63 | vzg          |     | 1   | 1     |
| 38 | skr  | 6   | 14  | 20    | 64 | vzm          |     | 2   | 2     |
| 39 | skʂ  | 9   | 2   | 11    | 65 | vzn          |     | 1   | 1     |
| 40 | skw  |     | 6   | 6     | 66 | vzn          |     | 1   | 1     |
| 41 | spj  |     | 2   | 2     | 67 | vzr          | 2   | 6   | 8     |
| 42 | ʃpj  | 2   |     | 2     | 68 | vzv          |     | 1   | 1     |
| 43 | spl  |     | 2   | 2     | 69 | zbj          |     | 7   | 7     |
| 44 | spr  | 3   | 26  | 29    | 70 | zbl          |     | 3   | 3     |
| 45 | spʂ  | 1   | 9   | 10    | 71 | zbr          | 1   | 4   | 5     |
| 46 | spw  |     | 4   | 4     | 72 | zdj          | 2   |     | 2     |
| 47 | stf  |     | 3   | 3     | 73 | zdr          | 7   |     | 7     |
| 48 | stj  | 2   |     | 2     | 74 | zgj          |     | 1   | 1     |
| 49 | str  | 32  | 5   | 37    | 75 | zgn          |     | 3   | 3     |
| 50 | stʂ  | 7   |     | 7     | 76 | zgr          | 2   | 2   | 4     |
| 51 | stw  |     | 2   | 2     | 77 | zgw          |     | 2   | 2     |
| 52 | sxr  |     | 2   | 2     | 78 | zgʐ          | 1   |     | 1     |
| 53 | tfj  | 2   |     | 2     | 79 | zmj          | 1   | 8   | 9     |
| 54 | tʃfj | 2   |     | 2     | 80 | zmp          |     | 2   | 2     |
| 55 | tkf  | 1   |     | 1     | 81 | zmr          |     | 1   | 1     |
| 56 | tkl  | 2   |     | 2     | 82 | zvj          | 3   | 9   | 12    |
| 57 | trf  | 3   |     | 3     | 83 | zvl          |     | 1   | 1     |
| 58 | tʂte | 2   |     | 2     | 84 | zvr          |     | 6   | 6     |
| 59 | vbr  | 1   |     | 1     | 85 | zvw          | 1   | 1   | 2     |
| 60 | vmj  |     | 1   | 1     |    | <b>total</b> | 189 | 192 | 381   |

Table 63. An alphabetical list of initial quadruple clusters with morphological divisions.

|   | CCCC         | lex | mor | total |
|---|--------------|-----|-----|-------|
| 1 | drgn         |     | 1   | 1     |
| 2 | fstr         |     | 3   | 3     |
| 3 | fstʂ         |     | 1   | 1     |
| 4 | gʐbj         | 2   |     | 2     |
| 5 | gzmj         | 1   |     | 1     |
| 6 | stfj         |     | 2   | 2     |
| 7 | vzdw         |     | 1   | 1     |
| 8 | vzgl         |     | 3   | 3     |
|   | <b>total</b> | 3   | 11  | 14    |

Table 64. The alphabetical list of word-final double and triple clusters ordered with morphological divisions and the NAD status.

|    | CC  | lex | mor | total |    | CC           | lex | mor | total |
|----|-----|-----|-----|-------|----|--------------|-----|-----|-------|
| 1  | ep  | 2   |     | 2     | 28 | ps           | 2   |     | 2     |
| 2  | ete | 82  | 45  | 127   | 29 | pş           | 2   |     | 2     |
| 3  | ft  | 1   |     | 1     | 30 | re           | 1   |     | 1     |
| 4  | ŷe  | 2   |     | 2     | 31 | rf           | 1   |     | 1     |
| 5  | jk  | 1   |     | 1     | 32 | rk           | 6   |     | 6     |
| 6  | jm  | 1   |     | 1     | 33 | rm           | 2   |     | 2     |
| 7  | jt  | 1   |     | 1     | 34 | rp           | 4   |     | 4     |
| 8  | kl  | 2   |     | 2     | 35 | rs           | 3   |     | 3     |
| 9  | ks  | 3   |     | 3     | 36 | rş           | 3   |     | 3     |
| 10 | kt  | 12  |     | 12    | 37 | rt           | 16  |     | 16    |
| 11 | lk  | 1   |     | 1     | 38 | rte          | 2   |     | 2     |
| 12 | lm  | 1   |     | 1     | 39 | sk           | 16  |     | 16    |
| 13 | lş  | 1   |     | 1     | 40 | st           | 19  |     | 19    |
| 14 | lt  | 3   |     | 3     | 41 | ş            | 1   |     | 1     |
| 15 | lts | 1   |     | 1     | 42 | şş           | 5   |     | 5     |
| 16 | mf  | 1   |     | 1     | 43 | sw           | 4   |     | 4     |
| 17 | mn  | 1   |     | 1     | 44 | şx           | 2   |     | 2     |
| 18 | mp  | 8   |     | 8     | 45 | tm           | 1   |     | 1     |
| 19 | mş  | 1   |     | 1     | 46 | tr           | 8   |     | 8     |
| 20 | ns  | 7   |     | 7     | 47 | wf           | 1   |     | 1     |
| 21 | nş  | 1   |     | 1     | 48 | wk           | 2   |     | 2     |
| 22 | nt  | 67  |     | 67    | 49 | wm           | 1   |     | 1     |
| 23 | ntş | 2   |     | 2     | 50 | ws           | 2   |     | 2     |
| 24 | nte | 6   | 164 | 170   | 51 | wş           | 3   |     | 3     |
| 25 | nts | 6   |     | 6     | 52 | wt           | 1   |     | 1     |
| 26 | ŋk  | 15  |     | 15    | 53 | wx           | 1   |     | 1     |
| 27 | pr  | 1   |     | 1     | 54 | zm           | 20  |     | 20    |
|    |     |     |     |       |    | <b>total</b> | 359 | 209 | 568   |

|    | CCC          | lex | mor | total |
|----|--------------|-----|-----|-------|
| 1  | jete         |     | 13  | 13    |
| 2  | ŷete         | 2   | 7   | 9     |
| 3  | kst          | 2   |     | 2     |
| 4  | ltr          | 1   |     | 1     |
| 5  | ntş          | 2   |     | 2     |
| 6  | ŋkt          | 2   |     | 2     |
| 7  | rete         | 2   |     | 2     |
| 8  | rşş          | 1   |     | 1     |
| 9  | str          | 1   |     | 1     |
| 10 | şş           | 1   |     | 1     |
|    | <b>total</b> | 14  | 20  | 34    |

### 5.2.4. Results: Hypothesis 3

Hypothesis 3 predicts that the degree of clusters preferability correlates with cluster frequency. In order to test hypothesis 3, we selected 5 most frequent initial and 5 most frequent final clusters in the corpus. Next, we analysed their morphological make-up in the paradigm and measured their NAD value. The most frequent initial clusters in the corpus comprise /pɕ, pr, st, mj, sp/. Clusters /pr/ and /mj/ are lexical and preferred in terms of the NAD principle. The remaining ones /pɕ, st, sp/ are dispreferred to a greater or lesser degree (/pɕ/ < /st/ < /sp/). The morphological analysis of clusters /st/ and /sp/ showed that they can be morphonotactic; however, /sp/ is realised morphologically more often than /st/. The consonant sequence /pɕ/, which is lexical but marked, owes its high frequency to prepositions *przez*, *przed*, *przy* as well as prefixes {*przy-*}, {*prze-*} and {*przed-*}, which, due to their high productivity, generate numerous words. Figure 32 shows the most frequent clusters and their lexical and morphonotactic realisation in the paradigm (cluster\_p) and the corpus (clusters\_c).

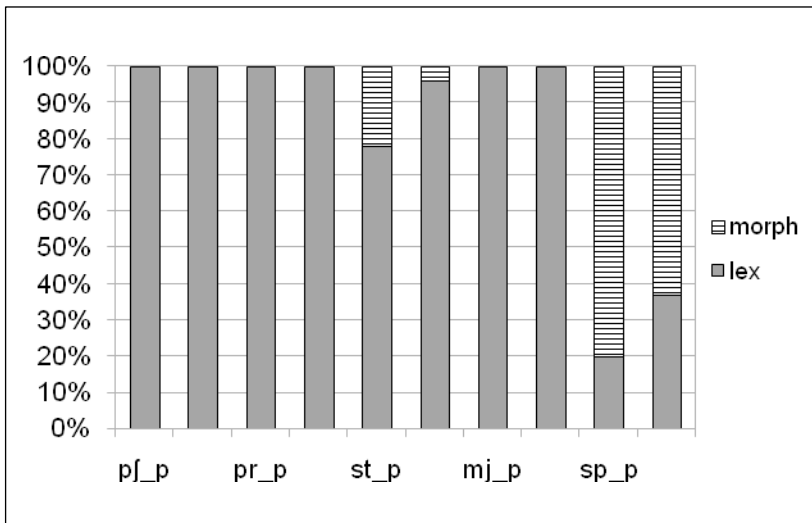


Figure 32. Hypothesis 3: Preferability vs. frequency (initial CC in the paradigm and corpus).

The most frequent consonant clusters in the word-final position are represented by /st, ɛtɕ, nt, nts, ŋk/, out of which /nt/ and /ŋk/ are preferred, whereas /st, ɛtɕ, nts/ are dispreferred. Cluster /st/ was monopolised by the

word *jest* ‘is’. Cluster / $\widehat{ct\acute{c}}$ / can be morphonotactic, e.g. in the words *iść* ‘to go’, *kraść* ‘to steal’, *paść* ‘to fall’, or lexical as within the suffix {-ość} in a numerous group of abstract nouns, e.g. *miłość* ‘love’, *radość* ‘joy’, *szczerłość* ‘sincerity’. The summary of the results for the lexical and morphonotactic realisation of the most frequent final clusters is provided in Figure 33.

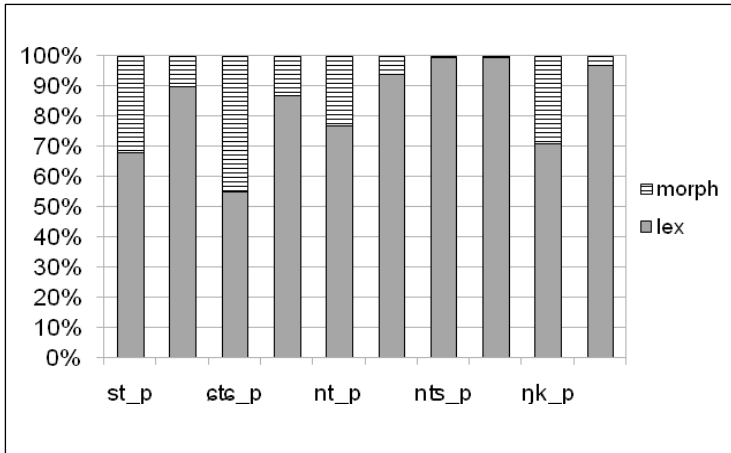


Figure 33. Hypothesis 3: Preferability vs. frequency (final CC in the paradigm and corpus).

**Summary.** In this section of the chapter we tested 3 hypotheses on the basis of the Polish cluster inventory. We were interested in observing the relationship between the size of a cluster and its morphological composition (hypothesis 1), preferability of a cluster and its morphological composition (hypothesis 2) as well as the frequency of a cluster and its preferred or dispreferred status (hypothesis 3). The analysis of our data showed that the first hypothesis was corroborated without any exception. The degree of morphological complexity was proved to increase along with the increase in the number of segments in a cluster. The second prediction turned out to be partially correct in that Polish morphonotactic CCs are strongly dispreferred, while lexical CCs tend to be mixed in terms of NAD. CCCs, in turn, were shown to be largely preferred. The data did not lend full support to hypothesis 3; 3 out of 5 clusters word-initially and word-finally were shown to be dispreferred.

The goal of the study is to provide a broader picture on phonotactics and morphonotactics from a cross-linguistic perspective. For this purpose, we se-

lected 2 languages, Polish and English, both of which are phonotactically and morphologically complex. Both allow sequences of several consonants, and both rely on suffixation in word-formation and inflection, which gives rise to morphonotactic clusters. The comparison of these typologically-different languages ensures an insightful comparison and a more reliable verification of the NAD principle. Therefore, we shall now proceed to the discussion of results obtained for English.

### 5.3. English

#### 5.3.1. Principles of morphological parsing

The following methodology of morphological division was applied to the English data. For the purpose of the study, we focused on the word-final position exclusively due to the absence of word-initial morphonotactics. A vast majority of word-final clusters were parsed automatically on the basis of graphemic divisions, i.e. base + {-ed} as in *work+ed*, or base + {-s} as in *cat+s*. The parsing posed no problems in the analysis as the orthography of the plural, Saxon Genitive, 3rd person singular forms as well as the past forms in English is regular and predictable. Additionally, a morphological boundary was placed before the suffix {-th}, forming adjectival nouns and ordinal numerals, as in *wid+th*, *dep+th*, *warm+th* and *ten+th*, *twelf+th*, *thousand+th*. Clusters emerging word-finally in irregular verbs, e.g. *went*, *kept*, *felt*, were grouped together with lexical clusters although they have a disputable morphonotactic status and are frequently placed on a continuum between morphologically simple or complex words. A handful of morphologically simplex words which might be accidentally treated by the automated parser as morphonotactic, e.g. *Thames*, *Wales*, *lens*, *backwards*, *James*, were analysed manually. A summary of all suffixes generating morphonotactic clusters is given below.

Table 65. A list of English suffixes generating morphonotactic clusters.

| function                 | pronunciation | examples |
|--------------------------|---------------|----------|
| plural {-s}              | /s/           | cats     |
|                          | /z/           | dogs     |
| possesive {-s}           | /s/           | Kate's   |
|                          | /z/           | John's   |
| 3rd person singular {-s} | /s/           | walks    |
|                          | /z/           | loves    |

|                       |            |                 |
|-----------------------|------------|-----------------|
| past simple {-ed}     | /t/<br>/d/ | worked<br>loved |
| past participle {-ed} | /t/<br>/d/ | worked<br>loved |
| ordinal {-th}         | /θ/        | sixth           |
| nominal {-th}         | /θ/        | depth           |

Let us now proceed to the presentation of the data extracted from the two sources: the word list and the corpus.

### 5.3.2. Results: Hypothesis 1

As discussed in Chapter 3, the analysis of the English data is based on two sources: the dictionary-based paradigm and the corpus. For every hypothesis, we first present the data extracted from the paradigm (cluster types and word types) and then discuss the corpus data (word tokens). Tables 66-74 as well as Figures 34-36 present the results for Hypothesis 1, which predicted that longer clusters are likely to be morphologically complex.

**Final cluster types.** In English, we found the total of 123 cluster types, among which there are 63 CC, 52 CCC and 8 CCCC types. As in the case of Polish, we introduced the distinction between clusters which are purely lexical, e.g. /ns/ in *dance*, /lk/ in *milk*, /mp/ in *lamp*; purely morphonotactic, e.g. /dʒd/ in *judge+d*, /sts/ in *last+s*, /lvz/ in *wolv+es*; and cluster types which occur in both contexts, e.g. /ft/ in *draft* vs. *laugh+ed*, /nd/ in *band* vs. *bann+ed*, /kst/ in *text* vs. *fax+ed*. The summary of the results is presented below. Percentage values were calculated for doubles, triples and quadruples separately.

Table 66. Hypothesis 1: Size vs. morphological composition  
(cluster types in the word list).

|      | lexical |    | morphonotactic |     | both |    |
|------|---------|----|----------------|-----|------|----|
|      | no.     | %  | no.            | %   | no.  | %  |
| CC   | 23      | 37 | 26             | 41  | 14   | 22 |
| CCC  | 3       | 6  | 38             | 73  | 11   | 21 |
| CCCC | 0       | 0  | 8              | 100 | 0    | 0  |

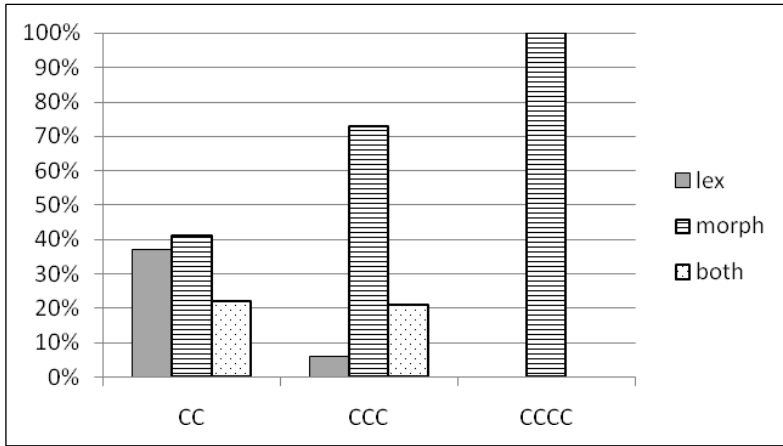


Figure 34. Hypothesis 1: Size vs. morphological composition (cluster types in the word list).

**Word types with final clusters.** The dictionary-based paradigm included 19 417 inflected word forms, out of which 4 181 contained lexical clusters and 15 236 were morphological. Unlike cluster types, individual clusters (cluster tokens) can be assigned either to a lexical or morphonotactic category. A summary of the results is provided in Table 67, while Tables 68-70 present the data for each cluster size separately, also shown in Figure 35.

Table 67. The total number of word types with phonotactic and morphonotactic final clusters (in the word list).

|             | lexical | mophonotactic | total  |
|-------------|---------|---------------|--------|
| words (no.) | 4 181   | 15 236        | 19 417 |
| words (%)   | 22%     | 78%           | 100%   |

Table 68. The number of word types with phonotactic and morphonotactic final CC clusters (in the word list).

|             | lexical | mophonotactic | total  |
|-------------|---------|---------------|--------|
| words (no.) | 4 136   | 12 928        | 17 064 |
| words (%)   | 24%     | 76%           | 100%   |

Table 69. The number of word types with phonotactic and morphonotactic final CCC clusters (in the word list).

|             | lexical | mophonotactic | total |
|-------------|---------|---------------|-------|
| words (no.) | 45      | 2 292         | 2 337 |
| words (%)   | 2%      | 98%           | 100%  |

Table 70. The number of word types with phonotactic and morphonotactic final CCCC clusters (in the word list).

|             | lexical | mophonotactic | total |
|-------------|---------|---------------|-------|
| words (no.) | 0       | 16            | 16    |
| words (%)   | 0%      | 100%          | 100%  |

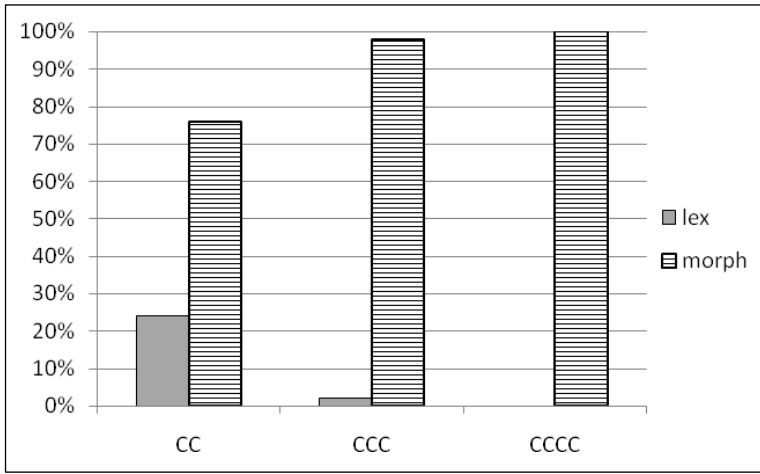


Figure 35. Hypothesis 1: Size vs. morphological composition (word types in the word list).

**Corpus frequency.** From the corpus we extracted all the words containing clusters together with their frequency of occurrence, the goal of which was to compare grammatically generated clusters in the paradigm with first-hand data. Therefore, the data in Table 71 and detailed data presented in Tables 72-74 and illustrated in 36 show the actual corpus frequency of clusters.

Table 71. The total number of word tokens with phonotactic and morphonotactic final clusters (in the corpus).

|             | lexical    | mophonotactic | total      |
|-------------|------------|---------------|------------|
| words (no.) | 32 790 335 | 24 963 022    | 57 753 357 |
| words (%)   | 57%        | 43%           | 100%       |

Table 72. The number of word tokens with phonotactic and morphonotactic final CC clusters (in the corpus).

|             | lexical    | mophonotactic | total      |
|-------------|------------|---------------|------------|
| words (no.) | 32 226 546 | 20 449 811    | 52 676 357 |
| words (%)   | 61%        | 39%           | 100%       |

Table 73. The number of word tokens with phonotactic and morphonotactic final CCC clusters (in the corpus).

|             | lexical | mophonotactic | total     |
|-------------|---------|---------------|-----------|
| words (no.) | 563 789 | 4 474 933     | 5 038 722 |
| words (%)   | 11%     | 89%           | 100%      |

Table 74. The number of word tokens with phonotactic and morphonotactic final CCCC clusters (in the corpus).

|             | lexical | mophonotactic | total  |
|-------------|---------|---------------|--------|
| words (no.) | 0       | 38 278        | 38 278 |
| words (%)   | 0%      | 100%          | 100%   |

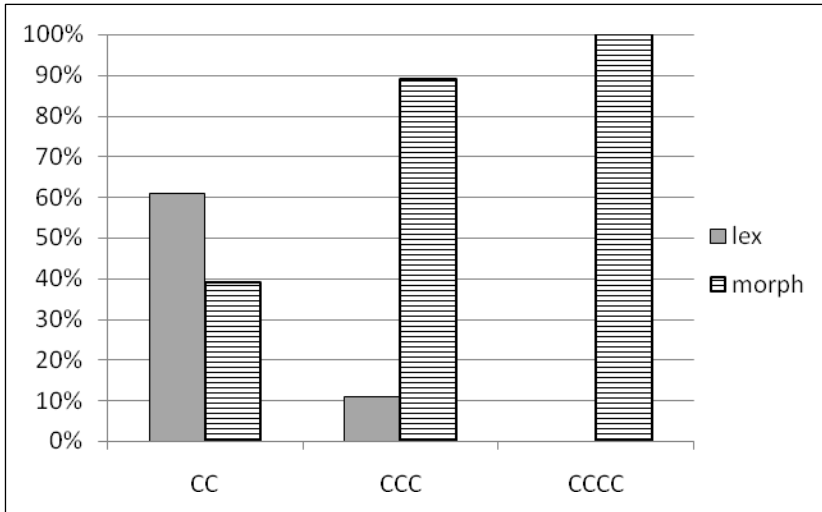


Figure 36. Hypothesis 1: Size vs. morphological composition (in the corpus).

**Summary.** The data confirm the prediction that the probability of a morphological boundary increases along with cluster length. This holds true for cluster types in the word list, word types in the word list (for the relation between CC and CCC clusters, the *p*-value amounts to 0) as well as word tokens in the corpus. 4-element clusters are morphonotactic with no exception. 3-element clusters are largely morphonotactic. Double clusters are mixed in terms of their lexical/morphonotactic status. Interestingly, in the word list lexical double clusters constitute 24% of all doubles, while in the corpus they amount to 61%.

### 5.3.3. Results: Hypothesis 2

We shall now turn to the discussion of the results obtained for Hypothesis 2. Tables 75-77 as well as Figures 37-39 present the results for Hypothesis 2. The results confirm the relationship between preferability in terms of NAD and the morphological composition of a cluster. Statistical testing was performed only for word types ( $p = 0$ ).

Table 75. Hypothesis 2: Morphological composition vs. preferability (cluster types in the word list).

|       | preferred |    | dispreferred |    |
|-------|-----------|----|--------------|----|
|       | no.       | %  | no.          | %  |
| lex   | 12        | 53 | 11           | 48 |
| morph | 4         | 15 | 22           | 85 |
| both  | 2         | 14 | 12           | 86 |

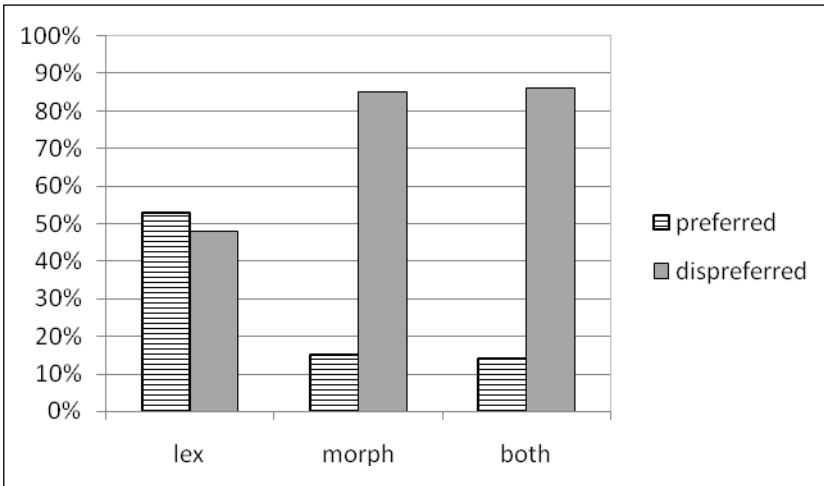


Figure 37. Hypothesis 2: Morphological composition vs. preferability (cluster types in the word list).

Table 76. Hypothesis 2: Morphological composition vs. preferability (word types in the word list).

|       | preferred |    | dispreferred |    |
|-------|-----------|----|--------------|----|
|       | no.       | %  | no.          | %  |
| lex   | 1441      | 35 | 2695         | 65 |
| morph | 876       | 7  | 12 052       | 93 |

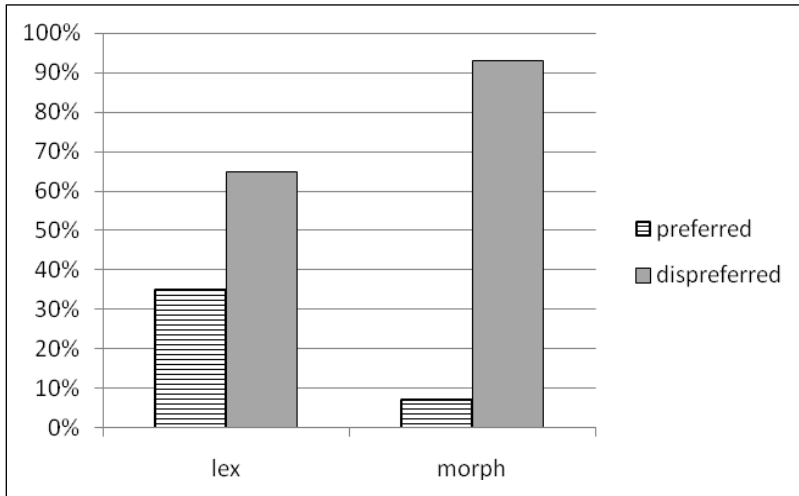


Figure 38. Hypothesis 2: Morphological composition vs. preferability (word types in the word list).

Table 77. Hypothesis 2: Morphological composition vs. preferability (in the corpus).

|       | preferred |    | dispreferred |    |
|-------|-----------|----|--------------|----|
|       | no.       | %  | no.          | %  |
| lex   | 8 596 785 | 26 | 23 629 761   | 74 |
| morph | 1 448 074 | 7  | 18 998 597   | 93 |

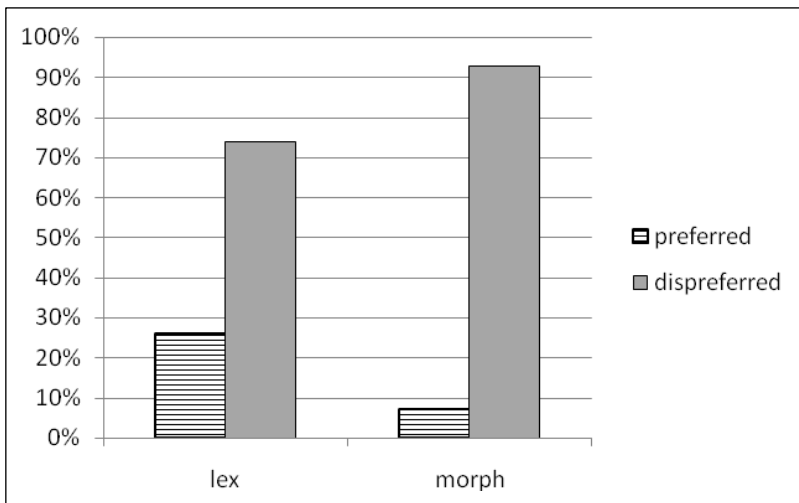


Figure 39. Hypothesis 2: Morphological composition vs. preferability (in the corpus).

### 5.3.4. Results: Hypothesis 3

In order to test Hypothesis 3, we selected five most frequent final double consonant clusters from the corpus, which included /nd, st, nt, nz, ts/. Out of these, only /nt/ was a preferred sequence according to NAD and, as Figure 40 demonstrates, it was an exclusively phonotactic sequence. The remaining 4 clusters were dispreferred according to NAD and all of them could occur across morpheme boundaries. /nz/ and /ts/ were heavily morphonotactic clusters; whereas /nd/ and /st/ could take either realisation (though in the corpus, they tended to occur intramorphemically).

As mentioned earlier, NAD is a scalar measure. Word-final /nt/ is borderline unmarked ( $3 = 3$  is true) and /nd/ is weakly marked ( $*3 < 2$  is not true). Also, as already noticed, voicing (Lx) is a disputable criterion. If we excluded it, and included the Sonorant vs. Obstruent difference instead (with the value of 1), then both /nt/ and /nd/ would be borderline unmarked (both are sequences of Sonorant-Sonorant-Obstruent). In future research, we aim to verify the criteria which account for phonotactics in the most optimal way.

/nz/ and /ts/ are both morphonotactic and marked. This is what we expected. As for /st/, it is well known that initial s+stop clusters are notoriously difficult to classify across models of phonotactic markedness. Sonority-based models generally admit final s+stops, since they include a slight sonority slope. NAD phonotactics disqualifies final s+stop clusters slightly less than initial ones: in both positions, these are strongly marked clusters. Thus, the NAD principle on its own cannot explain their occurrence. s+stops belong to the class of the so-called plateau clusters. These are generally problematic for phonotactic models.

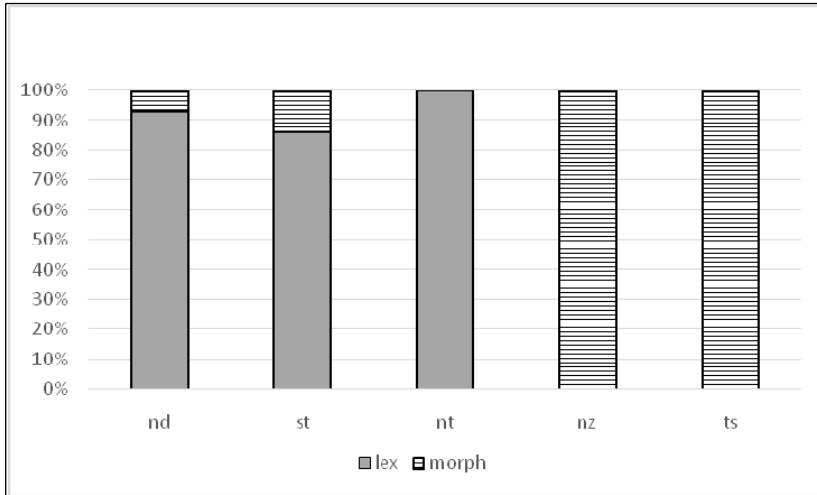


Figure 40. Hypothesis 3: Preferability vs. frequency.

It is interesting to compare a cluster ranking based on type frequency (word list) with a ranking based on token frequency (corpus), shown in Tables 78 and 79.

Table 78. Cluster raking based on type (word-list) frequency.

| type rank | cluster | example  |
|-----------|---------|----------|
| 1         | nz      | means    |
| 2         | ts      | its      |
| 3         | st      | just     |
| 4         | lz      | schools  |
| 5         | nt      | want     |
| 6         | ks      | makes    |
| 7         | nd      | and      |
| 8         | dz      | kids     |
| 9         | ld      | world    |
| 10        | nts     | students |

Table 79. Cluster raking based on token (corpus) frequency (\*copied from Table 78).

| corpus rank | cluster | top frequency examples            | corpus frequency | type rank* |
|-------------|---------|-----------------------------------|------------------|------------|
| 1           | nd      | and, around, found, find, end     | 26.33%           | 7          |
| 2           | st      | just, first, most, last, against  | 9.43%            | 3          |
| 3           | nt      | want, percent, president, student | 8.53%            | 5          |
| 4           | nz      | means, questions, ones, plans     | 5.28%            | 1          |
| 5           | ts      | its, states, minutes, nights      | 4.99%            | 2          |

|    |    |                                  |        |   |
|----|----|----------------------------------|--------|---|
| 6  | ld | world, old, told                 | 3.71%  | 9 |
| 7  | ns | since, once, sense               | 3.46%  | – |
| 8  | lz | schools, officials, miles, calls | 3.33%  | 4 |
| 9  | ks | makes, six, looks, weeks         | 2.92%  | 6 |
| 10 | kt | fact, looked, worked, effect     | 2.87%  | – |
|    |    |                                  | 70.85% |   |

As expected, the ranking of some clusters is influenced by the frequency of individual word forms which happen to be particularly frequent in texts, which reflects actual use. For example, the place of the /nd/ cluster in the token ranking is boosted by high frequency words, such as *and* (19.5% of all the wordforms considered!), *around* (0.5%), *found* (0.33%), etc. As a result, the cluster /nd/ moves from the seventh to the first place, and clusters /nz/ and /ts/ move from the top of the type ranking to the fourth and fifth places, respectively, in the token ranking. It is also important to note that the words with the /nd/ cluster represent as many as 26.32% of the total number of the words studied and that the words with the top ten clusters represent as many as 70% of the total number of the words studied.

**Summary.** The goal of this chapter was to examine Polish and English phonotactics according to three variables; morphological composition, preferability and frequency of clusters. In measuring cluster preferability, we followed the NAD principle presented in Chapter 1. We tested three hypotheses which were partially corroborated. Hypothesis 1 referring to the relation between cluster size and its morphological composition was fully supported by the data both in Polish and in English. Hypothesis 2, which concerned the relation between cluster preferability and its morphological composition turned out to be true for morphonotactic clusters. While double morphonotactic clusters were dispreferred to a large extent, lexical clusters showed a mixed character. As regards hypothesis 3 (preferability vs. frequency), the most frequent clusters turned out to include also numerous dispreferred clusters in both languages.

## Conclusions

The book had two aims: to test the adopted theoretical approach against the data and to present an exhaustive descriptive account of the latter. Three major hypotheses have been derived following the principles of the Beats-and-Binding phonology and natural morphonotactics. Although they have been corroborated to different degrees, the emerging generalizations concern four parameters or measures of cluster markedness. These are: size of a cluster, NAD (Net Auditory Distance), morphological complexity and frequency. All four measures define the naturalness (markedness) of a cluster. The greatest challenge consists in establishing their hierarchy. Size appears to be overwhelmingly the strongest parameter, i.e. the longer the cluster the more likely it is to be morphologically complex and, therefore, also phonologically marked. Consequently, morphological complexity comes second, and phonological naturalness measured by NAD comes third. Frequency, however, is a troublesome measure since it acts in a bidirectional or even circular fashion. On the one hand, the other three parameters have predictable frequency effects, i.e. clusters complex in size, morphological build-up and phonological structure (phonologically dispreferred ones) are expected to be less frequent than the respectively less complex ones. On the other hand, frequency may override all other criteria and end up on top of the hierarchy. This will not, however, make the most frequent cluster ‘unmarked’ or the most natural one: it will simply be the most used one. Therefore, the approach needs to be holistic, taking into account all the relevant criteria at all times, both those of structure and those of usage. The latter dichotomy went by a variety of different names in the history of linguistics, starting with competence vs. performance pair, via characteristics which are linguistic vs. extralinguistic, system-internal vs. external or structural vs. pragmatic. The holistic approach embraces the language and the speaker without dichotomizing them. The theory of Beats-and-Binding phonology and natural morphonotactics proved to serve well the analysis of consonantal clusters.

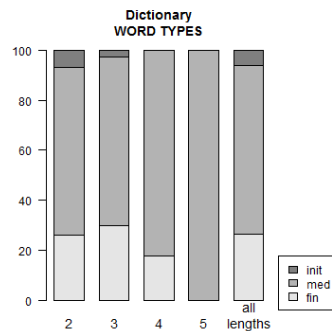
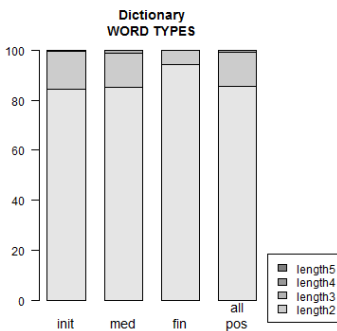
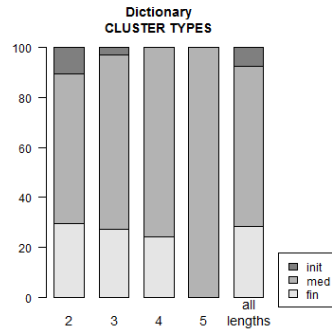
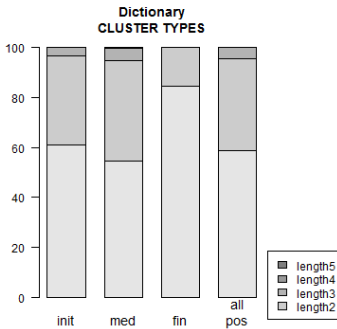
The second aim was to provide an extensive descriptive account of the consonantal phonotactics (and morphonotactics) of Polish as well as draw selected comparative analyses to English. Since we believe that the qual-

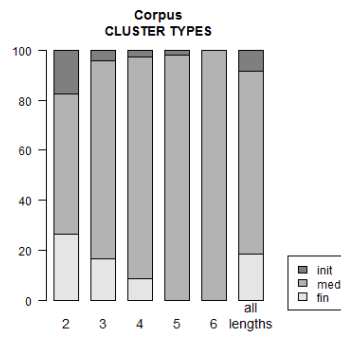
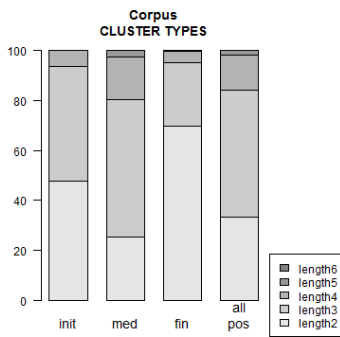
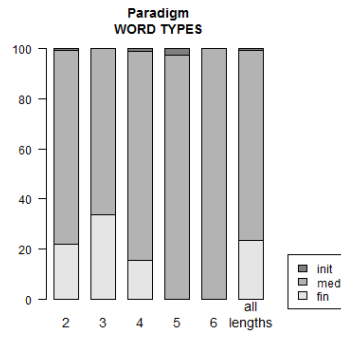
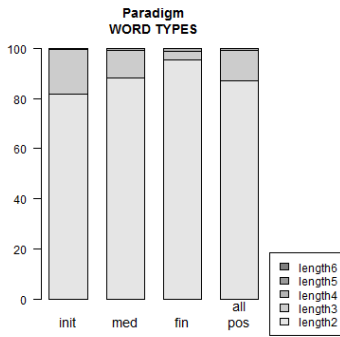
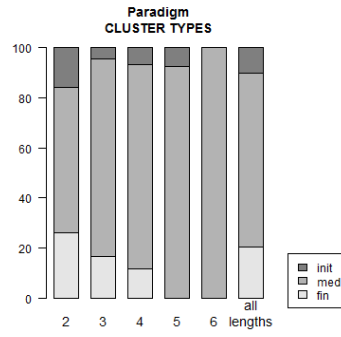
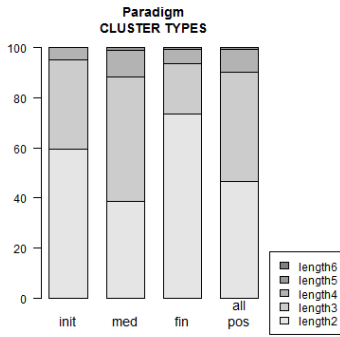
ity and quantity of the selected data largely determines the obtained overall picture of the examined structures, care has been taken to guarantee both as much as it was possible in the project. The resources for Polish consisted in three pools of data: a dictionary of the core vocabulary, a list of inflectional forms generated from the paradigms of the core lexical items, and a corpus of raw texts. In this way a representative pool of forms was created, which served as the source of frequency data for consonant clusters in three different modes: lexical frequency, paradigm frequency, and corpus frequency. Having all three values at the disposal made it possible to present a comprehensive picture of Polish phonotactics as well as to provide plausible explanations for both preferred (natural, unmarked) and dispreferred (marked) structures. Another important innovation of the present research was to rely on words and morphemes rather than syllables as domains of phonotactic constraints. We managed to describe word initial, word medial and word final phonotactics of undervived lexical items where all clusters were intramorphemic as well as the intermorphemic morphonotactics. This has been a pioneering approach across the studies of phonotactics to date. Comparisons to English provided additional verification of the applied methodology against a typologically different language.

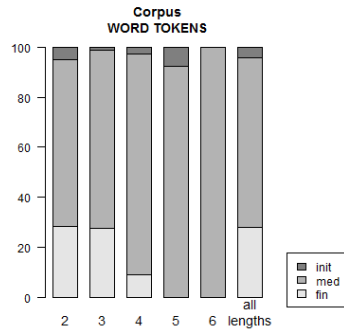
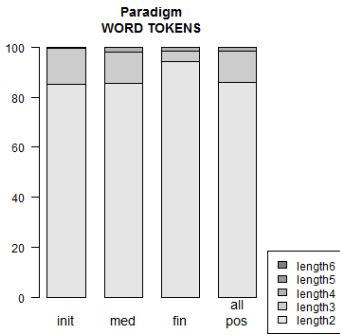
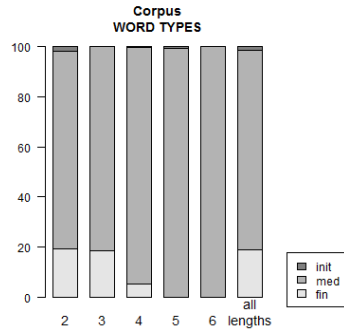
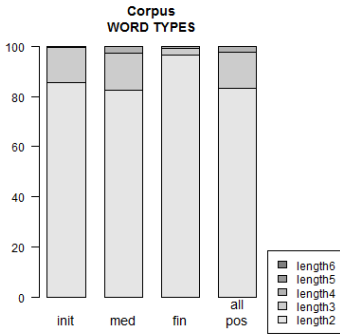
Plans for further research include, among others, a study of compounds. Clusters formed due to compounding are predicted to show the characteristics of lexical rather than morphonotactic clusters, due to their potential of becoming 'new words'. Spoken corpora need to be secured and analyzed, too, in order to test the predictive and explanatory power of the NAD Principle when applied to phonostylistic modifications of clusters. Clusters are supposed to improve their NAD parameter in casual speech; however, as we demonstrated elsewhere (Dressler et al. 2001), new marked clusters arise, too. Our goal is to provide and administer an open platform of resources for research on phonotactics (and morphonotactics) of typologically varied languages, starting with Polish, English, German and other languages we manage to analyze and making it available to others who might want to contribute. The NAD Phonotactic Calculator is already available online.

# Appendix A. Cluster distribution in Polish

Interpretation of the graphs: *cluster types* refers to the number of unique clusters, *word types* refers to the number of unique words in which clusters with specific characteristics were recorded, *word tokens* refers to the number of all occurrences of words in which clusters with specific characteristics were recorded. Column 1 in graph 1 below shows that among medial clusters recorded in the dictionary data: 61% are length 2, 35% are length 3, and 4% are length 4. Column 1 in graph 2 shows that among clusters of length 2 recorded in the dictionary data: approx. 30% are final, 60% are medial and 10% are initial.









## Appendix B. Polish cluster statistics

The following tables present Polish clusters in all positions and of all sizes with word type frequency statistics for the three resources studied and word token statistics for the newspaper corpus resource

### Legend

pos = position

len = length (size)

cl = cluster

dict = dictionary

para = paradigms

corp = corpus

wtyp = word types

wtok = words tokens

A value in the *dictwtyp* (dictionary word type) column provides the number of word types (unique words including inflected forms) featuring a particular cluster in one of the three positions in the dictionary resource. Values in the *paratwtyp* (paradigms word type) and *corpwtyp* (corpus word types) columns give similar statistics for the paradigms and corpus resources. A value in the *corpwtok* (corpus word token) column is the number of word tokens (all, possibly repeated, occurrences of words including inflected forms) featuring a particular cluster in one of the three positions in the corpus resource. A zero in a specific column means that no words with a given cluster in a particular position were recorded in the resource represented by the column. For example, this line

|   |   |    |   |    |    |     |
|---|---|----|---|----|----|-----|
| i | 2 | ɛp | 1 | 17 | 16 | 365 |
|---|---|----|---|----|----|-----|

says that one word with the cluster /ɛp/ (length 2) was recorded in the initial position in the dictionary resource, 17 unique words (word types) with the cluster /ɛp/ (length 2) were recorded in the initial position in the paradigms resource, 16 word types with the cluster /ɛp/ were recorded in the initial position in the corpus resource, 365 word tokens with the cluster /ɛp/ were recorded in the initial position in the corpus resource.

**Initial clusters**

| pos | len | cl              | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-----------------|----------|----------|----------|----------|
| i   | 2   | bj              | 26       | 349      | 1350     | 78009    |
| i   | 2   | bl              | 15       | 197      | 534      | 27778    |
| i   | 2   | br              | 29       | 479      | 1400     | 108783   |
| i   | 2   | bw              | 13       | 443      | 433      | 14699    |
| i   | 2   | bz              | 0        | 8        | 22       | 73       |
| i   | 2   | bz              | 0        | 1        | 4        | 28       |
| i   | 2   | bz <sub>z</sub> | 12       | 165      | 260      | 5387     |
| i   | 2   | cf              | 4        | 42       | 215      | 4737     |
| i   | 2   | cl              | 24       | 336      | 464      | 27795    |
| i   | 2   | em              | 2        | 21       | 57       | 1100     |
| i   | 2   | en              | 6        | 121      | 137      | 3454     |
| i   | 2   | ep              | 1        | 17       | 16       | 365      |
| i   | 2   | er              | 13       | 120      | 374      | 86028    |
| i   | 2   | es              | 0        | 0        | 1        | 1        |
| i   | 2   | ete             | 16       | 595      | 457      | 16264    |
| i   | 2   | db              | 1        | 59       | 47       | 2121     |
| i   | 2   | dj              | 9        | 94       | 425      | 9875     |
| i   | 2   | dl              | 3        | 4        | 9        | 172828   |
| i   | 2   | dm              | 2        | 130      | 57       | 302      |
| i   | 2   | dn              | 1        | 7        | 6        | 1887     |
| i   | 2   | dn              | 0        | 9        | 45       | 52888    |
| i   | 2   | dr              | 24       | 403      | 1278     | 116090   |
| i   | 2   | dv              | 4        | 43       | 1302     | 77361    |
| i   | 2   | dw              | 7        | 56       | 281      | 37457    |
| i   | 2   | dz              | 3        | 31       | 119      | 4308     |
| i   | 2   | dz <sub>b</sub> | 2        | 20       | 18       | 140      |
| i   | 2   | dz <sub>v</sub> | 3        | 79       | 75       | 1968     |
| i   | 2   | dz <sub>g</sub> | 0        | 0        | 7        | 8        |
| i   | 2   | dz <sub>v</sub> | 3        | 139      | 73       | 746      |
| i   | 2   | dz <sub>z</sub> | 0        | 9        | 14       | 71       |
| i   | 2   | fe              | 2        | 105      | 76       | 6461     |
| i   | 2   | fj              | 4        | 40       | 97       | 3665     |
| i   | 2   | fk              | 0        | 0        | 95       | 360      |
| i   | 2   | fl              | 5        | 99       | 412      | 5817     |
| i   | 2   | fp              | 6        | 299      | 420      | 12897    |
| i   | 2   | fr              | 7        | 73       | 748      | 55985    |
| i   | 2   | fs              | 2        | 130      | 116      | 451      |

|   |   |     |    |      |      |        |
|---|---|-----|----|------|------|--------|
| i | 2 | fş  | 8  | 42   | 240  | 106094 |
| i | 2 | ft  | 3  | 20   | 176  | 29169  |
| i | 2 | fts | 1  | 10   | 5    | 4218   |
| i | 2 | fte | 3  | 131  | 205  | 9890   |
| i | 2 | ftş | 6  | 41   | 170  | 61446  |
| i | 2 | fx  | 1  | 59   | 54   | 8968   |
| i | 2 | gb  | 0  | 0    | 10   | 30     |
| i | 2 | gd  | 3  | 7    | 94   | 92260  |
| i | 2 | gdz | 4  | 4    | 11   | 25280  |
| i | 2 | gj  | 1  | 11   | 161  | 34494  |
| i | 2 | gl  | 3  | 31   | 351  | 11087  |
| i | 2 | gm  | 3  | 30   | 50   | 22299  |
| i | 2 | gn  | 0  | 0    | 129  | 435    |
| i | 2 | gn  | 5  | 206  | 142  | 1936   |
| i | 2 | gr  | 49 | 749  | 1953 | 168826 |
| i | 2 | gv  | 8  | 136  | 308  | 16136  |
| i | 2 | gw  | 27 | 477  | 691  | 84713  |
| i | 2 | gz  | 0  | 0    | 13   | 58     |
| i | 2 | gz  | 0  | 0    | 3    | 5      |
| i | 2 | gz  | 9  | 151  | 394  | 15731  |
| i | 2 | kc  | 8  | 84   | 216  | 35306  |
| i | 2 | kf  | 17 | 233  | 593  | 64665  |
| i | 2 | kj  | 16 | 191  | 399  | 75143  |
| i | 2 | kl  | 35 | 643  | 1334 | 72306  |
| i | 2 | km  | 0  | 0    | 15   | 66     |
| i | 2 | kn  | 0  | 0    | 101  | 792    |
| i | 2 | kn  | 0  | 0    | 10   | 19     |
| i | 2 | kp  | 0  | 0    | 15   | 274    |
| i | 2 | kr  | 62 | 1057 | 2615 | 221134 |
| i | 2 | ks  | 0  | 0    | 112  | 986    |
| i | 2 | kş  | 10 | 151  | 549  | 26725  |
| i | 2 | kt  | 7  | 46   | 45   | 376132 |
| i | 2 | kts | 0  | 0    | 2    | 7      |
| i | 2 | kte | 0  | 0    | 7    | 129    |
| i | 2 | kw  | 10 | 360  | 392  | 12803  |
| i | 2 | lj  | 0  | 0    | 21   | 44     |
| i | 2 | ln  | 0  | 8    | 3    | 72     |
| i | 2 | ln  | 1  | 12   | 16   | 150    |
| i | 2 | lv  | 1  | 12   | 46   | 2613   |
| i | 2 | lz  | 2  | 12   | 28   | 394    |

| <b>pos</b> | <b>len</b> | <b>cl</b> | <b>dictwtyp</b> | <b>parawtyp</b> | <b>corpwtyp</b> | <b>corpwtok</b> |
|------------|------------|-----------|-----------------|-----------------|-----------------|-----------------|
| i          | 2          | mj        | 37              | 605             | 1622            | 363071          |
| i          | 2          | ml        | 3               | 20              | 80              | 2634            |
| i          | 2          | mn        | 2               | 79              | 62              | 3110            |
| i          | 2          | mp        | 2               | 2               | 84              | 47457           |
| i          | 2          | mr        | 11              | 303             | 237             | 4066            |
| i          | 2          | mş        | 0               | 0               | 29              | 1386            |
| i          | 2          | mts       | 0               | 0               | 1               | 2               |
| i          | 2          | mw        | 8               | 58              | 276             | 26552           |
| i          | 2          | mx        | 0               | 8               | 5               | 68              |
| i          | 2          | mz        | 0               | 0               | 16              | 120             |
| i          | 2          | pe        | 1               | 10              | 47              | 447             |
| i          | 2          | pj        | 68              | 724             | 2214            | 214786          |
| i          | 2          | pl        | 29              | 495             | 1116            | 66167           |
| i          | 2          | pn        | 0               | 0               | 32              | 252             |
| i          | 2          | pp        | 0               | 9               | 25              | 335             |
| i          | 2          | pr        | 136             | 2080            | 5480            | 1005968         |
| i          | 2          | ps        | 6               | 109             | 632             | 12864           |
| i          | 2          | pş        | 356             | 12081           | 16931           | 1222231         |
| i          | 2          | pt        | 3               | 30              | 73              | 1812            |
| i          | 2          | pte       | 0               | 0               | 4               | 22              |
| i          | 2          | pw        | 36              | 643             | 853             | 52695           |
| i          | 2          | px        | 1               | 71              | 54              | 1384            |
| i          | 2          | rd        | 0               | 0               | 1               | 1               |
| i          | 2          | rdz       | 3               | 21              | 56              | 509             |
| i          | 2          | rj        | 0               | 0               | 12              | 102             |
| i          | 2          | rt        | 0               | 0               | 6               | 115             |
| i          | 2          | rv        | 1               | 66              | 41              | 312             |
| i          | 2          | rz        | 0               | 0               | 17              | 131             |
| i          | 2          | se        | 1               | 11              | 15              | 39              |
| i          | 2          | sf        | 14              | 280             | 558             | 108813          |
| i          | 2          | sj        | 0               | 0               | 8               | 20              |
| i          | 2          | sk        | 43              | 1156            | 2575            | 140352          |
| i          | 2          | sl        | 0               | 0               | 94              | 1470            |
| i          | 2          | sm        | 14              | 295             | 476             | 7652            |
| i          | 2          | sn        | 4               | 100             | 143             | 1843            |
| i          | 2          | sp        | 91              | 2363            | 3062            | 359765          |
| i          | 2          | sr        | 3               | 21              | 133             | 3081            |
| i          | 2          | ss        | 2               | 130             | 112             | 447             |
| i          | 2          | sş        | 1               | 77              | 64              | 357             |

|   |   |      |     |      |      |        |
|---|---|------|-----|------|------|--------|
| i | 2 | st   | 110 | 1977 | 3812 | 442002 |
| i | 2 | sts  | 9   | 87   | 259  | 16481  |
| i | 2 | stş  | 0   | 0    | 16   | 41     |
| i | 2 | sw   | 54  | 651  | 1098 | 85852  |
| i | 2 | sx   | 13  | 378  | 382  | 7742   |
| i | 2 | şf   | 0   | 9    | 199  | 17796  |
| i | 2 | şk   | 10  | 132  | 337  | 46649  |
| i | 2 | şl   | 7   | 130  | 232  | 4758   |
| i | 2 | şm   | 5   | 51   | 137  | 642    |
| i | 2 | şn   | 3   | 29   | 63   | 455    |
| i | 2 | şp   | 8   | 141  | 289  | 12671  |
| i | 2 | şr   | 1   | 10   | 29   | 117    |
| i | 2 | şt   | 9   | 76   | 419  | 27687  |
| i | 2 | ştş  | 33  | 425  | 770  | 56754  |
| i | 2 | şw   | 0   | 14   | 10   | 899    |
| i | 2 | tf   | 14  | 173  | 246  | 32032  |
| i | 2 | tj   | 0   | 0    | 23   | 46     |
| i | 2 | tk   | 2   | 80   | 71   | 1863   |
| i | 2 | tl   | 2   | 21   | 45   | 3002   |
| i | 2 | tn   | 0   | 12   | 10   | 64     |
| i | 2 | tn   | 0   | 6    | 3    | 44     |
| i | 2 | tr   | 65  | 1292 | 2620 | 172517 |
| i | 2 | ts   | 0   | 0    | 1    | 25     |
| i | 2 | tş   | 14  | 485  | 960  | 94437  |
| i | 2 | ttş  | 0   | 0    | 11   | 183    |
| i | 2 | tw   | 14  | 310  | 288  | 15797  |
| i | 2 | tx   | 3   | 24   | 31   | 406    |
| i | 2 | tsf  | 0   | 0    | 51   | 285    |
| i | 2 | tsl  | 0   | 0    | 2    | 3      |
| i | 2 | tsm  | 2   | 22   | 50   | 2137   |
| i | 2 | tsn  | 0   | 0    | 24   | 490    |
| i | 2 | tsw  | 1   | 7    | 8    | 1946   |
| i | 2 | tef  | 2   | 70   | 65   | 1846   |
| i | 2 | tem  | 1   | 8    | 6    | 43     |
| i | 2 | tep  | 0   | 0    | 18   | 55     |
| i | 2 | tşf  | 2   | 19   | 104  | 15796  |
| i | 2 | tşk  | 0   | 0    | 6    | 45     |
| i | 2 | tşm  | 0   | 0    | 5    | 11     |
| i | 2 | tşt  | 0   | 0    | 522  | 21057  |
| i | 2 | tşte | 1   | 41   | 36   | 736    |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| i   | 2   | tst̥ | 0        | 33       | 27       | 154      |
| i   | 2   | tsw  | 2        | 17       | 100      | 49723    |
| i   | 2   | t̥sx | 0        | 0        | 6        | 20       |
| i   | 2   | vb   | 0        | 0        | 81       | 800      |
| i   | 2   | vd   | 3        | 31       | 117      | 1547     |
| i   | 2   | vdz  | 0        | 0        | 1        | 1        |
| i   | 2   | vdz  | 3        | 19       | 64       | 1358     |
| i   | 2   | vg   | 0        | 0        | 2        | 2        |
| i   | 2   | vj   | 69       | 1152     | 2761     | 357597   |
| i   | 2   | vl   | 2        | 136      | 174      | 1841     |
| i   | 2   | vm   | 2        | 130      | 99       | 515      |
| i   | 2   | vn   | 4        | 34       | 116      | 6420     |
| i   | 2   | vp   | 1        | 9        | 171      | 35160    |
| i   | 2   | vr   | 15       | 311      | 467      | 57012    |
| i   | 2   | vv   | 0        | 0        | 27       | 143      |
| i   | 2   | vw   | 20       | 338      | 665      | 160291   |
| i   | 2   | vz   | 8        | 160      | 172      | 12619    |
| i   | 2   | vz   | 1        | 48       | 53       | 7972     |
| i   | 2   | vz̥  | 8        | 236      | 211      | 17901    |
| i   | 2   | wb   | 0        | 8        | 8        | 81       |
| i   | 2   | wg   | 0        | 0        | 16       | 39       |
| i   | 2   | wk   | 0        | 0        | 10       | 20       |
| i   | 2   | wz   | 1        | 8        | 27       | 653      |
| i   | 2   | wz   | 0        | 1        | 0        | 0        |
| i   | 2   | wz̥  | 0        | 0        | 2        | 8        |
| i   | 2   | xf   | 7        | 232      | 190      | 16240    |
| i   | 2   | xj   | 0        | 0        | 81       | 1223     |
| i   | 2   | xl   | 1        | 10       | 170      | 1988     |
| i   | 2   | xm   | 3        | 32       | 35       | 679      |
| i   | 2   | xr   | 3        | 139      | 393      | 9825     |
| i   | 2   | x̥s  | 3        | 33       | 109      | 4809     |
| i   | 2   | xt   | 0        | 0        | 1        | 1        |
| i   | 2   | xt̥s | 0        | 25       | 19       | 34032    |
| i   | 2   | xt̥e | 2        | 56       | 55       | 20096    |
| i   | 2   | xw   | 9        | 155      | 270      | 9432     |
| i   | 2   | zb   | 12       | 453      | 707      | 37171    |
| i   | 2   | zd   | 17       | 766      | 1559     | 99892    |
| i   | 2   | zdz  | 1        | 60       | 166      | 3123     |
| i   | 2   | zg   | 17       | 770      | 395      | 50944    |

|   |   |      |    |     |      |        |
|---|---|------|----|-----|------|--------|
| i | 2 | zj   | 5  | 236 | 296  | 18318  |
| i | 2 | zl   | 4  | 144 | 411  | 13056  |
| i | 2 | zm   | 15 | 672 | 915  | 19833  |
| i | 2 | zn   | 19 | 486 | 637  | 129922 |
| i | 2 | zp   | 12 | 604 | 584  | 16507  |
| i | 2 | zr   | 14 | 533 | 1049 | 50468  |
| i | 2 | zv   | 21 | 813 | 981  | 58138  |
| i | 2 | zw   | 14 | 384 | 571  | 65375  |
| i | 2 | zz   | 1  | 1   | 3    | 930    |
| i | 2 | zz   | 0  | 0   | 12   | 34     |
| i | 2 | zz   | 2  | 78  | 238  | 3992   |
| i | 2 | zdz  | 0  | 0   | 3    | 5      |
| i | 2 | zl   | 1  | 2   | 3    | 3278   |
| i | 2 | zr   | 6  | 58  | 58   | 11516  |
| i | 2 | zb   | 0  | 0   | 11   | 119    |
| i | 2 | zd   | 0  | 0   | 2    | 3      |
| i | 2 | zg   | 0  | 0   | 6    | 24     |
| i | 2 | zl   | 0  | 0   | 6    | 13     |
| i | 2 | zm   | 1  | 10  | 59   | 570    |
| i | 2 | zn   | 2  | 20  | 28   | 445    |
| i | 2 | zr   | 0  | 0   | 11   | 40     |
| i | 2 | zv   | 0  | 0   | 43   | 320    |
| i | 2 | zw   | 0  | 0   | 37   | 158    |
| i | 3 | brd  | 0  | 0   | 5    | 39     |
| i | 3 | brdz | 0  | 0   | 1    | 1      |
| i | 3 | brn  | 0  | 0   | 20   | 180    |
| i | 3 | brp  | 0  | 0   | 6    | 69     |
| i | 3 | brv  | 0  | 1   | 9    | 89     |
| i | 3 | bzd  | 0  | 0   | 21   | 336    |
| i | 3 | bzdz | 0  | 0   | 1    | 1      |
| i | 3 | bzd  | 0  | 0   | 17   | 25     |
| i | 3 | bzm  | 0  | 0   | 2    | 1309   |
| i | 3 | efj  | 28 | 348 | 718  | 120651 |
| i | 3 | emj  | 14 | 238 | 252  | 19545  |
| i | 3 | epj  | 6  | 184 | 158  | 4419   |
| i | 3 | dmj  | 0  | 0   | 1    | 4      |
| i | 3 | drj  | 1  | 59  | 20   | 194    |
| i | 3 | drj  | 0  | 0   | 2    | 2      |
| i | 3 | drv  | 0  | 0   | 27   | 238    |
| i | 3 | drz  | 0  | 0   | 33   | 297    |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| i   | 3   | dvj  | 0        | 0        | 3        | 12371    |
| i   | 3   | dzv  | 1        | 1        | 5        | 2560     |
| i   | 3   | dzgn | 0        | 0        | 6        | 19       |
| i   | 3   | dzgn | 0        | 0        | 1        | 1        |
| i   | 3   | dzvj | 4        | 32       | 60       | 2073     |
| i   | 3   | fel  | 0        | 0        | 18       | 43       |
| i   | 3   | fer  | 1        | 1        | 1        | 16332    |
| i   | 3   | fete | 1        | 11       | 57       | 729      |
| i   | 3   | flk  | 1        | 12       | 46       | 66       |
| i   | 3   | fkr  | 1        | 1        | 101      | 6897     |
| i   | 3   | fkş  | 0        | 0        | 1        | 4        |
| i   | 3   | fkw  | 2        | 80       | 59       | 4998     |
| i   | 3   | fpj  | 0        | 0        | 21       | 147      |
| i   | 3   | fpl  | 0        | 0        | 58       | 228      |
| i   | 3   | fpr  | 6        | 270      | 161      | 35973    |
| i   | 3   | fpş  | 0        | 0        | 19       | 47       |
| i   | 3   | fpw  | 7        | 268      | 154      | 30218    |
| i   | 3   | fsk  | 5        | 248      | 156      | 27467    |
| i   | 3   | fsp  | 11       | 202      | 1140     | 82576    |
| i   | 3   | fsş  | 0        | 1        | 0        | 0        |
| i   | 3   | fst  | 14       | 379      | 284      | 11480    |
| i   | 3   | fsw  | 0        | 0        | 44       | 285      |
| i   | 3   | fsx  | 3        | 83       | 85       | 11901    |
| i   | 3   | fşş  | 0        | 0        | 96       | 4514     |
| i   | 3   | ftv  | 0        | 0        | 79       | 613      |
| i   | 3   | ftw  | 0        | 0        | 27       | 81       |
| i   | 3   | fxw  | 0        | 0        | 44       | 465      |
| i   | 3   | grd  | 0        | 0        | 2        | 3        |
| i   | 3   | gvd  | 0        | 0        | 3        | 7        |
| i   | 3   | gvj  | 3        | 31       | 89       | 6314     |
| i   | 3   | gzm  | 1        | 21       | 17       | 59       |
| i   | 3   | kfj  | 6        | 61       | 105      | 15699    |
| i   | 3   | kln  | 0        | 13       | 3        | 21       |
| i   | 3   | klp  | 0        | 10       | 2        | 21       |
| i   | 3   | klv  | 0        | 0        | 1        | 2        |
| i   | 3   | kmj  | 0        | 0        | 4        | 24       |
| i   | 3   | kpj  | 0        | 0        | 26       | 93       |
| i   | 3   | krf  | 2        | 21       | 76       | 3204     |
| i   | 3   | krj  | 0        | 0        | 14       | 42       |

|   |   |      |    |     |     |       |
|---|---|------|----|-----|-----|-------|
| i | 3 | krn  | 0  | 0   | 14  | 64    |
| i | 3 | krt  | 2  | 18  | 3   | 51    |
| i | 3 | kst  | 2  | 80  | 114 | 8391  |
| i | 3 | kstş | 0  | 0   | 2   | 2     |
| i | 3 | lep  | 0  | 0   | 22  | 151   |
| i | 3 | lgn  | 0  | 0   | 8   | 18    |
| i | 3 | lgn  | 0  | 0   | 1   | 2     |
| i | 3 | lvj  | 0  | 6   | 7   | 95    |
| i | 3 | mete | 1  | 37  | 37  | 224   |
| i | 3 | mdl  | 1  | 60  | 17  | 40    |
| i | 3 | mdw  | 2  | 11  | 13  | 58    |
| i | 3 | mgj  | 0  | 1   | 6   | 38    |
| i | 3 | mgl  | 1  | 13  | 14  | 260   |
| i | 3 | mgn  | 0  | 0   | 4   | 52    |
| i | 3 | mgw  | 1  | 8   | 15  | 297   |
| i | 3 | mkn  | 0  | 0   | 16  | 37    |
| i | 3 | mkp  | 0  | 0   | 2   | 17    |
| i | 3 | mst  | 0  | 0   | 3   | 10    |
| i | 3 | mstş | 0  | 33  | 12  | 99    |
| i | 3 | plv  | 0  | 0   | 4   | 8     |
| i | 3 | prj  | 0  | 0   | 30  | 1252  |
| i | 3 | psk  | 0  | 0   | 10  | 38    |
| i | 3 | pst  | 0  | 0   | 4   | 8     |
| i | 3 | pstş | 2  | 20  | 60  | 522   |
| i | 3 | pwte | 2  | 18  | 18  | 969   |
| i | 3 | pxl  | 1  | 10  | 6   | 42    |
| i | 3 | pxn  | 1  | 30  | 10  | 177   |
| i | 3 | pxp  | 0  | 30  | 12  | 118   |
| i | 3 | pxw  | 1  | 8   | 4   | 45    |
| i | 3 | rvj  | 0  | 4   | 2   | 28    |
| i | 3 | rzp  | 0  | 0   | 5   | 17    |
| i | 3 | rzp  | 0  | 0   | 6   | 11    |
| i | 3 | sfj  | 0  | 0   | 13  | 97    |
| i | 3 | sfl  | 0  | 0   | 6   | 7     |
| i | 3 | sfr  | 0  | 0   | 15  | 209   |
| i | 3 | skf  | 0  | 0   | 59  | 578   |
| i | 3 | skj  | 1  | 60  | 58  | 5818  |
| i | 3 | skl  | 4  | 88  | 148 | 9038  |
| i | 3 | skn  | 0  | 0   | 9   | 15    |
| i | 3 | skr  | 21 | 707 | 615 | 16388 |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| i   | 3   | sks  | 0        | 0        | 10       | 17       |
| i   | 3   | skş  | 12       | 223      | 295      | 5765     |
| i   | 3   | skw  | 7        | 177      | 315      | 42476    |
| i   | 3   | smr  | 0        | 0        | 21       | 118      |
| i   | 3   | spæ  | 0        | 0        | 5        | 6        |
| i   | 3   | spj  | 2        | 48       | 158      | 1556     |
| i   | 3   | spl  | 2        | 117      | 125      | 640      |
| i   | 3   | spr  | 29       | 819      | 902      | 136070   |
| i   | 3   | spş  | 10       | 316      | 466      | 80082    |
| i   | 3   | spw  | 4        | 199      | 294      | 8590     |
| i   | 3   | sst  | 0        | 0        | 26       | 195      |
| i   | 3   | stf  | 3        | 130      | 135      | 11222    |
| i   | 3   | stj  | 0        | 0        | 7        | 15       |
| i   | 3   | str  | 37       | 623      | 1061     | 107047   |
| i   | 3   | stş  | 7        | 294      | 351      | 10812    |
| i   | 3   | stw  | 2        | 120      | 83       | 420      |
| i   | 3   | stx  | 0        | 0        | 6        | 8        |
| i   | 3   | stsj | 0        | 0        | 16       | 35       |
| i   | 3   | sxf  | 0        | 0        | 29       | 311      |
| i   | 3   | sxj  | 0        | 0        | 10       | 27       |
| i   | 3   | sxl  | 0        | 0        | 40       | 167      |
| i   | 3   | sxn  | 0        | 0        | 7        | 17       |
| i   | 3   | sxp  | 0        | 0        | 2        | 9        |
| i   | 3   | sxr  | 2        | 68       | 42       | 1650     |
| i   | 3   | sxş  | 0        | 0        | 5        | 6        |
| i   | 3   | sxw  | 0        | 0        | 40       | 176      |
| i   | 3   | şfj  | 0        | 1        | 1        | 3        |
| i   | 3   | şkf  | 0        | 0        | 2        | 14       |
| i   | 3   | şkj  | 1        | 10       | 27       | 634      |
| i   | 3   | şkl  | 2        | 21       | 86       | 2286     |
| i   | 3   | şkr  | 0        | 0        | 3        | 6        |
| i   | 3   | şkw  | 0        | 0        | 5        | 1240     |
| i   | 3   | şpj  | 2        | 23       | 43       | 1241     |
| i   | 3   | şpr  | 0        | 0        | 37       | 133      |
| i   | 3   | ştr  | 0        | 0        | 8        | 19       |
| i   | 3   | ştsf | 0        | 0        | 2        | 3        |
| i   | 3   | tfj  | 2        | 67       | 46       | 22824    |
| i   | 3   | tkf  | 1        | 37       | 14       | 872      |
| i   | 3   | tkl  | 2        | 13       | 10       | 24       |

|   |   |       |   |     |     |       |
|---|---|-------|---|-----|-----|-------|
| i | 3 | tkn   | 0 | 0   | 9   | 64    |
| i | 3 | tkp   | 0 | 0   | 12  | 24    |
| i | 3 | trf   | 3 | 69  | 88  | 21651 |
| i | 3 | trj   | 3 | 22  | 117 | 1765  |
| i | 3 | t̄sn  | 0 | 0   | 4   | 17    |
| i | 3 | t̄ste | 2 | 21  | 40  | 686   |
| i | 3 | txn   | 0 | 0   | 12  | 60    |
| i | 3 | txp   | 0 | 0   | 7   | 67    |
| i | 3 | tsfj  | 0 | 0   | 3   | 8     |
| i | 3 | tskl  | 0 | 0   | 13  | 32    |
| i | 3 | tskp  | 0 | 0   | 1   | 2     |
| i | 3 | t̄cfj | 2 | 16  | 92  | 2879  |
| i | 3 | temj  | 0 | 1   | 11  | 65    |
| i | 3 | t̄skp | 0 | 0   | 1   | 1     |
| i | 3 | vbj   | 0 | 0   | 24  | 128   |
| i | 3 | vbr   | 1 | 1   | 1   | 3263  |
| i | 3 | vdm   | 0 | 0   | 3   | 3     |
| i | 3 | vdr   | 0 | 0   | 95  | 2478  |
| i | 3 | vgj   | 0 | 0   | 1   | 1     |
| i | 3 | vgl   | 0 | 0   | 5   | 461   |
| i | 3 | vgn   | 0 | 0   | 12  | 30    |
| i | 3 | vgr   | 0 | 0   | 21  | 33    |
| i | 3 | vgw   | 0 | 0   | 20  | 33    |
| i | 3 | vgz   | 0 | 0   | 1   | 1     |
| i | 3 | vmj   | 1 | 60  | 19  | 49    |
| i | 3 | vvj   | 0 | 0   | 18  | 72    |
| i | 3 | vzb   | 2 | 130 | 168 | 3725  |
| i | 3 | vzd   | 1 | 59  | 24  | 162   |
| i | 3 | vzg   | 1 | 8   | 38  | 1044  |
| i | 3 | vzl   | 0 | 0   | 19  | 104   |
| i | 3 | vzm   | 2 | 130 | 132 | 4770  |
| i | 3 | vzn   | 1 | 70  | 89  | 3336  |
| i | 3 | vzp   | 1 | 60  | 79  | 1231  |
| i | 3 | vzr   | 8 | 288 | 133 | 43653 |
| i | 3 | vzv   | 1 | 1   | 5   | 286   |
| i | 3 | vzd   | 0 | 0   | 1   | 2     |
| i | 3 | wbj   | 0 | 1   | 1   | 12    |
| i | 3 | xfj   | 0 | 0   | 31  | 183   |
| i | 3 | xmj   | 0 | 0   | 55  | 936   |
| i | 3 | x̄st  | 0 | 0   | 5   | 163   |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| i   | 3   | xstē | 0        | 0        | 26       | 152      |
| i   | 3   | xstš | 0        | 0        | 11       | 15       |
| i   | 3   | zbj  | 7        | 131      | 260      | 16467    |
| i   | 3   | zbl  | 3        | 190      | 111      | 5063     |
| i   | 3   | zbr  | 5        | 40       | 134      | 9167     |
| i   | 3   | zbw  | 0        | 0        | 20       | 43       |
| i   | 3   | zbz  | 0        | 0        | 4        | 4        |
| i   | 3   | zbz̄ | 0        | 0        | 10       | 14       |
| i   | 3   | zdj  | 2        | 48       | 54       | 9086     |
| i   | 3   | zdm  | 0        | 0        | 10       | 24       |
| i   | 3   | zdr  | 7        | 162      | 247      | 24856    |
| i   | 3   | zdv  | 0        | 0        | 8        | 30       |
| i   | 3   | zdw  | 0        | 0        | 19       | 66       |
| i   | 3   | zdz̄ | 0        | 0        | 5        | 15       |
| i   | 3   | zdzv | 0        | 0        | 3        | 3        |
| i   | 3   | zgj  | 1        | 48       | 37       | 468      |
| i   | 3   | zgl  | 0        | 0        | 14       | 63       |
| i   | 3   | zgn  | 0        | 0        | 13       | 25       |
| i   | 3   | zgn  | 3        | 116      | 51       | 191      |
| i   | 3   | zgr  | 4        | 81       | 170      | 10168    |
| i   | 3   | zgv  | 0        | 0        | 24       | 433      |
| i   | 3   | zgw  | 2        | 130      | 177      | 14235    |
| i   | 3   | zgz̄ | 1        | 10       | 67       | 415      |
| i   | 3   | zmj  | 9        | 431      | 361      | 75361    |
| i   | 3   | zmn  | 2        | 130      | 71       | 13458    |
| i   | 3   | zmr  | 1        | 9        | 32       | 281      |
| i   | 3   | zmw  | 0        | 0        | 1        | 1        |
| i   | 3   | zvj  | 12       | 433      | 565      | 82604    |
| i   | 3   | zvl  | 1        | 70       | 24       | 435      |
| i   | 3   | zvr  | 6        | 171      | 136      | 24403    |
| i   | 3   | zvw  | 2        | 11       | 12       | 13414    |
| i   | 3   | zdzb | 0        | 0        | 1        | 1        |
| i   | 3   | zgn  | 0        | 0        | 1        | 1        |
| i   | 4   | brvj | 0        | 4        | 7        | 31       |
| i   | 4   | bzmj | 0        | 0        | 38       | 3537     |
| i   | 4   | drgn | 1        | 30       | 9        | 73       |
| i   | 4   | drgn | 0        | 19       | 3        | 20       |
| i   | 4   | drvj | 0        | 0        | 8        | 38       |
| i   | 4   | dzyj | 0        | 3        | 12       | 1310     |

|   |   |       |   |     |     |       |
|---|---|-------|---|-----|-----|-------|
| i | 4 | fskr  | 0 | 0   | 1   | 66    |
| i | 4 | fskş  | 0 | 0   | 33  | 153   |
| i | 4 | fspj  | 0 | 0   | 62  | 4405  |
| i | 4 | fstr  | 3 | 23  | 16  | 138   |
| i | 4 | fstş  | 1 | 10  | 181 | 5758  |
| i | 4 | gzbj  | 2 | 20  | 20  | 331   |
| i | 4 | gzmj  | 1 | 49  | 14  | 91    |
| i | 4 | krfj  | 0 | 5   | 63  | 769   |
| i | 4 | pstr  | 0 | 0   | 38  | 191   |
| i | 4 | skfj  | 0 | 0   | 8   | 13    |
| i | 4 | skrf  | 0 | 0   | 6   | 8     |
| i | 4 | stfj  | 2 | 130 | 74  | 22170 |
| i | 4 | ştj   | 0 | 0   | 3   | 17    |
| i | 4 | tkfj  | 0 | 22  | 16  | 362   |
| i | 4 | tşmj  | 0 | 0   | 4   | 15    |
| i | 4 | tşpj  | 0 | 0   | 6   | 8     |
| i | 4 | vzbj  | 0 | 0   | 10  | 24    |
| i | 4 | vzbr  | 1 | 11  | 21  | 120   |
| i | 4 | vzdr  | 0 | 0   | 16  | 35    |
| i | 4 | vzdw  | 1 | 1   | 6   | 1118  |
| i | 4 | vzgl  | 3 | 22  | 29  | 18389 |
| i | 4 | vzmj  | 0 | 0   | 22  | 475   |
| i | 4 | zdzbl | 0 | 0   | 1   | 2     |
| i | 4 | zdzbw | 0 | 0   | 3   | 22    |

**Medial clusters**

| pos | len | cl  | dietwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-----|----------|----------|----------|----------|
| m   | 2   | bb  | 0        | 0        | 104      | 1629     |
| m   | 2   | bd  | 0        | 0        | 137      | 1243     |
| m   | 2   | bdz | 0        | 0        | 29       | 63       |
| m   | 2   | bg  | 0        | 0        | 10       | 12       |
| m   | 2   | bj  | 57       | 1989     | 3694     | 195187   |
| m   | 2   | bl  | 28       | 652      | 1620     | 128500   |
| m   | 2   | bm  | 2        | 158      | 74       | 270      |
| m   | 2   | bn  | 15       | 134      | 576      | 38127    |
| m   | 2   | bn  | 7        | 155      | 541      | 38473    |
| m   | 2   | br  | 54       | 1617     | 2429     | 151677   |
| m   | 2   | bv  | 0        | 0        | 230      | 5024     |
| m   | 2   | bw  | 5        | 179      | 410      | 3670     |
| m   | 2   | bz  | 0        | 0        | 13       | 70       |
| m   | 2   | bz  | 0        | 0        | 4        | 16       |
| m   | 2   | bz  | 5        | 89       | 505      | 20709    |
| m   | 2   | ef  | 0        | 0        | 30       | 123      |
| m   | 2   | ek  | 0        | 0        | 92       | 561      |
| m   | 2   | el  | 41       | 1900     | 1833     | 141631   |
| m   | 2   | em  | 3        | 8649     | 3148     | 48531    |
| m   | 2   | en  | 6        | 52       | 277      | 6349     |
| m   | 2   | en  | 23       | 864      | 1541     | 120904   |
| m   | 2   | ep  | 0        | 0        | 17       | 89       |
| m   | 2   | er  | 9        | 64       | 192      | 31452    |
| m   | 2   | et  | 0        | 0        | 37       | 198      |
| m   | 2   | ets | 0        | 0        | 11       | 21       |
| m   | 2   | etc | 69       | 10247    | 6375     | 538049   |
| m   | 2   | ez  | 0        | 48       | 0        | 0        |
| m   | 2   | db  | 11       | 567      | 609      | 31858    |
| m   | 2   | dd  | 9        | 467      | 444      | 19428    |
| m   | 2   | ddz | 7        | 164      | 171      | 15071    |
| m   | 2   | dg  | 0        | 0        | 230      | 6265     |
| m   | 2   | dj  | 16       | 377      | 1230     | 67738    |
| m   | 2   | dl  | 19       | 323      | 1674     | 58367    |
| m   | 2   | dm  | 8        | 182      | 510      | 32076    |
| m   | 2   | dn  | 53       | 736      | 2574     | 319117   |
| m   | 2   | dn  | 77       | 1524     | 3552     | 297480   |
| m   | 2   | dr  | 30       | 736      | 2072     | 55421    |

|   |   |     |    |     |      |       |
|---|---|-----|----|-----|------|-------|
| m | 2 | dv  | 17 | 254 | 1121 | 57061 |
| m | 2 | dw  | 28 | 930 | 1442 | 87270 |
| m | 2 | dz  | 3  | 189 | 222  | 22309 |
| m | 2 | dż  | 1  | 13  | 37   | 2247  |
| m | 2 | dź  | 5  | 159 | 351  | 8979  |
| m | 2 | dźb | 0  | 0   | 12   | 126   |
| m | 2 | dźm | 0  | 16  | 7    | 503   |
| m | 2 | dźv | 1  | 49  | 79   | 1326  |
| m | 2 | dźz | 0  | 8   | 0    | 0     |
| m | 2 | dźb | 1  | 5   | 4    | 4465  |
| m | 2 | dźg | 0  | 0   | 7    | 13    |
| m | 2 | dźm | 0  | 187 | 47   | 2135  |
| m | 2 | dźv | 0  | 0   | 37   | 238   |
| m | 2 | dźz | 0  | 95  | 1    | 1     |
| m | 2 | dźb | 3  | 20  | 48   | 20921 |
| m | 2 | dźd | 0  | 0   | 9    | 98    |
| m | 2 | dźm | 0  | 0   | 3    | 28    |
| m | 2 | dźr | 0  | 0   | 2    | 3     |
| m | 2 | fe  | 1  | 12  | 48   | 396   |
| m | 2 | fj  | 13 | 312 | 563  | 27700 |
| m | 2 | fk  | 49 | 424 | 1988 | 56392 |
| m | 2 | fl  | 9  | 120 | 316  | 3633  |
| m | 2 | fm  | 0  | 2   | 10   | 124   |
| m | 2 | fn  | 3  | 50  | 83   | 2936  |
| m | 2 | fj  | 2  | 36  | 49   | 1450  |
| m | 2 | fp  | 0  | 0   | 63   | 895   |
| m | 2 | fr  | 3  | 29  | 566  | 7628  |
| m | 2 | fs  | 0  | 9   | 32   | 323   |
| m | 2 | fş  | 7  | 984 | 901  | 38381 |
| m | 2 | ft  | 6  | 169 | 285  | 12130 |
| m | 2 | fts | 10 | 191 | 1993 | 88991 |
| m | 2 | fte | 3  | 117 | 90   | 1227  |
| m | 2 | ftş | 12 | 134 | 849  | 44853 |
| m | 2 | fx  | 0  | 0   | 21   | 75    |
| m | 2 | gb  | 1  | 5   | 33   | 4360  |
| m | 2 | gd  | 3  | 22  | 141  | 15090 |
| m | 2 | gdz | 2  | 2   | 3    | 947   |
| m | 2 | gj  | 18 | 221 | 1102 | 69306 |
| m | 2 | gl  | 20 | 616 | 1030 | 41964 |
| m | 2 | gm  | 2  | 21  | 279  | 8780  |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 2   | gn    | 26       | 659      | 1258     | 36475    |
| m   | 2   | gn    | 8        | 385      | 500      | 21894    |
| m   | 2   | gr    | 62       | 1431     | 3636     | 206764   |
| m   | 2   | gv    | 0        | 0        | 149      | 2830     |
| m   | 2   | gw    | 25       | 892      | 1327     | 87858    |
| m   | 2   | gz    | 4        | 42       | 320      | 16013    |
| m   | 2   | gz    | 6        | 89       | 371      | 80057    |
| m   | 2   | jb    | 3        | 22       | 94       | 16384    |
| m   | 2   | je    | 0        | 10       | 170      | 2526     |
| m   | 2   | jd    | 2        | 118      | 425      | 24174    |
| m   | 2   | jd̄z  | 0        | 0        | 1        | 16       |
| m   | 2   | jd̄z  | 0        | 61       | 89       | 11914    |
| m   | 2   | jd̄z  | 0        | 0        | 14       | 385      |
| m   | 2   | jf    | 0        | 0        | 43       | 269      |
| m   | 2   | jg    | 0        | 6        | 77       | 3858     |
| m   | 2   | jj    | 0        | 0        | 19       | 69       |
| m   | 2   | jk    | 12       | 124      | 758      | 16693    |
| m   | 2   | jl    | 0        | 1        | 109      | 20212    |
| m   | 2   | jm    | 9        | 3206     | 1264     | 65965    |
| m   | 2   | jn    | 103      | 1039     | 5508     | 225541   |
| m   | 2   | jñ    | 22       | 150      | 831      | 20997    |
| m   | 2   | jp    | 0        | 28       | 186      | 4742     |
| m   | 2   | jr    | 0        | 18       | 80       | 596      |
| m   | 2   | js    | 0        | 16       | 84       | 787      |
| m   | 2   | js̄   | 11       | 213      | 2669     | 91037    |
| m   | 2   | jt    | 0        | 8        | 153      | 4206     |
| m   | 2   | jts   | 1        | 27       | 279      | 17083    |
| m   | 2   | jte   | 1        | 2624     | 471      | 16141    |
| m   | 2   | jt̄s̄ | 5        | 44       | 458      | 15944    |
| m   | 2   | jv    | 1        | 12       | 165      | 26623    |
| m   | 2   | jw    | 0        | 0        | 40       | 635      |
| m   | 2   | jx    | 0        | 0        | 36       | 933      |
| m   | 2   | jz    | 0        | 4        | 102      | 1230     |
| m   | 2   | jz    | 0        | 1        | 8        | 22       |
| m   | 2   | jz̄   | 15       | 1819     | 443      | 17999    |
| m   | 2   | jē   | 8        | 141      | 206      | 7957     |
| m   | 2   | js̄   | 0        | 0        | 47       | 3979     |
| m   | 2   | jz̄   | 4        | 81       | 143      | 12441    |
| m   | 2   | jz̄   | 0        | 66       | 0        | 0        |

|   |   |                  |     |      |      |        |
|---|---|------------------|-----|------|------|--------|
| m | 2 | ke               | 0   | 3    | 173  | 3403   |
| m | 2 | kf               | 12  | 385  | 862  | 29062  |
| m | 2 | kj               | 20  | 388  | 1984 | 128863 |
| m | 2 | kk               | 5   | 29   | 104  | 4694   |
| m | 2 | kl               | 18  | 379  | 1736 | 57996  |
| m | 2 | km               | 0   | 0    | 26   | 173    |
| m | 2 | kn               | 18  | 368  | 248  | 7734   |
| m | 2 | kñ               | 3   | 424  | 245  | 4804   |
| m | 2 | kp               | 0   | 0    | 50   | 258    |
| m | 2 | kr               | 62  | 2018 | 4107 | 212187 |
| m | 2 | ks               | 5   | 67   | 1301 | 56026  |
| m | 2 | kş               | 2   | 62   | 485  | 6635   |
| m | 2 | kt               | 67  | 1071 | 3617 | 276412 |
| m | 2 | kts              | 15  | 275  | 845  | 40351  |
| m | 2 | k <sup>te</sup>  | 0   | 29   | 98   | 11981  |
| m | 2 | k <sup>ş</sup>   | 0   | 0    | 30   | 113    |
| m | 2 | kw               | 29  | 836  | 1548 | 101575 |
| m | 2 | kx               | 0   | 0    | 37   | 1031   |
| m | 2 | lb               | 5   | 41   | 267  | 36445  |
| m | 2 | le               | 0   | 1    | 19   | 135    |
| m | 2 | ld               | 3   | 198  | 376  | 11826  |
| m | 2 | ldz              | 0   | 1    | 5    | 90     |
| m | 2 | ldz              | 0   | 1    | 19   | 182    |
| m | 2 | ldz <sub>l</sub> | 0   | 0    | 1    | 1      |
| m | 2 | lf               | 3   | 22   | 183  | 4331   |
| m | 2 | lg               | 6   | 42   | 263  | 12688  |
| m | 2 | lj               | 5   | 43   | 1203 | 41335  |
| m | 2 | lk               | 40  | 330  | 1856 | 213487 |
| m | 2 | ll               | 2   | 20   | 670  | 16004  |
| m | 2 | lm               | 2   | 115  | 397  | 21721  |
| m | 2 | ln               | 116 | 1289 | 5865 | 415876 |
| m | 2 | lñ               | 82  | 753  | 2250 | 137002 |
| m | 2 | lp               | 0   | 0    | 171  | 1851   |
| m | 2 | lr               | 0   | 0    | 8    | 19     |
| m | 2 | ls               | 0   | 0    | 172  | 4970   |
| m | 2 | lş               | 1   | 10   | 249  | 13573  |
| m | 2 | lt               | 9   | 99   | 952  | 36175  |
| m | 2 | lts              | 0   | 82   | 396  | 13831  |
| m | 2 | l <sup>te</sup>  | 0   | 89   | 28   | 259    |
| m | 2 | l <sup>ş</sup>   | 6   | 146  | 526  | 21460  |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| m   | 2   | lv   | 3        | 30       | 326      | 8563     |
| m   | 2   | lx   | 1        | 10       | 78       | 2472     |
| m   | 2   | lz   | 0        | 0        | 36       | 177      |
| m   | 2   | lż   | 0        | 0        | 9        | 48       |
| m   | 2   | lż̇  | 1        | 103      | 76       | 471      |
| m   | 2   | mb   | 27       | 395      | 1489     | 37307    |
| m   | 2   | mę   | 0        | 0        | 10       | 27       |
| m   | 2   | mdź  | 0        | 0        | 133      | 2115     |
| m   | 2   | mf   | 2        | 20       | 232      | 3842     |
| m   | 2   | mj   | 87       | 2442     | 5097     | 301749   |
| m   | 2   | mk   | 12       | 111      | 409      | 13975    |
| m   | 2   | ml   | 1        | 1        | 101      | 1652     |
| m   | 2   | mm   | 0        | 24       | 146      | 2628     |
| m   | 2   | mn   | 25       | 295      | 943      | 38068    |
| m   | 2   | mp   | 17       | 318      | 865      | 36426    |
| m   | 2   | mp   | 47       | 1351     | 2507     | 148248   |
| m   | 2   | mr   | 1        | 23       | 187      | 1615     |
| m   | 2   | ms   | 0        | 0        | 59       | 679      |
| m   | 2   | mş   | 1        | 18       | 40       | 90       |
| m   | 2   | mt   | 6        | 24       | 37       | 6438     |
| m   | 2   | mts  | 0        | 27       | 110      | 21355    |
| m   | 2   | mte  | 0        | 25       | 6        | 100      |
| m   | 2   | mtş  | 5        | 34       | 138      | 20259    |
| m   | 2   | mv   | 2        | 20       | 34       | 1024     |
| m   | 2   | mw   | 0        | 0        | 33       | 83       |
| m   | 2   | mx   | 0        | 0        | 9        | 42       |
| m   | 2   | mz   | 0        | 0        | 10       | 139      |
| m   | 2   | mz   | 0        | 0        | 1        | 1        |
| m   | 2   | mż̇  | 1        | 25       | 45       | 1268     |
| m   | 2   | nb   | 0        | 0        | 61       | 777      |
| m   | 2   | ne   | 0        | 8        | 68       | 2716     |
| m   | 2   | nd   | 71       | 1425     | 4391     | 430608   |
| m   | 2   | ndź  | 19       | 1262     | 1537     | 150390   |
| m   | 2   | ndź̇ | 1        | 7        | 107      | 4314     |
| m   | 2   | ndź̇ | 0        | 0        | 15       | 85       |
| m   | 2   | nf   | 13       | 246      | 760      | 80943    |
| m   | 2   | nj   | 0        | 0        | 1        | 20       |
| m   | 2   | nl   | 0        | 0        | 28       | 4294     |
| m   | 2   | nm   | 0        | 0        | 6        | 28       |

|   |   |     |     |       |       |        |
|---|---|-----|-----|-------|-------|--------|
| m | 2 | nn  | 60  | 589   | 2164  | 211445 |
| m | 2 | np  | 25  | 325   | 765   | 57188  |
| m | 2 | np  | 0   | 0     | 2     | 52     |
| m | 2 | nr  | 0   | 0     | 54    | 5237   |
| m | 2 | ns  | 27  | 294   | 2241  | 131117 |
| m | 2 | nş  | 0   | 0     | 103   | 3332   |
| m | 2 | nt  | 109 | 3369  | 12215 | 544810 |
| m | 2 | nts | 43  | 12092 | 20158 | 564556 |
| m | 2 | nte | 3   | 71    | 429   | 34096  |
| m | 2 | ntş | 36  | 1635  | 1675  | 52108  |
| m | 2 | nv  | 8   | 85    | 643   | 72330  |
| m | 2 | nx  | 0   | 0     | 123   | 2136   |
| m | 2 | nz  | 7   | 73    | 327   | 20539  |
| m | 2 | nż  | 3   | 42    | 220   | 12452  |
| m | 2 | nb  | 0   | 0     | 13    | 211    |
| m | 2 | ndz | 25  | 731   | 675   | 144371 |
| m | 2 | nk  | 0   | 8     | 144   | 1484   |
| m | 2 | nm  | 0   | 128   | 29    | 265    |
| m | 2 | nts | 4   | 103   | 602   | 52710  |
| m | 2 | nte | 20  | 1898  | 2705  | 107291 |
| m | 2 | ntş | 12  | 398   | 1084  | 48710  |
| m | 2 | ng  | 17  | 736   | 2645  | 106689 |
| m | 2 | nk  | 91  | 1431  | 4739  | 335182 |
| m | 2 | pe  | 0   | 3     | 55    | 1189   |
| m | 2 | pf  | 2   | 14    | 43    | 723    |
| m | 2 | pj  | 43  | 1594  | 3281  | 117292 |
| m | 2 | pk  | 20  | 207   | 775   | 23337  |
| m | 2 | pl  | 10  | 300   | 1337  | 25912  |
| m | 2 | pm  | 0   | 57    | 4     | 15     |
| m | 2 | pn  | 10  | 229   | 295   | 8016   |
| m | 2 | pn  | 7   | 352   | 381   | 12296  |
| m | 2 | pp  | 0   | 0     | 93    | 3555   |
| m | 2 | pr  | 52  | 2377  | 3998  | 177510 |
| m | 2 | ps  | 7   | 156   | 737   | 18412  |
| m | 2 | pş  | 34  | 929   | 2011  | 83440  |
| m | 2 | pt  | 17  | 262   | 1137  | 23556  |
| m | 2 | pts | 8   | 138   | 332   | 27651  |
| m | 2 | pte | 3   | 172   | 257   | 11436  |
| m | 2 | ptş | 1   | 52    | 154   | 1894   |
| m | 2 | pw  | 26  | 913   | 1293  | 72781  |

| pos | len | cl  | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-----|----------|----------|----------|----------|
| m   | 2   | px  | 1        | 70       | 148      | 4006     |
| m   | 2   | pż  | 0        | 28       | 0        | 0        |
| m   | 2   | rb  | 9        | 223      | 1067     | 50228    |
| m   | 2   | re  | 1        | 20       | 57       | 4778     |
| m   | 2   | rd  | 19       | 280      | 1522     | 45413    |
| m   | 2   | rdz | 6        | 330      | 422      | 70139    |
| m   | 2   | rdz | 7        | 170      | 401      | 73556    |
| m   | 2   | rf  | 1        | 5        | 318      | 2166     |
| m   | 2   | rg  | 19       | 296      | 1427     | 119629   |
| m   | 2   | rj  | 42       | 477      | 1979     | 126807   |
| m   | 2   | rk  | 68       | 890      | 3298     | 103611   |
| m   | 2   | rl  | 3        | 121      | 737      | 41075    |
| m   | 2   | rm  | 30       | 543      | 3038     | 242846   |
| m   | 2   | rn  | 53       | 630      | 2927     | 125303   |
| m   | 2   | rn  | 40       | 431      | 2120     | 59661    |
| m   | 2   | rp  | 3        | 109      | 625      | 12421    |
| m   | 2   | rr  | 1        | 10       | 440      | 13516    |
| m   | 2   | rs  | 9        | 159      | 905      | 49208    |
| m   | 2   | rş  | 5        | 126      | 465      | 82567    |
| m   | 2   | rt  | 60       | 955      | 4076     | 286827   |
| m   | 2   | rts | 13       | 249      | 1184     | 68292    |
| m   | 2   | rte | 8        | 205      | 524      | 50816    |
| m   | 2   | rţş | 15       | 471      | 1240     | 86586    |
| m   | 2   | rv  | 29       | 818      | 1504     | 61785    |
| m   | 2   | rw  | 1        | 259      | 343      | 9726     |
| m   | 2   | rx  | 6        | 49       | 583      | 15587    |
| m   | 2   | rż  | 6        | 263      | 522      | 35450    |
| m   | 2   | se  | 0        | 0        | 84       | 669      |
| m   | 2   | sf  | 8        | 184      | 416      | 7344     |
| m   | 2   | sj  | 7        | 66       | 653      | 95651    |
| m   | 2   | sk  | 104      | 2629     | 5771     | 250518   |
| m   | 2   | sl  | 0        | 0        | 164      | 6317     |
| m   | 2   | sm  | 14       | 288      | 581      | 16071    |
| m   | 2   | sn  | 42       | 672      | 1069     | 82006    |
| m   | 2   | sn  | 0        | 0        | 1        | 1        |
| m   | 2   | sp  | 42       | 1272     | 2824     | 192378   |
| m   | 2   | sr  | 0        | 1        | 26       | 48       |
| m   | 2   | ss  | 5        | 153      | 237      | 3285     |
| m   | 2   | sş  | 2        | 137      | 162      | 4705     |

|   |   |     |     |      |       |         |
|---|---|-----|-----|------|-------|---------|
| m | 2 | st  | 259 | 5257 | 16227 | 1142579 |
| m | 2 | sts | 2   | 37   | 540   | 28792   |
| m | 2 | ste | 5   | 318  | 144   | 874     |
| m | 2 | stş | 3   | 21   | 152   | 2343    |
| m | 2 | sw  | 45  | 1559 | 2783  | 176435  |
| m | 2 | sx  | 1   | 71   | 172   | 3589    |
| m | 2 | şc  | 0   | 4    | 11    | 329     |
| m | 2 | şf  | 1   | 8    | 24    | 964     |
| m | 2 | şk  | 33  | 768  | 1765  | 103458  |
| m | 2 | şl  | 0   | 87   | 214   | 5285    |
| m | 2 | şm  | 0   | 58   | 87    | 1125    |
| m | 2 | şn  | 9   | 83   | 298   | 5857    |
| m | 2 | şp  | 6   | 33   | 238   | 5401    |
| m | 2 | şr  | 4   | 140  | 178   | 16018   |
| m | 2 | şr  | 0   | 0    | 3     | 4       |
| m | 2 | şs  | 1   | 6    | 68    | 2914    |
| m | 2 | şş  | 4   | 40   | 291   | 72321   |
| m | 2 | şt  | 14  | 315  | 557   | 72597   |
| m | 2 | sts | 0   | 13   | 95    | 1733    |
| m | 2 | ste | 2   | 245  | 60    | 8009    |
| m | 2 | stş | 37  | 1921 | 3335  | 146848  |
| m | 2 | şw  | 8   | 240  | 222   | 52920   |
| m | 2 | şx  | 0   | 16   | 73    | 1573    |
| m | 2 | şz  | 0   | 29   | 1     | 1       |
| m | 2 | te  | 2   | 19   | 183   | 39011   |
| m | 2 | tf  | 21  | 358  | 1022  | 65251   |
| m | 2 | tj  | 1   | 8    | 161   | 2949    |
| m | 2 | tk  | 80  | 888  | 3732  | 290209  |
| m | 2 | tl  | 12  | 395  | 941   | 18108   |
| m | 2 | tm  | 4   | 43   | 167   | 6011    |
| m | 2 | tn  | 34  | 415  | 1435  | 72541   |
| m | 2 | tn  | 47  | 493  | 2101  | 150946  |
| m | 2 | tp  | 23  | 774  | 1361  | 93144   |
| m | 2 | tr  | 36  | 1079 | 3447  | 104296  |
| m | 2 | ts  | 4   | 260  | 483   | 15698   |
| m | 2 | tş  | 37  | 1440 | 1714  | 116473  |
| m | 2 | tt  | 2   | 2    | 386   | 25863   |
| m | 2 | tts | 6   | 103  | 250   | 22075   |
| m | 2 | tte | 3   | 29   | 353   | 6154    |
| m | 2 | ttş | 13  | 337  | 738   | 80104   |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| m   | 2   | tw   | 6        | 425      | 467      | 6979     |
| m   | 2   | tx   | 2        | 118      | 309      | 5975     |
| m   | 2   | tsb  | 0        | 0        | 1        | 1        |
| m   | 2   | tsd  | 0        | 0        | 1        | 1        |
| m   | 2   | tsf  | 0        | 0        | 10       | 43       |
| m   | 2   | tsj  | 84       | 755      | 5564     | 618007   |
| m   | 2   | tsk  | 25       | 165      | 2670     | 79608    |
| m   | 2   | tsl  | 1        | 9        | 74       | 1726     |
| m   | 2   | tsm  | 0        | 0        | 12       | 15       |
| m   | 2   | tsn  | 8        | 72       | 296      | 34949    |
| m   | 2   | tsp  | 7        | 289      | 292      | 36559    |
| m   | 2   | tsp  | 0        | 0        | 22       | 112      |
| m   | 2   | tsr  | 0        | 0        | 36       | 348      |
| m   | 2   | tst  | 1        | 18       | 26       | 259      |
| m   | 2   | tsts | 0        | 21       | 151      | 2058     |
| m   | 2   | tste | 0        | 17       | 8        | 58       |
| m   | 2   | tstş | 0        | 0        | 24       | 198      |
| m   | 2   | tsw  | 0        | 0        | 112      | 15569    |
| m   | 2   | tsx  | 0        | 0        | 5        | 89       |
| m   | 2   | tef  | 1        | 60       | 29       | 97       |
| m   | 2   | tek  | 0        | 0        | 33       | 268      |
| m   | 2   | tem  | 0        | 95       | 39       | 1607     |
| m   | 2   | tep  | 0        | 0        | 13       | 16       |
| m   | 2   | tets | 1        | 31       | 40       | 2263     |
| m   | 2   | tete | 0        | 278      | 47       | 182      |
| m   | 2   | tetş | 0        | 0        | 9        | 25       |
| m   | 2   | tez  | 0        | 46       | 0        | 0        |
| m   | 2   | tsf  | 0        | 0        | 30       | 104      |
| m   | 2   | tsj  | 0        | 0        | 8        | 38       |
| m   | 2   | tsk  | 51       | 484      | 1952     | 29657    |
| m   | 2   | tsl  | 3        | 25       | 65       | 1266     |
| m   | 2   | tsm  | 0        | 128      | 51       | 2253     |
| m   | 2   | tsn  | 177      | 1768     | 8615     | 406234   |
| m   | 2   | tsp  | 92       | 478      | 1760     | 122048   |
| m   | 2   | tsp  | 1        | 7        | 4        | 9391     |
| m   | 2   | tst  | 4        | 40       | 102      | 6665     |
| m   | 2   | tsts | 0        | 47       | 133      | 1365     |
| m   | 2   | tste | 3        | 181      | 104      | 4148     |
| m   | 2   | tstş | 0        | 22       | 9        | 219      |

|   |   |      |     |      |       |        |
|---|---|------|-----|------|-------|--------|
| m | 2 | t̄sw | 0   | 0    | 38    | 65     |
| m | 2 | t̄sz | 0   | 64   | 0     | 0      |
| m | 2 | vb   | 0   | 0    | 44    | 324    |
| m | 2 | vd   | 6   | 80   | 144   | 30145  |
| m | 2 | v̄dz | 1   | 92   | 148   | 8033   |
| m | 2 | v̄dz | 4   | 118  | 211   | 21232  |
| m | 2 | vg   | 0   | 0    | 52    | 3587   |
| m | 2 | vj   | 151 | 5982 | 10830 | 752508 |
| m | 2 | vl   | 5   | 50   | 295   | 14819  |
| m | 2 | vm   | 0   | 90   | 33    | 369    |
| m | 2 | vn   | 86  | 1109 | 3133  | 242193 |
| m | 2 | vp   | 67  | 848  | 2732  | 227199 |
| m | 2 | vr   | 12  | 484  | 682   | 17836  |
| m | 2 | vv   | 0   | 0    | 19    | 135    |
| m | 2 | vw   | 0   | 0    | 441   | 9926   |
| m | 2 | vz   | 1   | 1    | 25    | 762    |
| m | 2 | vz   | 0   | 0    | 60    | 675    |
| m | 2 | v̄z  | 2   | 101  | 67    | 1293   |
| m | 2 | wb   | 2   | 5978 | 1980  | 19044  |
| m | 2 | wē   | 0   | 0    | 6     | 15     |
| m | 2 | wd   | 3   | 29   | 350   | 30242  |
| m | 2 | w̄dz | 0   | 0    | 1     | 8      |
| m | 2 | w̄dz | 2   | 21   | 79    | 18965  |
| m | 2 | wf   | 0   | 0    | 68    | 2098   |
| m | 2 | wg   | 0   | 7    | 250   | 9060   |
| m | 2 | wj   | 0   | 0    | 2     | 2      |
| m | 2 | wk   | 39  | 405  | 1447  | 112072 |
| m | 2 | wl   | 0   | 12   | 189   | 2100   |
| m | 2 | wm   | 1   | 20   | 194   | 3428   |
| m | 2 | wn   | 9   | 75   | 380   | 29950  |
| m | 2 | wp   | 15  | 613  | 751   | 52071  |
| m | 2 | wp   | 2   | 12   | 118   | 488    |
| m | 2 | wr   | 4   | 40   | 317   | 11196  |
| m | 2 | ws   | 0   | 0    | 72    | 2174   |
| m | 2 | wş   | 4   | 143  | 133   | 4126   |
| m | 2 | wt   | 21  | 224  | 1380  | 66526  |
| m | 2 | w̄ts | 2   | 72   | 337   | 19015  |
| m | 2 | w̄te | 2   | 70   | 130   | 3445   |
| m | 2 | w̄tş | 6   | 49   | 147   | 9071   |
| m | 2 | wv   | 1   | 11   | 97    | 839    |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| m   | 2   | ww   | 0        | 32       | 56       | 858      |
| m   | 2   | wx   | 0        | 0        | 58       | 1054     |
| m   | 2   | wz   | 2        | 68       | 163      | 2864     |
| m   | 2   | wż  | 0        | 0        | 15       | 94       |
| m   | 2   | wż  | 2        | 23       | 159      | 7966     |
| m   | 2   | ŵn   | 0        | 0        | 1        | 10       |
| m   | 2   | ŵs   | 3        | 60       | 292      | 9111     |
| m   | 2   | ŵş   | 0        | 0        | 53       | 173      |
| m   | 2   | ŵv   | 0        | 0        | 15       | 171      |
| m   | 2   | ŵx   | 3        | 148      | 82       | 331      |
| m   | 2   | ŵz   | 20       | 644      | 971      | 111279   |
| m   | 2   | ŵż  | 24       | 642      | 1360     | 42195    |
| m   | 2   | xb   | 0        | 0        | 3        | 45       |
| m   | 2   | xd   | 0        | 0        | 11       | 789      |
| m   | 2   | xđz  | 0        | 0        | 4        | 27       |
| m   | 2   | xf   | 9        | 361      | 528      | 19649    |
| m   | 2   | xj   | 0        | 0        | 57       | 2968     |
| m   | 2   | xk   | 0        | 2        | 5        | 1089     |
| m   | 2   | xl   | 3        | 41       | 284      | 1476     |
| m   | 2   | xm   | 3        | 22       | 132      | 850      |
| m   | 2   | xn   | 8        | 204      | 284      | 20055    |
| m   | 2   | xṅ  | 13       | 553      | 524      | 24942    |
| m   | 2   | xp   | 0        | 0        | 15       | 43       |
| m   | 2   | xr   | 5        | 101      | 377      | 27770    |
| m   | 2   | xs   | 0        | 0        | 12       | 16       |
| m   | 2   | xş   | 0        | 0        | 102      | 245      |
| m   | 2   | xt   | 1        | 8        | 177      | 1554     |
| m   | 2   | xts  | 1        | 10       | 22       | 978      |
| m   | 2   | xte  | 0        | 1        | 91       | 950      |
| m   | 2   | xtş  | 2        | 12       | 16       | 16822    |
| m   | 2   | xw   | 4        | 255      | 286      | 3292     |
| m   | 2   | xx   | 0        | 0        | 3        | 5        |
| m   | 2   | xż  | 0        | 0        | 4        | 164      |
| m   | 2   | zb   | 9        | 385      | 752      | 36548    |
| m   | 2   | zd   | 17       | 452      | 969      | 38520    |
| m   | 2   | zḋz | 2        | 70       | 167      | 10205    |
| m   | 2   | zg   | 5        | 223      | 483      | 10938    |
| m   | 2   | zj   | 12       | 125      | 919      | 88962    |
| m   | 2   | zl   | 4        | 117      | 394      | 12042    |

|   |   |      |    |      |      |        |
|---|---|------|----|------|------|--------|
| m | 2 | zm   | 9  | 493  | 1667 | 51338  |
| m | 2 | zn   | 36 | 1297 | 2141 | 166922 |
| m | 2 | zp   | 0  | 11   | 55   | 170    |
| m | 2 | zr   | 9  | 257  | 539  | 22894  |
| m | 2 | zv   | 29 | 980  | 1261 | 107035 |
| m | 2 | zw   | 3  | 342  | 467  | 11579  |
| m | 2 | zz   | 0  | 0    | 62   | 710    |
| m | 2 | zz   | 0  | 0    | 12   | 90     |
| m | 2 | zz   | 4  | 144  | 126  | 907    |
| m | 2 | zb   | 5  | 99   | 47   | 5105   |
| m | 2 | zdz  | 4  | 86   | 183  | 18223  |
| m | 2 | zl   | 3  | 76   | 134  | 4380   |
| m | 2 | zm   | 0  | 60   | 15   | 695    |
| m | 2 | zn   | 10 | 76   | 183  | 9037   |
| m | 2 | zp   | 15 | 383  | 744  | 52174  |
| m | 2 | zr   | 0  | 23   | 17   | 33     |
| m | 2 | zv   | 2  | 12   | 37   | 1102   |
| m | 2 | zz   | 0  | 30   | 0    | 0      |
| m | 2 | zb   | 6  | 46   | 116  | 14972  |
| m | 2 | zd   | 1  | 10   | 30   | 38057  |
| m | 2 | zdz  | 0  | 0    | 1    | 1      |
| m | 2 | zdz  | 8  | 443  | 461  | 7391   |
| m | 2 | zg   | 0  | 0    | 27   | 153    |
| m | 2 | zl   | 11 | 316  | 316  | 50088  |
| m | 2 | zm   | 0  | 183  | 84   | 778    |
| m | 2 | zn   | 20 | 228  | 751  | 132666 |
| m | 2 | zp   | 34 | 1090 | 1097 | 57668  |
| m | 2 | zr   | 0  | 0    | 12   | 28     |
| m | 2 | zv   | 1  | 8    | 23   | 174    |
| m | 2 | zw   | 0  | 0    | 25   | 40     |
| m | 2 | zz   | 0  | 93   | 0    | 0      |
| m | 3 | bbj  | 0  | 0    | 1    | 5      |
| m | 3 | bbr  | 0  | 0    | 1    | 1      |
| m | 3 | bdr  | 0  | 0    | 5    | 9      |
| m | 3 | bdzv | 0  | 0    | 9    | 15     |
| m | 3 | bgr  | 0  | 0    | 12   | 17     |
| m | 3 | blj  | 4  | 42   | 100  | 3546   |
| m | 3 | bmj  | 0  | 0    | 9    | 17     |
| m | 3 | brd  | 0  | 0    | 2    | 2      |
| m | 3 | brj  | 0  | 0    | 25   | 1045   |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 3   | brk   | 0        | 0        | 1        | 1        |
| m   | 3   | brn   | 2        | 12       | 48       | 1843     |
| m   | 3   | brp   | 0        | 1        | 17       | 87       |
| m   | 3   | bvj   | 0        | 0        | 87       | 893      |
| m   | 3   | bwb   | 0        | 0        | 2        | 2        |
| m   | 3   | bwk   | 2        | 17       | 14       | 1282     |
| m   | 3   | bwtş  | 0        | 0        | 1        | 1        |
| m   | 3   | bzb   | 0        | 0        | 1        | 1        |
| m   | 3   | bzd   | 0        | 0        | 8        | 9        |
| m   | 3   | bzn   | 0        | 0        | 1        | 1        |
| m   | 3   | bzm   | 0        | 10       | 3        | 107      |
| m   | 3   | cfj   | 19       | 785      | 934      | 39736    |
| m   | 3   | ekj   | 0        | 0        | 20       | 201      |
| m   | 3   | elm   | 0        | 22       | 6        | 66       |
| m   | 3   | eln   | 1        | 10       | 159      | 2178     |
| m   | 3   | elp   | 4        | 26       | 62       | 1693     |
| m   | 3   | elte  | 0        | 22       | 4        | 24       |
| m   | 3   | elz   | 0        | 11       | 0        | 0        |
| m   | 3   | emj   | 6        | 334      | 725      | 10184    |
| m   | 3   | epj   | 6        | 264      | 181      | 2370     |
| m   | 3   | epte  | 0        | 4        | 0        | 0        |
| m   | 3   | etem  | 0        | 40       | 16       | 349      |
| m   | 3   | etes  | 0        | 0        | 12       | 95       |
| m   | 3   | etets | 1        | 10       | 25       | 369      |
| m   | 3   | etete | 0        | 38       | 8        | 14       |
| m   | 3   | etez  | 0        | 18       | 0        | 0        |
| m   | 3   | dbj   | 3        | 99       | 223      | 9738     |
| m   | 3   | dbl   | 0        | 0        | 46       | 325      |
| m   | 3   | dbr   | 1        | 10       | 25       | 76       |
| m   | 3   | dbw   | 0        | 0        | 10       | 24       |
| m   | 3   | dbz   | 0        | 0        | 20       | 100      |
| m   | 3   | ddp   | 0        | 0        | 7        | 12       |
| m   | 3   | ddr   | 0        | 0        | 1        | 2        |
| m   | 3   | ddv   | 0        | 0        | 57       | 147      |
| m   | 3   | ddw   | 0        | 0        | 33       | 318      |
| m   | 3   | ddzv  | 0        | 0        | 11       | 33       |
| m   | 3   | dgd   | 0        | 0        | 2        | 2        |
| m   | 3   | dgj   | 0        | 0        | 3        | 7        |
| m   | 3   | dgl   | 0        | 0        | 52       | 389      |

|   |   |      |   |     |     |       |
|---|---|------|---|-----|-----|-------|
| m | 3 | dgm  | 0 | 0   | 4   | 11    |
| m | 3 | dgn  | 0 | 0   | 7   | 11    |
| m | 3 | dgr  | 1 | 11  | 185 | 2114  |
| m | 3 | dgv  | 0 | 0   | 10  | 41    |
| m | 3 | dgw  | 1 | 10  | 13  | 285   |
| m | 3 | dgz  | 2 | 136 | 93  | 406   |
| m | 3 | djm  | 0 | 0   | 5   | 87    |
| m | 3 | dln  | 0 | 0   | 1   | 13    |
| m | 3 | dlm  | 0 | 0   | 1   | 1     |
| m | 3 | dlp  | 0 | 0   | 5   | 10    |
| m | 3 | dlts | 0 | 0   | 5   | 432   |
| m | 3 | dlte | 0 | 0   | 2   | 3     |
| m | 3 | dlv  | 0 | 0   | 2   | 2     |
| m | 3 | dmj  | 9 | 83  | 497 | 34219 |
| m | 3 | dml  | 0 | 0   | 1   | 1     |
| m | 3 | dmr  | 2 | 136 | 30  | 97    |
| m | 3 | dmw  | 0 | 0   | 34  | 147   |
| m | 3 | drdz | 0 | 0   | 3   | 3     |
| m | 3 | drj  | 0 | 0   | 13  | 33    |
| m | 3 | drj  | 0 | 0   | 31  | 715   |
| m | 3 | drv  | 0 | 0   | 9   | 31    |
| m | 3 | drz  | 0 | 0   | 9   | 39    |
| m | 3 | dvdz | 2 | 108 | 17  | 72    |
| m | 3 | dvj  | 6 | 153 | 241 | 11886 |
| m | 3 | dvl  | 0 | 0   | 34  | 205   |
| m | 3 | dvr  | 4 | 142 | 128 | 3913  |
| m | 3 | dvw  | 1 | 11  | 17  | 669   |
| m | 3 | dvz  | 0 | 0   | 56  | 211   |
| m | 3 | dvz  | 0 | 0   | 2   | 6     |
| m | 3 | dwb  | 0 | 117 | 47  | 196   |
| m | 3 | dwk  | 0 | 0   | 1   | 1     |
| m | 3 | dwş  | 0 | 34  | 14  | 36    |
| m | 3 | dzg  | 0 | 0   | 1   | 1     |
| m | 3 | dzj  | 0 | 0   | 3   | 3     |
| m | 3 | dzl  | 0 | 0   | 4   | 6     |
| m | 3 | dzn  | 2 | 130 | 55  | 1411  |
| m | 3 | dzv  | 4 | 142 | 26  | 4042  |
| m | 3 | dzp  | 0 | 0   | 1   | 2     |
| m | 3 | dzv  | 0 | 0   | 1   | 2     |
| m | 3 | dzbr | 0 | 0   | 1   | 15    |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 3   | dzmj  | 0        | 0        | 7        | 31       |
| m   | 3   | dzvj  | 3        | 29       | 68       | 990      |
| m   | 3   | dzbj  | 0        | 1        | 1        | 1259     |
| m   | 3   | fel   | 0        | 0        | 3        | 9        |
| m   | 3   | fep   | 0        | 0        | 1        | 1        |
| m   | 3   | fete  | 0        | 0        | 44       | 372      |
| m   | 3   | fkj   | 0        | 4        | 24       | 199      |
| m   | 3   | fl    | 0        | 1        | 0        | 0        |
| m   | 3   | fkr   | 0        | 0        | 2        | 2        |
| m   | 3   | fkw   | 0        | 0        | 2        | 3        |
| m   | 3   | fpj   | 0        | 0        | 4        | 26       |
| m   | 3   | fpl   | 0        | 0        | 1        | 1        |
| m   | 3   | fpr   | 0        | 0        | 28       | 64       |
| m   | 3   | fpş   | 0        | 0        | 2        | 2        |
| m   | 3   | fpw   | 0        | 0        | 12       | 24       |
| m   | 3   | fsk   | 12       | 84       | 3391     | 89672    |
| m   | 3   | fsp   | 0        | 0        | 50       | 387      |
| m   | 3   | fst   | 7        | 196      | 204      | 27175    |
| m   | 3   | fsts  | 0        | 12       | 169      | 980      |
| m   | 3   | fsw   | 0        | 0        | 6        | 54       |
| m   | 3   | fsx   | 0        | 0        | 31       | 590      |
| m   | 3   | fşk   | 0        | 0        | 1        | 2        |
| m   | 3   | fşt   | 1        | 10       | 10       | 56       |
| m   | 3   | fştş  | 0        | 0        | 17       | 56       |
| m   | 3   | ftj   | 0        | 0        | 2        | 22       |
| m   | 3   | ftk   | 0        | 0        | 1        | 1        |
| m   | 3   | fr    | 0        | 0        | 2        | 3        |
| m   | 3   | ftete | 0        | 2        | 2        | 7        |
| m   | 3   | glj   | 0        | 0        | 2        | 4        |
| m   | 3   | gmj   | 0        | 0        | 5        | 20       |
| m   | 3   | grts  | 0        | 0        | 2        | 2        |
| m   | 3   | gvj   | 0        | 0        | 51       | 262      |
| m   | 3   | gwb   | 0        | 51       | 16       | 3571     |
| m   | 3   | gwş   | 0        | 12       | 3        | 7        |
| m   | 3   | gzm   | 0        | 10       | 4        | 7        |
| m   | 3   | jbj   | 0        | 1        | 16       | 497      |
| m   | 3   | jbl   | 0        | 0        | 12       | 13157    |
| m   | 3   | jbr   | 0        | 0        | 19       | 41       |
| m   | 3   | jbw   | 0        | 0        | 8        | 12       |

|   |   |      |    |     |     |       |
|---|---|------|----|-----|-----|-------|
| m | 3 | jbz  | 0  | 0   | 6   | 13    |
| m | 3 | jcl  | 0  | 0   | 5   | 8     |
| m | 3 | jete | 8  | 132 | 150 | 24505 |
| m | 3 | jdr  | 0  | 0   | 40  | 1401  |
| m | 3 | jdw  | 0  | 0   | 10  | 1003  |
| m | 3 | jdzm | 0  | 30  | 5   | 62    |
| m | 3 | jdzz | 0  | 15  | 0   | 0     |
| m | 3 | jfj  | 0  | 0   | 2   | 74    |
| m | 3 | jfs  | 0  | 0   | 8   | 13    |
| m | 3 | jftş | 0  | 1   | 10  | 758   |
| m | 3 | jgj  | 0  | 2   | 5   | 102   |
| m | 3 | jgl  | 0  | 0   | 4   | 13    |
| m | 3 | jgr  | 0  | 0   | 24  | 471   |
| m | 3 | jgv  | 0  | 0   | 7   | 19    |
| m | 3 | jgw  | 0  | 0   | 39  | 645   |
| m | 3 | jgz  | 0  | 0   | 1   | 2     |
| m | 3 | jkj  | 0  | 3   | 17  | 324   |
| m | 3 | jkl  | 0  | 0   | 2   | 2     |
| m | 3 | jkr  | 0  | 0   | 11  | 903   |
| m | 3 | kw   | 0  | 0   | 1   | 1     |
| m | 3 | jlb  | 0  | 0   | 2   | 2     |
| m | 3 | jlz  | 0  | 0   | 8   | 42    |
| m | 3 | jmj  | 1  | 34  | 83  | 9736  |
| m | 3 | jmk  | 0  | 0   | 2   | 39    |
| m | 3 | jmm  | 0  | 1   | 0   | 0     |
| m | 3 | jmn  | 2  | 2   | 14  | 19843 |
| m | 3 | jmr  | 0  | 0   | 3   | 4     |
| m | 3 | jmtc | 0  | 1   | 0   | 0     |
| m | 3 | jmw  | 0  | 1   | 12  | 1223  |
| m | 3 | jmz  | 0  | 1   | 0   | 0     |
| m | 3 | jnf  | 0  | 0   | 2   | 32    |
| m | 3 | npj  | 1  | 3   | 27  | 7091  |
| m | 3 | jpg  | 0  | 0   | 6   | 57    |
| m | 3 | jpl  | 0  | 1   | 0   | 0     |
| m | 3 | jpr  | 0  | 11  | 45  | 3263  |
| m | 3 | jps  | 0  | 10  | 47  | 147   |
| m | 3 | jpts | 0  | 0   | 1   | 26    |
| m | 3 | jpw  | 0  | 2   | 7   | 18    |
| m | 3 | jsf  | 0  | 1   | 3   | 6     |
| m | 3 | jsk  | 10 | 76  | 970 | 56055 |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 3   | jsm   | 0        | 2        | 43       | 173      |
| m   | 3   | jsn   | 0        | 1        | 0        | 0        |
| m   | 3   | jsp   | 0        | 2        | 11       | 46       |
| m   | 3   | jsr   | 0        | 0        | 6        | 14       |
| m   | 3   | jst   | 2        | 9        | 67       | 2086     |
| m   | 3   | jsts  | 6        | 43       | 170      | 49230    |
| m   | 3   | jsw   | 0        | 4        | 51       | 1468     |
| m   | 3   | jsx   | 0        | 0        | 3        | 8        |
| m   | 3   | jsk   | 0        | 1        | 1        | 2        |
| m   | 3   | jsl   | 0        | 1        | 9        | 44       |
| m   | 3   | jsm   | 0        | 2        | 1        | 1        |
| m   | 3   | jst   | 0        | 1        | 8        | 86       |
| m   | 3   | jste  | 0        | 4        | 2        | 15       |
| m   | 3   | jstę  | 0        | 6        | 25       | 113      |
| m   | 3   | jsz   | 0        | 1        | 0        | 0        |
| m   | 3   | jtf   | 0        | 1        | 7        | 31       |
| m   | 3   | jtk   | 1        | 4        | 21       | 256      |
| m   | 3   | jtl   | 0        | 0        | 3        | 3        |
| m   | 3   | jtn   | 0        | 0        | 5        | 13       |
| m   | 3   | jtr   | 0        | 4        | 34       | 1315     |
| m   | 3   | jtę   | 0        | 1        | 2        | 2        |
| m   | 3   | jttę  | 0        | 0        | 9        | 94       |
| m   | 3   | jtw   | 0        | 1        | 9        | 21       |
| m   | 3   | jtx   | 0        | 1        | 0        | 0        |
| m   | 3   | jtsk  | 0        | 0        | 5        | 34       |
| m   | 3   | jtsn  | 0        | 0        | 1        | 1        |
| m   | 3   | jtek  | 0        | 0        | 1        | 2        |
| m   | 3   | jtete | 0        | 30       | 10       | 30       |
| m   | 3   | jtęte | 0        | 0        | 2        | 2        |
| m   | 3   | jvdz  | 0        | 1        | 3        | 5        |
| m   | 3   | jvj   | 0        | 2        | 28       | 29862    |
| m   | 3   | jvr   | 0        | 0        | 5        | 5        |
| m   | 3   | jvw   | 0        | 1        | 9        | 93       |
| m   | 3   | jxl   | 0        | 0        | 10       | 17       |
| m   | 3   | jxr   | 0        | 0        | 7        | 94       |
| m   | 3   | jxs   | 0        | 0        | 7        | 170      |
| m   | 3   | jxw   | 0        | 0        | 6        | 10       |
| m   | 3   | jzb   | 0        | 0        | 20       | 164      |
| m   | 3   | jzd   | 0        | 0        | 10       | 162      |

|   |   |      |    |     |      |       |
|---|---|------|----|-----|------|-------|
| m | 3 | jzg  | 0  | 1   | 0    | 0     |
| m | 3 | jzj  | 0  | 0   | 2    | 2     |
| m | 3 | jzl  | 0  | 0   | 1    | 9     |
| m | 3 | jzn  | 0  | 2   | 21   | 120   |
| m | 3 | jzr  | 0  | 1   | 2    | 6     |
| m | 3 | jzv  | 0  | 2   | 18   | 116   |
| m | 3 | jzw  | 0  | 1   | 2    | 3     |
| m | 3 | jzl  | 0  | 0   | 13   | 214   |
| m | 3 | jzm  | 0  | 1   | 2    | 90    |
| m | 3 | jzv  | 0  | 0   | 2    | 2     |
| m | 3 | jzz  | 0  | 1   | 0    | 0     |
| m | 3 | jel  | 2  | 28  | 121  | 3206  |
| m | 3 | jem  | 0  | 9   | 1    | 2     |
| m | 3 | jep  | 1  | 19  | 79   | 1089  |
| m | 3 | jete | 4  | 50  | 122  | 39032 |
| m | 3 | jez  | 0  | 4   | 0    | 0     |
| m | 3 | jsk  | 15 | 105 | 3212 | 76804 |
| m | 3 | jst  | 0  | 0   | 11   | 41    |
| m | 3 | jsts | 0  | 15  | 189  | 2901  |
| m | 3 | jstš | 0  | 0   | 47   | 308   |
| m | 3 | jzb  | 0  | 0   | 3    | 26    |
| m | 3 | jzl  | 0  | 1   | 22   | 295   |
| m | 3 | jzm  | 0  | 3   | 1    | 2     |
| m | 3 | jzp  | 1  | 20  | 38   | 2067  |
| m | 3 | jzz  | 0  | 1   | 0    | 0     |
| m | 3 | kete | 0  | 2   | 14   | 2003  |
| m | 3 | kfj  | 0  | 0   | 48   | 222   |
| m | 3 | kfl  | 0  | 0   | 2    | 2     |
| m | 3 | kfr  | 0  | 0   | 3    | 52    |
| m | 3 | kkj  | 0  | 4   | 11   | 976   |
| m | 3 | klj  | 0  | 0   | 2    | 2     |
| m | 3 | kpj  | 0  | 0   | 6    | 11    |
| m | 3 | krf  | 0  | 0   | 56   | 196   |
| m | 3 | ksb  | 0  | 0   | 8    | 29    |
| m | 3 | ksd  | 0  | 0   | 4    | 4     |
| m | 3 | ksdz | 0  | 0   | 1    | 1     |
| m | 3 | ksf  | 0  | 0   | 18   | 292   |
| m | 3 | ksj  | 1  | 9   | 73   | 1575  |
| m | 3 | ksk  | 0  | 0   | 30   | 86    |
| m | 3 | ksl  | 0  | 0   | 43   | 2195  |

| pos | len | cl                | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------------------|----------|----------|----------|----------|
| m   | 3   | k <sup>m</sup>    | 0        | 0        | 66       | 760      |
| m   | 3   | k <sup>n</sup>    | 0        | 0        | 6        | 12       |
| m   | 3   | k <sup>ɲ</sup>    | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>p</sup>    | 3        | 31       | 407      | 30843    |
| m   | 3   | k <sup>r</sup>    | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>s</sup>    | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>ʂ</sup>    | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>t</sup>    | 0        | 16       | 180      | 6700     |
| m   | 3   | k <sup>t̃s</sup>  | 0        | 0        | 79       | 685      |
| m   | 3   | k <sup>t̃ʂ</sup>  | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>v</sup>    | 0        | 0        | 5        | 41       |
| m   | 3   | k <sup>x</sup>    | 0        | 0        | 40       | 292      |
| m   | 3   | k <sup>z</sup>    | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>ẓ</sup>   | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>ɸ</sup>    | 0        | 0        | 5        | 6        |
| m   | 3   | k <sup>t</sup>    | 3        | 60       | 284      | 11682    |
| m   | 3   | k <sup>tj</sup>   | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>tk</sup>   | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>tr</sup>   | 3        | 26       | 470      | 23510    |
| m   | 3   | k <sup>t̃</sup>   | 0        | 0        | 1        | 3        |
| m   | 3   | k <sup>tts</sup>  | 0        | 0        | 1        | 1        |
| m   | 3   | k <sup>t̃sj</sup> | 13       | 117      | 646      | 149828   |
| m   | 3   | k <sup>wb</sup>   | 0        | 39       | 10       | 28       |
| m   | 3   | k <sup>wʂ</sup>   | 0        | 9        | 2        | 2        |
| m   | 3   | l <sup>bj</sup>   | 0        | 0        | 71       | 660      |
| m   | 3   | l <sup>bl</sup>   | 0        | 0        | 27       | 1121     |
| m   | 3   | l <sup>br</sup>   | 0        | 0        | 21       | 662      |
| m   | 3   | l <sup>bw</sup>   | 0        | 0        | 14       | 222      |
| m   | 3   | l <sup>bẓ</sup>  | 0        | 0        | 54       | 1688     |
| m   | 3   | l <sup>ɸj</sup>   | 0        | 0        | 42       | 252      |
| m   | 3   | l <sup>db</sup>   | 0        | 0        | 15       | 283      |
| m   | 3   | l <sup>dj</sup>   | 0        | 0        | 7        | 125      |
| m   | 3   | l <sup>dm</sup>   | 0        | 0        | 4        | 19       |
| m   | 3   | l <sup>dv</sup>   | 0        | 0        | 1        | 1        |
| m   | 3   | l <sup>d̃zb</sup> | 0        | 0        | 13       | 303      |
| m   | 3   | l <sup>fj</sup>   | 0        | 0        | 12       | 128      |
| m   | 3   | l <sup>fk</sup>   | 0        | 0        | 1        | 1        |
| m   | 3   | l <sup>fr</sup>   | 0        | 0        | 10       | 585      |
| m   | 3   | l <sup>fʂ</sup>   | 0        | 0        | 1        | 1        |

|   |   |       |    |    |     |        |
|---|---|-------|----|----|-----|--------|
| m | 3 | lgj   | 0  | 0  | 44  | 4246   |
| m | 3 | lgn   | 0  | 0  | 7   | 57     |
| m | 3 | lgn   | 0  | 0  | 2   | 5      |
| m | 3 | lgr   | 0  | 0  | 17  | 1205   |
| m | 3 | lgw   | 0  | 0  | 1   | 2      |
| m | 3 | lgz   | 0  | 0  | 54  | 1681   |
| m | 3 | lkj   | 0  | 12 | 49  | 20671  |
| m | 3 | lkl   | 2  | 27 | 41  | 584    |
| m | 3 | lkm   | 0  | 0  | 6   | 12     |
| m | 3 | lkn   | 1  | 30 | 11  | 45     |
| m | 3 | lkn   | 0  | 19 | 6   | 40     |
| m | 3 | lkr   | 0  | 0  | 6   | 12     |
| m | 3 | lkw   | 0  | 16 | 12  | 70     |
| m | 3 | lmj   | 0  | 1  | 25  | 2969   |
| m | 3 | lmt   | 0  | 0  | 8   | 76     |
| m | 3 | lpl   | 0  | 0  | 9   | 58     |
| m | 3 | lpr   | 0  | 0  | 4   | 374    |
| m | 3 | lsj   | 0  | 0  | 13  | 181    |
| m | 3 | lsk   | 11 | 62 | 895 | 129016 |
| m | 3 | lsn   | 0  | 0  | 2   | 32     |
| m | 3 | lst   | 0  | 0  | 23  | 308    |
| m | 3 | lsts  | 0  | 9  | 52  | 41714  |
| m | 3 | lšk   | 0  | 0  | 2   | 4      |
| m | 3 | lšt   | 0  | 0  | 43  | 5021   |
| m | 3 | lštš  | 1  | 10 | 64  | 1183   |
| m | 3 | lte   | 0  | 0  | 1   | 1      |
| m | 3 | ltf   | 0  | 0  | 1   | 2      |
| m | 3 | ltk   | 0  | 0  | 4   | 17     |
| m | 3 | ltp   | 0  | 0  | 1   | 1      |
| m | 3 | ltr   | 0  | 9  | 238 | 1004   |
| m | 3 | lts   | 0  | 0  | 5   | 21     |
| m | 3 | ltš   | 0  | 1  | 1   | 4      |
| m | 3 | ltt   | 0  | 0  | 3   | 15     |
| m | 3 | ltsk  | 0  | 0  | 6   | 93     |
| m | 3 | ltsl  | 0  | 0  | 3   | 17     |
| m | 3 | ltsx  | 0  | 0  | 3   | 241    |
| m | 3 | ltsk  | 0  | 0  | 5   | 20     |
| m | 3 | ltšm  | 0  | 4  | 4   | 11     |
| m | 3 | ltšte | 0  | 4  | 2   | 4      |
| m | 3 | ltsz  | 0  | 2  | 0   | 0      |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| m   | 3   | lvj  | 7        | 33       | 54       | 8560     |
| m   | 3   | lxj  | 0        | 0        | 3        | 69       |
| m   | 3   | lxn  | 0        | 0        | 8        | 32       |
| m   | 3   | lxp  | 0        | 0        | 5        | 5        |
| m   | 3   | lzb  | 0        | 0        | 4        | 49       |
| m   | 3   | lzn  | 0        | 0        | 9        | 67       |
| m   | 3   | lzp  | 0        | 0        | 1        | 16       |
| m   | 3   | mbd  | 0        | 0        | 2        | 5        |
| m   | 3   | mbj  | 5        | 258      | 454      | 39650    |
| m   | 3   | mbi  | 0        | 0        | 66       | 593      |
| m   | 3   | mbm  | 0        | 11       | 0        | 0        |
| m   | 3   | mbn  | 0        | 8        | 74       | 6279     |
| m   | 3   | mbp  | 0        | 1        | 113      | 2003     |
| m   | 3   | mbr  | 0        | 0        | 125      | 2750     |
| m   | 3   | mbw  | 0        | 0        | 9        | 26       |
| m   | 3   | mbz  | 0        | 6        | 15       | 30       |
| m   | 3   | metc | 1        | 39       | 14       | 236      |
| m   | 3   | mdl  | 1        | 49       | 28       | 106      |
| m   | 3   | mdw  | 0        | 1        | 0        | 0        |
| m   | 3   | mfj  | 0        | 1        | 2        | 29       |
| m   | 3   | mfl  | 0        | 0        | 11       | 38       |
| m   | 3   | mgl  | 0        | 1        | 15       | 41       |
| m   | 3   | mgp  | 0        | 0        | 4        | 33       |
| m   | 3   | mkf  | 0        | 0        | 1        | 1        |
| m   | 3   | mkj  | 0        | 3        | 26       | 381      |
| m   | 3   | mkn  | 1        | 30       | 69       | 2224     |
| m   | 3   | mkp  | 0        | 30       | 70       | 7895     |
| m   | 3   | mmj  | 0        | 0        | 4        | 7        |
| m   | 3   | mpj  | 5        | 313      | 377      | 17550    |
| m   | 3   | mpk  | 3        | 39       | 117      | 1147     |
| m   | 3   | mpl  | 8        | 192      | 480      | 12197    |
| m   | 3   | mpm  | 0        | 22       | 5        | 9        |
| m   | 3   | mpn  | 7        | 90       | 127      | 20199    |
| m   | 3   | mpo  | 4        | 33       | 125      | 10110    |
| m   | 3   | mpr  | 3        | 140      | 315      | 11182    |
| m   | 3   | mps  | 0        | 0        | 8        | 62       |
| m   | 3   | mpş  | 0        | 0        | 21       | 798      |
| m   | 3   | mpt  | 0        | 0        | 23       | 558      |
| m   | 3   | mpts | 1        | 13       | 54       | 9748     |

|   |   |       |    |     |     |       |
|---|---|-------|----|-----|-----|-------|
| m | 3 | mptē  | 0  | 33  | 3   | 3     |
| m | 3 | mptš  | 1  | 10  | 63  | 5210  |
| m | 3 | mpz   | 0  | 11  | 0   | 0     |
| m | 3 | msk   | 3  | 21  | 242 | 4890  |
| m | 3 | mst   | 1  | 8   | 40  | 3568  |
| m | 3 | mst̄s | 0  | 3   | 18  | 205   |
| m | 3 | mst̄  | 0  | 0   | 1   | 1     |
| m | 3 | mst̄š | 0  | 11  | 23  | 55    |
| m | 3 | ndg   | 0  | 0   | 25  | 267   |
| m | 3 | ndj   | 2  | 17  | 98  | 13590 |
| m | 3 | ndl   | 4  | 112 | 189 | 26061 |
| m | 3 | ndn   | 11 | 140 | 333 | 20740 |
| m | 3 | ndŋ   | 11 | 217 | 256 | 18473 |
| m | 3 | ndr   | 3  | 28  | 552 | 12129 |
| m | 3 | ndv   | 0  | 0   | 12  | 100   |
| m | 3 | ndw   | 0  | 32  | 14  | 62    |
| m | 3 | ndz   | 1  | 5   | 133 | 23183 |
| m | 3 | ndzb  | 0  | 0   | 6   | 8     |
| m | 3 | ndzl  | 0  | 9   | 14  | 229   |
| m | 3 | ndzm  | 0  | 2   | 1   | 1105  |
| m | 3 | ndzn  | 0  | 0   | 10  | 96    |
| m | 3 | ndzŋ  | 0  | 0   | 13  | 56    |
| m | 3 | nfl   | 2  | 20  | 124 | 14101 |
| m | 3 | nfr   | 0  | 0   | 95  | 4192  |
| m | 3 | npr   | 0  | 0   | 7   | 85    |
| m | 3 | nsb   | 0  | 0   | 1   | 9     |
| m | 3 | nsd   | 0  | 0   | 2   | 4     |
| m | 3 | nsf   | 1  | 10  | 119 | 3072  |
| m | 3 | nsg   | 0  | 0   | 13  | 129   |
| m | 3 | nsj   | 4  | 38  | 117 | 5170  |
| m | 3 | nsk   | 0  | 0   | 8   | 378   |
| m | 3 | nsl   | 0  | 0   | 28  | 61    |
| m | 3 | nsm   | 2  | 21  | 85  | 3188  |
| m | 3 | nsn   | 0  | 0   | 5   | 15    |
| m | 3 | nsp   | 5  | 53  | 357 | 19782 |
| m | 3 | nss   | 0  | 0   | 10  | 37    |
| m | 3 | nst   | 9  | 84  | 461 | 45457 |
| m | 3 | nst̄s | 0  | 0   | 85  | 1383  |
| m | 3 | nsv   | 0  | 0   | 11  | 25    |
| m | 3 | nsx   | 0  | 0   | 1   | 1     |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 3   | nşf   | 0        | 0        | 3        | 22       |
| m   | 3   | nşl   | 0        | 0        | 3        | 5        |
| m   | 3   | nşp   | 0        | 0        | 2        | 2        |
| m   | 3   | nşt   | 0        | 0        | 36       | 277      |
| m   | 3   | nşte  | 0        | 0        | 1        | 6        |
| m   | 3   | nte   | 0        | 0        | 4        | 4        |
| m   | 3   | ntf   | 0        | 0        | 30       | 374      |
| m   | 3   | ntj   | 0        | 0        | 44       | 997      |
| m   | 3   | ntk   | 23       | 229      | 979      | 60867    |
| m   | 3   | ntl   | 0        | 0        | 71       | 443      |
| m   | 3   | ntn   | 14       | 144      | 706      | 19037    |
| m   | 3   | ntņ   | 10       | 66       | 286      | 9955     |
| m   | 3   | ntp   | 1        | 32       | 22       | 419      |
| m   | 3   | ntr   | 13       | 109      | 1238     | 82371    |
| m   | 3   | nts   | 0        | 0        | 2        | 11       |
| m   | 3   | ntş   | 1        | 8        | 172      | 3107     |
| m   | 3   | ntt   | 0        | 0        | 6        | 32       |
| m   | 3   | ntts  | 0        | 13       | 57       | 2178     |
| m   | 3   | nttş  | 0        | 0        | 50       | 635      |
| m   | 3   | ntw   | 0        | 0        | 15       | 173      |
| m   | 3   | ntx   | 0        | 0        | 2        | 145      |
| m   | 3   | ntsf  | 0        | 0        | 6        | 11       |
| m   | 3   | ntsj  | 17       | 153      | 908      | 111041   |
| m   | 3   | ntsk  | 5        | 29       | 391      | 7965     |
| m   | 3   | ntsl  | 0        | 0        | 55       | 3011     |
| m   | 3   | ntsts | 0        | 4        | 24       | 121      |
| m   | 3   | ntsw  | 0        | 0        | 5        | 34       |
| m   | 3   | ntşk  | 4        | 59       | 180      | 2013     |
| m   | 3   | ntşl  | 0        | 0        | 1        | 1        |
| m   | 3   | ntşm  | 0        | 40       | 6        | 18       |
| m   | 3   | ntşn  | 9        | 82       | 588      | 25344    |
| m   | 3   | ntşņ  | 10       | 58       | 200      | 25219    |
| m   | 3   | ntsts | 0        | 3        | 7        | 101      |
| m   | 3   | ntste | 0        | 40       | 6        | 36       |
| m   | 3   | ntşz  | 0        | 22       | 0        | 0        |
| m   | 3   | nvj   | 0        | 0        | 3        | 79       |
| m   | 3   | nrx   | 0        | 0        | 51       | 196      |
| m   | 3   | nzb   | 0        | 0        | 2        | 5        |
| m   | 3   | nzej  | 1        | 9        | 10       | 692      |

|   |   |       |    |     |     |       |
|---|---|-------|----|-----|-----|-------|
| m | 3 | nbj   | 0  | 0   | 20  | 63    |
| m | 3 | ndzb  | 0  | 0   | 2   | 2     |
| m | 3 | ndzm  | 0  | 62  | 8   | 104   |
| m | 3 | ndzv  | 0  | 0   | 1   | 2     |
| m | 3 | ndzz  | 0  | 31  | 0   | 0     |
| m | 3 | nkj   | 0  | 1   | 18  | 176   |
| m | 3 | ntedz | 0  | 0   | 150 | 2701  |
| m | 3 | ntem  | 0  | 20  | 4   | 7     |
| m | 3 | ntes  | 0  | 0   | 24  | 247   |
| m | 3 | ntete | 0  | 82  | 9   | 43    |
| m | 3 | ntez  | 0  | 10  | 0   | 0     |
| m | 3 | ntsk  | 0  | 0   | 1   | 1     |
| m | 3 | ntsm  | 0  | 12  | 4   | 21    |
| m | 3 | ntste | 0  | 12  | 1   | 1     |
| m | 3 | ntsz  | 0  | 6   | 0   | 0     |
| m | 3 | ngj   | 1  | 9   | 236 | 11341 |
| m | 3 | ngl   | 3  | 26  | 364 | 14451 |
| m | 3 | ngm   | 0  | 0   | 4   | 65    |
| m | 3 | ngn   | 10 | 287 | 294 | 14270 |
| m | 3 | ngp   | 2  | 293 | 352 | 16346 |
| m | 3 | ngr   | 0  | 0   | 121 | 7786  |
| m | 3 | ngv   | 0  | 0   | 48  | 256   |
| m | 3 | ngw   | 4  | 32  | 108 | 5738  |
| m | 3 | ngz   | 0  | 0   | 12  | 1443  |
| m | 3 | nke   | 0  | 1   | 3   | 494   |
| m | 3 | nkf   | 0  | 0   | 52  | 2189  |
| m | 3 | nkj   | 3  | 70  | 355 | 21337 |
| m | 3 | kk    | 2  | 7   | 27  | 1600  |
| m | 3 | kl    | 0  | 12  | 103 | 1324  |
| m | 3 | kn    | 7  | 120 | 100 | 7460  |
| m | 3 | kpn   | 1  | 80  | 106 | 2259  |
| m | 3 | kp    | 0  | 0   | 21  | 94    |
| m | 3 | kr    | 2  | 13  | 201 | 8578  |
| m | 3 | kks   | 0  | 0   | 6   | 25    |
| m | 3 | kks   | 5  | 258 | 215 | 86650 |
| m | 3 | kt    | 5  | 52  | 271 | 21059 |
| m | 3 | kts   | 0  | 4   | 47  | 482   |
| m | 3 | kte   | 1  | 12  | 15  | 817   |
| m | 3 | kts   | 0  | 0   | 52  | 144   |
| m | 3 | kw    | 0  | 32  | 39  | 172   |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 3   | pkx   | 0        | 0        | 14       | 386      |
| m   | 3   | pel   | 0        | 0        | 4        | 4        |
| m   | 3   | pete  | 0        | 0        | 6        | 8        |
| m   | 3   | pkj   | 0        | 16       | 56       | 3695     |
| m   | 3   | pkl   | 0        | 0        | 1        | 1        |
| m   | 3   | pkw   | 0        | 0        | 6        | 7        |
| m   | 3   | plj   | 0        | 0        | 1        | 2        |
| m   | 3   | pln   | 0        | 0        | 15       | 652      |
| m   | 3   | plp   | 0        | 0        | 1        | 1        |
| m   | 3   | prj   | 0        | 0        | 12       | 102      |
| m   | 3   | psj   | 0        | 0        | 12       | 107      |
| m   | 3   | psk   | 1        | 7        | 190      | 3799     |
| m   | 3   | psl   | 0        | 0        | 21       | 76       |
| m   | 3   | psm   | 0        | 0        | 9        | 11       |
| m   | 3   | psn   | 0        | 0        | 1        | 2        |
| m   | 3   | psp   | 0        | 0        | 1        | 2        |
| m   | 3   | pst   | 2        | 21       | 130      | 3439     |
| m   | 3   | psts  | 0        | 1        | 31       | 180      |
| m   | 3   | psw   | 5        | 161      | 89       | 7832     |
| m   | 3   | psk   | 0        | 0        | 1        | 2        |
| m   | 3   | psm   | 0        | 2        | 0        | 0        |
| m   | 3   | psn   | 1        | 12       | 4        | 7        |
| m   | 3   | psp   | 1        | 4        | 3        | 3        |
| m   | 3   | pst   | 0        | 0        | 18       | 87       |
| m   | 3   | pste  | 0        | 2        | 1        | 2        |
| m   | 3   | psts  | 0        | 0        | 9        | 18       |
| m   | 3   | psz   | 0        | 1        | 0        | 0        |
| m   | 3   | ptk   | 0        | 0        | 3        | 8        |
| m   | 3   | ptr   | 0        | 0        | 18       | 66       |
| m   | 3   | ptw   | 0        | 0        | 4        | 4        |
| m   | 3   | ptsj  | 1        | 9        | 124      | 11292    |
| m   | 3   | ptsm  | 0        | 0        | 2        | 2        |
| m   | 3   | ptsm  | 0        | 6        | 0        | 0        |
| m   | 3   | ptste | 0        | 6        | 0        | 0        |
| m   | 3   | ptsz  | 0        | 3        | 0        | 0        |
| m   | 3   | pwte  | 0        | 0        | 5        | 5        |
| m   | 3   | pxl   | 0        | 0        | 2        | 3        |
| m   | 3   | pxn   | 2        | 60       | 53       | 386      |
| m   | 3   | pxp   | 0        | 60       | 75       | 307      |

|   |   |      |   |    |     |       |
|---|---|------|---|----|-----|-------|
| m | 3 | rbj  | 0 | 5  | 76  | 1434  |
| m | 3 | rbl  | 0 | 0  | 11  | 86    |
| m | 3 | rbn  | 0 | 0  | 3   | 4     |
| m | 3 | rbp  | 0 | 0  | 22  | 541   |
| m | 3 | rbr  | 0 | 0  | 1   | 1     |
| m | 3 | ref  | 0 | 0  | 4   | 6     |
| m | 3 | rem  | 0 | 0  | 1   | 2     |
| m | 3 | rep  | 0 | 0  | 1   | 1     |
| m | 3 | rer  | 0 | 0  | 2   | 12    |
| m | 3 | rete | 1 | 21 | 30  | 549   |
| m | 3 | rdb  | 0 | 0  | 5   | 32    |
| m | 3 | rdj  | 0 | 0  | 87  | 1465  |
| m | 3 | rdl  | 3 | 26 | 23  | 818   |
| m | 3 | rdn  | 0 | 0  | 4   | 135   |
| m | 3 | rdp  | 0 | 0  | 26  | 175   |
| m | 3 | rdr  | 0 | 0  | 16  | 35    |
| m | 3 | rdv  | 0 | 0  | 6   | 59    |
| m | 3 | rdw  | 2 | 18 | 25  | 374   |
| m | 3 | rdzm | 0 | 10 | 2   | 8     |
| m | 3 | rdzz | 0 | 5  | 0   | 0     |
| m | 3 | rfe  | 0 | 1  | 2   | 856   |
| m | 3 | rfj  | 0 | 0  | 3   | 19    |
| m | 3 | rfl  | 0 | 0  | 2   | 6     |
| m | 3 | rfs  | 1 | 10 | 97  | 72855 |
| m | 3 | rft  | 0 | 0  | 3   | 18    |
| m | 3 | rfts | 1 | 19 | 13  | 14423 |
| m | 3 | rftş | 0 | 0  | 1   | 1     |
| m | 3 | rgj  | 2 | 17 | 123 | 4037  |
| m | 3 | rgl  | 0 | 0  | 15  | 23    |
| m | 3 | rgm  | 0 | 0  | 5   | 136   |
| m | 3 | rgn  | 0 | 0  | 14  | 285   |
| m | 3 | rgp  | 1 | 11 | 23  | 145   |
| m | 3 | rgr  | 0 | 0  | 16  | 61    |
| m | 3 | rgw  | 0 | 0  | 6   | 6     |
| m | 3 | rgz  | 0 | 0  | 1   | 1     |
| m | 3 | rke  | 2 | 22 | 29  | 235   |
| m | 3 | rkf  | 0 | 0  | 2   | 430   |
| m | 3 | rkj  | 1 | 36 | 233 | 6787  |
| m | 3 | rkl  | 0 | 0  | 14  | 35    |
| m | 3 | rkm  | 0 | 0  | 14  | 158   |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| m   | 3   | rkn  | 1        | 30       | 17       | 74       |
| m   | 3   | rkn  | 0        | 19       | 11       | 16       |
| m   | 3   | rkr  | 0        | 0        | 1        | 1        |
| m   | 3   | rks  | 0        | 0        | 13       | 127      |
| m   | 3   | rkt  | 0        | 0        | 35       | 405      |
| m   | 3   | rkw  | 0        | 0        | 1        | 1        |
| m   | 3   | rkx  | 0        | 0        | 1        | 1        |
| m   | 3   | rlb  | 0        | 0        | 1        | 38       |
| m   | 3   | rlj  | 0        | 0        | 3        | 37       |
| m   | 3   | rls  | 0        | 0        | 2        | 6        |
| m   | 3   | rmj  | 1        | 69       | 178      | 15068    |
| m   | 3   | rmk  | 0        | 0        | 3        | 6        |
| m   | 3   | rml  | 0        | 0        | 2        | 2        |
| m   | 3   | rmm  | 0        | 4        | 0        | 0        |
| m   | 3   | rmp  | 0        | 0        | 4        | 8        |
| m   | 3   | rmtc | 0        | 4        | 0        | 0        |
| m   | 3   | rmz  | 0        | 2        | 0        | 0        |
| m   | 3   | rns  | 0        | 0        | 3        | 37       |
| m   | 3   | rnts | 0        | 0        | 23       | 99       |
| m   | 3   | rnv  | 0        | 0        | 5        | 46       |
| m   | 3   | rnx  | 0        | 0        | 10       | 97       |
| m   | 3   | rng  | 0        | 0        | 3        | 58       |
| m   | 3   | rnk  | 1        | 14       | 20       | 253      |
| m   | 3   | rpe  | 0        | 0        | 2        | 2        |
| m   | 3   | rpj  | 2        | 103      | 165      | 5103     |
| m   | 3   | rpk  | 1        | 6        | 34       | 422      |
| m   | 3   | rpl  | 2        | 13       | 57       | 1412     |
| m   | 3   | rpm  | 0        | 6        | 0        | 0        |
| m   | 3   | rpn  | 1        | 30       | 28       | 177      |
| m   | 3   | rpn  | 1        | 50       | 70       | 13001    |
| m   | 3   | rpr  | 0        | 0        | 135      | 6000     |
| m   | 3   | rps  | 0        | 0        | 18       | 86       |
| m   | 3   | rpt  | 0        | 0        | 2        | 2        |
| m   | 3   | rpts | 0        | 1        | 18       | 157      |
| m   | 3   | rpte | 0        | 6        | 0        | 0        |
| m   | 3   | rpts | 0        | 0        | 3        | 5        |
| m   | 3   | rpw  | 0        | 0        | 4        | 9        |
| m   | 3   | rpz  | 0        | 3        | 0        | 0        |
| m   | 3   | rrj  | 0        | 0        | 1        | 6        |

|   |   |         |    |     |      |       |
|---|---|---------|----|-----|------|-------|
| m | 3 | rsf     | 0  | 0   | 31   | 176   |
| m | 3 | rsj     | 1  | 9   | 96   | 8717  |
| m | 3 | rsk     | 46 | 304 | 1931 | 37651 |
| m | 3 | rsn     | 0  | 0   | 5    | 8     |
| m | 3 | rsp     | 3  | 22  | 34   | 4403  |
| m | 3 | rst     | 1  | 11  | 27   | 192   |
| m | 3 | rst̄s   | 0  | 43  | 64   | 421   |
| m | 3 | rsw     | 0  | 0   | 1    | 1     |
| m | 3 | r̄sf    | 0  | 0   | 2    | 19    |
| m | 3 | r̄sk    | 0  | 0   | 1    | 1     |
| m | 3 | r̄sl    | 0  | 0   | 3    | 8     |
| m | 3 | r̄sm    | 0  | 2   | 1    | 1     |
| m | 3 | r̄sn    | 0  | 0   | 1    | 2     |
| m | 3 | r̄sr    | 0  | 0   | 5    | 22    |
| m | 3 | r̄st    | 4  | 40  | 77   | 2752  |
| m | 3 | r̄st̄s  | 0  | 0   | 3    | 3     |
| m | 3 | r̄st̄e  | 0  | 8   | 0    | 0     |
| m | 3 | r̄st̄ş  | 2  | 128 | 81   | 364   |
| m | 3 | r̄sz    | 0  | 1   | 0    | 0     |
| m | 3 | rte     | 0  | 0   | 2    | 4     |
| m | 3 | rtf     | 5  | 85  | 108  | 5387  |
| m | 3 | rtj     | 3  | 29  | 28   | 7377  |
| m | 3 | rtk     | 4  | 41  | 111  | 5345  |
| m | 3 | rtl     | 0  | 0   | 6    | 441   |
| m | 3 | rtm     | 1  | 10  | 20   | 361   |
| m | 3 | rtn     | 0  | 0   | 83   | 8124  |
| m | 3 | rt̄n    | 0  | 0   | 32   | 327   |
| m | 3 | rtr     | 3  | 79  | 136  | 2830  |
| m | 3 | rt̄ş    | 0  | 0   | 16   | 87    |
| m | 3 | rtts    | 0  | 3   | 4    | 196   |
| m | 3 | rt̄t̄ş  | 0  | 0   | 4    | 154   |
| m | 3 | rtw     | 0  | 0   | 8    | 1334  |
| m | 3 | rt̄sj   | 4  | 32  | 153  | 10827 |
| m | 3 | rt̄sk   | 0  | 0   | 42   | 276   |
| m | 3 | rt̄sm   | 0  | 0   | 6    | 18    |
| m | 3 | rt̄sp   | 0  | 0   | 1    | 1     |
| m | 3 | rt̄st   | 0  | 0   | 3    | 27    |
| m | 3 | rt̄st̄s | 0  | 0   | 1    | 1     |
| m | 3 | rt̄sv   | 0  | 0   | 5    | 13    |
| m | 3 | rt̄sx   | 0  | 0   | 4    | 6     |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 3   | rtd   | 0        | 0        | 2        | 4        |
| m   | 3   | rtef  | 0        | 0        | 22       | 983      |
| m   | 3   | rtek  | 0        | 0        | 2        | 2        |
| m   | 3   | rtel  | 0        | 0        | 1        | 1        |
| m   | 3   | rtem  | 0        | 4        | 3        | 5        |
| m   | 3   | rtep  | 0        | 0        | 1        | 2        |
| m   | 3   | rter  | 0        | 0        | 2        | 3        |
| m   | 3   | rtet  | 0        | 0        | 3        | 5        |
| m   | 3   | rtete | 0        | 14       | 0        | 0        |
| m   | 3   | rtez  | 0        | 2        | 0        | 0        |
| m   | 3   | rtsk  | 0        | 0        | 15       | 67       |
| m   | 3   | rtsl  | 0        | 0        | 4        | 4        |
| m   | 3   | rtsm  | 0        | 10       | 16       | 193      |
| m   | 3   | rtsn  | 0        | 0        | 16       | 85       |
| m   | 3   | rtsp  | 0        | 0        | 1        | 1        |
| m   | 3   | rtsts | 0        | 0        | 1        | 4        |
| m   | 3   | rtste | 0        | 10       | 0        | 0        |
| m   | 3   | rtsz  | 0        | 5        | 0        | 0        |
| m   | 3   | rvj   | 3        | 61       | 196      | 3117     |
| m   | 3   | rvn   | 2        | 20       | 52       | 820      |
| m   | 3   | rvp   | 0        | 2        | 36       | 218      |
| m   | 3   | rvr   | 0        | 0        | 1        | 1        |
| m   | 3   | rvz   | 0        | 0        | 1        | 1        |
| m   | 3   | rwb   | 0        | 45       | 13       | 41       |
| m   | 3   | rwş   | 0        | 13       | 10       | 18       |
| m   | 3   | rxf   | 0        | 1        | 2        | 50       |
| m   | 3   | rxj   | 0        | 0        | 24       | 487      |
| m   | 3   | rxl   | 0        | 0        | 9        | 78       |
| m   | 3   | rxn   | 0        | 0        | 2        | 4        |
| m   | 3   | rxp   | 0        | 0        | 2        | 3        |
| m   | 3   | rzb   | 0        | 0        | 15       | 813      |
| m   | 3   | rzl   | 0        | 18       | 8        | 30       |
| m   | 3   | rzn   | 3        | 46       | 14       | 50       |
| m   | 3   | rzn   | 0        | 57       | 47       | 175      |
| m   | 3   | rzw   | 0        | 48       | 17       | 39       |
| m   | 3   | rzm   | 0        | 6        | 1        | 1        |
| m   | 3   | rzp   | 0        | 0        | 34       | 91       |
| m   | 3   | rzp   | 0        | 0        | 48       | 136      |
| m   | 3   | rzq   | 0        | 3        | 0        | 0        |

|   |   |      |    |     |     |       |
|---|---|------|----|-----|-----|-------|
| m | 3 | sef  | 0  | 0   | 1   | 1     |
| m | 3 | sej  | 0  | 0   | 5   | 10    |
| m | 3 | ser  | 0  | 0   | 1   | 1     |
| m | 3 | sete | 0  | 0   | 8   | 11    |
| m | 3 | ske  | 0  | 0   | 3   | 6     |
| m | 3 | skdz | 0  | 0   | 4   | 4     |
| m | 3 | skf  | 1  | 10  | 121 | 5106  |
| m | 3 | skj  | 0  | 69  | 552 | 29902 |
| m | 3 | skl  | 2  | 14  | 114 | 375   |
| m | 3 | skm  | 0  | 0   | 2   | 5     |
| m | 3 | skr  | 3  | 90  | 426 | 3952  |
| m | 3 | skş  | 0  | 1   | 95  | 305   |
| m | 3 | skw  | 3  | 90  | 110 | 1399  |
| m | 3 | smj  | 0  | 1   | 1   | 6     |
| m | 3 | smr  | 0  | 0   | 1   | 1     |
| m | 3 | snk  | 1  | 18  | 15  | 125   |
| m | 3 | spj  | 11 | 235 | 516 | 67024 |
| m | 3 | spl  | 2  | 138 | 62  | 545   |
| m | 3 | spp  | 0  | 0   | 1   | 2     |
| m | 3 | spr  | 6  | 270 | 623 | 18932 |
| m | 3 | spş  | 0  | 0   | 197 | 2624  |
| m | 3 | spw  | 3  | 84  | 102 | 3592  |
| m | 3 | ssk  | 0  | 0   | 13  | 1051  |
| m | 3 | ssm  | 0  | 0   | 10  | 14    |
| m | 3 | ssn  | 0  | 0   | 2   | 2     |
| m | 3 | ssp  | 0  | 0   | 14  | 266   |
| m | 3 | sst  | 2  | 104 | 123 | 1489  |
| m | 3 | ssw  | 0  | 0   | 32  | 87    |
| m | 3 | sşf  | 0  | 0   | 1   | 1     |
| m | 3 | sşk  | 0  | 0   | 5   | 67    |
| m | 3 | sşm  | 0  | 0   | 1   | 11    |
| m | 3 | sşn  | 0  | 0   | 1   | 1     |
| m | 3 | sştş | 0  | 0   | 38  | 100   |
| m | 3 | stb  | 0  | 0   | 6   | 16    |
| m | 3 | std  | 0  | 0   | 6   | 12    |
| m | 3 | stf  | 5  | 38  | 275 | 22617 |
| m | 3 | stj  | 4  | 38  | 262 | 14633 |
| m | 3 | stk  | 21 | 191 | 832 | 83620 |
| m | 3 | stl  | 0  | 0   | 26  | 96    |
| m | 3 | stm  | 0  | 0   | 48  | 1175  |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 3   | stn   | 4        | 40       | 157      | 7580     |
| m   | 3   | stñ   | 5        | 87       | 357      | 41004    |
| m   | 3   | stp   | 0        | 0        | 31       | 65       |
| m   | 3   | str   | 24       | 298      | 2241     | 120285   |
| m   | 3   | sts   | 0        | 0        | 55       | 814      |
| m   | 3   | stş   | 23       | 874      | 1304     | 51062    |
| m   | 3   | stt   | 0        | 0        | 6        | 9        |
| m   | 3   | stts  | 0        | 14       | 55       | 1188     |
| m   | 3   | stw   | 0        | 0        | 3        | 4        |
| m   | 3   | stx   | 0        | 0        | 4        | 4        |
| m   | 3   | stz   | 0        | 0        | 6        | 21       |
| m   | 3   | stż  | 0        | 0        | 96       | 1893     |
| m   | 3   | stsf  | 0        | 0        | 2        | 2        |
| m   | 3   | stsw  | 0        | 0        | 8        | 335      |
| m   | 3   | stşw  | 0        | 0        | 12       | 31       |
| m   | 3   | swb   | 0        | 75       | 25       | 245      |
| m   | 3   | swş   | 0        | 21       | 5        | 6        |
| m   | 3   | sxf   | 0        | 0        | 10       | 54       |
| m   | 3   | sxl   | 0        | 12       | 5        | 6        |
| m   | 3   | sxm   | 0        | 0        | 10       | 19       |
| m   | 3   | sxn   | 2        | 32       | 6        | 20       |
| m   | 3   | sxp   | 0        | 38       | 32       | 92       |
| m   | 3   | sxr   | 0        | 0        | 1        | 1        |
| m   | 3   | sxw   | 0        | 32       | 31       | 106      |
| m   | 3   | şdżb | 0        | 0        | 1        | 2        |
| m   | 3   | şfj   | 0        | 1        | 1        | 13       |
| m   | 3   | şkf   | 0        | 0        | 8        | 51       |
| m   | 3   | şkj   | 0        | 21       | 113      | 2182     |
| m   | 3   | şkl   | 0        | 0        | 32       | 276      |
| m   | 3   | şkp   | 0        | 0        | 20       | 96       |
| m   | 3   | şlñ   | 0        | 0        | 1        | 1        |
| m   | 3   | şmj   | 0        | 0        | 2        | 3        |
| m   | 3   | şpj   | 0        | 0        | 4        | 8        |
| m   | 3   | şpl   | 0        | 0        | 2        | 2        |
| m   | 3   | şpr   | 0        | 0        | 4        | 4        |
| m   | 3   | ştk   | 1        | 8        | 10       | 442      |
| m   | 3   | ştr   | 0        | 0        | 7        | 25       |
| m   | 3   | stş   | 0        | 0        | 1        | 8        |
| m   | 3   | stts  | 0        | 1        | 0        | 0        |

|   |   |       |   |     |     |       |
|---|---|-------|---|-----|-----|-------|
| m | 3 | stsk  | 1 | 8   | 43  | 286   |
| m | 3 | stsm  | 0 | 18  | 2   | 3     |
| m | 3 | ststs | 0 | 1   | 2   | 4     |
| m | 3 | stste | 0 | 20  | 4   | 4     |
| m | 3 | stsz  | 0 | 9   | 0   | 0     |
| m | 3 | sxl   | 0 | 0   | 3   | 17    |
| m | 3 | sxn   | 0 | 0   | 4   | 37    |
| m | 3 | sxp   | 4 | 41  | 67  | 8416  |
| m | 3 | sxw   | 0 | 0   | 10  | 65    |
| m | 3 | tef   | 0 | 0   | 4   | 10    |
| m | 3 | tel   | 0 | 0   | 5   | 12    |
| m | 3 | ten   | 0 | 0   | 31  | 76    |
| m | 3 | ter   | 0 | 0   | 8   | 31    |
| m | 3 | tete  | 0 | 0   | 5   | 7     |
| m | 3 | tfj   | 7 | 520 | 454 | 28355 |
| m | 3 | tfm   | 0 | 4   | 2   | 2     |
| m | 3 | tfr   | 0 | 0   | 13  | 49    |
| m | 3 | tfş   | 0 | 0   | 1   | 1     |
| m | 3 | tftē  | 0 | 4   | 0   | 0     |
| m | 3 | tftş  | 2 | 12  | 19  | 832   |
| m | 3 | tfz   | 0 | 2   | 0   | 0     |
| m | 3 | tkf   | 1 | 38  | 15  | 113   |
| m | 3 | tkj   | 4 | 75  | 245 | 11307 |
| m | 3 | tkl   | 0 | 0   | 67  | 874   |
| m | 3 | tkn   | 2 | 60  | 88  | 2141  |
| m | 3 | tkp   | 0 | 60  | 95  | 1632  |
| m | 3 | tkp   | 0 | 0   | 2   | 2     |
| m | 3 | tkr   | 5 | 259 | 282 | 17930 |
| m | 3 | tkş   | 0 | 0   | 9   | 16    |
| m | 3 | tkw   | 1 | 70  | 139 | 1377  |
| m | 3 | tlm   | 0 | 4   | 0   | 0     |
| m | 3 | tln   | 1 | 10  | 14  | 369   |
| m | 3 | tlp   | 0 | 1   | 0   | 0     |
| m | 3 | tlte  | 0 | 4   | 1   | 1     |
| m | 3 | tlz   | 0 | 2   | 0   | 0     |
| m | 3 | tmj   | 0 | 1   | 8   | 190   |
| m | 3 | tpj   | 1 | 48  | 82  | 448   |
| m | 3 | tpl   | 0 | 0   | 12  | 21    |
| m | 3 | tpr   | 4 | 148 | 272 | 4699  |
| m | 3 | tps   | 0 | 0   | 1   | 3     |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 3   | tpş   | 1        | 11       | 64       | 303      |
| m   | 3   | tpw   | 3        | 118      | 171      | 4696     |
| m   | 3   | trf   | 6        | 250      | 229      | 7480     |
| m   | 3   | trj   | 5        | 43       | 188      | 3110     |
| m   | 3   | trk   | 0        | 0        | 15       | 749      |
| m   | 3   | tsf   | 0        | 0        | 4        | 36       |
| m   | 3   | tsk   | 0        | 0        | 85       | 769      |
| m   | 3   | tsm   | 0        | 0        | 23       | 151      |
| m   | 3   | tsp   | 0        | 0        | 8        | 145      |
| m   | 3   | tst   | 13       | 335      | 503      | 85303    |
| m   | 3   | tsts  | 0        | 0        | 3        | 5        |
| m   | 3   | tsw   | 2        | 138      | 169      | 1761     |
| m   | 3   | tşk   | 3        | 27       | 52       | 10169    |
| m   | 3   | tşl   | 0        | 0        | 1        | 3        |
| m   | 3   | tşm   | 0        | 8        | 5        | 64       |
| m   | 3   | tşn   | 2        | 22       | 56       | 1763     |
| m   | 3   | tşp   | 0        | 2        | 12       | 66       |
| m   | 3   | tşq   | 0        | 0        | 4        | 18       |
| m   | 3   | tşt   | 0        | 0        | 7        | 18       |
| m   | 3   | tşte  | 0        | 8        | 2        | 56       |
| m   | 3   | tstş  | 0        | 0        | 31       | 59       |
| m   | 3   | tşz   | 0        | 4        | 0        | 0        |
| m   | 3   | ttf   | 2        | 130      | 132      | 2829     |
| m   | 3   | ttj   | 0        | 0        | 7        | 83       |
| m   | 3   | ttl   | 0        | 0        | 5        | 21       |
| m   | 3   | ttr   | 0        | 0        | 76       | 151      |
| m   | 3   | ttş   | 2        | 130      | 105      | 1947     |
| m   | 3   | ttw   | 0        | 0        | 8        | 46       |
| m   | 3   | ttşm  | 0        | 4        | 0        | 0        |
| m   | 3   | ttşt  | 0        | 0        | 19       | 29       |
| m   | 3   | ttşte | 0        | 4        | 0        | 0        |
| m   | 3   | ttşw  | 0        | 0        | 12       | 30       |
| m   | 3   | ttşz  | 0        | 2        | 0        | 0        |
| m   | 3   | twb   | 0        | 30       | 2        | 2        |
| m   | 3   | twş   | 0        | 7        | 0        | 0        |
| m   | 3   | txf   | 0        | 0        | 37       | 145      |
| m   | 3   | txl   | 0        | 0        | 7        | 10       |
| m   | 3   | txn   | 1        | 30       | 18       | 204      |
| m   | 3   | txp   | 1        | 27       | 23       | 271      |

|   |   |      |    |    |      |       |
|---|---|------|----|----|------|-------|
| m | 3 | txr  | 0  | 0  | 1    | 1     |
| m | 3 | txş  | 0  | 0  | 9    | 16    |
| m | 3 | txw  | 0  | 0  | 12   | 111   |
| m | 3 | tzw  | 0  | 0  | 9    | 15    |
| m | 3 | tsgr | 0  | 0  | 4    | 9     |
| m | 3 | tskj | 0  | 86 | 1481 | 47577 |
| m | 3 | tskn | 0  | 0  | 7    | 41    |
| m | 3 | tskñ | 0  | 0  | 3    | 8     |
| m | 3 | tspj | 0  | 0  | 5    | 8     |
| m | 3 | tspr | 0  | 0  | 5    | 25    |
| m | 3 | tspş | 0  | 0  | 4    | 102   |
| m | 3 | tssw | 0  | 0  | 5    | 48    |
| m | 3 | tssp | 0  | 0  | 3    | 6     |
| m | 3 | tstf | 12 | 84 | 631  | 62562 |
| m | 3 | tstr | 0  | 0  | 2    | 43    |
| m | 3 | tszn | 0  | 0  | 1    | 1     |
| m | 3 | tefj | 0  | 0  | 21   | 54    |
| m | 3 | tekj | 0  | 0  | 2    | 7     |
| m | 3 | temj | 0  | 0  | 24   | 125   |
| m | 3 | tşfn | 0  | 0  | 1    | 1     |
| m | 3 | tşkj | 0  | 11 | 86   | 463   |
| m | 3 | tşmj | 0  | 0  | 1    | 44    |
| m | 3 | tştm | 0  | 0  | 1    | 1     |
| m | 3 | vbj  | 0  | 0  | 3    | 4     |
| m | 3 | vbw  | 0  | 0  | 1    | 1     |
| m | 3 | vdj  | 0  | 0  | 2    | 17    |
| m | 3 | vdr  | 0  | 0  | 3    | 3     |
| m | 3 | vdzm | 0  | 2  | 2    | 36    |
| m | 3 | vdzz | 0  | 1  | 0    | 0     |
| m | 3 | vgn  | 0  | 0  | 1    | 1     |
| m | 3 | vgr  | 0  | 0  | 8    | 15    |
| m | 3 | vgz  | 0  | 0  | 2    | 2     |
| m | 3 | vmj  | 0  | 0  | 3    | 4     |
| m | 3 | vvj  | 0  | 0  | 5    | 6     |
| m | 3 | vvw  | 0  | 0  | 3    | 7     |
| m | 3 | vvz  | 0  | 0  | 1    | 1     |
| m | 3 | vzb  | 0  | 0  | 1    | 1     |
| m | 3 | vzd  | 0  | 0  | 1    | 2     |
| m | 3 | vzg  | 0  | 0  | 1    | 1     |
| m | 3 | vzm  | 0  | 0  | 6    | 7     |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| m   | 3   | vzn  | 0        | 0        | 2        | 2        |
| m   | 3   | vzn  | 0        | 11       | 12       | 23       |
| m   | 3   | vzr  | 0        | 0        | 54       | 270      |
| m   | 3   | vzd  | 0        | 0        | 1        | 4        |
| m   | 3   | wbj  | 0        | 0        | 15       | 18       |
| m   | 3   | wbr  | 0        | 0        | 6        | 29       |
| m   | 3   | wbz  | 0        | 0        | 1        | 2        |
| m   | 3   | wbz  | 0        | 0        | 15       | 677      |
| m   | 3   | wen  | 0        | 0        | 1        | 1        |
| m   | 3   | wer  | 0        | 0        | 11       | 124      |
| m   | 3   | wetē | 0        | 0        | 6        | 172      |
| m   | 3   | wdj  | 0        | 0        | 68       | 1341     |
| m   | 3   | wdn  | 0        | 0        | 2        | 4        |
| m   | 3   | wdr  | 1        | 8        | 7        | 55       |
| m   | 3   | wdw  | 0        | 0        | 4        | 15       |
| m   | 3   | wdz  | 0        | 1        | 1        | 10       |
| m   | 3   | wfe  | 0        | 0        | 1        | 1        |
| m   | 3   | wfm  | 0        | 0        | 4        | 61       |
| m   | 3   | wfr  | 0        | 0        | 2        | 2        |
| m   | 3   | wfs  | 0        | 0        | 1        | 1        |
| m   | 3   | wgj  | 0        | 1        | 3        | 30       |
| m   | 3   | wgr  | 0        | 0        | 18       | 102      |
| m   | 3   | wgw  | 2        | 21       | 16       | 46       |
| m   | 3   | wke  | 0        | 0        | 6        | 41       |
| m   | 3   | wkj  | 0        | 21       | 91       | 3175     |
| m   | 3   | wkl  | 0        | 0        | 2        | 2        |
| m   | 3   | wkn  | 1        | 30       | 21       | 113      |
| m   | 3   | wkn  | 0        | 30       | 21       | 70       |
| m   | 3   | wkr  | 0        | 0        | 8        | 21       |
| m   | 3   | wks  | 0        | 0        | 1        | 1        |
| m   | 3   | wkts | 0        | 0        | 15       | 821      |
| m   | 3   | wkw  | 0        | 0        | 13       | 68       |
| m   | 3   | wlj  | 0        | 0        | 6        | 17       |
| m   | 3   | wmg  | 0        | 0        | 2        | 10       |
| m   | 3   | wmj  | 0        | 1        | 47       | 495      |
| m   | 3   | wmk  | 0        | 0        | 2        | 4        |
| m   | 3   | wmn  | 0        | 0        | 3        | 81       |
| m   | 3   | wmp  | 0        | 0        | 1        | 87       |
| m   | 3   | wmr  | 0        | 0        | 4        | 66       |

|   |   |       |   |    |     |       |
|---|---|-------|---|----|-----|-------|
| m | 3 | wmz   | 0 | 0  | 8   | 30    |
| m | 3 | wjm   | 0 | 8  | 0   | 0     |
| m | 3 | wjtc  | 0 | 8  | 0   | 0     |
| m | 3 | wjz   | 0 | 4  | 0   | 0     |
| m | 3 | wpj   | 0 | 6  | 20  | 66    |
| m | 3 | wpk   | 0 | 0  | 4   | 24    |
| m | 3 | wpl   | 0 | 0  | 11  | 19    |
| m | 3 | wpr   | 1 | 10 | 118 | 18486 |
| m | 3 | wps   | 0 | 0  | 58  | 559   |
| m | 3 | wpts  | 0 | 0  | 1   | 1     |
| m | 3 | wpw   | 0 | 0  | 4   | 4     |
| m | 3 | wsk   | 0 | 0  | 22  | 286   |
| m | 3 | wsn   | 0 | 0  | 1   | 1     |
| m | 3 | wsp   | 0 | 0  | 13  | 141   |
| m | 3 | wst   | 0 | 0  | 71  | 1295  |
| m | 3 | wsts  | 0 | 0  | 7   | 73    |
| m | 3 | wsw   | 0 | 0  | 13  | 18    |
| m | 3 | wsf   | 0 | 0  | 3   | 3     |
| m | 3 | wsl   | 0 | 0  | 3   | 14    |
| m | 3 | wst   | 0 | 0  | 1   | 3     |
| m | 3 | wsts  | 0 | 0  | 3   | 27    |
| m | 3 | wtf   | 0 | 0  | 43  | 846   |
| m | 3 | wtk   | 2 | 17 | 13  | 91    |
| m | 3 | wtn   | 0 | 0  | 30  | 62    |
| m | 3 | wtn   | 0 | 0  | 2   | 3     |
| m | 3 | wtr   | 0 | 0  | 2   | 2     |
| m | 3 | wts   | 0 | 0  | 1   | 3     |
| m | 3 | wtw   | 0 | 0  | 3   | 4     |
| m | 3 | wtsj  | 1 | 9  | 6   | 651   |
| m | 3 | wtsk  | 0 | 0  | 9   | 49    |
| m | 3 | wts   | 0 | 0  | 1   | 1     |
| m | 3 | wtem  | 0 | 4  | 0   | 0     |
| m | 3 | wtete | 0 | 4  | 0   | 0     |
| m | 3 | wtez  | 0 | 2  | 0   | 0     |
| m | 3 | wtsf  | 0 | 0  | 1   | 1     |
| m | 3 | wtsw  | 0 | 0  | 1   | 2     |
| m | 3 | wvj   | 1 | 18 | 55  | 506   |
| m | 3 | wwv   | 0 | 0  | 18  | 1877  |
| m | 3 | wwb   | 0 | 6  | 0   | 0     |
| m | 3 | wws   | 0 | 1  | 0   | 0     |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| m   | 3   | wxl  | 0        | 0        | 1        | 1        |
| m   | 3   | wxw  | 0        | 0        | 1        | 1        |
| m   | 3   | wzl  | 0        | 6        | 1        | 1        |
| m   | 3   | wzn  | 1        | 26       | 32       | 137      |
| m   | 3   | wzɲ  | 0        | 19       | 2        | 5        |
| m   | 3   | wzw  | 0        | 16       | 13       | 43       |
| m   | 3   | wzł  | 0        | 6        | 0        | 0        |
| m   | 3   | wzɲ  | 0        | 19       | 0        | 0        |
| m   | 3   | ẃfş  | 0        | 139      | 50       | 1110     |
| m   | 3   | ẃsk  | 9        | 80       | 269      | 54506    |
| m   | 3   | ẃsn  | 1        | 10       | 61       | 2125     |
| m   | 3   | ẃst  | 7        | 163      | 145      | 17990    |
| m   | 3   | ẃsts | 0        | 6        | 20       | 2440     |
| m   | 3   | ẃsw  | 3        | 62       | 42       | 172      |
| m   | 3   | ẃşe  | 0        | 0        | 3        | 7        |
| m   | 3   | ẃşk  | 5        | 34       | 92       | 16208    |
| m   | 3   | ẃşp  | 0        | 0        | 4        | 19       |
| m   | 3   | ẃşş  | 0        | 0        | 40       | 613      |
| m   | 3   | ẃşts | 0        | 3        | 17       | 1575     |
| m   | 3   | ẃşte | 0        | 22       | 1        | 1        |
| m   | 3   | ẃştş | 1        | 11       | 110      | 13051    |
| m   | 3   | ẃvz  | 0        | 0        | 21       | 4838     |
| m   | 3   | ẃxl  | 0        | 0        | 3        | 13       |
| m   | 3   | ẃxw  | 0        | 0        | 9        | 19       |
| m   | 3   | ẃzn  | 0        | 0        | 9        | 26       |
| m   | 3   | ẃzv  | 0        | 0        | 2        | 2        |
| m   | 3   | ẃzw  | 1        | 18       | 48       | 797      |
| m   | 3   | ẃzł  | 0        | 0        | 29       | 823      |
| m   | 3   | ẃzɲ  | 0        | 23       | 2        | 22       |
| m   | 3   | ẃzɲ  | 6        | 51       | 166      | 11383    |
| m   | 3   | ẃzɲ  | 2        | 11       | 123      | 1123     |
| m   | 3   | ẃzz  | 0        | 11       | 0        | 0        |
| m   | 3   | xfj  | 1        | 66       | 68       | 519      |
| m   | 3   | xfw  | 0        | 0        | 17       | 98       |
| m   | 3   | xmj  | 1        | 1        | 15       | 5042     |
| m   | 3   | xpş  | 0        | 0        | 1        | 1        |
| m   | 3   | xsk  | 0        | 0        | 8        | 228      |
| m   | 3   | xst  | 0        | 0        | 11       | 1045     |
| m   | 3   | xsts | 0        | 0        | 1        | 1        |

|   |   |      |    |     |     |       |
|---|---|------|----|-----|-----|-------|
| m | 3 | xsw  | 0  | 0   | 1   | 1     |
| m | 3 | xst  | 0  | 0   | 17  | 64    |
| m | 3 | xste | 0  | 0   | 12  | 87    |
| m | 3 | xstş | 0  | 0   | 24  | 98    |
| m | 3 | xtj  | 0  | 0   | 12  | 21    |
| m | 3 | xtr  | 0  | 0   | 3   | 29    |
| m | 3 | xtsk | 0  | 0   | 1   | 1     |
| m | 3 | xtşn | 0  | 0   | 1   | 1     |
| m | 3 | xwb  | 0  | 15  | 2   | 2     |
| m | 3 | xwş  | 0  | 2   | 0   | 0     |
| m | 3 | zbj  | 3  | 100 | 192 | 3383  |
| m | 3 | zbr  | 2  | 12  | 182 | 3070  |
| m | 3 | zbw  | 0  | 0   | 38  | 300   |
| m | 3 | zbz  | 0  | 0   | 7   | 12    |
| m | 3 | zdb  | 0  | 0   | 2   | 3     |
| m | 3 | zdj  | 0  | 0   | 6   | 7     |
| m | 3 | zdm  | 0  | 0   | 24  | 48    |
| m | 3 | zdn  | 1  | 10  | 60  | 563   |
| m | 3 | zdñ  | 1  | 10  | 35  | 1706  |
| m | 3 | zdr  | 11 | 346 | 281 | 3508  |
| m | 3 | zdv  | 0  | 0   | 14  | 65    |
| m | 3 | zdz  | 0  | 0   | 30  | 67    |
| m | 3 | zdzv | 0  | 0   | 5   | 35    |
| m | 3 | zgd  | 0  | 0   | 2   | 3     |
| m | 3 | zgj  | 0  | 2   | 7   | 112   |
| m | 3 | zgl  | 1  | 59  | 28  | 261   |
| m | 3 | zgn  | 1  | 30  | 19  | 53    |
| m | 3 | zgñ  | 1  | 90  | 34  | 63    |
| m | 3 | zgr  | 1  | 70  | 158 | 4301  |
| m | 3 | zgv  | 0  | 0   | 2   | 2     |
| m | 3 | zgw  | 0  | 0   | 69  | 1490  |
| m | 3 | zgz  | 2  | 136 | 105 | 551   |
| m | 3 | zmj  | 1  | 49  | 365 | 5312  |
| m | 3 | zml  | 0  | 0   | 1   | 1     |
| m | 3 | zmn  | 2  | 130 | 38  | 223   |
| m | 3 | zmñ  | 0  | 0   | 3   | 4     |
| m | 3 | zmr  | 0  | 0   | 19  | 38    |
| m | 3 | zvj  | 5  | 347 | 301 | 22288 |
| m | 3 | zvl  | 0  | 0   | 20  | 48    |
| m | 3 | zvr  | 0  | 0   | 12  | 18    |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 3   | zvw   | 0        | 0        | 102      | 2135     |
| m   | 3   | zvz   | 0        | 0        | 2        | 6        |
| m   | 3   | zwb   | 0        | 27       | 8        | 94       |
| m   | 3   | zws   | 0        | 7        | 2        | 21       |
| m   | 3   | zzw   | 0        | 0        | 12       | 35       |
| m   | 3   | zby   | 2        | 80       | 67       | 820      |
| m   | 3   | zby   | 0        | 4        | 0        | 0        |
| m   | 3   | zbz   | 0        | 2        | 0        | 0        |
| m   | 3   | zdzdz | 0        | 0        | 80       | 1041     |
| m   | 3   | zdzm  | 0        | 3        | 1        | 2        |
| m   | 3   | zdzs  | 0        | 1        | 0        | 0        |
| m   | 3   | zmj   | 0        | 4        | 12       | 1233     |
| m   | 3   | zjm   | 0        | 2        | 0        | 0        |
| m   | 3   | zjte  | 0        | 2        | 0        | 0        |
| m   | 3   | zjz   | 0        | 1        | 0        | 0        |
| m   | 3   | zvj   | 0        | 2        | 56       | 517      |
| m   | 3   | zby   | 0        | 2        | 13       | 1415     |
| m   | 3   | zbn   | 0        | 0        | 6        | 13       |
| m   | 3   | zbn   | 0        | 0        | 2        | 6        |
| m   | 3   | zdzm  | 0        | 2        | 0        | 0        |
| m   | 3   | zdzs  | 0        | 1        | 0        | 0        |
| m   | 3   | zgn   | 0        | 0        | 8        | 17       |
| m   | 3   | zgn   | 0        | 0        | 5        | 10       |
| m   | 3   | zmj   | 0        | 0        | 19       | 42       |
| m   | 3   | zvj   | 0        | 1        | 32       | 379      |
| m   | 4   | brsk  | 0        | 0        | 1        | 4        |
| m   | 4   | brvj  | 0        | 0        | 1        | 1        |
| m   | 4   | bwkj  | 0        | 1        | 1        | 8        |
| m   | 4   | bzdr  | 0        | 0        | 1        | 1        |
| m   | 4   | bzmj  | 1        | 39       | 71       | 454      |
| m   | 4   | ddzvj | 0        | 0        | 22       | 104      |
| m   | 4   | dgvj  | 0        | 0        | 2        | 3        |
| m   | 4   | drvj  | 0        | 0        | 3        | 3        |
| m   | 4   | dvzg  | 0        | 0        | 1        | 3        |
| m   | 4   | dvzr  | 0        | 0        | 1        | 1        |
| m   | 4   | dzbj  | 0        | 0        | 3        | 5        |
| m   | 4   | dzvj  | 0        | 0        | 48       | 1030     |
| m   | 4   | dzvr  | 0        | 0        | 5        | 16       |
| m   | 4   | dzvj  | 0        | 0        | 33       | 167      |

|   |   |       |    |     |      |       |
|---|---|-------|----|-----|------|-------|
| m | 4 | ffsk  | 0  | 0   | 8    | 161   |
| m | 4 | fskj  | 0  | 48  | 2028 | 49144 |
| m | 4 | fspj  | 0  | 0   | 2    | 2     |
| m | 4 | fstf  | 1  | 7   | 88   | 2065  |
| m | 4 | fstr  | 0  | 0   | 3    | 6     |
| m | 4 | fstş  | 0  | 0   | 82   | 1879  |
| m | 4 | fşkj  | 0  | 0   | 1    | 5     |
| m | 4 | gzbj  | 0  | 0   | 1    | 1     |
| m | 4 | gzmj  | 1  | 39  | 4    | 17    |
| m | 4 | jbgv  | 0  | 0   | 1    | 1     |
| m | 4 | jefj  | 0  | 0   | 28   | 307   |
| m | 4 | jemj  | 0  | 0   | 15   | 110   |
| m | 4 | jfsp  | 0  | 0   | 14   | 187   |
| m | 4 | jfst  | 0  | 0   | 2    | 2     |
| m | 4 | jkrf  | 0  | 0   | 8    | 22    |
| m | 4 | jmkj  | 0  | 0   | 1    | 4     |
| m | 4 | jpfr  | 0  | 0   | 5    | 13    |
| m | 4 | jskf  | 0  | 0   | 1    | 1     |
| m | 4 | jskj  | 0  | 33  | 545  | 37449 |
| m | 4 | jskr  | 0  | 2   | 19   | 73    |
| m | 4 | jspr  | 0  | 3   | 18   | 60    |
| m | 4 | jstf  | 1  | 7   | 63   | 5036  |
| m | 4 | jstr  | 0  | 11  | 74   | 486   |
| m | 4 | jstş  | 0  | 2   | 1    | 3     |
| m | 4 | jsxl  | 0  | 0   | 1    | 1     |
| m | 4 | jtkj  | 0  | 0   | 5    | 78    |
| m | 4 | jtkl  | 0  | 1   | 4    | 11    |
| m | 4 | jtrf  | 0  | 1   | 8    | 41    |
| m | 4 | jtskj | 0  | 0   | 3    | 19    |
| m | 4 | jvzn  | 0  | 0   | 4    | 6     |
| m | 4 | jzbr  | 0  | 0   | 1    | 1     |
| m | 4 | jzdr  | 0  | 1   | 9    | 28    |
| m | 4 | jzgr  | 0  | 1   | 1    | 1     |
| m | 4 | şskj  | 0  | 60  | 1920 | 45969 |
| m | 4 | şstf  | 13 | 144 | 381  | 87282 |
| m | 4 | krfj  | 0  | 0   | 22   | 108   |
| m | 4 | krsk  | 0  | 0   | 3    | 4     |
| m | 4 | ksbr  | 0  | 0   | 1    | 1     |
| m | 4 | kske  | 0  | 0   | 1    | 1     |
| m | 4 | kskl  | 0  | 0   | 18   | 393   |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 4   | kskr  | 0        | 0        | 4        | 14       |
| m   | 4   | kspj  | 0        | 0        | 9        | 43       |
| m   | 4   | kspl  | 0        | 0        | 133      | 4677     |
| m   | 4   | kspr  | 2        | 22       | 101      | 2176     |
| m   | 4   | kspş  | 0        | 0        | 3        | 5        |
| m   | 4   | kssp  | 0        | 0        | 1        | 1        |
| m   | 4   | kstr  | 0        | 0        | 160      | 2580     |
| m   | 4   | ksvj  | 0        | 0        | 5        | 6        |
| m   | 4   | ksxw  | 0        | 0        | 1        | 1        |
| m   | 4   | lbmj  | 0        | 0        | 3        | 8        |
| m   | 4   | lgzd  | 0        | 0        | 4        | 10       |
| m   | 4   | lkwb  | 0        | 3        | 0        | 0        |
| m   | 4   | lkwş  | 0        | 1        | 0        | 0        |
| m   | 4   | lmsk  | 0        | 0        | 8        | 33       |
| m   | 4   | lskj  | 0        | 36       | 481      | 61242    |
| m   | 4   | lstf  | 1        | 7        | 99       | 3134     |
| m   | 4   | lstr  | 0        | 0        | 3        | 51       |
| m   | 4   | ltkr  | 0        | 0        | 1        | 2        |
| m   | 4   | ltst  | 0        | 0        | 2        | 3        |
| m   | 4   | ltskj | 0        | 0        | 4        | 47       |
| m   | 4   | ltskj | 0        | 0        | 2        | 3        |
| m   | 4   | lzdr  | 0        | 0        | 1        | 1        |
| m   | 4   | lzbj  | 0        | 0        | 14       | 2425     |
| m   | 4   | mblj  | 0        | 0        | 3        | 4        |
| m   | 4   | mbrj  | 0        | 0        | 29       | 370      |
| m   | 4   | mbrn  | 0        | 0        | 12       | 61       |
| m   | 4   | mbrn  | 0        | 0        | 2        | 3        |
| m   | 4   | mpkj  | 0        | 2        | 8        | 65       |
| m   | 4   | mpsk  | 0        | 0        | 33       | 752      |
| m   | 4   | mpsts | 0        | 0        | 4        | 13       |
| m   | 4   | mptsj | 0        | 0        | 11       | 715      |
| m   | 4   | mskj  | 0        | 12       | 151      | 3584     |
| m   | 4   | mskn  | 0        | 0        | 3        | 3        |
| m   | 4   | mskp  | 0        | 0        | 1        | 1        |
| m   | 4   | mstf  | 2        | 14       | 24       | 1164     |
| m   | 4   | ndgr  | 0        | 0        | 4        | 18       |
| m   | 4   | ndrj  | 0        | 0        | 19       | 502      |
| m   | 4   | ndrk  | 0        | 0        | 10       | 16       |
| m   | 4   | ndrn  | 0        | 0        | 9        | 40       |

|   |   |       |   |    |     |       |
|---|---|-------|---|----|-----|-------|
| m | 4 | ndrp  | 0 | 0  | 10  | 15    |
| m | 4 | ndrts | 0 | 0  | 10  | 215   |
| m | 4 | ndvw  | 0 | 0  | 4   | 11    |
| m | 4 | ndwb  | 0 | 6  | 0   | 0     |
| m | 4 | ndwş  | 0 | 1  | 0   | 0     |
| m | 4 | nsgr  | 0 | 0  | 11  | 188   |
| m | 4 | nskj  | 0 | 0  | 1   | 3     |
| m | 4 | nskr  | 0 | 0  | 23  | 178   |
| m | 4 | nsml  | 0 | 0  | 1   | 1     |
| m | 4 | nspl  | 0 | 0  | 30  | 257   |
| m | 4 | nspr  | 0 | 0  | 3   | 6     |
| m | 4 | nstr  | 6 | 62 | 470 | 15256 |
| m | 4 | nstw  | 0 | 0  | 1   | 1     |
| m | 4 | ntfj  | 0 | 0  | 19  | 39    |
| m | 4 | ntkj  | 0 | 13 | 37  | 5674  |
| m | 4 | ntpj  | 0 | 11 | 35  | 1278  |
| m | 4 | ntpl  | 1 | 6  | 28  | 10813 |
| m | 4 | ntpm  | 0 | 2  | 0   | 0     |
| m | 4 | ntpte | 0 | 2  | 0   | 0     |
| m | 4 | ntpz  | 0 | 1  | 0   | 0     |
| m | 4 | ntrd  | 0 | 0  | 6   | 8     |
| m | 4 | ntrk  | 0 | 0  | 19  | 159   |
| m | 4 | ntrl  | 0 | 0  | 4   | 16    |
| m | 4 | ntrm  | 0 | 0  | 7   | 12    |
| m | 4 | ntrn  | 0 | 0  | 3   | 5     |
| m | 4 | ntrp  | 0 | 0  | 12  | 15    |
| m | 4 | ntrr  | 0 | 0  | 21  | 92    |
| m | 4 | ntrs  | 0 | 0  | 2   | 2     |
| m | 4 | ntrts | 0 | 0  | 2   | 2     |
| m | 4 | ntrv  | 0 | 0  | 17  | 353   |
| m | 4 | ntrw  | 0 | 0  | 2   | 3     |
| m | 4 | ntrz  | 0 | 0  | 1   | 18    |
| m | 4 | ntşb  | 0 | 0  | 2   | 2     |
| m | 4 | ntşdz | 0 | 0  | 1   | 2     |
| m | 4 | ntşf  | 0 | 0  | 2   | 5     |
| m | 4 | ntşg  | 0 | 0  | 4   | 4     |
| m | 4 | ntşk  | 0 | 0  | 21  | 31    |
| m | 4 | ntşl  | 0 | 0  | 1   | 1     |
| m | 4 | ntşm  | 0 | 0  | 1   | 1     |
| m | 4 | ntşn  | 2 | 20 | 28  | 11721 |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| m   | 4   | ntşn  | 0        | 2        | 32       | 309      |
| m   | 4   | ntşp  | 0        | 0        | 21       | 76       |
| m   | 4   | ntşr  | 0        | 0        | 5        | 5        |
| m   | 4   | ntşs  | 0        | 0        | 4        | 5        |
| m   | 4   | ntşt  | 0        | 0        | 1        | 1        |
| m   | 4   | ntşts | 0        | 0        | 2        | 2        |
| m   | 4   | ntştş | 0        | 0        | 3        | 5        |
| m   | 4   | ntşx  | 0        | 0        | 1        | 1        |
| m   | 4   | ntşz  | 0        | 0        | 9        | 19       |
| m   | 4   | ntşzı | 0        | 0        | 3        | 4        |
| m   | 4   | ntskj | 0        | 16       | 209      | 4339     |
| m   | 4   | ntskr | 0        | 0        | 6        | 32       |
| m   | 4   | ntstf | 0        | 0        | 34       | 91       |
| m   | 4   | ntşzw | 0        | 0        | 2        | 2        |
| m   | 4   | ntşkj | 0        | 3        | 6        | 7        |
| m   | 4   | ntşmj | 2        | 25       | 12       | 1424     |
| m   | 4   | ndzvj | 0        | 0        | 7        | 9        |
| m   | 4   | nglj  | 0        | 0        | 5        | 1012     |
| m   | 4   | ngrj  | 0        | 0        | 2        | 3        |
| m   | 4   | ngvj  | 0        | 0        | 5        | 31       |
| m   | 4   | ngwb  | 0        | 0        | 1        | 1        |
| m   | 4   | ηkkj  | 0        | 4        | 3        | 260      |
| m   | 4   | ηkpl  | 0        | 0        | 2        | 4        |
| m   | 4   | ηksj  | 0        | 0        | 1        | 2        |
| m   | 4   | ηkst  | 0        | 0        | 29       | 1166     |
| m   | 4   | ηkşm  | 0        | 4        | 0        | 0        |
| m   | 4   | ηkşts | 0        | 0        | 2        | 2        |
| m   | 4   | ηkşte | 0        | 4        | 0        | 0        |
| m   | 4   | ηkşzı | 0        | 2        | 0        | 0        |
| m   | 4   | ηktsj | 2        | 68       | 182      | 23769    |
| m   | 4   | ηkwb  | 0        | 6        | 0        | 0        |
| m   | 4   | ηkwş  | 0        | 1        | 1        | 1        |
| m   | 4   | pemj  | 0        | 0        | 11       | 16       |
| m   | 4   | pskj  | 0        | 4        | 88       | 1690     |
| m   | 4   | pskr  | 0        | 0        | 30       | 554      |
| m   | 4   | pstf  | 2        | 14       | 39       | 684      |
| m   | 4   | pstr  | 2        | 21       | 65       | 820      |
| m   | 4   | pstş  | 0        | 0        | 12       | 35       |
| m   | 4   | ptrj  | 0        | 0        | 1        | 1        |

|   |   |       |    |     |      |       |
|---|---|-------|----|-----|------|-------|
| m | 4 | refj  | 0  | 0   | 2    | 2     |
| m | 4 | rdzv  | 0  | 0   | 1    | 1     |
| m | 4 | rdzvj | 0  | 0   | 1    | 1     |
| m | 4 | rfsk  | 0  | 0   | 1    | 1     |
| m | 4 | rgvj  | 0  | 0   | 9    | 23    |
| m | 4 | rkfj  | 0  | 0   | 4    | 118   |
| m | 4 | rmkj  | 0  | 0   | 1    | 1     |
| m | 4 | rmsk  | 0  | 0   | 1    | 3     |
| m | 4 | rηkj  | 0  | 2   | 2    | 6     |
| m | 4 | rpkj  | 0  | 4   | 5    | 63    |
| m | 4 | rpsk  | 0  | 0   | 21   | 630   |
| m | 4 | rpsts | 0  | 0   | 1    | 34    |
| m | 4 | rptsj | 0  | 0   | 4    | 20    |
| m | 4 | rskj  | 0  | 172 | 1139 | 27916 |
| m | 4 | rskl  | 0  | 0   | 3    | 5     |
| m | 4 | rskn  | 0  | 0   | 3    | 7     |
| m | 4 | rskη  | 0  | 0   | 1    | 1     |
| m | 4 | rskr  | 0  | 0   | 1    | 1     |
| m | 4 | rspr  | 0  | 0   | 3    | 3     |
| m | 4 | rstf  | 10 | 79  | 560  | 43198 |
| m | 4 | rstk  | 3  | 15  | 25   | 307   |
| m | 4 | rstr  | 0  | 0   | 3    | 5     |
| m | 4 | rstş  | 0  | 0   | 3    | 11    |
| m | 4 | rstts | 0  | 2   | 2    | 17    |
| m | 4 | rşpj  | 0  | 0   | 2    | 10    |
| m | 4 | rşşk  | 1  | 8   | 10   | 49    |
| m | 4 | rşşm  | 0  | 4   | 0    | 0     |
| m | 4 | rşşts | 0  | 1   | 0    | 0     |
| m | 4 | rşşte | 0  | 4   | 0    | 0     |
| m | 4 | rşşz  | 0  | 2   | 0    | 0     |
| m | 4 | rtfj  | 0  | 44  | 45   | 536   |
| m | 4 | rtfm  | 0  | 4   | 1    | 7     |
| m | 4 | rtfte | 0  | 4   | 1    | 2     |
| m | 4 | rtfz  | 0  | 2   | 0    | 0     |
| m | 4 | rtkj  | 0  | 1   | 5    | 32    |
| m | 4 | rtsm  | 0  | 0   | 8    | 19    |
| m | 4 | rtskj | 0  | 0   | 28   | 209   |
| m | 4 | rtstf | 0  | 0   | 1    | 1     |
| m | 4 | rtepr | 0  | 0   | 4    | 4     |
| m | 4 | rtevj | 0  | 0   | 7    | 129   |

| pos | len | cl      | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|---------|----------|----------|----------|----------|
| m   | 4   | rt̄smj  | 0        | 0        | 1        | 47       |
| m   | 4   | rxfj    | 0        | 5        | 5        | 14       |
| m   | 4   | rzwb    | 0        | 9        | 0        | 0        |
| m   | 4   | rzw̄s   | 0        | 2        | 0        | 0        |
| m   | 4   | sefj    | 0        | 0        | 29       | 76       |
| m   | 4   | semj    | 0        | 0        | 21       | 76       |
| m   | 4   | sepj    | 0        | 0        | 9        | 19       |
| m   | 4   | sfete   | 0        | 0        | 20       | 113      |
| m   | 4   | sfkw    | 0        | 0        | 1        | 1        |
| m   | 4   | sfst    | 0        | 0        | 15       | 68       |
| m   | 4   | skfj    | 0        | 0        | 25       | 2376     |
| m   | 4   | skrf    | 0        | 0        | 15       | 52       |
| m   | 4   | sk̄st   | 0        | 0        | 8        | 16       |
| m   | 4   | s̄nkj   | 0        | 1        | 1        | 42       |
| m   | 4   | spw̄te  | 0        | 0        | 5        | 11       |
| m   | 4   | ssk̄s   | 0        | 0        | 1        | 1        |
| m   | 4   | sskw    | 0        | 0        | 3        | 19       |
| m   | 4   | ssp̄s   | 0        | 0        | 10       | 97       |
| m   | 4   | sstr    | 1        | 9        | 40       | 661      |
| m   | 4   | sst̄s   | 4        | 260      | 137      | 6999     |
| m   | 4   | sskj    | 0        | 0        | 2        | 2        |
| m   | 4   | stfj    | 0        | 4        | 95       | 1706     |
| m   | 4   | stfr    | 0        | 0        | 2        | 2        |
| m   | 4   | stkj    | 0        | 8        | 19       | 20865    |
| m   | 4   | stkl    | 0        | 0        | 11       | 14       |
| m   | 4   | stmj    | 0        | 0        | 1        | 8        |
| m   | 4   | stpj    | 0        | 0        | 2        | 2        |
| m   | 4   | stpr    | 0        | 0        | 4        | 5        |
| m   | 4   | stp̄s   | 0        | 0        | 1        | 3        |
| m   | 4   | strf    | 0        | 0        | 17       | 69       |
| m   | 4   | strj    | 0        | 0        | 29       | 141      |
| m   | 4   | stst    | 0        | 0        | 2        | 2        |
| m   | 4   | st̄sl   | 0        | 0        | 1        | 1        |
| m   | 4   | st̄sm   | 0        | 4        | 0        | 0        |
| m   | 4   | st̄ste  | 0        | 4        | 1        | 1        |
| m   | 4   | st̄sz   | 0        | 2        | 0        | 0        |
| m   | 4   | sttr    | 0        | 0        | 4        | 5        |
| m   | 4   | st̄t̄sw | 0        | 0        | 1        | 4        |
| m   | 4   | stxn    | 1        | 30       | 7        | 58       |

|   |   |       |   |     |     |      |
|---|---|-------|---|-----|-----|------|
| m | 4 | stxp  | 0 | 19  | 10  | 72   |
| m | 4 | stxş  | 0 | 0   | 2   | 2    |
| m | 4 | stefj | 0 | 0   | 2   | 2    |
| m | 4 | sxfj  | 0 | 0   | 17  | 97   |
| m | 4 | sxwş  | 0 | 2   | 0   | 0    |
| m | 4 | strj  | 0 | 0   | 1   | 1    |
| m | 4 | stşkj | 0 | 0   | 1   | 3    |
| m | 4 | stşkn | 0 | 0   | 7   | 39   |
| m | 4 | stşkp | 0 | 0   | 3   | 10   |
| m | 4 | tefj  | 0 | 0   | 115 | 876  |
| m | 4 | temj  | 0 | 0   | 13  | 29   |
| m | 4 | tepj  | 0 | 0   | 29  | 137  |
| m | 4 | tfsk  | 0 | 0   | 1   | 2    |
| m | 4 | tfst  | 0 | 0   | 6   | 222  |
| m | 4 | tkfj  | 0 | 22  | 6   | 8    |
| m | 4 | tkşt  | 0 | 0   | 21  | 59   |
| m | 4 | tskj  | 0 | 0   | 5   | 677  |
| m | 4 | tskr  | 0 | 0   | 2   | 2    |
| m | 4 | tspr  | 0 | 0   | 1   | 1    |
| m | 4 | tspş  | 1 | 7   | 47  | 673  |
| m | 4 | tstr  | 0 | 0   | 83  | 675  |
| m | 4 | tstş  | 0 | 0   | 24  | 169  |
| m | 4 | tşkl  | 0 | 0   | 3   | 8    |
| m | 4 | txmj  | 0 | 0   | 8   | 24   |
| m | 4 | tstfj | 0 | 12  | 121 | 7629 |
| m | 4 | testf | 0 | 0   | 7   | 58   |
| m | 4 | vdrj  | 0 | 0   | 1   | 1    |
| m | 4 | vmgj  | 0 | 0   | 3   | 36   |
| m | 4 | vmgw  | 0 | 0   | 3   | 14   |
| m | 4 | vzgl  | 2 | 130 | 90  | 9792 |
| m | 4 | wbz̄m | 0 | 0   | 1   | 14   |
| m | 4 | wefj  | 0 | 0   | 11  | 87   |
| m | 4 | wdzvj | 0 | 0   | 1   | 1    |
| m | 4 | wgzb  | 0 | 0   | 2   | 5    |
| m | 4 | wkrf  | 0 | 0   | 1   | 11   |
| m | 4 | wkşt  | 0 | 0   | 5   | 9    |
| m | 4 | wktsj | 0 | 0   | 9   | 2722 |
| m | 4 | wmsk  | 0 | 0   | 6   | 77   |
| m | 4 | wmsts | 0 | 0   | 1   | 3    |
| m | 4 | wmştş | 0 | 0   | 3   | 9    |

| pos | len | cl     | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|--------|----------|----------|----------|----------|
| m   | 4   | wnts̄j | 0        | 0        | 1        | 1        |
| m   | 4   | wskj   | 0        | 0        | 8        | 59       |
| m   | 4   | wspr   | 0        | 0        | 14       | 85       |
| m   | 4   | wstr   | 0        | 0        | 62       | 9348     |
| m   | 4   | wtkj   | 0        | 1        | 1        | 5        |
| m   | 4   | wts̄kj | 0        | 0        | 5        | 41       |
| m   | 4   | wzwb   | 0        | 3        | 0        | 0        |
| m   | 4   | ŵskj   | 0        | 10       | 44       | 7254     |
| m   | 4   | ŵskn   | 1        | 10       | 16       | 559      |
| m   | 4   | ŵskp   | 2        | 108      | 55       | 472      |
| m   | 4   | ŵstf   | 1        | 7        | 31       | 4859     |
| m   | 4   | ŵstk   | 1        | 8        | 31       | 523      |
| m   | 4   | ŵstn   | 0        | 0        | 3        | 3        |
| m   | 4   | ŵstp   | 0        | 0        | 18       | 41       |
| m   | 4   | ŵstŝ   | 0        | 0        | 28       | 581      |
| m   | 4   | ŵstts  | 0        | 1        | 3        | 7        |
| m   | 4   | ŵswb   | 0        | 6        | 0        | 0        |
| m   | 4   | ŵswŝ   | 0        | 1        | 0        | 0        |
| m   | 4   | ŵŝkj   | 0        | 4        | 15       | 1882     |
| m   | 4   | ŵxfj   | 0        | 0        | 2        | 2        |
| m   | 4   | xefj   | 1        | 10       | 11       | 354      |
| m   | 4   | xfst   | 0        | 0        | 24       | 216      |
| m   | 4   | xskj   | 0        | 0        | 5        | 113      |
| m   | 4   | xstr   | 0        | 0        | 23       | 460      |
| m   | 4   | xtkl   | 0        | 0        | 4        | 6        |
| m   | 4   | xzvj   | 0        | 0        | 2        | 2        |
| m   | 4   | zbz̄m  | 0        | 0        | 1        | 1        |
| m   | 4   | zd̄zvj | 0        | 0        | 11       | 94       |
| m   | 4   | zgvj   | 0        | 0        | 7        | 26       |
| m   | 4   | zzbj   | 0        | 0        | 2        | 10       |
| m   | 4   | zzvj   | 0        | 0        | 1        | 1        |
| m   | 4   | zzvr   | 0        | 0        | 8        | 159      |
| m   | 4   | zzvw   | 0        | 0        | 11       | 193      |
| m   | 5   | brskj  | 0        | 0        | 2        | 5        |
| m   | 5   | ddzvj  | 0        | 0        | 1        | 2        |
| m   | 5   | ffstŝ  | 0        | 0        | 2        | 3        |
| m   | 5   | fstfj  | 0        | 1        | 17       | 268      |
| m   | 5   | jfstr  | 0        | 1        | 1        | 1        |
| m   | 5   | jŝskj  | 0        | 0        | 1        | 1        |

|   |   |        |   |    |     |      |
|---|---|--------|---|----|-----|------|
| m | 5 | jstfj  | 0 | 1  | 11  | 335  |
| m | 5 | jvzgl  | 0 | 0  | 2   | 2    |
| m | 5 | ĵstfj  | 0 | 12 | 43  | 5112 |
| m | 5 | krskj  | 0 | 0  | 1   | 2    |
| m | 5 | lfştr  | 0 | 0  | 3   | 4    |
| m | 5 | lmskj  | 0 | 0  | 4   | 52   |
| m | 5 | lstfj  | 0 | 1  | 15  | 221  |
| m | 5 | mpskj  | 0 | 0  | 23  | 417  |
| m | 5 | mpstf  | 2 | 14 | 35  | 8771 |
| m | 5 | mstfj  | 0 | 2  | 3   | 51   |
| m | 5 | ntrpr  | 0 | 0  | 16  | 38   |
| m | 5 | ntrpş  | 0 | 0  | 2   | 4    |
| m | 5 | ntrvp  | 0 | 0  | 1   | 1    |
| m | 5 | ntskj  | 0 | 0  | 1   | 2    |
| m | 5 | ntşbz  | 0 | 0  | 1   | 1    |
| m | 5 | ntşer  | 0 | 0  | 1   | 1    |
| m | 5 | ntşgr  | 0 | 0  | 1   | 1    |
| m | 5 | ntşkl  | 0 | 0  | 6   | 7    |
| m | 5 | ntşkr  | 0 | 0  | 4   | 6    |
| m | 5 | ntşmj  | 0 | 0  | 1   | 1    |
| m | 5 | ntşpl  | 0 | 0  | 1   | 1    |
| m | 5 | ntşpr  | 0 | 0  | 4   | 5    |
| m | 5 | ntşpw  | 0 | 0  | 1   | 1    |
| m | 5 | ntşsp  | 0 | 0  | 6   | 23   |
| m | 5 | ntşst  | 0 | 0  | 2   | 2    |
| m | 5 | ntşşk  | 0 | 0  | 5   | 15   |
| m | 5 | ntşşp  | 0 | 0  | 2   | 3    |
| m | 5 | ntşvw  | 0 | 0  | 2   | 3    |
| m | 5 | ntşxş  | 0 | 0  | 2   | 3    |
| m | 5 | ntstfj | 0 | 0  | 3   | 6    |
| m | 5 | pstfj  | 0 | 2  | 6   | 141  |
| m | 5 | rmskj  | 0 | 0  | 2   | 5    |
| m | 5 | rmstr  | 0 | 0  | 6   | 453  |
| m | 5 | rpskj  | 0 | 0  | 14  | 552  |
| m | 5 | rstfj  | 0 | 8  | 107 | 4726 |
| m | 5 | rstkj  | 0 | 4  | 4   | 14   |
| m | 5 | stzvj  | 0 | 0  | 4   | 9    |
| m | 5 | testfj | 0 | 0  | 2   | 90   |
| m | 5 | wbzımj | 0 | 0  | 17  | 33   |
| m | 5 | wmskj  | 0 | 0  | 4   | 99   |

| <b>pos</b> | <b>len</b> | <b>cl</b> | <b>dictwtyp</b> | <b>parawtyp</b> | <b>corpwtyp</b> | <b>corpwtok</b> |
|------------|------------|-----------|-----------------|-----------------|-----------------|-----------------|
| m          | 5          | wstrj     | 0               | 0               | 34              | 8373            |
| m          | 5          | ŵstfj     | 0               | 1               | 3               | 666             |
| m          | 5          | zbz̄mj    | 0               | 0               | 15              | 138             |
| m          | 5          | zvz̄gl    | 2               | 26              | 20              | 1763            |
| m          | 6          | mpstfj    | 0               | 2               | 6               | 772             |
| m          | 6          | nt̄şfsp   | 0               | 0               | 4               | 5               |
| m          | 6          | nt̄şzvj   | 0               | 0               | 3               | 5               |

**Final clusters**

| pos | len | cl    | dietwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| f   | 2   | bl    | 0        | 0        | 3        | 5        |
| f   | 2   | br    | 1        | 2        | 7        | 1571     |
| f   | 2   | bw    | 0        | 0        | 4        | 193      |
| f   | 2   | el    | 0        | 11       | 14       | 2577     |
| f   | 2   | em    | 0        | 1        | 1        | 114      |
| f   | 2   | en    | 2        | 2        | 6        | 354      |
| f   | 2   | ep    | 0        | 4        | 5        | 526      |
| f   | 2   | ete   | 124      | 144      | 1957     | 210241   |
| f   | 2   | dl    | 0        | 0        | 2        | 14       |
| f   | 2   | dm    | 0        | 0        | 3        | 36       |
| f   | 2   | dr    | 0        | 2        | 6        | 947      |
| f   | 2   | dw    | 0        | 39       | 82       | 15426    |
| f   | 2   | dzm   | 0        | 0        | 1        | 6        |
| f   | 2   | fl    | 0        | 0        | 1        | 236      |
| f   | 2   | fr    | 0        | 1        | 2        | 124      |
| f   | 2   | ft    | 1        | 4        | 14       | 732      |
| f   | 2   | fts   | 0        | 0        | 1        | 83       |
| f   | 2   | fte   | 0        | 1        | 1        | 108      |
| f   | 2   | gm    | 0        | 0        | 1        | 1        |
| f   | 2   | gw    | 0        | 17       | 47       | 12005    |
| f   | 2   | je    | 0        | 2        | 4        | 1359     |
| f   | 2   | jf    | 0        | 0        | 1        | 44       |
| f   | 2   | jk    | 1        | 1        | 3        | 1185     |
| f   | 2   | jl    | 0        | 0        | 2        | 53       |
| f   | 2   | jm    | 1        | 1        | 3        | 5152     |
| f   | 2   | jn    | 0        | 0        | 2        | 80       |
| f   | 2   | jp    | 0        | 0        | 2        | 25       |
| f   | 2   | js    | 0        | 0        | 4        | 361      |
| f   | 2   | js̄   | 0        | 2        | 3        | 36       |
| f   | 2   | jt    | 1        | 1        | 13       | 1634     |
| f   | 2   | jts   | 0        | 0        | 2        | 4        |
| f   | 2   | jte   | 0        | 15       | 11       | 101      |
| f   | 2   | jt̄s̄ | 0        | 0        | 2        | 5        |
| f   | 2   | jē   | 2        | 9        | 9        | 1648     |
| f   | 2   | kf    | 0        | 0        | 1        | 1        |
| f   | 2   | kl    | 2        | 2        | 10       | 2571     |
| f   | 2   | kr    | 0        | 0        | 3        | 63       |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| f   | 2   | ks   | 3        | 3        | 83       | 15291    |
| f   | 2   | kş   | 0        | 0        | 2        | 7        |
| f   | 2   | kt   | 12       | 12       | 68       | 44299    |
| f   | 2   | kw   | 0        | 13       | 41       | 2522     |
| f   | 2   | lf   | 0        | 0        | 17       | 1101     |
| f   | 2   | lk   | 1        | 4        | 16       | 2931     |
| f   | 2   | lm   | 1        | 1        | 20       | 9087     |
| f   | 2   | lp   | 0        | 8        | 2        | 601      |
| f   | 2   | lp   | 0        | 0        | 5        | 123      |
| f   | 2   | ls   | 0        | 0        | 6        | 918      |
| f   | 2   | lş   | 1        | 1        | 2        | 87       |
| f   | 2   | lt   | 3        | 4        | 57       | 5256     |
| f   | 2   | lts  | 1        | 1        | 8        | 622      |
| f   | 2   | lts̄ | 0        | 2        | 4        | 110      |
| f   | 2   | lx   | 0        | 0        | 1        | 16       |
| f   | 2   | me   | 0        | 4        | 6        | 3835     |
| f   | 2   | mf   | 1        | 1        | 4        | 241      |
| f   | 2   | ml   | 0        | 0        | 1        | 274      |
| f   | 2   | mn   | 1        | 4        | 4        | 323      |
| f   | 2   | mp   | 8        | 32       | 71       | 8575     |
| f   | 2   | ms   | 0        | 0        | 3        | 23       |
| f   | 2   | mş   | 1        | 1        | 1        | 1        |
| f   | 2   | mtē  | 0        | 0        | 1        | 3        |
| f   | 2   | mts̄ | 0        | 0        | 1        | 3        |
| f   | 2   | nf   | 0        | 0        | 1        | 4        |
| f   | 2   | nn   | 0        | 1        | 12       | 271      |
| f   | 2   | ns   | 7        | 8        | 67       | 10880    |
| f   | 2   | nş   | 1        | 1        | 7        | 1723     |
| f   | 2   | nt   | 67       | 87       | 558      | 178070   |
| f   | 2   | nts̄ | 6        | 1089     | 2976     | 142580   |
| f   | 2   | nts̄ | 2        | 20       | 28       | 4048     |
| f   | 2   | nx   | 0        | 0        | 1        | 2        |
| f   | 2   | nts̄ | 0        | 1        | 1        | 3        |
| f   | 2   | nte  | 170      | 390      | 553      | 60677    |
| f   | 2   | nts̄ | 0        | 7        | 7        | 24       |
| f   | 2   | nk   | 15       | 21       | 246      | 77194    |
| f   | 2   | pn   | 0        | 0        | 1        | 23       |
| f   | 2   | pr   | 0        | 0        | 5        | 591      |
| f   | 2   | ps   | 2        | 2        | 21       | 828      |

|   |   |     |    |    |     |        |
|---|---|-----|----|----|-----|--------|
| f | 2 | pş  | 2  | 3  | 4   | 752    |
| f | 2 | pt  | 0  | 1  | 15  | 1000   |
| f | 2 | pte | 0  | 0  | 1   | 19     |
| f | 2 | ptş | 0  | 3  | 0   | 0      |
| f | 2 | pw  | 0  | 0  | 5   | 10     |
| f | 2 | re  | 1  | 1  | 1   | 69     |
| f | 2 | rf  | 2  | 5  | 20  | 9236   |
| f | 2 | rk  | 6  | 8  | 76  | 11286  |
| f | 2 | rl  | 0  | 0  | 5   | 283    |
| f | 2 | rm  | 2  | 8  | 33  | 26219  |
| f | 2 | rn  | 0  | 2  | 32  | 4725   |
| f | 2 | rñ  | 0  | 7  | 10  | 373    |
| f | 2 | rp  | 4  | 9  | 17  | 7355   |
| f | 2 | rs  | 3  | 3  | 38  | 27275  |
| f | 2 | rş  | 3  | 7  | 16  | 998    |
| f | 2 | rt  | 16 | 22 | 196 | 55680  |
| f | 2 | rts | 0  | 2  | 12  | 508    |
| f | 2 | rte | 2  | 22 | 14  | 4597   |
| f | 2 | rtş | 0  | 6  | 10  | 437    |
| f | 2 | rw  | 0  | 15 | 49  | 5875   |
| f | 2 | rx  | 0  | 0  | 3   | 75     |
| f | 2 | sf  | 0  | 1  | 4   | 548    |
| f | 2 | sk  | 16 | 28 | 156 | 19252  |
| f | 2 | sm  | 0  | 3  | 4   | 1417   |
| f | 2 | sp  | 0  | 3  | 5   | 1124   |
| f | 2 | st  | 18 | 27 | 151 | 426348 |
| f | 2 | sw  | 4  | 30 | 53  | 17359  |
| f | 2 | şf  | 0  | 1  | 0   | 0      |
| f | 2 | şp  | 0  | 2  | 6   | 2526   |
| f | 2 | şş  | 0  | 0  | 1   | 1      |
| f | 2 | şt  | 1  | 2  | 9   | 5349   |
| f | 2 | ştş | 5  | 16 | 37  | 2860   |
| f | 2 | şx  | 2  | 2  | 2   | 184    |
| f | 2 | tf  | 0  | 7  | 7   | 209    |
| f | 2 | tl  | 0  | 2  | 2   | 16     |
| f | 2 | tm  | 1  | 1  | 4   | 512    |
| f | 2 | tr  | 8  | 9  | 46  | 14896  |
| f | 2 | tş  | 0  | 4  | 5   | 1121   |
| f | 2 | ttş | 0  | 2  | 1   | 1      |
| f | 2 | tw  | 0  | 10 | 14  | 72     |

| pos | len | cl    | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-------|----------|----------|----------|----------|
| f   | 2   | t̄sp  | 0        | 1        | 1        | 313      |
| f   | 2   | t̄st  | 0        | 1        | 2        | 43       |
| f   | 2   | vr    | 0        | 0        | 2        | 207      |
| f   | 2   | wf    | 1        | 1        | 1        | 32       |
| f   | 2   | wk    | 2        | 2        | 9        | 388      |
| f   | 2   | wl    | 0        | 0        | 2        | 70       |
| f   | 2   | wm    | 1        | 1        | 4        | 132      |
| f   | 2   | wn    | 0        | 0        | 4        | 243      |
| f   | 2   | wɲ    | 0        | 4        | 0        | 0        |
| f   | 2   | wp    | 0        | 1        | 2        | 137      |
| f   | 2   | wr    | 0        | 0        | 9        | 109      |
| f   | 2   | ws    | 0        | 0        | 6        | 535      |
| f   | 2   | wʂ    | 0        | 0        | 3        | 94       |
| f   | 2   | wt    | 1        | 3        | 26       | 6703     |
| f   | 2   | wte   | 0        | 2        | 2        | 23       |
| f   | 2   | wt̄s  | 0        | 0        | 1        | 22       |
| f   | 2   | w̄s   | 2        | 5        | 14       | 521      |
| f   | 2   | w̄ʂ   | 3        | 15       | 20       | 9985     |
| f   | 2   | w̄x   | 1        | 1        | 3        | 19       |
| f   | 2   | xε    | 0        | 2        | 3        | 702      |
| f   | 2   | xm    | 0        | 0        | 1        | 62       |
| f   | 2   | xr    | 0        | 0        | 1        | 4        |
| f   | 2   | xʂ    | 0        | 0        | 1        | 1        |
| f   | 2   | xt    | 0        | 1        | 19       | 687      |
| f   | 2   | xw    | 0        | 5        | 9        | 377      |
| f   | 2   | zm    | 20       | 20       | 473      | 9889     |
| f   | 2   | zn    | 0        | 6        | 14       | 5452     |
| f   | 2   | zw    | 0        | 9        | 33       | 4256     |
| f   | 2   | zɲ    | 0        | 1        | 4        | 528      |
| f   | 3   | fsk   | 0        | 0        | 2        | 8        |
| f   | 3   | jete  | 13       | 14       | 17       | 9854     |
| f   | 3   | jsk   | 0        | 1        | 2        | 1400     |
| f   | 3   | jst̄s | 0        | 1        | 1        | 5102     |
| f   | 3   | jte   | 9        | 11       | 18       | 18081    |
| f   | 3   | jsk   | 0        | 0        | 6        | 3294     |
| f   | 3   | kst   | 2        | 2        | 9        | 4049     |
| f   | 3   | lkw   | 0        | 1        | 3        | 44       |
| f   | 3   | lsk   | 0        | 0        | 5        | 29       |
| f   | 3   | ltr   | 1        | 1        | 2        | 59       |

|   |   |      |   |    |    |      |
|---|---|------|---|----|----|------|
| f | 3 | mbl  | 0 | 0  | 1  | 11   |
| f | 3 | mbr  | 0 | 0  | 1  | 6    |
| f | 3 | mpt  | 0 | 0  | 1  | 30   |
| f | 3 | msk  | 0 | 0  | 3  | 8    |
| f | 3 | mst  | 0 | 1  | 1  | 1    |
| f | 3 | ndw  | 0 | 2  | 1  | 1    |
| f | 3 | nst  | 0 | 0  | 1  | 80   |
| f | 3 | nte  | 0 | 0  | 2  | 45   |
| f | 3 | ntf  | 0 | 0  | 2  | 2    |
| f | 3 | ntn  | 0 | 1  | 0  | 0    |
| f | 3 | ntp  | 0 | 1  | 1  | 1    |
| f | 3 | ntr  | 0 | 1  | 3  | 42   |
| f | 3 | ntş  | 2 | 3  | 3  | 3723 |
| f | 3 | ngw  | 0 | 0  | 7  | 23   |
| f | 3 | ñks  | 0 | 0  | 1  | 45   |
| f | 3 | ñkş  | 0 | 2  | 1  | 1    |
| f | 3 | ñkt  | 2 | 2  | 6  | 2499 |
| f | 3 | ñkw  | 0 | 2  | 10 | 88   |
| f | 3 | psk  | 0 | 0  | 4  | 525  |
| f | 3 | rete | 2 | 2  | 3  | 160  |
| f | 3 | rks  | 0 | 0  | 2  | 304  |
| f | 3 | rpw  | 0 | 0  | 1  | 2    |
| f | 3 | rsk  | 0 | 0  | 4  | 227  |
| f | 3 | rst  | 0 | 0  | 3  | 145  |
| f | 3 | rst  | 0 | 0  | 1  | 13   |
| f | 3 | rştş | 1 | 3  | 2  | 103  |
| f | 3 | rtf  | 0 | 2  | 2  | 23   |
| f | 3 | rzw  | 0 | 3  | 5  | 21   |
| f | 3 | stf  | 0 | 4  | 15 | 4949 |
| f | 3 | stm  | 0 | 0  | 1  | 1    |
| f | 3 | str  | 1 | 3  | 8  | 1419 |
| f | 3 | stş  | 1 | 3  | 27 | 4211 |
| f | 3 | tstf | 0 | 12 | 31 | 3105 |
| f | 3 | wkw  | 0 | 0  | 2  | 2    |
| f | 3 | wst  | 0 | 0  | 6  | 396  |
| f | 3 | wzw  | 0 | 1  | 3  | 4    |
| f | 3 | ŵsk  | 0 | 1  | 3  | 1983 |
| f | 3 | ŵst  | 0 | 0  | 1  | 1    |
| f | 3 | ŵsw  | 0 | 2  | 2  | 15   |
| f | 3 | ŵşş  | 0 | 0  | 1  | 6    |

| <b>pos</b> | <b>len</b> | <b>cl</b>                      | <b>dictwtyp</b> | <b>parawtyp</b> | <b>corpwtyp</b> | <b>corpwtok</b> |
|------------|------------|--------------------------------|-----------------|-----------------|-----------------|-----------------|
| f          | 3          | $\tilde{w}\tilde{s}t\tilde{s}$ | 0               | 0               | 3               | 55              |
| f          | 3          | $\tilde{w}zw$                  | 0               | 0               | 3               | 17              |
| f          | 3          | xtr                            | 0               | 0               | 1               | 25              |
| f          | 4          | fstf                           | 0               | 1               | 2               | 16              |
| f          | 4          | jstf                           | 0               | 1               | 9               | 525             |
| f          | 4          | $\tilde{j}stf$                 | 0               | 12              | 22              | 7011            |
| f          | 4          | lstf                           | 0               | 1               | 4               | 203             |
| f          | 4          | mstf                           | 0               | 2               | 1               | 153             |
| f          | 4          | ntstf                          | 0               | 0               | 1               | 3               |
| f          | 4          | pstf                           | 0               | 2               | 6               | 61              |
| f          | 4          | rstf                           | 0               | 8               | 23              | 9585            |
| f          | 4          | $\tilde{w}stf$                 | 0               | 1               | 2               | 676             |
| f          | 5          | mpstf                          | 0               | 2               | 5               | 2706            |

## Appendix C. Polish cluster examples

The following tables present representative examples of words featuring Polish initial, medial and final clusters in all positions. Wherever possible, a high frequency example word from the newspaper corpus resource is shown. If a cluster was not present in the corpus resource (which is indicated by a '0' in the *corp* column), a random example from one of the other resources is given. The tables can also be used as a quick reference resource showing in which of the three resources words with particular clusters in a specific position were present.

### Initial clusters

| cl  | len | dict | para | corp | example    |
|-----|-----|------|------|------|------------|
| bj  | 2   | 1    | 1    | 1    | biura      |
| bl  | 2   | 1    | 1    | 1    | blisko     |
| br  | 2   | 1    | 1    | 1    | brak       |
| bw  | 2   | 1    | 1    | 1    | błąd       |
| bz  | 2   | 0    | 1    | 1    | bzu        |
| bz  | 2   | 0    | 1    | 1    | bzik       |
| bz  | 2   | 1    | 1    | 1    | brzegu     |
| ef  | 2   | 1    | 1    | 1    | świderek   |
| el  | 2   | 1    | 1    | 1    | śledztwo   |
| em  | 2   | 1    | 1    | 1    | śmigłowców |
| en  | 2   | 1    | 1    | 1    | śniegu     |
| ep  | 2   | 1    | 1    | 1    | spi        |
| er  | 2   | 1    | 1    | 1    | środków    |
| eś  | 2   | 0    | 0    | 1    | śrzodek    |
| ete | 2   | 1    | 1    | 1    | ściśle     |
| db  | 2   | 1    | 1    | 1    | dbać       |
| dj  | 2   | 1    | 1    | 1    | dialogu    |
| dl  | 2   | 1    | 1    | 1    | dla        |
| dm  | 2   | 1    | 1    | 1    | dmuchać    |
| dn  | 2   | 1    | 1    | 1    | dna        |
| dɲ  | 2   | 0    | 1    | 1    | dni        |
| dr  | 2   | 1    | 1    | 1    | drugiej    |
| dv  | 2   | 1    | 1    | 1    | dwa        |
| dw  | 2   | 1    | 1    | 1    | długo      |

| cl     | len | dict | para | corp | example      |
|--------|-----|------|------|------|--------------|
| dz     | 2   | 1    | 1    | 1    | drzew        |
| dz̄b   | 2   | 1    | 1    | 1    | dzbanek      |
| dz̄v   | 2   | 1    | 1    | 1    | dzwoni       |
| dz̄g   | 2   | 0    | 0    | 1    | dźgalo       |
| dz̄v   | 2   | 1    | 1    | 1    | dźwigala     |
| dz̄dz̄ | 2   | 0    | 1    | 1    | dźdźownic    |
| fe     | 2   | 1    | 1    | 1    | wsi          |
| fj     | 2   | 1    | 1    | 1    | fiat         |
| fk     | 2   | 0    | 0    | 1    | wkalkulowane |
| fl     | 2   | 1    | 1    | 1    | florydzie    |
| fp     | 2   | 1    | 1    | 1    | wpisu        |
| fr     | 2   | 1    | 1    | 1    | francja      |
| fs     | 2   | 1    | 1    | 1    | wsadzili     |
| f̄s    | 2   | 1    | 1    | 1    | wszystkich   |
| ft     | 2   | 1    | 1    | 1    | wtedy        |
| f̄ts   | 2   | 1    | 1    | 1    | wcale        |
| f̄te   | 2   | 1    | 1    | 1    | wciąż        |
| f̄t̄s  | 2   | 1    | 1    | 1    | wczoraj      |
| fx     | 2   | 1    | 1    | 1    | wchodzi      |
| gb     | 2   | 0    | 0    | 1    | gbur         |
| gd     | 2   | 1    | 1    | 1    | gdym         |
| gdz    | 2   | 1    | 1    | 1    | gdzie        |
| gj     | 2   | 1    | 1    | 1    | giełda       |
| gl     | 2   | 1    | 1    | 1    | globalne     |
| gm     | 2   | 1    | 1    | 1    | gminy        |
| gn     | 2   | 0    | 0    | 1    | gnębi        |
| gn     | 2   | 1    | 1    | 1    | gniew        |
| gr     | 2   | 1    | 1    | 1    | grupy        |
| gv     | 2   | 1    | 1    | 1    | gwarancji    |
| gw     | 2   | 1    | 1    | 1    | głównie      |
| gz     | 2   | 0    | 0    | 1    | gzymy        |
| gz     | 2   | 0    | 0    | 1    | gzik         |
| gz̄    | 2   | 1    | 1    | 1    | grzegorz     |
| ke     | 2   | 1    | 1    | 1    | książki      |
| kf     | 2   | 1    | 1    | 1    | kwoty        |
| kj     | 2   | 1    | 1    | 1    | kiedy        |
| kl     | 2   | 1    | 1    | 1    | klientów     |
| km     | 2   | 0    | 0    | 1    | kmicica      |
| kn     | 2   | 0    | 0    | 1    | knajpy       |

|     |   |   |   |   |              |
|-----|---|---|---|---|--------------|
| kn  | 2 | 0 | 0 | 1 | kniaż        |
| kp  | 2 | 0 | 0 | 1 | kpi          |
| kr  | 2 | 1 | 1 | 1 | kraju        |
| ks  | 2 | 0 | 0 | 1 | ksefobii     |
| kş  | 2 | 1 | 1 | 1 | krzysztof    |
| kt  | 2 | 1 | 1 | 1 | które        |
| kts | 2 | 0 | 0 | 1 | kcynia       |
| kte | 2 | 0 | 0 | 1 | kciuki       |
| kw  | 2 | 1 | 1 | 1 | kłopoty      |
| lj  | 2 | 0 | 0 | 1 | lian         |
| ln  | 2 | 0 | 1 | 1 | lnu          |
| lp  | 2 | 1 | 1 | 1 | lniane       |
| lv  | 2 | 1 | 1 | 1 | lwowie       |
| lz  | 2 | 1 | 1 | 1 | lżejsze      |
| mj  | 2 | 1 | 1 | 1 | między       |
| ml  | 2 | 1 | 1 | 1 | mleka        |
| mn  | 2 | 1 | 1 | 1 | mną          |
| mp  | 2 | 1 | 1 | 1 | mnie         |
| mr  | 2 | 1 | 1 | 1 | mróz         |
| mş  | 2 | 0 | 0 | 1 | mszy         |
| mts | 2 | 0 | 0 | 1 | mceńskiego   |
| mw  | 2 | 1 | 1 | 1 | młodych      |
| mx  | 2 | 0 | 1 | 1 | mchu         |
| mz  | 2 | 0 | 0 | 1 | mrzonki      |
| pe  | 2 | 1 | 1 | 1 | psie         |
| pj  | 2 | 1 | 1 | 1 | pierwszy     |
| pl  | 2 | 1 | 1 | 1 | plus         |
| pn  | 2 | 0 | 0 | 1 | pneumatyczny |
| pp  | 2 | 0 | 1 | 1 | pnie         |
| pr  | 2 | 1 | 1 | 1 | proc         |
| ps  | 2 | 1 | 1 | 1 | psa          |
| pş  | 2 | 1 | 1 | 1 | przez        |
| pt  | 2 | 1 | 1 | 1 | ptaków       |
| pte | 2 | 0 | 0 | 1 | pcimiu       |
| pw  | 2 | 1 | 1 | 1 | płacić       |
| px  | 2 | 1 | 1 | 1 | pchać        |
| rd  | 2 | 0 | 0 | 1 | rdestowa     |
| rdz | 2 | 1 | 1 | 1 | rdzenia      |
| rj  | 2 | 0 | 0 | 1 | rial         |
| rt  | 2 | 0 | 0 | 1 | rtęci        |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| rv        | 2          | 1           | 1           | 1           | rwaniu         |
| rż        | 2          | 0           | 0           | 1           | rżany          |
| se        | 2          | 1           | 1           | 1           | zsiadłe        |
| sf        | 2          | 1           | 1           | 1           | swoje          |
| sj        | 2          | 0           | 0           | 1           | sjesty         |
| sk        | 2          | 1           | 1           | 1           | skarbu         |
| sl        | 2          | 0           | 0           | 1           | slobodana      |
| sm        | 2          | 1           | 1           | 1           | smalec         |
| sn        | 2          | 1           | 1           | 1           | snu            |
| sp        | 2          | 1           | 1           | 1           | spółki         |
| sr        | 2          | 1           | 1           | 1           | srebrny        |
| ss        | 2          | 1           | 1           | 1           | zsumować       |
| sş        | 2          | 1           | 1           | 1           | zszedł         |
| st        | 2          | 1           | 1           | 1           | stanie         |
| sts       | 2          | 1           | 1           | 1           | scenariusz     |
| stş       | 2          | 0           | 0           | 1           | sczezną        |
| sw        | 2          | 1           | 1           | 1           | służby         |
| sx        | 2          | 1           | 1           | 1           | schab          |
| şf        | 2          | 0           | 1           | 1           | szwecja        |
| şk        | 2          | 1           | 1           | 1           | szkoły         |
| şl        | 2          | 1           | 1           | 1           | szlachetnych   |
| şm        | 2          | 1           | 1           | 1           | szmaty         |
| şn        | 2          | 1           | 1           | 1           | sznur          |
| şp        | 2          | 1           | 1           | 1           | szpitala       |
| şr        | 2          | 1           | 1           | 1           | szranki        |
| şt        | 2          | 1           | 1           | 1           | sztuki         |
| ştş       | 2          | 1           | 1           | 1           | szczególnie    |
| şw        | 2          | 0           | 1           | 1           | szło           |
| tf        | 2          | 1           | 1           | 1           | twórcy         |
| tj        | 2          | 0           | 0           | 1           | tiananmen      |
| tk        | 2          | 1           | 1           | 1           | tkanki         |
| tl        | 2          | 1           | 1           | 1           | tle            |
| tn        | 2          | 0           | 1           | 1           | tną            |
| tp        | 2          | 0           | 1           | 1           | tnie           |
| tr        | 2          | 1           | 1           | 1           | trudno         |
| ts        | 2          | 0           | 0           | 1           | tsunami        |
| tş        | 2          | 1           | 1           | 1           | trzeba         |
| ttş       | 2          | 0           | 0           | 1           | tczewie        |
| tw        | 2          | 1           | 1           | 1           | tłumaczy       |

|      |   |   |   |   |              |
|------|---|---|---|---|--------------|
| tx   | 2 | 1 | 1 | 1 | tchem        |
| tsf  | 2 | 0 | 0 | 1 | cwaniaków    |
| tsl  | 2 | 0 | 0 | 1 | clenia       |
| tsm  | 2 | 1 | 1 | 1 | cmentarzu    |
| tsn  | 2 | 0 | 0 | 1 | cnoty        |
| tsw  | 2 | 1 | 1 | 1 | cła          |
| tef  | 2 | 1 | 1 | 1 | ćwiczenia    |
| tem  | 2 | 1 | 1 | 1 | éma          |
| tep  | 2 | 0 | 0 | 1 | épa          |
| tşf  | 2 | 1 | 1 | 1 | czwartek     |
| tşk  | 2 | 0 | 0 | 1 | czkawka      |
| tşm  | 2 | 0 | 0 | 1 | czmychnął    |
| tşt  | 2 | 0 | 0 | 1 | cztery       |
| tşte | 2 | 1 | 1 | 1 | czci         |
| tştş | 2 | 0 | 1 | 1 | czczą        |
| tşw  | 2 | 1 | 1 | 1 | członków     |
| tşx  | 2 | 0 | 0 | 1 | czchowie     |
| vb   | 2 | 0 | 0 | 1 | wbudowany    |
| vd   | 2 | 1 | 1 | 1 | wdowa        |
| vdz  | 2 | 0 | 0 | 1 | wdzydzki     |
| vdz  | 2 | 1 | 1 | 1 | wdzięczności |
| vg   | 2 | 0 | 0 | 1 | wgapiać      |
| vj   | 2 | 1 | 1 | 1 | więc         |
| vl   | 2 | 1 | 1 | 1 | wlicza       |
| vm   | 2 | 1 | 1 | 1 | wmówić       |
| vn   | 2 | 1 | 1 | 1 | wnętrza      |
| vp   | 2 | 1 | 1 | 1 | wniosek      |
| vr   | 2 | 1 | 1 | 1 | wraz         |
| vv   | 2 | 0 | 0 | 1 | wwozu        |
| vw   | 2 | 1 | 1 | 1 | właśnie      |
| vz   | 2 | 1 | 1 | 1 | wzór         |
| vz   | 2 | 1 | 1 | 1 | wziąć        |
| vz   | 2 | 1 | 1 | 1 | września     |
| wb   | 2 | 0 | 1 | 1 | łby          |
| wg   | 2 | 0 | 0 | 1 | łgarzem      |
| wk   | 2 | 0 | 0 | 1 | łkając       |
| wz   | 2 | 1 | 1 | 1 | łzy          |
| wz   | 2 | 0 | 1 | 0 | łzie         |
| wz   | 2 | 0 | 0 | 1 | łze          |
| xf   | 2 | 1 | 1 | 1 | chwili       |

| cl   | len | dict | para | corp | example       |
|------|-----|------|------|------|---------------|
| xj   | 2   | 0    | 0    | 1    | hierarchii    |
| xl   | 2   | 1    | 1    | 1    | chleb         |
| xm   | 2   | 1    | 1    | 1    | chmury        |
| xr   | 2   | 1    | 1    | 1    | chronić       |
| xş   | 2   | 1    | 1    | 1    | chrześcijan   |
| xt   | 2   | 0    | 0    | 1    | chtonicznych  |
| xts  | 2   | 0    | 1    | 1    | chce          |
| xte  | 2   | 1    | 1    | 1    | chciał        |
| xw   | 2   | 1    | 1    | 1    | chłopca       |
| zb   | 2   | 1    | 1    | 1    | zbyt          |
| zd   | 2   | 1    | 1    | 1    | zdaniem       |
| zdż  | 2   | 1    | 1    | 1    | zdziśław      |
| zg   | 2   | 1    | 1    | 1    | zgodnie       |
| zj   | 2   | 1    | 1    | 1    | zjednoczonych |
| zl   | 2   | 1    | 1    | 1    | zlecenia      |
| zm   | 2   | 1    | 1    | 1    | zmarł         |
| zn   | 2   | 1    | 1    | 1    | znacznie      |
| zp   | 2   | 1    | 1    | 1    | zniszczenia   |
| zr   | 2   | 1    | 1    | 1    | zresztą       |
| zv   | 2   | 1    | 1    | 1    | zwykle        |
| zw   | 2   | 1    | 1    | 1    | złotych       |
| zz   | 2   | 1    | 1    | 1    | zza           |
| zz   | 2   | 0    | 0    | 1    | zziębnięty    |
| zz   | 2   | 1    | 1    | 1    | zrzeszenia    |
| zdż  | 2   | 0    | 0    | 1    | ździebko      |
| zl   | 2   | 1    | 1    | 1    | źle           |
| zr   | 2   | 1    | 1    | 1    | źródło        |
| zb   | 2   | 0    | 0    | 1    | źbikowski     |
| zd   | 2   | 0    | 0    | 1    | źdana         |
| zg   | 2   | 0    | 0    | 1    | źrgów         |
| zl   | 2   | 0    | 0    | 1    | źlebie        |
| zm   | 2   | 1    | 1    | 1    | źmijewski     |
| zp   | 2   | 1    | 1    | 1    | źniwa         |
| zr   | 2   | 0    | 0    | 1    | źrący         |
| zv   | 2   | 0    | 0    | 1    | źwiru         |
| zw   | 2   | 0    | 0    | 1    | źłobków       |
| brd  | 3   | 0    | 0    | 1    | brdą          |
| brdż | 3   | 0    | 0    | 1    | brdzie        |
| brn  | 3   | 0    | 0    | 1    | brno          |

|       |   |   |   |   |            |
|-------|---|---|---|---|------------|
| brp   | 3 | 0 | 0 | 1 | brnie      |
| brv   | 3 | 0 | 1 | 1 | brwi       |
| bzd   | 3 | 0 | 0 | 1 | bzdura     |
| bzdź  | 3 | 0 | 0 | 1 | bździny    |
| bzd   | 3 | 0 | 0 | 1 | brzdąc     |
| bzm   | 3 | 0 | 0 | 1 | brzmi      |
| efj   | 3 | 1 | 1 | 1 | świata     |
| emj   | 3 | 1 | 1 | 1 | śmierci    |
| epj   | 3 | 1 | 1 | 1 | śpiewa     |
| dmj   | 3 | 0 | 0 | 1 | dmie       |
| dr̄g  | 3 | 1 | 1 | 1 | drgań      |
| drj   | 3 | 0 | 0 | 1 | driada     |
| drv   | 3 | 0 | 0 | 1 | drwi       |
| drz̄  | 3 | 0 | 0 | 1 | drzą       |
| dvj   | 3 | 0 | 0 | 1 | dwie       |
| dzy   | 3 | 1 | 1 | 1 | drzwi      |
| dzgn  | 3 | 0 | 0 | 1 | dźgnął     |
| dzgn  | 3 | 0 | 0 | 1 | dźgnięta   |
| dzv̄j | 3 | 1 | 1 | 1 | dźwięku    |
| fel   | 3 | 0 | 0 | 1 | wślizguje  |
| fer   | 3 | 1 | 1 | 1 | wśród      |
| fete  | 3 | 1 | 1 | 1 | wściekły   |
| fl̄k  | 3 | 1 | 1 | 1 | wkleśle    |
| fkr   | 3 | 1 | 1 | 1 | wkrótce    |
| fk̄s  | 3 | 0 | 0 | 1 | wkrze      |
| fk̄w  | 3 | 1 | 1 | 1 | wkład      |
| fpj   | 3 | 0 | 0 | 1 | wpierw     |
| fpl   | 3 | 0 | 0 | 1 | wplecione  |
| fpr   | 3 | 1 | 1 | 1 | wprawdzie  |
| fp̄s  | 3 | 0 | 0 | 1 | wprzód     |
| fpw   | 3 | 1 | 1 | 1 | wpływ      |
| fsk   | 3 | 1 | 1 | 1 | wskazuje   |
| fsp   | 3 | 1 | 1 | 1 | współpracy |
| fs̄s  | 3 | 0 | 1 | 0 | wszedłszy  |
| fst   | 3 | 1 | 1 | 1 | wstępne    |
| fsw   | 3 | 0 | 0 | 1 | wślawił    |
| fsx   | 3 | 1 | 1 | 1 | wschodniej |
| fst̄s | 3 | 0 | 0 | 1 | wszczęcia  |
| ft̄r  | 3 | 0 | 0 | 1 | wtrysku    |
| ftw   | 3 | 0 | 0 | 1 | wtłoczyć   |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| fxw       | 3          | 0           | 0           | 1           | wchłonać       |
| grd       | 3          | 0           | 0           | 1           | grdykę         |
| gvd       | 3          | 0           | 0           | 1           | gwdy           |
| gvj       | 3          | 1           | 1           | 1           | gwiazdy        |
| gzm       | 3          | 1           | 1           | 1           | grzmi          |
| kfj       | 3          | 1           | 1           | 1           | kwietnia       |
| kln       | 3          | 0           | 1           | 1           | klną           |
| klp       | 3          | 0           | 1           | 1           | klnie          |
| klv       | 3          | 0           | 0           | 1           | klwów          |
| kmj       | 3          | 0           | 0           | 1           | kmieć          |
| kpj       | 3          | 0           | 0           | 1           | kpią           |
| krf       | 3          | 1           | 1           | 1           | krwi           |
| krj       | 3          | 0           | 0           | 1           | krioterapii    |
| krn       | 3          | 0           | 0           | 1           | krnąbrnych     |
| krt       | 3          | 1           | 1           | 1           | krtani         |
| kst       | 3          | 1           | 1           | 1           | kształcenia    |
| kstş      | 3          | 0           | 0           | 1           | krzczonów      |
| lep       | 3          | 0           | 0           | 1           | łśni           |
| lgn       | 3          | 0           | 0           | 1           | lgną           |
| lgn       | 3          | 0           | 0           | 1           | lgnie          |
| lvj       | 3          | 0           | 1           | 1           | lwią           |
| mete      | 3          | 1           | 1           | 1           | mści           |
| mdl       | 3          | 1           | 1           | 1           | mdleją         |
| mdw       | 3          | 1           | 1           | 1           | mdłości        |
| mgj       | 3          | 0           | 1           | 1           | mgiełką        |
| mgl       | 3          | 1           | 1           | 1           | mgle           |
| mgp       | 3          | 0           | 0           | 1           | mgnieniu       |
| mgw       | 3          | 1           | 1           | 1           | mgła           |
| mkn       | 3          | 0           | 0           | 1           | mkną           |
| mkp       | 3          | 0           | 0           | 1           | mknie          |
| mst       | 3          | 0           | 0           | 1           | mstowska       |
| mstş      | 3          | 0           | 1           | 1           | mszczonowie    |
| plv       | 3          | 0           | 0           | 1           | plwociny       |
| prj       | 3          | 0           | 0           | 1           | priorytetów    |
| psk       | 3          | 0           | 0           | 1           | pskowa         |
| pst       | 3          | 0           | 0           | 1           | prztyczek      |
| pstş      | 3          | 1           | 1           | 1           | pszczół        |
| pwte      | 3          | 1           | 1           | 1           | płci           |
| pxl       | 3          | 1           | 1           | 1           | pchlich        |

|      |   |   |   |   |                   |
|------|---|---|---|---|-------------------|
| pxn  | 3 | 1 | 1 | 1 | pchnął            |
| pxp  | 3 | 0 | 1 | 1 | pchnięcie         |
| pxw  | 3 | 1 | 1 | 1 | pchły             |
| rvj  | 3 | 0 | 1 | 1 | rwie              |
| rzn  | 3 | 0 | 0 | 1 | rznąć             |
| rzn  | 3 | 0 | 0 | 1 | rźnie             |
| sfj  | 3 | 0 | 0 | 1 | swietłana         |
| sfl  | 3 | 0 | 0 | 1 | sflaczały         |
| sfr  | 3 | 0 | 0 | 1 | sfrustrowany      |
| skf  | 3 | 0 | 0 | 1 | skwapliwie        |
| skj  | 3 | 1 | 1 | 1 | skierował         |
| skl  | 3 | 1 | 1 | 1 | sklepów           |
| skn  | 3 | 0 | 0 | 1 | sknery            |
| skr  | 3 | 1 | 1 | 1 | skreślony         |
| sks  | 3 | 0 | 0 | 1 | skserowane        |
| skş  | 3 | 1 | 1 | 1 | skrzydła          |
| skw  | 3 | 1 | 1 | 1 | składki           |
| smr  | 3 | 0 | 0 | 1 | smród             |
| spe  | 3 | 0 | 0 | 1 | spiało            |
| spj  | 3 | 1 | 1 | 1 | spieszyc          |
| spl  | 3 | 1 | 1 | 1 | splot             |
| spr  | 3 | 1 | 1 | 1 | sprawie           |
| spş  | 3 | 1 | 1 | 1 | sprzedaży         |
| spw  | 3 | 1 | 1 | 1 | spląty            |
| sst  | 3 | 0 | 0 | 1 | zstępnych         |
| stf  | 3 | 1 | 1 | 1 | stworzyć          |
| stj  | 3 | 0 | 0 | 1 | stiuki            |
| str  | 3 | 1 | 1 | 1 | strony            |
| stş  | 3 | 1 | 1 | 1 | strzelił          |
| stw  | 3 | 1 | 1 | 1 | stłuczki          |
| stx  | 3 | 0 | 0 | 1 | stchórzyłem       |
| stsj | 3 | 0 | 0 | 1 | scjentologicznego |
| sxf  | 3 | 0 | 0 | 1 | schwytyany        |
| sxj  | 3 | 0 | 0 | 1 | zhierarchizowane  |
| sxl  | 3 | 0 | 0 | 1 | schludnie         |
| sxn  | 3 | 0 | 0 | 1 | schną             |
| sxp  | 3 | 0 | 0 | 1 | schnie            |
| sxr  | 3 | 1 | 1 | 1 | schroniska        |
| sxş  | 3 | 0 | 0 | 1 | schrzanili        |
| sxw  | 3 | 0 | 0 | 1 | schładzania       |

| cl    | len | dict | para | corp | example      |
|-------|-----|------|------|------|--------------|
| sfj   | 3   | 0    | 1    | 1    | szwie        |
| skf   | 3   | 0    | 0    | 1    | szkwał       |
| skj   | 3   | 1    | 1    | 1    | szkielet     |
| skl   | 3   | 1    | 1    | 1    | szklarski    |
| skr   | 3   | 0    | 0    | 1    | szkrab       |
| skw   | 3   | 0    | 0    | 1    | szkła        |
| spj   | 3   | 1    | 1    | 1    | szpiegostwo  |
| spr   | 3   | 0    | 0    | 1    | szprotawie   |
| str   | 3   | 0    | 0    | 1    | sztrasburgu  |
| stsf  | 3   | 0    | 0    | 1    | szczwanych   |
| tfj   | 3   | 1    | 1    | 1    | twierdzi     |
| tkf   | 3   | 1    | 1    | 1    | tkwi         |
| tkl   | 3   | 1    | 1    | 1    | tkliwości    |
| tkn   | 3   | 0    | 0    | 1    | tknąć        |
| tkp   | 3   | 0    | 0    | 1    | tknie        |
| trf   | 3   | 1    | 1    | 1    | trwa         |
| trj   | 3   | 1    | 1    | 1    | triumf       |
| tspn  | 3   | 0    | 0    | 1    | trznadel     |
| tste  | 3   | 1    | 1    | 1    | trzciniński  |
| txn   | 3   | 0    | 0    | 1    | tchnąć       |
| txp   | 3   | 0    | 0    | 1    | tchnie       |
| tsfj  | 3   | 0    | 0    | 1    | cwietajewej  |
| tskl  | 3   | 0    | 0    | 1    | ckliwy       |
| tskp  | 3   | 0    | 0    | 1    | ckni         |
| tefj  | 3   | 1    | 1    | 1    | ćwierci      |
| temj  | 3   | 0    | 1    | 1    | ćmielowie    |
| tspkp | 3   | 0    | 0    | 1    | czknieć      |
| vbj   | 3   | 0    | 0    | 1    | wbiegł       |
| vbr   | 3   | 1    | 1    | 1    | wbrew        |
| vdm   | 3   | 0    | 0    | 1    | wdmuchiwane  |
| vdr   | 3   | 0    | 0    | 1    | wdrożenie    |
| vgj   | 3   | 0    | 0    | 1    | wgięcie      |
| vgl   | 3   | 0    | 0    | 1    | wglądu       |
| vgn   | 3   | 0    | 0    | 1    | wgniecenia   |
| vgr   | 3   | 0    | 0    | 1    | wgryźć       |
| vgw   | 3   | 0    | 0    | 1    | wgłębienie   |
| vgz   | 3   | 0    | 0    | 1    | wgrzebywania |
| vmj   | 3   | 1    | 1    | 1    | wmieszał     |
| vvj   | 3   | 0    | 0    | 1    | wwieźć       |

|      |   |   |   |   |              |
|------|---|---|---|---|--------------|
| vzb  | 3 | 1 | 1 | 1 | wzbudza      |
| vzd  | 3 | 1 | 1 | 1 | wzdycha      |
| vzg  | 3 | 1 | 1 | 1 | wzgórza      |
| vzl  | 3 | 0 | 0 | 1 | wzloty       |
| vzm  | 3 | 1 | 1 | 1 | wzmocnienie  |
| vzn  | 3 | 1 | 1 | 1 | wznowienie   |
| vzn  | 3 | 1 | 1 | 1 | wzniesiono   |
| vzr  | 3 | 1 | 1 | 1 | wzrost       |
| vzv  | 3 | 1 | 1 | 1 | wzwyż        |
| vzd  | 3 | 0 | 0 | 1 | wżdy         |
| wbj  | 3 | 0 | 1 | 1 | łbie         |
| xfj  | 3 | 0 | 0 | 1 | chwiejność   |
| xmj  | 3 | 0 | 0 | 1 | chmielewski  |
| xst  | 3 | 0 | 0 | 1 | chrztu       |
| xste | 3 | 0 | 0 | 1 | chrzciciela  |
| xstę | 3 | 0 | 0 | 1 | chrzczenie   |
| zbj  | 3 | 1 | 1 | 1 | zbiorów      |
| zbl  | 3 | 1 | 1 | 1 | zbliża       |
| zbr  | 3 | 1 | 1 | 1 | zbrodni      |
| zbw  | 3 | 0 | 0 | 1 | zbląkany     |
| zbz  | 3 | 0 | 0 | 1 | zbykowani    |
| zbz  | 3 | 0 | 0 | 1 | zbrzydła     |
| zdz  | 3 | 1 | 1 | 1 | zdjęcia      |
| zdm  | 3 | 0 | 0 | 1 | zdmuchnął    |
| zdr  | 3 | 1 | 1 | 1 | zdrowia      |
| zdv  | 3 | 0 | 0 | 1 | zdwojoną     |
| zdw  | 3 | 0 | 0 | 1 | zdlawić      |
| zdz  | 3 | 0 | 0 | 1 | zdrzemnąć    |
| zdzv | 3 | 0 | 0 | 1 | zdzwaniamy   |
| zgj  | 3 | 1 | 1 | 1 | zgierzu      |
| zgl  | 3 | 0 | 0 | 1 | zgliszczach  |
| zgn  | 3 | 0 | 0 | 1 | zgnębiony    |
| zgn  | 3 | 1 | 1 | 1 | zgniły       |
| zgr  | 3 | 1 | 1 | 1 | zgromadzenia |
| zgv  | 3 | 0 | 0 | 1 | zgwalcenie   |
| zgw  | 3 | 1 | 1 | 1 | zgłoszenia   |
| zgz  | 3 | 1 | 1 | 1 | zgrzewalna   |
| zmj  | 3 | 1 | 1 | 1 | zmiany       |
| zmp  | 3 | 1 | 1 | 1 | zmniejszenie |
| zmr  | 3 | 1 | 1 | 1 | zmroku       |

| cl    | len | dict | para | corp | example     |
|-------|-----|------|------|------|-------------|
| zmw   | 3   | 0    | 0    | 1    | zmłócenia   |
| zvj   | 3   | 1    | 1    | 1    | związku     |
| zvl   | 3   | 1    | 1    | 1    | zwleka      |
| zvr   | 3   | 1    | 1    | 1    | zwrotu      |
| zvw   | 3   | 1    | 1    | 1    | zwłaszcza   |
| zdzb  | 3   | 0    | 0    | 1    | żdźbeł      |
| zgn   | 3   | 0    | 0    | 1    | żgnął       |
| brvj  | 4   | 0    | 1    | 1    | brwiowy     |
| bzmj  | 4   | 0    | 0    | 1    | brzmienie   |
| drgn  | 4   | 1    | 1    | 1    | drgnęło     |
| drgn  | 4   | 0    | 1    | 1    | drgnie      |
| drvj  | 4   | 0    | 0    | 1    | drwią       |
| dzyj  | 4   | 0    | 1    | 1    | drzwiami    |
| fskr  | 4   | 0    | 0    | 1    | wskroś      |
| fskş  | 4   | 0    | 0    | 1    | wskrzesić   |
| fspj  | 4   | 0    | 0    | 1    | wspierania  |
| fstr  | 4   | 1    | 1    | 1    | wstręt      |
| fstş  | 4   | 1    | 1    | 1    | wstrzymanie |
| gzbj  | 4   | 1    | 1    | 1    | grzbiecie   |
| gzmj  | 4   | 1    | 1    | 1    | grzmiął     |
| krfj  | 4   | 0    | 1    | 1    | krwią       |
| pstr  | 4   | 0    | 0    | 1    | pstrągi     |
| skfj  | 4   | 0    | 0    | 1    | skwierczą   |
| skrf  | 4   | 0    | 0    | 1    | skrwawionym |
| stfj  | 4   | 1    | 1    | 1    | stwierdził  |
| stfj  | 4   | 0    | 0    | 1    | sztwiertnia |
| tkfj  | 4   | 0    | 1    | 1    | tkwią       |
| tşmj  | 4   | 0    | 0    | 1    | trzmiele    |
| tşpj  | 4   | 0    | 0    | 1    | trzpiotka   |
| vzbj  | 4   | 0    | 0    | 1    | wzbiera     |
| vzbr  | 4   | 1    | 1    | 1    | wzbroniony  |
| vzdr  | 4   | 0    | 0    | 1    | wzdraga     |
| vzdw  | 4   | 1    | 1    | 1    | wzdłuż      |
| vzgl  | 4   | 1    | 1    | 1    | względu     |
| vzmj  | 4   | 0    | 0    | 1    | wzmianki    |
| zdzbl | 4   | 0    | 0    | 1    | żdźble      |
| zdzbw | 4   | 0    | 0    | 1    | żdźbła      |

**Medial clusters**

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b>  |
|-----------|------------|-------------|-------------|-------------|-----------------|
| bb        | 2          | 0           | 0           | 1           | lobby           |
| bd        | 2          | 0           | 0           | 1           | obdarzony       |
| bdz       | 2          | 0           | 0           | 1           | obdzielić       |
| bg        | 2          | 0           | 0           | 1           | obgadywać       |
| bj        | 2          | 1           | 1           | 1           | sobie           |
| bl        | 2          | 1           | 1           | 1           | publicznych     |
| bm        | 2          | 1           | 1           | 1           | zróbmy          |
| bn        | 2          | 1           | 1           | 1           | podobne         |
| bn        | 2          | 1           | 1           | 1           | podobnie        |
| br        | 2          | 1           | 1           | 1           | obroty          |
| bv        | 2          | 0           | 0           | 1           | subwencji       |
| bw        | 2          | 1           | 1           | 1           | jabłoński       |
| bz        | 2          | 0           | 0           | 1           | łobzów          |
| bz        | 2          | 0           | 0           | 1           | łobzie          |
| bz        | 2          | 1           | 1           | 1           | dobrze          |
| ef        | 2          | 0           | 0           | 1           | prześwit        |
| ek        | 2          | 0           | 0           | 1           | grześkowiak     |
| el        | 2          | 1           | 1           | 1           | jeśli           |
| em        | 2          | 1           | 1           | 1           | jesteśmy        |
| en        | 2          | 1           | 1           | 1           | głośno          |
| en        | 2          | 1           | 1           | 1           | wcześniej       |
| ep        | 2          | 0           | 0           | 1           | uśpić           |
| er        | 2          | 1           | 1           | 1           | bezpośrednio    |
| et        | 2          | 0           | 0           | 1           | huśtawce        |
| ets       | 2          | 0           | 0           | 1           | krztuścowi      |
| ete       | 2          | 1           | 1           | 1           | wartości        |
| ez        | 2          | 0           | 1           | 0           | gaśże           |
| db        | 2          | 1           | 1           | 1           | odbędzie        |
| dd        | 2          | 1           | 1           | 1           | oddać           |
| ddz       | 2          | 1           | 1           | 1           | oddział         |
| dg        | 2          | 0           | 0           | 1           | bydgoszczy      |
| dj        | 2          | 1           | 1           | 1           | media           |
| dl        | 2          | 1           | 1           | 1           | sprawiedliwości |
| dm        | 2          | 1           | 1           | 1           | administracji   |
| dn        | 2          | 1           | 1           | 1           | jednak          |
| dn        | 2          | 1           | 1           | 1           | zgodnie         |
| dr        | 2          | 1           | 1           | 1           | podróży         |

| cl   | len | dict | para | corp | example           |
|------|-----|------|------|------|-------------------|
| dv   | 2   | 1    | 1    | 1    | podwyżki          |
| dw   | 2   | 1    | 1    | 1    | według            |
| dz   | 2   | 1    | 1    | 1    | nadzorczej        |
| dz   | 2   | 1    | 1    | 1    | podziemnych       |
| dz   | 2   | 1    | 1    | 1    | odrzucił          |
| dz̄b | 2   | 0    | 0    | 1    | lidzbarku         |
| dzm  | 2   | 0    | 1    | 1    | powiedzmy         |
| dzv  | 2   | 1    | 1    | 1    | zadzwoił          |
| dz̄z | 2   | 0    | 1    | 0    | zapowiedzże       |
| dz̄b | 2   | 1    | 1    | 1    | choćby            |
| dz̄g | 2   | 0    | 0    | 1    | zadz̄gana         |
| dzm  | 2   | 0    | 1    | 1    | ludźmi            |
| dzv  | 2   | 0    | 0    | 1    | udźwignąć         |
| dz̄z | 2   | 0    | 1    | 1    | wsadźże           |
| dz̄b | 2   | 1    | 1    | 1    | liczba            |
| dz̄d | 2   | 0    | 0    | 1    | poczdamie         |
| dzm  | 2   | 0    | 0    | 1    | menedźmentu       |
| dzr  | 2   | 0    | 0    | 1    | hidzra            |
| fe   | 2   | 1    | 1    | 1    | owskiak           |
| fj   | 2   | 1    | 1    | 1    | ofiar             |
| fk   | 2   | 1    | 1    | 1    | przeciwko         |
| fl   | 2   | 1    | 1    | 1    | refleksji         |
| fm   | 2   | 0    | 1    | 1    | hofman            |
| fn   | 2   | 1    | 1    | 1    | nieufności        |
| fn   | 2   | 1    | 1    | 1    | cofnięcie         |
| fp   | 2   | 0    | 0    | 1    | przeciwpożarowej  |
| fr   | 2   | 1    | 1    | 1    | afryki            |
| fs   | 2   | 0    | 1    | 1    | owsa              |
| fş   | 2   | 1    | 1    | 1    | zawsze            |
| ft   | 2   | 1    | 1    | 1    | naftowej          |
| fts  | 2   | 1    | 1    | 1    | pracodawcy        |
| fte  | 2   | 1    | 1    | 1    | dowcip            |
| ftş  | 2   | 1    | 1    | 1    | wówczas           |
| fx   | 2   | 0    | 0    | 1    | przeciwchemicznej |
| gb   | 2   | 1    | 1    | 1    | jakby             |
| gd   | 2   | 1    | 1    | 1    | nigdy             |
| gdz  | 2   | 1    | 1    | 1    | nigdzie           |
| gj   | 2   | 1    | 1    | 1    | drugiej           |
| gl   | 2   | 1    | 1    | 1    | mogli             |

|      |   |   |   |   |                 |
|------|---|---|---|---|-----------------|
| gm   | 2 | 1 | 1 | 1 | fragmenty       |
| gn   | 2 | 1 | 1 | 1 | prognozy        |
| gn   | 2 | 1 | 1 | 1 | zbigniew        |
| gr   | 2 | 1 | 1 | 1 | zagranicznych   |
| gv   | 2 | 0 | 0 | 1 | zagwarantować   |
| gw   | 2 | 1 | 1 | 1 | ubiegłym        |
| gz   | 2 | 1 | 1 | 1 | egzaminu        |
| gz   | 2 | 1 | 1 | 1 | także           |
| jb   | 2 | 1 | 1 | 1 | najbardziej     |
| je   | 2 | 0 | 1 | 1 | najaktywniejsi  |
| jd   | 2 | 1 | 1 | 1 | znajduje        |
| jd̄z | 2 | 0 | 0 | 1 | tajdze          |
| jd̄z | 2 | 0 | 1 | 1 | dojdzie         |
| jd̄z | 2 | 0 | 0 | 1 | azerbejdżan     |
| jf   | 2 | 0 | 0 | 1 | sejfu           |
| jg   | 2 | 0 | 1 | 1 | najgorsze       |
| jj   | 2 | 0 | 0 | 1 | najjaśniejszego |
| jk   | 2 | 1 | 1 | 1 | kolejka         |
| jl   | 2 | 0 | 1 | 1 | najlepiej       |
| jm   | 2 | 1 | 1 | 1 | sejmu           |
| jn   | 2 | 1 | 1 | 1 | kolejne         |
| jñ   | 2 | 1 | 1 | 1 | wojnie          |
| jp   | 2 | 0 | 1 | 1 | najpóźniej      |
| jr   | 2 | 0 | 1 | 1 | najróżniejszych |
| js   | 2 | 0 | 1 | 1 | rejsu           |
| jś   | 2 | 1 | 1 | 1 | najważniejsze   |
| jt   | 2 | 0 | 1 | 1 | wojtasiński     |
| jts  | 2 | 1 | 1 | 1 | szwajcaria      |
| jte  | 2 | 1 | 1 | 1 | wojciech        |
| jts̄ | 2 | 1 | 1 | 1 | najczęściej     |
| jv   | 2 | 1 | 1 | 1 | najważniejsze   |
| jw   | 2 | 0 | 0 | 1 | najłatwiej      |
| jx   | 2 | 0 | 0 | 1 | najchętniej     |
| jz   | 2 | 0 | 1 | 1 | pejzaż          |
| jz   | 2 | 0 | 1 | 1 | najzimniej      |
| jż   | 2 | 1 | 1 | 1 | obejrzeć        |
| je   | 2 | 1 | 1 | 1 | sąsiadów        |
| js̄  | 2 | 0 | 0 | 1 | tańsze          |
| jz̄  | 2 | 1 | 1 | 1 | więzienia       |
| jż̄  | 2 | 0 | 1 | 0 | stańże          |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| ke        | 2          | 0           | 1           | 1           | kodeksie       |
| kf        | 2          | 1           | 1           | 1           | likwidacji     |
| kj        | 2          | 1           | 1           | 1           | takie          |
| kk        | 2          | 1           | 1           | 1           | lekki          |
| kl        | 2          | 1           | 1           | 1           | zwykle         |
| km        | 2          | 0           | 0           | 1           | bukmacherzy    |
| kn        | 2          | 1           | 1           | 1           | uniknąć        |
| kp        | 2          | 1           | 1           | 1           | zabraknie      |
| kp        | 2          | 0           | 0           | 1           | kokpit         |
| kr        | 2          | 1           | 1           | 1           | okresie        |
| ks        | 2          | 1           | 1           | 1           | kodeksu        |
| kś        | 2          | 1           | 1           | 1           | świętokrzyskie |
| kt        | 2          | 1           | 1           | 1           | dyrektor       |
| ks̄       | 2          | 1           | 1           | 1           | sukces         |
| kte       | 2          | 0           | 1           | 1           | trakcie        |
| ktś       | 2          | 0           | 0           | 1           | nikczemny      |
| kw        | 2          | 1           | 1           | 1           | przykład       |
| kx        | 2          | 0           | 0           | 1           | sztokholmie    |
| lb        | 2          | 1           | 1           | 1           | albo           |
| le        | 2          | 0           | 1           | 1           | pulsie         |
| ld        | 2          | 1           | 1           | 1           | holding        |
| ldz       | 2          | 0           | 1           | 1           | uldze          |
| ldz       | 2          | 0           | 1           | 1           | saldzie        |
| ldz̄      | 2          | 0           | 0           | 1           | seldzuckiej    |
| lf        | 2          | 1           | 1           | 1           | polfa          |
| lg        | 2          | 1           | 1           | 1           | ulgi           |
| lj        | 2          | 1           | 1           | 1           | milionów       |
| lk        | 2          | 1           | 1           | 1           | tylko          |
| ll        | 2          | 1           | 1           | 1           | ballady        |
| lm        | 2          | 1           | 1           | 1           | filmu          |
| ln        | 2          | 1           | 1           | 1           | działalności   |
| lp        | 2          | 1           | 1           | 1           | szczególnie    |
| lp        | 2          | 0           | 0           | 1           | kolporter      |
| lr        | 2          | 0           | 0           | 1           | ulrycha        |
| ls        | 2          | 0           | 0           | 1           | polsat         |
| lś        | 2          | 1           | 1           | 1           | dalsze         |
| lt        | 2          | 1           | 1           | 1           | kultury        |
| lts̄      | 2          | 0           | 1           | 1           | walce          |
| lte       | 2          | 0           | 1           | 1           | malcie         |

|      |   |   |   |   |                  |
|------|---|---|---|---|------------------|
| lt̂s | 2 | 1 | 1 | 1 | walczyć          |
| lv   | 2 | 1 | 1 | 1 | absolwentów      |
| lx   | 2 | 1 | 1 | 1 | filharmonii      |
| lz   | 2 | 0 | 0 | 1 | alzacji          |
| lz   | 2 | 0 | 0 | 1 | zaolzia          |
| lẑ  | 2 | 1 | 1 | 1 | niemalże         |
| mb   | 2 | 1 | 1 | 1 | luksemburg       |
| mc   | 2 | 0 | 0 | 1 | stłamsić         |
| mdz  | 2 | 0 | 0 | 1 | osiemdziesiątych |
| mf   | 2 | 1 | 1 | 1 | komfort          |
| mj   | 2 | 1 | 1 | 1 | natomiast        |
| mk   | 2 | 1 | 1 | 1 | bramki           |
| ml   | 2 | 1 | 1 | 1 | kremla           |
| mm   | 2 | 0 | 1 | 1 | immunitetu       |
| mn   | 2 | 1 | 1 | 1 | ogromne          |
| mp   | 2 | 1 | 1 | 1 | przypomnijmy     |
| mp   | 2 | 1 | 1 | 1 | postępowania     |
| mr   | 2 | 1 | 1 | 1 | zamrożenie       |
| ms   | 2 | 0 | 0 | 1 | samsung          |
| mş   | 2 | 1 | 1 | 1 | dymsza           |
| mt   | 2 | 1 | 1 | 1 | tamtych          |
| mt̂s | 2 | 0 | 1 | 1 | niemcy           |
| mt̂c | 2 | 0 | 1 | 1 | tamci            |
| mt̂ş | 2 | 1 | 1 | 1 | tymczasem        |
| mv   | 2 | 1 | 1 | 1 | tramwaju         |
| mw   | 2 | 0 | 0 | 1 | miłomłyn         |
| mx   | 2 | 0 | 0 | 1 | symcha           |
| mz   | 2 | 0 | 0 | 1 | keramzyt         |
| mz   | 2 | 0 | 0 | 1 | giemie           |
| mẑ  | 2 | 1 | 1 | 1 | łomzy            |
| nb   | 2 | 0 | 0 | 1 | edynburgu        |
| nc   | 2 | 0 | 1 | 1 | sensie           |
| nd   | 2 | 1 | 1 | 1 | będą             |
| ndz  | 2 | 1 | 1 | 1 | między           |
| ndẑ | 2 | 1 | 1 | 1 | indziej          |
| ndẑ | 2 | 0 | 0 | 1 | kilimandżaro     |
| nf   | 2 | 1 | 1 | 1 | informacji       |
| nj   | 2 | 0 | 0 | 1 | tanjug           |
| nl   | 2 | 0 | 0 | 1 | finlandia        |
| nm   | 2 | 0 | 0 | 1 | tiananmen        |

| cl  | len | dict | para | corp | example     |
|-----|-----|------|------|------|-------------|
| nn  | 2   | 1    | 1    | 1    | innych      |
| ηη  | 2   | 1    | 1    | 1    | inni        |
| np  | 2   | 0    | 0    | 1    | greenpeace  |
| nr  | 2   | 0    | 0    | 1    | henryk      |
| ns  | 2   | 1    | 1    | 1    | finansów    |
| nş  | 2   | 0    | 0    | 1    | czynszu     |
| nt  | 2   | 1    | 1    | 1    | prezydenta  |
| nts | 2   | 1    | 1    | 1    | więcej      |
| nte | 2   | 1    | 1    | 1    | momencie    |
| ntş | 2   | 1    | 1    | 1    | połączenia  |
| nv  | 2   | 1    | 1    | 1    | inwestorów  |
| nx  | 2   | 0    | 0    | 1    | manhattan   |
| nz  | 2   | 1    | 1    | 1    | transakcji  |
| nż  | 2   | 1    | 1    | 1    | branży      |
| ηb  | 2   | 0    | 0    | 1    | hańba       |
| ηdz | 2   | 1    | 1    | 1    | będzie      |
| ηk  | 2   | 0    | 1    | 1    | wańkowicza  |
| ηm  | 2   | 0    | 1    | 1    | końmi       |
| ηts | 2   | 1    | 1    | 1    | końca       |
| ηte | 2   | 1    | 1    | 1    | pięciu      |
| ηtş | 2   | 1    | 1    | 1    | zakończeniu |
| ηg  | 2   | 1    | 1    | 1    | ciągu       |
| ηk  | 2   | 1    | 1    | 1    | rynku       |
| pe  | 2   | 0    | 1    | 1    | najlepsi    |
| pf  | 2   | 1    | 1    | 1    | obficie     |
| pj  | 2   | 1    | 1    | 1    | dopiero     |
| pk  | 2   | 1    | 1    | 1    | szybko      |
| pl  | 2   | 1    | 1    | 1    | zaplanowano |
| pm  | 2   | 0    | 1    | 1    | kupmy       |
| pn  | 2   | 1    | 1    | 1    | kupno       |
| pη  | 2   | 1    | 1    | 1    | stopniu     |
| pp  | 2   | 0    | 0    | 1    | lepper      |
| pr  | 2   | 1    | 1    | 1    | oprócz      |
| ps  | 2   | 1    | 1    | 1    | absolwentów |
| pş  | 2   | 1    | 1    | 1    | poprzez     |
| pt  | 2   | 1    | 1    | 1    | optimus     |
| pts | 2   | 1    | 1    | 1    | lipca       |
| pte | 2   | 1    | 1    | 1    | szybciej    |
| ptş | 2   | 1    | 1    | 1    | ropczyce    |

|     |   |   |   |   |               |
|-----|---|---|---|---|---------------|
| pw  | 2 | 1 | 1 | 1 | opłaty        |
| px  | 2 | 1 | 1 | 1 | obchodzi      |
| pz  | 2 | 0 | 1 | 0 | nasypże       |
| rb  | 2 | 1 | 1 | 1 | skarbu        |
| re  | 2 | 1 | 1 | 1 | konkursie     |
| rd  | 2 | 1 | 1 | 1 | ordynacji     |
| rdz | 2 | 1 | 1 | 1 | bardzo        |
| rdz | 2 | 1 | 1 | 1 | bardziej      |
| rf  | 2 | 1 | 1 | 1 | interfax      |
| rg  | 2 | 1 | 1 | 1 | organizacji   |
| rj  | 2 | 1 | 1 | 1 | akcjonariuszy |
| rk  | 2 | 1 | 1 | 1 | gospodarki    |
| rl  | 2 | 1 | 1 | 1 | berlin        |
| rm  | 2 | 1 | 1 | 1 | firmy         |
| rn  | 2 | 1 | 1 | 1 | internet      |
| rn  | 2 | 1 | 1 | 1 | października  |
| rp  | 2 | 1 | 1 | 1 | korporacji    |
| rr  | 2 | 1 | 1 | 1 | terrorystów   |
| rs  | 2 | 1 | 1 | 1 | kursu         |
| rş  | 2 | 1 | 1 | 1 | warszawie     |
| rt  | 2 | 1 | 1 | 1 | wartości      |
| rts | 2 | 1 | 1 | 1 | marca         |
| rte | 2 | 1 | 1 | 1 | śmierci       |
| rtş | 2 | 1 | 1 | 1 | gospodarczej  |
| rv  | 2 | 1 | 1 | 1 | norwegia      |
| rw  | 2 | 1 | 1 | 1 | zmarłego      |
| rx  | 2 | 1 | 1 | 1 | archiwum      |
| rz  | 2 | 1 | 1 | 1 | marża         |
| se  | 2 | 0 | 0 | 1 | rozsiane      |
| sf  | 2 | 1 | 1 | 1 | atmosfera     |
| sj  | 2 | 1 | 1 | 1 | komisji       |
| sk  | 2 | 1 | 1 | 1 | środowiska    |
| sl  | 2 | 0 | 0 | 1 | islamu        |
| sm  | 2 | 1 | 1 | 1 | pisma         |
| sn  | 2 | 1 | 1 | 1 | własnych      |
| sn  | 2 | 0 | 0 | 1 | krosnie       |
| sp  | 2 | 1 | 1 | 1 | gospodarki    |
| sr  | 2 | 0 | 1 | 1 | osram         |
| ss  | 2 | 1 | 1 | 1 | rozsądek      |
| sş  | 2 | 1 | 1 | 1 | rozszerzenia  |

| cl  | len | dict | para | corp | example          |
|-----|-----|------|------|------|------------------|
| st  | 2   | 1    | 1    | 1    | ustawy           |
| sts | 2   | 1    | 1    | 1    | wszyscy          |
| ste | 2   | 1    | 1    | 1    | rozciąga         |
| stş | 2   | 1    | 1    | 1    | rozczarowanie    |
| sw  | 2   | 1    | 1    | 1    | usług            |
| sx  | 2   | 1    | 1    | 1    | mitsubishi       |
| şe  | 2   | 0    | 1    | 1    | najbliżsi        |
| şf  | 2   | 1    | 1    | 1    | kruszwica        |
| şk  | 2   | 1    | 1    | 1    | mieszkania       |
| şl  | 2   | 0    | 1    | 1    | przyszli         |
| şm  | 2   | 0    | 1    | 1    | koszmar          |
| şn  | 2   | 1    | 1    | 1    | słuszne          |
| şp  | 2   | 1    | 1    | 1    | słusznie         |
| şp  | 2   | 1    | 1    | 1    | hiszpania        |
| şr  | 2   | 0    | 0    | 1    | oszronione       |
| şs  | 2   | 1    | 1    | 1    | tożsamości       |
| şş  | 2   | 1    | 1    | 1    | najbliższych     |
| st  | 2   | 1    | 1    | 1    | koszty           |
| sts | 2   | 0    | 1    | 1    | podwyżce         |
| ste | 2   | 1    | 1    | 1    | wreszcie         |
| stş | 2   | 1    | 1    | 1    | jeszcze          |
| şw  | 2   | 1    | 1    | 1    | doszło           |
| şx  | 2   | 0    | 1    | 1    | zwierzchowski    |
| şz  | 2   | 0    | 1    | 1    | napiszże         |
| te  | 2   | 1    | 1    | 1    | przedsiębiorstwa |
| tf  | 2   | 1    | 1    | 1    | łatwo            |
| tj  | 2   | 1    | 1    | 1    | sympatię         |
| tk  | 2   | 1    | 1    | 1    | przypadku        |
| tl  | 2   | 1    | 1    | 1    | świetle          |
| tm  | 2   | 1    | 1    | 1    | hetman           |
| tn  | 2   | 1    | 1    | 1    | prywatnych       |
| tñ  | 2   | 1    | 1    | 1    | ostatnich        |
| tp  | 2   | 1    | 1    | 1    | odpowiednio      |
| tr  | 2   | 1    | 1    | 1    | zatrudnienia     |
| ts  | 2   | 1    | 1    | 1    | odsetek          |
| tş  | 2   | 1    | 1    | 1    | potrzeby         |
| tt  | 2   | 1    | 1    | 1    | netto            |
| tts | 2   | 1    | 1    | 1    | wkrótce          |
| tte | 2   | 1    | 1    | 1    | odcinku          |

|      |   |   |   |   |                   |
|------|---|---|---|---|-------------------|
| tṡ   | 2 | 1 | 1 | 1 | podczas           |
| tw   | 2 | 1 | 1 | 1 | światło           |
| tx   | 2 | 1 | 1 | 1 | odchodzi          |
| tsb  | 2 | 0 | 0 | 1 | specbajki         |
| tsd  | 2 | 0 | 0 | 1 | specdelegację     |
| tsf  | 2 | 0 | 0 | 1 | gucwiński         |
| tsj  | 2 | 1 | 1 | 1 | informacji        |
| tsk  | 2 | 1 | 1 | 1 | dziecka           |
| tsl  | 2 | 1 | 1 | 1 | nocleg            |
| tsm  | 2 | 0 | 0 | 1 | policmajstra      |
| tsn  | 2 | 1 | 1 | 1 | mocno             |
| tsp  | 2 | 1 | 1 | 1 | obecnie           |
| tsp  | 2 | 0 | 0 | 1 | kacper            |
| tsr  | 2 | 0 | 0 | 1 | socrealizmu       |
| tst  | 2 | 1 | 1 | 1 | octem             |
| tsts | 2 | 0 | 1 | 1 | niemieccy         |
| tste | 2 | 0 | 1 | 1 | powiedzcie        |
| tstṡ | 2 | 0 | 0 | 1 | białostoczczyźnie |
| tsw  | 2 | 0 | 0 | 1 | wrocław           |
| tsx  | 2 | 0 | 0 | 1 | icchak            |
| tef  | 2 | 1 | 1 | 1 | poćwiczyć         |
| tek  | 2 | 0 | 0 | 1 | maćkowiak         |
| tem  | 2 | 0 | 1 | 1 | dziećmi           |
| tep  | 2 | 0 | 0 | 1 | zaćpani           |
| tets | 2 | 1 | 1 | 1 | uchodźców         |
| tete | 2 | 0 | 1 | 1 | chodźcie          |
| tetṡ | 2 | 0 | 0 | 1 | uchodźczych       |
| tez  | 2 | 0 | 1 | 0 | gniećze           |
| tṡf  | 2 | 0 | 0 | 1 | poczwardka        |
| tṡj  | 2 | 0 | 0 | 1 | bertolucciego     |
| tṡk  | 2 | 1 | 1 | 1 | pożyczki          |
| tṡl  | 2 | 1 | 1 | 1 | życzliwości       |
| tṡm  | 2 | 0 | 1 | 1 | kaczmarek         |
| tṡn  | 2 | 1 | 1 | 1 | zagranicznych     |
| tṡp  | 2 | 1 | 1 | 1 | stycznia          |
| tṡp  | 2 | 1 | 1 | 1 | rzeczpospolitej   |
| tṡt  | 2 | 1 | 1 | 1 | poczty            |
| tṡts | 2 | 0 | 1 | 1 | ucieczce          |
| tṡte | 2 | 1 | 1 | 1 | nieuczciwej       |
| tṡtṡ | 2 | 0 | 1 | 1 | uczczenia         |

| cl    | len | dict | para | corp | example        |
|-------|-----|------|------|------|----------------|
| t̄sw  | 2   | 0    | 0    | 1    | uczłowieczyć   |
| t̄sz  | 2   | 0    | 1    | 0    | syczące        |
| vb    | 2   | 0    | 0    | 1    | kowboj         |
| vd    | 2   | 1    | 1    | 1    | prawdopodobnie |
| v̄dz  | 2   | 1    | 1    | 1    | sprawdza       |
| v̄dz  | 2   | 1    | 1    | 1    | wprawdzie      |
| vg    | 2   | 0    | 0    | 1    | afganistanie   |
| vj    | 2   | 1    | 1    | 1    | powiedział     |
| vl    | 2   | 1    | 1    | 1    | budowlanych    |
| vm    | 2   | 0    | 1    | 1    | zastanówmy     |
| vn    | 2   | 1    | 1    | 1    | zarówno        |
| v̄n   | 2   | 1    | 1    | 1    | również        |
| vr    | 2   | 1    | 1    | 1    | powrót         |
| vv    | 2   | 0    | 0    | 1    | przeciwwagą    |
| vw    | 2   | 0    | 0    | 1    | pawła          |
| vz    | 2   | 1    | 1    | 1    | nawzajem       |
| vz    | 2   | 0    | 0    | 1    | powzięcia      |
| v̄z   | 2   | 1    | 1    | 1    | zawrzeć        |
| wb    | 2   | 1    | 1    | 1    | byłby          |
| wc    | 2   | 0    | 0    | 1    | półsierota     |
| wd    | 2   | 1    | 1    | 1    | giełda         |
| w̄dz  | 2   | 0    | 0    | 1    | wółdze         |
| w̄dz  | 2   | 1    | 1    | 1    | giełdzie       |
| wf    | 2   | 0    | 0    | 1    | półfinale      |
| wg    | 2   | 0    | 1    | 1    | małgorzata     |
| wj    | 2   | 0    | 0    | 1    | półjawne       |
| wk    | 2   | 1    | 1    | 1    | spółki         |
| wl    | 2   | 0    | 1    | 1    | paula          |
| wm    | 2   | 1    | 1    | 1    | muzułmanów     |
| wn    | 2   | 1    | 1    | 1    | pełne          |
| w̄n   | 2   | 1    | 1    | 1    | zupełnie       |
| wp    | 2   | 1    | 1    | 1    | małpy          |
| wr    | 2   | 1    | 1    | 1    | półroczu       |
| ws    | 2   | 0    | 0    | 1    | piłsudskiego   |
| w̄s   | 2   | 1    | 1    | 1    | falszywe       |
| wt    | 2   | 1    | 1    | 1    | autor          |
| w̄ts  | 2   | 1    | 1    | 1    | spółce         |
| w̄te  | 2   | 1    | 1    | 1    | kształcie      |
| w̄t̄s | 2   | 1    | 1    | 1    | współczesnej   |

|                  |   |   |   |   |                |
|------------------|---|---|---|---|----------------|
| wv               | 2 | 1 | 1 | 1 | półwyspie      |
| ww               | 2 | 0 | 1 | 1 | mułła          |
| wx               | 2 | 0 | 0 | 1 | bełchatów      |
| wz               | 2 | 1 | 1 | 1 | klauzuli       |
| wz̄              | 2 | 0 | 0 | 1 | wokółziemskich |
| wz̄ <sub>l</sub> | 2 | 1 | 1 | 1 | małżonków      |
| wn̄              | 2 | 0 | 0 | 1 | paręnaście     |
| ws̄              | 2 | 1 | 1 | 1 | mięsa          |
| wś̄              | 2 | 0 | 0 | 1 | ukąszenia      |
| wv̄              | 2 | 0 | 0 | 1 | wąwozu         |
| wx̄              | 2 | 1 | 1 | 1 | pęcherza       |
| wz̄              | 2 | 1 | 1 | 1 | związane       |
| wz̄ <sub>l</sub> | 2 | 1 | 1 | 1 | książek        |
| xb               | 2 | 0 | 0 | 1 | niechby        |
| xd               | 2 | 0 | 0 | 1 | bohdan         |
| xdz̄             | 2 | 0 | 0 | 1 | bohdziewiczza  |
| xf               | 2 | 1 | 1 | 1 | uchwały        |
| xj               | 2 | 0 | 0 | 1 | monachium      |
| xk               | 2 | 0 | 1 | 1 | jakichkolwiek  |
| xl               | 2 | 1 | 1 | 1 | bachleda       |
| xm               | 2 | 1 | 1 | 1 | ahmed          |
| xn               | 2 | 1 | 1 | 1 | technologii    |
| xn̄              | 2 | 1 | 1 | 1 | powszechnie    |
| xp               | 2 | 0 | 0 | 1 | wszechpolskiej |
| xr               | 2 | 1 | 1 | 1 | ochrony        |
| xs               | 2 | 0 | 0 | 1 | trzechsetnym   |
| xś̄              | 2 | 0 | 0 | 1 | bundeswehrze   |
| xt               | 2 | 1 | 1 | 1 | jachtu         |
| xts̄             | 2 | 1 | 1 | 1 | zechce         |
| xtē             | 2 | 0 | 1 | 1 | zechciał       |
| xtś̄             | 2 | 1 | 1 | 1 | dotychczas     |
| xw               | 2 | 1 | 1 | 1 | wybuchła       |
| xx               | 2 | 0 | 0 | 1 | wahhabitów     |
| xz̄              | 2 | 0 | 0 | 1 | tychże         |
| zb               | 2 | 1 | 1 | 1 | pozbawienia    |
| zd               | 2 | 1 | 1 | 1 | pojazdów       |
| zdz̄             | 2 | 1 | 1 | 1 | rozdział       |
| zḡ              | 2 | 1 | 1 | 1 | mózgu          |
| zj               | 2 | 1 | 1 | 1 | decyzji        |
| zl               | 2 | 1 | 1 | 1 | rozliczeń      |

| cl               | len | dict | para | corp | example        |
|------------------|-----|------|------|------|----------------|
| zm               | 2   | 1    | 1    | 1    | rozmowy        |
| zn               | 2   | 1    | 1    | 1    | oznacza        |
| z <sub>n</sub>   | 2   | 0    | 1    | 1    | niezniszczalny |
| zr               | 2   | 1    | 1    | 1    | bezrobocia     |
| zv               | 2   | 1    | 1    | 1    | rozwoju        |
| zw               | 2   | 1    | 1    | 1    | znalazły       |
| zz               | 2   | 0    | 0    | 1    | bezzasadne     |
| z <sub>z</sub>   | 2   | 0    | 0    | 1    | rozziew        |
| z <sub>z</sub>   | 2   | 1    | 1    | 1    | rozrządu       |
| zb               | 2   | 1    | 1    | 1    | prośbą         |
| z <sub>dz</sub>  | 2   | 1    | 1    | 1    | października   |
| z <sub>l</sub>   | 2   | 1    | 1    | 1    | znaleźli       |
| zm               | 2   | 0    | 1    | 1    | koźmiński      |
| zn               | 2   | 1    | 1    | 1    | późno          |
| z <sub>n</sub>   | 2   | 1    | 1    | 1    | później        |
| zr               | 2   | 0    | 1    | 1    | przezroczyste  |
| zv               | 2   | 1    | 1    | 1    | trzeźwości     |
| z <sub>z</sub>   | 2   | 0    | 1    | 0    | zarażze        |
| z <sub>b</sub>   | 2   | 1    | 1    | 1    | służby         |
| z <sub>d</sub>   | 2   | 1    | 1    | 1    | każdy          |
| z <sub>dz</sub>  | 2   | 0    | 0    | 1    | każdziutkiemu  |
| z <sub>dz</sub>  | 2   | 1    | 1    | 1    | przyjeżdża     |
| z <sub>g</sub>   | 2   | 0    | 0    | 1    | wierzgacz      |
| z <sub>l</sub>   | 2   | 1    | 1    | 1    | możliwości     |
| z <sub>m</sub>   | 2   | 0    | 1    | 1    | założmy        |
| z <sub>n</sub>   | 2   | 1    | 1    | 1    | można          |
| z <sub>n</sub>   | 2   | 1    | 1    | 1    | najważniejsze  |
| z <sub>r</sub>   | 2   | 0    | 0    | 1    | pożre          |
| z <sub>v</sub>   | 2   | 1    | 1    | 1    | łyżwach        |
| z <sub>w</sub>   | 2   | 0    | 0    | 1    | wyżłobionych   |
| z <sub>z</sub>   | 2   | 0    | 1    | 0    | strzeże        |
| bbj              | 3   | 0    | 0    | 1    | abbie          |
| bbr              | 3   | 0    | 0    | 1    | subbranże      |
| bdr              | 3   | 0    | 0    | 1    | obdrapane      |
| b <sub>dzv</sub> | 3   | 0    | 0    | 1    | obdzwoiłam     |
| bgr              | 3   | 0    | 0    | 1    | obgryza        |
| blj              | 3   | 1    | 1    | 1    | biblioteki     |
| bmj              | 3   | 0    | 0    | 1    | obmiaru        |
| brd              | 3   | 0    | 0    | 1    | ubrdalam       |

|       |   |   |   |   |                  |
|-------|---|---|---|---|------------------|
| brj   | 3 | 0 | 0 | 1 | gabriel          |
| brk   | 3 | 0 | 0 | 1 | bobrku           |
| brn   | 3 | 1 | 1 | 1 | srebrny          |
| br̄n  | 3 | 0 | 1 | 1 | srebrników       |
| bvj   | 3 | 0 | 0 | 1 | obwieszczenia    |
| bwb   | 3 | 0 | 0 | 1 | osłabłby         |
| bwk   | 3 | 1 | 1 | 1 | jabłka           |
| bwt̄s | 3 | 0 | 0 | 1 | jabłczanymi      |
| bzb   | 3 | 0 | 0 | 1 | antyhabsburskich |
| bzd   | 3 | 0 | 0 | 1 | nabzdycza        |
| bzn   | 3 | 0 | 0 | 1 | obznajomienia    |
| bzm   | 3 | 0 | 1 | 1 | zabrzmi          |
| efj   | 3 | 1 | 1 | 1 | oświadczył       |
| ekj   | 3 | 0 | 0 | 1 | miśkiewicz       |
| elm   | 3 | 0 | 1 | 1 | podkreślmy       |
| eln   | 3 | 1 | 1 | 1 | nieumyślne       |
| eln   | 3 | 1 | 1 | 1 | pomyślnie        |
| elte  | 3 | 0 | 1 | 1 | pomyślcie        |
| elz   | 3 | 0 | 1 | 0 | namyślże         |
| emj   | 3 | 1 | 1 | 1 | piśmie           |
| epj   | 3 | 1 | 1 | 1 | pośpiechu        |
| epte  | 3 | 0 | 1 | 0 | rzeźbie          |
| etem  | 3 | 0 | 1 | 1 | gośćmi           |
| etes  | 3 | 0 | 0 | 1 | sześćset         |
| etets | 3 | 1 | 1 | 1 | jeźdźców         |
| etete | 3 | 0 | 1 | 1 | puścicie         |
| etez  | 3 | 0 | 1 | 0 | puścże           |
| dbj   | 3 | 1 | 1 | 1 | odbiorców        |
| dbl   | 3 | 0 | 0 | 1 | odblokowanie     |
| dbr   | 3 | 1 | 1 | 1 | podbródek        |
| dbw   | 3 | 0 | 0 | 1 | śródbłonek       |
| dbz   | 3 | 0 | 0 | 1 | nadbrzeżnych     |
| dd̄n  | 3 | 0 | 0 | 1 | naddnieprzańska  |
| ddr   | 3 | 0 | 0 | 1 | poddruku         |
| ddv   | 3 | 0 | 0 | 1 | ponaddwukrotnie  |
| ddw   | 3 | 0 | 0 | 1 | oddłużenie       |
| ddzv  | 3 | 0 | 0 | 1 | oddzwonić        |
| dgd   | 3 | 0 | 0 | 1 | podgdańskich     |
| dgj   | 3 | 0 | 0 | 1 | odgiąć           |
| dgl   | 3 | 0 | 0 | 1 | podglądać        |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b>    |
|-----------|------------|-------------|-------------|-------------|-------------------|
| dgm       | 3          | 0           | 0           | 1           | ponadgminnym      |
| dgn       | 3          | 0           | 0           | 1           | nadgniłych        |
| dgr       | 3          | 1           | 1           | 1           | odgrywa           |
| dgv       | 3          | 0           | 0           | 1           | odgwizdał         |
| dgw       | 3          | 1           | 1           | 1           | odgłosy           |
| dgz       | 3          | 1           | 1           | 1           | podgrzewane       |
| djm       | 3          | 0           | 0           | 1           | tudjmana          |
| dln       | 3          | 0           | 0           | 1           | módlmy            |
| dln       | 3          | 0           | 0           | 1           | nieprzywiedlne    |
| dlp       | 3          | 0           | 0           | 1           | modlnicze         |
| dlts      | 3          | 0           | 0           | 1           | siedlce           |
| dlte      | 3          | 0           | 0           | 1           | módlcie           |
| dlv       | 3          | 0           | 0           | 1           | podlwowskich      |
| dmj       | 3          | 1           | 1           | 1           | przedmiotem       |
| dml       | 3          | 0           | 0           | 1           | nadmleczarz       |
| dmr       | 3          | 1           | 1           | 1           | odmrożenia        |
| dmw       | 3          | 0           | 0           | 1           | odmłodzić         |
| drdz      | 3          | 0           | 0           | 1           | odrdzewiające     |
| drg       | 3          | 0           | 0           | 1           | rozedrganie       |
| drj       | 3          | 0           | 0           | 1           | adriana           |
| drv       | 3          | 0           | 0           | 1           | wydrwigroszy      |
| drz       | 3          | 0           | 0           | 1           | zadrzy            |
| dvdz      | 3          | 1           | 1           | 1           | odwdzięczyć       |
| dvj       | 3          | 1           | 1           | 1           | zaledwie          |
| dvl       | 3          | 0           | 0           | 1           | odwlec            |
| dvr       | 3          | 1           | 1           | 1           | odwrotnie         |
| dvw       | 3          | 1           | 1           | 1           | podwładnych       |
| dvz       | 3          | 0           | 0           | 1           | odwzajemnia       |
| dvz       | 3          | 0           | 0           | 1           | przedwrześniowego |
| dwb       | 3          | 0           | 1           | 1           | wszedłby          |
| dwk       | 3          | 0           | 0           | 1           | mydłka            |
| dwş       | 3          | 0           | 1           | 1           | wyszedłszy        |
| dzg       | 3          | 0           | 0           | 1           | przedzgonne       |
| dzj       | 3          | 0           | 0           | 1           | przedjazdowym     |
| dzl       | 3          | 0           | 0           | 1           | podzleceniodawcy  |
| dzn       | 3          | 1           | 1           | 1           | odznaczenia       |
| dzv       | 3          | 1           | 1           | 1           | nadzwyczajne      |
| dzp       | 3          | 0           | 0           | 1           | przedźniwne       |
| dzv       | 3          | 0           | 0           | 1           | niedrzwica        |

|       |   |   |   |   |                          |
|-------|---|---|---|---|--------------------------|
| dzbr  | 3 | 0 | 0 | 1 | szpicbródka              |
| dzmj  | 3 | 0 | 0 | 1 | ludźmierzu               |
| dzvj  | 3 | 1 | 1 | 1 | niedźwiedzia             |
| dzbj  | 3 | 0 | 1 | 1 | liczbie                  |
| fel   | 3 | 0 | 0 | 1 | przeciwślizgowych        |
| fep   | 3 | 0 | 0 | 1 | przeciwnieżne            |
| fete  | 3 | 0 | 0 | 1 | powściągliwość           |
| fkj   | 3 | 0 | 1 | 1 | ołówkiem                 |
| fkl   | 3 | 0 | 1 | 0 | niewklęsto               |
| fkr   | 3 | 0 | 0 | 1 | przeciwkradzieżowego     |
| fkw   | 3 | 0 | 0 | 1 | powkłada                 |
| fpj   | 3 | 0 | 0 | 1 | przeciwpiechotnych       |
| fpl   | 3 | 0 | 0 | 1 | przeciwpleśniowe         |
| fpr   | 3 | 0 | 0 | 1 | niewprowadzenia          |
| fps   | 3 | 0 | 0 | 1 | przeciwprzeziębieniowych |
| fpw   | 3 | 0 | 0 | 1 | niewpłacenia             |
| fsk   | 3 | 1 | 1 | 1 | warszawskim              |
| fsp   | 3 | 0 | 0 | 1 | niewspółmiernie          |
| fst   | 3 | 1 | 1 | 1 | powstanie                |
| fsts  | 3 | 0 | 1 | 1 | warszawscy               |
| fsw   | 3 | 0 | 0 | 1 | przeciwśłoneczne         |
| fsx   | 3 | 0 | 0 | 1 | środkowowschodniej       |
| fšk   | 3 | 0 | 0 | 1 | wawrzka                  |
| fšt   | 3 | 1 | 1 | 1 | befsztyk                 |
| fštš  | 3 | 0 | 0 | 1 | dąbrowszczaków           |
| ftj   | 3 | 0 | 0 | 1 | muftiego                 |
| ftk   | 3 | 0 | 0 | 1 | haftkę                   |
| fr    | 3 | 0 | 0 | 1 | niewtrącanie             |
| ftete | 3 | 0 | 1 | 1 | sprawdźcie               |
| glj   | 3 | 0 | 0 | 1 | intaglio                 |
| gmj   | 3 | 0 | 0 | 1 | dogmie                   |
| grts  | 3 | 0 | 0 | 1 | igrców                   |
| gvj   | 3 | 0 | 0 | 1 | trzygwiazdkowym          |
| gwb   | 3 | 0 | 1 | 1 | mógłby                   |
| gwš   | 3 | 0 | 1 | 1 | sposzrzęłszy             |
| gzm   | 3 | 0 | 1 | 1 | wodogrzmotów             |
| bj    | 3 | 0 | 1 | 1 | najbiedniejszych         |
| jbl   | 3 | 0 | 0 | 1 | najbliższych             |
| jbr   | 3 | 0 | 0 | 1 | najbrutalniejszych       |
| jbw   | 3 | 0 | 0 | 1 | najbłyskotliwszym        |

| cl   | len | dict | para | corp | example                |
|------|-----|------|------|------|------------------------|
| jbz  | 3   | 0    | 0    | 1    | najbrzydszym           |
| jel  | 3   | 0    | 0    | 1    | najśliczniejsze        |
| jete | 3   | 1    | 1    | 1    | wejścia                |
| jdr  | 3   | 0    | 0    | 1    | najdroższe             |
| jdw  | 3   | 0    | 0    | 1    | najdłużej              |
| jdzm | 3   | 0    | 1    | 1    | przejdźmy              |
| jdzz | 3   | 0    | 1    | 0    | nadejdźże              |
| jfj  | 3   | 0    | 0    | 1    | sejfie                 |
| jfs  | 3   | 0    | 0    | 1    | najwszechstronniejszym |
| jftś | 3   | 0    | 1    | 1    | najwcześniej           |
| jgj  | 3   | 0    | 1    | 1    | dwojgiem               |
| jgl  | 3   | 0    | 0    | 1    | breughla               |
| jgr  | 3   | 0    | 0    | 1    | najgroźniejszych       |
| jgv  | 3   | 0    | 0    | 1    | najgwałtowniejsze      |
| jgw  | 3   | 0    | 0    | 1    | najgłośniejszych       |
| jgz  | 3   | 0    | 0    | 1    | najgrzeczniej          |
| jkj  | 3   | 0    | 1    | 1    | strajkiem              |
| jkl  | 3   | 0    | 0    | 1    | najklasycyjszych       |
| jkr  | 3   | 0    | 0    | 1    | najkrócej              |
| kw   | 3   | 0    | 0    | 1    | najkłopotliwszy        |
| jlb  | 3   | 0    | 0    | 1    | koktajlbarem           |
| jlz  | 3   | 0    | 0    | 1    | najlżejszy             |
| jmj  | 3   | 1    | 1    | 1    | sejmie                 |
| jmk  | 3   | 0    | 0    | 1    | dejmka                 |
| jmm  | 3   | 0    | 1    | 0    | zdejmy                 |
| jmn  | 3   | 1    | 1    | 1    | najmniej               |
| jmr  | 3   | 0    | 0    | 1    | najmroczniejszej       |
| jme  | 3   | 0    | 1    | 0    | zdejmcież              |
| mw   | 3   | 0    | 1    | 1    | najmłodszych           |
| jmz  | 3   | 0    | 1    | 0    | zdejme                 |
| jnf  | 3   | 0    | 0    | 1    | szejnfeld              |
| npj  | 3   | 1    | 1    | 1    | najpierw               |
| pk   | 3   | 0    | 0    | 1    | knajpki                |
| jpl  | 3   | 0    | 1    | 0    | najplastycyjszej       |
| jpr  | 3   | 0    | 1    | 1    | najprawdopodobniej     |
| ps   | 3   | 0    | 1    | 1    | najprzeróżniejszych    |
| jpts | 3   | 0    | 0    | 1    | knajpce                |
| pw   | 3   | 0    | 1    | 1    | najpłynniejszych       |
| jsf  | 3   | 0    | 1    | 1    | najswobodniej          |

|                             |   |   |   |   |                      |
|-----------------------------|---|---|---|---|----------------------|
| j <sub>sk</sub>             | 3 | 1 | 1 | 1 | europejskich         |
| j <sub>sm</sub>             | 3 | 0 | 1 | 1 | sejsmiczne           |
| j <sub>sn</sub>             | 3 | 0 | 1 | 0 | najsnobistyczniej    |
| j <sub>sp</sub>             | 3 | 0 | 1 | 1 | najspokojniej        |
| j <sub>sr</sub>             | 3 | 0 | 0 | 1 | najsrozsza           |
| j <sub>st</sub>             | 3 | 1 | 1 | 1 | najstarszych         |
| j <sub>sts</sub>            | 3 | 1 | 1 | 1 | miejsce              |
| j <sub>sw</sub>             | 3 | 0 | 1 | 1 | najslyniejszych      |
| j <sub>sx</sub>             | 3 | 0 | 0 | 1 | ajschylosa           |
| j <sub>sk</sub>             | 3 | 0 | 1 | 1 | najszkodliwszy       |
| j <sub>sl</sub>             | 3 | 0 | 1 | 1 | najszlachetniejszych |
| j <sub>sm</sub>             | 3 | 0 | 1 | 1 | zmniejszy            |
| j <sub>st</sub>             | 3 | 0 | 1 | 1 | gieysztora           |
| j <sub>ste</sub>            | 3 | 0 | 1 | 1 | spójrzcie            |
| j <sub>st<sub>s</sub></sub> | 3 | 0 | 1 | 1 | najszczęśliwszym     |
| j <sub>sz</sub>             | 3 | 0 | 1 | 0 | zmniejszże           |
| j <sub>tf</sub>             | 3 | 0 | 1 | 1 | najtwardszych        |
| j <sub>tk</sub>             | 3 | 1 | 1 | 1 | wojtka               |
| j <sub>tl</sub>             | 3 | 0 | 0 | 1 | zabajtlować          |
| j <sub>tn</sub>             | 3 | 0 | 0 | 1 | lejtnant             |
| j <sub>tr</sub>             | 3 | 0 | 1 | 1 | najtrudniejszych     |
| j <sub>ts</sub>             | 3 | 0 | 1 | 1 | najtrzęwieszego      |
| j <sub>tt<sub>s</sub></sub> | 3 | 0 | 0 | 1 | wojtczak             |
| j <sub>tw</sub>             | 3 | 0 | 1 | 1 | najtłumniej          |
| j <sub>tx</sub>             | 3 | 0 | 1 | 0 | najtchórzliwiej      |
| j <sub>tsk</sub>            | 3 | 0 | 0 | 1 | kuwejcki             |
| j <sub>tsn</sub>            | 3 | 0 | 0 | 1 | najcnotliwszym       |
| j <sub>tek</sub>            | 3 | 0 | 0 | 1 | pójdźki              |
| j <sub>tete</sub>           | 3 | 0 | 1 | 1 | pójdźcie             |
| j <sub>tste</sub>           | 3 | 0 | 0 | 1 | najczcigodniejszy    |
| j <sub>vdz</sub>            | 3 | 0 | 1 | 1 | najwdzięczniejsze    |
| j <sub>vj</sub>             | 3 | 0 | 1 | 1 | największych         |
| j <sub>vr</sub>             | 3 | 0 | 0 | 1 | najwrażliwszej       |
| j <sub>vw</sub>             | 3 | 0 | 1 | 1 | najwłaściwszym       |
| j <sub>xl</sub>             | 3 | 0 | 0 | 1 | trójchloroetan       |
| j <sub>xr</sub>             | 3 | 0 | 0 | 1 | majchrowski          |
| j <sub>x<sub>s</sub></sub>  | 3 | 0 | 0 | 1 | majchrzak            |
| j <sub>xw</sub>             | 3 | 0 | 0 | 1 | najchłodniej         |
| j <sub>zb</sub>             | 3 | 0 | 0 | 1 | bejsbolowy           |
| j <sub>zd</sub>             | 3 | 0 | 0 | 1 | najzdolniejszych     |

| cl    | len | dict | para | corp | example            |
|-------|-----|------|------|------|--------------------|
| jzg   | 3   | 0    | 1    | 0    | najzgodniej        |
| jzj   | 3   | 0    | 0    | 1    | najzjadliwszy      |
| jzl   | 3   | 0    | 0    | 1    | bajzlu             |
| jzn   | 3   | 0    | 1    | 1    | najznakomitszych   |
| jzr   | 3   | 0    | 1    | 1    | najzręczniejszej   |
| jzv   | 3   | 0    | 1    | 1    | najzwyczajniejszej |
| jzw   | 3   | 0    | 1    | 1    | najzłośliwszy      |
| jzł   | 3   | 0    | 0    | 1    | podejrzliwie       |
| jzm   | 3   | 0    | 1    | 1    | spójrzmy           |
| jzv   | 3   | 0    | 0    | 1    | najzwawiej         |
| jzz   | 3   | 0    | 1    | 0    | spójrzże           |
| ~jel  | 3   | 1    | 1    | 1    | szczęśliwy         |
| ~jem  | 3   | 0    | 1    | 1    | któraśmy           |
| ~jep  | 3   | 1    | 1    | 1    | mięśni             |
| ~jete | 3   | 1    | 1    | 1    | części             |
| ~jez  | 3   | 0    | 1    | 0    | trząśże            |
| ~jsk  | 3   | 1    | 1    | 1    | amerykański        |
| ~jst  | 3   | 0    | 0    | 1    | państwko           |
| ~jsts | 3   | 0    | 1    | 1    | amerykańscy        |
| ~jstę | 3   | 0    | 0    | 1    | wileńszczyźnie     |
| ~jzb  | 3   | 0    | 0    | 1    | więzby             |
| ~jzl  | 3   | 0    | 1    | 1    | wstrzemięźliwość   |
| ~jzm  | 3   | 0    | 1    | 1    | gałęźmi            |
| ~jzp  | 3   | 1    | 1    | 1    | więźniów           |
| ~jzz  | 3   | 0    | 1    | 0    | uwięźże            |
| kete  | 3   | 0    | 1    | 1    | kontekście         |
| kfj   | 3   | 0    | 0    | 1    | relikwie           |
| kfl   | 3   | 0    | 0    | 1    | stagflacją         |
| kfr   | 3   | 0    | 0    | 1    | zygfryd            |
| kkj   | 3   | 0    | 1    | 1    | lekkiej            |
| klj   | 3   | 0    | 0    | 1    | herakliusz         |
| kpj   | 3   | 0    | 0    | 1    | wykpienia          |
| krf   | 3   | 0    | 0    | 1    | zakrwawiony        |
| ksb   | 3   | 0    | 0    | 1    | seksbomba          |
| ksd   | 3   | 0    | 0    | 1    | eksdyrektor        |
| ksdz  | 3   | 0    | 0    | 1    | eksdzokeja         |
| ksf   | 3   | 0    | 0    | 1    | oxford             |
| ksj   | 3   | 1    | 1    | 1    | refleksji          |
| ksk   | 3   | 0    | 0    | 1    | ekskomuniką        |

|      |   |   |   |   |                 |
|------|---|---|---|---|-----------------|
| ksl  | 3 | 0 | 0 | 1 | weksli          |
| ksm  | 3 | 0 | 0 | 1 | eksmisji        |
| ksn  | 3 | 0 | 0 | 1 | eksner          |
| ksp  | 3 | 0 | 0 | 1 | kuksnięciem     |
| ksp  | 3 | 1 | 1 | 1 | eksportu        |
| ksr  | 3 | 0 | 0 | 1 | eksrekordzista  |
| kss  | 3 | 0 | 0 | 1 | ekssekreтарz    |
| kss  | 3 | 0 | 0 | 1 | eksszefa        |
| kst  | 3 | 0 | 1 | 1 | tekstu          |
| ksts | 3 | 0 | 0 | 1 | ekscesy         |
| kstś | 3 | 0 | 0 | 1 | eksczempionami  |
| ksv  | 3 | 0 | 0 | 1 | ekswicepremiera |
| ksx  | 3 | 0 | 0 | 1 | ekshumacji      |
| ksz  | 3 | 0 | 0 | 1 | ekszawodnikowi  |
| ksz  | 3 | 0 | 0 | 1 | ekszony         |
| ksp  | 3 | 0 | 0 | 1 | bukszpanowa     |
| kst  | 3 | 1 | 1 | 1 | przekształcenia |
| ktj  | 3 | 0 | 0 | 1 | traktierni      |
| ktk  | 3 | 0 | 0 | 1 | architektka     |
| ktr  | 3 | 1 | 1 | 1 | elektrim        |
| ktś  | 3 | 0 | 0 | 1 | elektrze        |
| ktts | 3 | 0 | 0 | 1 | architektce     |
| ktj  | 3 | 1 | 1 | 1 | akcji           |
| kwb  | 3 | 0 | 1 | 1 | rzekłbym        |
| kwś  | 3 | 0 | 1 | 1 | wyrzekłszy      |
| lbj  | 3 | 0 | 0 | 1 | uwielbiam       |
| lbl  | 3 | 0 | 0 | 1 | elblągu         |
| lbr  | 3 | 0 | 0 | 1 | olbrot          |
| lbw  | 3 | 0 | 0 | 1 | wielbłąd        |
| lbz  | 3 | 0 | 0 | 1 | olbrzymie       |
| leñ  | 3 | 0 | 0 | 1 | olśniewające    |
| ldb  | 3 | 0 | 0 | 1 | goldberger      |
| ldj  | 3 | 0 | 0 | 1 | vivaldiego      |
| ldm  | 3 | 0 | 0 | 1 | feldmarszałka   |
| ldv  | 3 | 0 | 0 | 1 | szyldwacha      |
| ldzb | 3 | 0 | 0 | 1 | salzburgu       |
| lfj  | 3 | 0 | 0 | 1 | szelfie         |
| lfk  | 3 | 0 | 0 | 1 | sylwka          |
| lfr  | 3 | 0 | 0 | 1 | alfred          |
| lfś  | 3 | 0 | 0 | 1 | belfrze         |

| cl     | len | dict | para | corp | example       |
|--------|-----|------|------|------|---------------|
| lgj    | 3   | 0    | 0    | 1    | belgia        |
| lgn    | 3   | 0    | 0    | 1    | przyłgnęła    |
| lgn    | 3   | 0    | 0    | 1    | przyłgnie     |
| lgr    | 3   | 0    | 0    | 1    | belgradzie    |
| lgw    | 3   | 0    | 0    | 1    | zawilgły      |
| lgz    | 3   | 0    | 0    | 1    | pielgrzymki   |
| lkj    | 3   | 0    | 1    | 1    | wielkiej      |
| lkl    | 3   | 1    | 1    | 1    | folkloru      |
| lkm    | 3   | 0    | 0    | 1    | tolknicka     |
| lkn    | 3   | 1    | 1    | 1    | milkną        |
| lkn    | 3   | 0    | 1    | 1    | milknie       |
| lkr    | 3   | 0    | 0    | 1    | szylkretowego |
| lkw    | 3   | 0    | 1    | 1    | zamilkły      |
| lmj    | 3   | 0    | 1    | 1    | filmie        |
| lmt    | 3   | 0    | 0    | 1    | palmtopów     |
| lpl    | 3   | 0    | 0    | 1    | pelplinie     |
| lpr    | 3   | 0    | 0    | 1    | stalprodukt   |
| lsj    | 3   | 0    | 0    | 1    | celsjusza     |
| lsk    | 3   | 1    | 1    | 1    | polski        |
| lsn    | 3   | 0    | 0    | 1    | elsner        |
| lst    | 3   | 0    | 0    | 1    | ulsteru       |
| lsts   | 3   | 0    | 1    | 1    | polsce        |
| lšk    | 3   | 0    | 0    | 1    | halszka       |
| łst    | 3   | 0    | 0    | 1    | olsztyńska    |
| łstš   | 3   | 1    | 1    | 1    | opolszczyźnie |
| lte    | 3   | 0    | 0    | 1    | reynoldsie    |
| ltf    | 3   | 0    | 0    | 1    | feldfebla     |
| ltk    | 3   | 0    | 0    | 1    | waldka        |
| ltn    | 3   | 0    | 0    | 1    | jurgieltnika  |
| ltr    | 3   | 0    | 1    | 1    | filtry        |
| lts    | 3   | 0    | 0    | 1    | reynoldsa     |
| ltš    | 3   | 0    | 1    | 1    | filtre        |
| ltt    | 3   | 0    | 0    | 1    | oldtimerów    |
| łtsk   | 3   | 0    | 0    | 1    | grunwaldzka   |
| łtsl   | 3   | 0    | 0    | 1    | helcla        |
| łtsx   | 3   | 0    | 0    | 1    | alzheimera    |
| łtšk   | 3   | 0    | 0    | 1    | wilczki       |
| łtsm   | 3   | 0    | 1    | 1    | walczmy       |
| łtszte | 3   | 0    | 1    | 1    | walczcie      |

|        |   |   |   |   |                  |
|--------|---|---|---|---|------------------|
| łt̄sz̄ | 3 | 0 | 1 | 0 | walcze           |
| lvj    | 3 | 1 | 1 | 1 | jakichkolwiek    |
| lxj    | 3 | 0 | 0 | 1 | melchiora        |
| lxn    | 3 | 0 | 0 | 1 | pulchna          |
| lxp    | 3 | 0 | 0 | 1 | spulchniacze     |
| lzb    | 3 | 0 | 0 | 1 | himilsbach       |
| lzn    | 3 | 0 | 0 | 1 | pilzno           |
| lzn̄   | 3 | 0 | 0 | 1 | piłźnie          |
| mbd    | 3 | 0 | 0 | 1 | lambda           |
| mbj    | 3 | 1 | 1 | 1 | przedsiębiorstwa |
| mb̄l   | 3 | 0 | 0 | 1 | wembley          |
| mbm    | 3 | 0 | 1 | 0 | porąbmyż         |
| mbn    | 3 | 0 | 1 | 1 | odrębnych        |
| mb̄n̄  | 3 | 0 | 1 | 1 | odrębnie         |
| mbr    | 3 | 0 | 0 | 1 | dąbrowski        |
| mbw    | 3 | 0 | 0 | 1 | oziębło          |
| mbz̄   | 3 | 0 | 1 | 1 | zembrzuski       |
| mete   | 3 | 1 | 1 | 1 | zemścić          |
| mdl    | 3 | 1 | 1 | 1 | zemdlał          |
| mdw    | 3 | 0 | 1 | 0 | niemdło          |
| mfj    | 3 | 0 | 1 | 1 | triumfie         |
| mfl    | 3 | 0 | 0 | 1 | pamfletu         |
| mgl    | 3 | 0 | 1 | 1 | zamglone         |
| mgn̄   | 3 | 0 | 0 | 1 | okamgnienie      |
| mkf    | 3 | 0 | 0 | 1 | kumkwatem        |
| mkj    | 3 | 0 | 1 | 1 | zamkiem          |
| mkn    | 3 | 1 | 1 | 1 | zamknąć          |
| mk̄n̄  | 3 | 0 | 1 | 1 | zamknięcia       |
| mmj    | 3 | 0 | 0 | 1 | emmie            |
| mpj    | 3 | 1 | 1 | 1 | wystąpienia      |
| mpk    | 3 | 1 | 1 | 1 | zębki            |
| mpl    | 3 | 1 | 1 | 1 | egzemplarzy      |
| mpm    | 3 | 0 | 1 | 1 | wstąpmy          |
| mpn    | 3 | 1 | 1 | 1 | następnego       |
| mp̄n̄  | 3 | 1 | 1 | 1 | następnie        |
| mpr    | 3 | 1 | 1 | 1 | imprezy          |
| mps    | 3 | 0 | 0 | 1 | simpson          |
| mp̄s̄  | 3 | 0 | 0 | 1 | głębsze          |
| mpt̄   | 3 | 0 | 0 | 1 | southampton      |
| mpts   | 3 | 1 | 1 | 1 | zastępca         |

| cl                      | len | dict | para | corp | example           |
|-------------------------|-----|------|------|------|-------------------|
| mpt $\bar{c}$ e         | 3   | 0    | 1    | 1    | wystąpcie         |
| mpt $\bar{s}$           | 3   | 1    | 1    | 1    | przestępczości    |
| mpz $\bar{z}$           | 3   | 0    | 1    | 0    | wykapże           |
| msk                     | 3   | 1    | 1    | 1    | islamskich        |
| mst                     | 3   | 1    | 1    | 1    | białymstoku       |
| mst $\bar{s}$           | 3   | 0    | 1    | 1    | islamscy          |
| m $\bar{s}$ t           | 3   | 0    | 0    | 1    | rumsztyku         |
| m $\bar{s}$ t $\bar{s}$ | 3   | 0    | 1    | 1    | radomszczańskiego |
| ndg                     | 3   | 0    | 0    | 1    | montgomery        |
| ndj                     | 3   | 1    | 1    | 1    | holandia          |
| ndl                     | 3   | 1    | 1    | 1    | handlu            |
| ndn                     | 3   | 1    | 1    | 1    | oszczędności      |
| nd $\bar{n}$            | 3   | 1    | 1    | 1    | urzędników        |
| ndr                     | 3   | 1    | 1    | 1    | aleksandra        |
| ndv                     | 3   | 0    | 0    | 1    | połudwica         |
| ndw                     | 3   | 0    | 1    | 1    | żądło             |
| nd $\bar{z}$            | 3   | 1    | 1    | 1    | andrzej           |
| ndzb                    | 3   | 0    | 0    | 1    | tancbuda          |
| nd $\bar{z}$ l          | 3   | 0    | 1    | 1    | pędzla            |
| nd $\bar{z}$ m          | 3   | 0    | 1    | 1    | pieniężmi         |
| nd $\bar{z}$ n          | 3   | 0    | 0    | 1    | nędzne            |
| nd $\bar{z}$ $\bar{n}$  | 3   | 0    | 0    | 1    | nędznicy          |
| nfl                     | 3   | 1    | 1    | 1    | inflacji          |
| nfr                     | 3   | 0    | 0    | 1    | infrastruktury    |
| npr                     | 3   | 0    | 0    | 1    | lumpenproletariat |
| nsb                     | 3   | 0    | 0    | 1    | transbud          |
| nsd                     | 3   | 0    | 0    | 1    | transdukcji       |
| nsf                     | 3   | 1    | 1    | 1    | transformacji     |
| nsg                     | 3   | 0    | 0    | 1    | transgenicznych   |
| nsj                     | 3   | 1    | 1    | 1    | pensji            |
| nsk                     | 3   | 0    | 0    | 1    | skanska           |
| ns $\bar{l}$            | 3   | 0    | 0    | 1    | transliteracji    |
| nsm                     | 3   | 1    | 1    | 1    | transmisji        |
| nsn                     | 3   | 0    | 0    | 1    | transnarodowych   |
| nsp                     | 3   | 1    | 1    | 1    | transportu        |
| nss                     | 3   | 0    | 0    | 1    | transsystem       |
| nst                     | 3   | 1    | 1    | 1    | instytucji        |
| nst $\bar{s}$           | 3   | 0    | 0    | 1    | inscenizacji      |
| nsv                     | 3   | 0    | 0    | 1    | transwestytów     |

|       |   |   |   |   |                   |
|-------|---|---|---|---|-------------------|
| nsx   | 3 | 0 | 0 | 1 | transhoryzontalny |
| nşf   | 3 | 0 | 0 | 1 | brunszwiku        |
| nşl   | 3 | 0 | 0 | 1 | anszłus           |
| nşp   | 3 | 0 | 0 | 1 | grynszpanowego    |
| nşt   | 3 | 0 | 0 | 1 | kunsztu           |
| nşte  | 3 | 0 | 0 | 1 | kunszcie          |
| nte   | 3 | 0 | 0 | 1 | najzaciętsi       |
| ntf   | 3 | 0 | 0 | 1 | antwerpii         |
| ntj   | 3 | 0 | 0 | 1 | santiago          |
| ntk   | 3 | 1 | 1 | 1 | początku          |
| ntl   | 3 | 0 | 0 | 1 | pętli             |
| ntn   | 3 | 1 | 1 | 1 | umiejętności      |
| ntp   | 3 | 1 | 1 | 1 | chętnie           |
| ntp   | 3 | 1 | 1 | 1 | wątpić            |
| ntr   | 3 | 1 | 1 | 1 | centrum           |
| nts   | 3 | 0 | 0 | 1 | amundsena         |
| ntş   | 3 | 1 | 1 | 1 | wnętrza           |
| ntt   | 3 | 0 | 0 | 1 | landtagu          |
| ntts  | 3 | 0 | 1 | 1 | zarządcy          |
| nttş  | 3 | 0 | 0 | 1 | irlandczycy       |
| ntw   | 3 | 0 | 0 | 1 | wątle             |
| ntx   | 3 | 0 | 0 | 1 | trondheim         |
| ntşf  | 3 | 0 | 0 | 1 | bęcwał            |
| ntşj  | 3 | 1 | 1 | 1 | francja           |
| ntşk  | 3 | 1 | 1 | 1 | prezydenckich     |
| ntşl  | 3 | 0 | 0 | 1 | tysiąclecia       |
| ntşts | 3 | 0 | 1 | 1 | arogancy          |
| ntşw  | 3 | 0 | 0 | 1 | więclawska        |
| ntşk  | 3 | 1 | 1 | 1 | gorączka          |
| ntşl  | 3 | 0 | 0 | 1 | męczliwość        |
| ntşm  | 3 | 0 | 1 | 1 | jęczmyk           |
| ntşn  | 3 | 1 | 1 | 1 | miesięczne        |
| ntşp  | 3 | 1 | 1 | 1 | wyłącznie         |
| ntşts | 3 | 0 | 1 | 1 | gorączce          |
| ntşte | 3 | 0 | 1 | 1 | łącznie           |
| ntşz  | 3 | 0 | 1 | 0 | wyłączze          |
| nvj   | 3 | 0 | 0 | 1 | kanwie            |
| nrx   | 3 | 0 | 0 | 1 | synchroniczne     |
| nzb   | 3 | 0 | 0 | 1 | ginsberg          |
| nzj   | 3 | 1 | 1 | 1 | recenzje          |

| cl    | len | dict | para | corp | example             |
|-------|-----|------|------|------|---------------------|
| nbj   | 3   | 0    | 0    | 1    | hańbie              |
| ndzb  | 3   | 0    | 0    | 1    | ględźbę             |
| ndzm  | 3   | 0    | 1    | 1    | baǳmy               |
| ndzv  | 3   | 0    | 0    | 1    | lędźwi              |
| ndzz  | 3   | 0    | 1    | 0    | sporządźze          |
| nkj   | 3   | 0    | 1    | 1    | małeńkie            |
| ntedz | 3   | 0    | 0    | 1    | dziewięćdziesiątych |
| ntem  | 3   | 0    | 1    | 1    | poświęćmy           |
| ntes  | 3   | 0    | 0    | 1    | pięćset             |
| ntete | 3   | 0    | 1    | 1    | baǳcie              |
| ntez  | 3   | 0    | 1    | 0    | zachećże            |
| ntšk  | 3   | 0    | 0    | 1    | pomarańczką         |
| ntšm  | 3   | 0    | 1    | 1    | skończmy            |
| ntšte | 3   | 0    | 1    | 1    | skończie            |
| ntšz  | 3   | 0    | 1    | 0    | końcże              |
| ngj   | 3   | 1    | 1    | 1    | angielski           |
| ngl   | 3   | 1    | 1    | 1    | ciągle              |
| ngm   | 3   | 0    | 0    | 1    | ingmara             |
| ngn   | 3   | 1    | 1    | 1    | osiągnąć            |
| ngp   | 3   | 1    | 1    | 1    | osiągnięcia         |
| ngr   | 3   | 0    | 0    | 1    | węgry               |
| ngv   | 3   | 0    | 0    | 1    | chorągwi            |
| ngw   | 3   | 1    | 1    | 1    | ciągle              |
| ngz   | 3   | 0    | 0    | 1    | węgrzech            |
| ņke   | 3   | 0    | 1    | 1    | najwięksi           |
| ņkf   | 3   | 0    | 0    | 1    | frankfurcie         |
| ņkj   | 3   | 1    | 1    | 1    | warunkiem           |
| ņkk   | 3   | 1    | 1    | 1    | hongkongu           |
| ņkl   | 3   | 0    | 1    | 1    | konkluzji           |
| ņkn   | 3   | 1    | 1    | 1    | piękna              |
| ņkp   | 3   | 1    | 1    | 1    | pięknie             |
| ņkp   | 3   | 0    | 0    | 1    | pingpongista        |
| ņkr   | 3   | 1    | 1    | 1    | konkretnych         |
| ņks   | 3   | 0    | 0    | 1    | sfinksa             |
| ņkš   | 3   | 1    | 1    | 1    | większość           |
| ņkt   | 3   | 1    | 1    | 1    | punktów             |
| ņkts  | 3   | 0    | 1    | 1    | sankcyjne           |
| ņkte  | 3   | 1    | 1    | 1    | punkcie             |
| ņktš  | 3   | 0    | 0    | 1    | dziękczynienia      |

|       |   |   |   |   |                |
|-------|---|---|---|---|----------------|
| ηkw   | 3 | 0 | 1 | 1 | pęła           |
| ηkx   | 3 | 0 | 0 | 1 | szanghaju      |
| pel   | 3 | 0 | 0 | 1 | obślinionym    |
| pete  | 3 | 0 | 0 | 1 | obściskiwał    |
| pkj   | 3 | 0 | 1 | 1 | szybkiego      |
| pkł   | 3 | 0 | 0 | 1 | obkleić        |
| pkw   | 3 | 0 | 0 | 1 | obkłada        |
| plj   | 3 | 0 | 0 | 1 | panoplium      |
| pln   | 3 | 0 | 0 | 1 | cieplnej       |
| plp   | 3 | 0 | 0 | 1 | cieplnie       |
| prj   | 3 | 0 | 0 | 1 | cypriana       |
| psj   | 3 | 0 | 0 | 1 | autopsji       |
| psk   | 3 | 1 | 1 | 1 | arabskich      |
| psl   | 3 | 0 | 0 | 1 | bobsleje       |
| psm   | 3 | 0 | 0 | 1 | obsmażanymi    |
| psn   | 3 | 0 | 0 | 1 | wypsnęło       |
| psp   | 3 | 0 | 0 | 1 | wypsnie        |
| pst   | 3 | 1 | 1 | 1 | substancji     |
| psts  | 3 | 0 | 1 | 1 | arabscy        |
| psw   | 3 | 1 | 1 | 1 | obsługi        |
| pšk   | 3 | 0 | 0 | 1 | wieprzka       |
| pšm   | 3 | 0 | 1 | 0 | ulepszymż      |
| pšn   | 3 | 1 | 1 | 1 | pieprzny       |
| pšp   | 3 | 1 | 1 | 1 | pieprznie      |
| pšt   | 3 | 0 | 0 | 1 | kapsztadzie    |
| pšte  | 3 | 0 | 1 | 1 | odpieprzcie    |
| pštš  | 3 | 0 | 0 | 1 | kaszubszczyzny |
| pšz   | 3 | 0 | 1 | 0 | ulepszące      |
| ptk   | 3 | 0 | 0 | 1 | adeptką        |
| ptr   | 3 | 0 | 0 | 1 | globtroterów   |
| ptw   | 3 | 0 | 0 | 1 | obtłuczoną     |
| ptsj  | 3 | 1 | 1 | 1 | korupcji       |
| ptsm  | 3 | 0 | 0 | 1 | obcmokiwanych  |
| ptšm  | 3 | 0 | 1 | 0 | depczmy        |
| ptšte | 3 | 0 | 1 | 0 | podepczcież    |
| ptšz  | 3 | 0 | 1 | 0 | podepczże      |
| pwte  | 3 | 0 | 0 | 1 | dwupłciowa     |
| pxl   | 3 | 0 | 0 | 1 | zapchlonego    |
| pxn   | 3 | 1 | 1 | 1 | zepchnąć       |
| pxp   | 3 | 0 | 1 | 1 | zepchnięte     |

| cl   | len | dict | para | corp | example          |
|------|-----|------|------|------|------------------|
| rbj  | 3   | 0    | 1    | 1    | skarbiec         |
| rbl  | 3   | 0    | 0    | 1    | norblina         |
| rbn  | 3   | 0    | 0    | 1    | jednogarbne      |
| rbn  | 3   | 0    | 0    | 1    | skarbnik         |
| rbr  | 3   | 0    | 0    | 1    | superbroni       |
| ref  | 3   | 0    | 0    | 1    | morświnów        |
| rem  | 3   | 0    | 0    | 1    | superśmigłowcami |
| rep  | 3   | 0    | 0    | 1    | napiersnik       |
| rer  | 3   | 0    | 0    | 1    | superśredniej    |
| rete | 3   | 1    | 1    | 1    | pierścień        |
| rdb  | 3   | 0    | 0    | 1    | wartburg         |
| rdj  | 3   | 0    | 0    | 1    | verdiego         |
| rdl  | 3   | 1    | 1    | 1    | podgardle        |
| rdn  | 3   | 0    | 0    | 1    | szewardnadze     |
| rdn  | 3   | 0    | 0    | 1    | stwardnienie     |
| rdr  | 3   | 0    | 0    | 1    | hardrockowych    |
| rdv  | 3   | 0    | 0    | 1    | śniardwy         |
| rdw  | 3   | 1    | 1    | 1    | gardła           |
| rdzm | 3   | 0    | 1    | 1    | stwierdźmy       |
| rdzz | 3   | 0    | 1    | 0    | potwierdźże      |
| rfe  | 3   | 0    | 1    | 1    | pierwsi          |
| rfj  | 3   | 0    | 0    | 1    | harfie           |
| rfl  | 3   | 0    | 0    | 1    | interflower      |
| rfş  | 3   | 1    | 1    | 1    | pierwszy         |
| rft  | 3   | 0    | 0    | 1    | superwtorek      |
| rfts | 3   | 1    | 1    | 1    | czerwca          |
| rftş | 3   | 0    | 0    | 1    | czerwczyk        |
| rgj  | 3   | 1    | 1    | 1    | energię          |
| rgl  | 3   | 0    | 0    | 1    | superglina       |
| rgm  | 3   | 0    | 0    | 1    | bergmana         |
| rgn  | 3   | 0    | 0    | 1    | wtargnęli        |
| rgn  | 3   | 1    | 1    | 1    | wtargnięcie      |
| rgr  | 3   | 0    | 0    | 1    | supergrupy       |
| rgw  | 3   | 0    | 0    | 1    | czwórgłosie      |
| rgz  | 3   | 0    | 0    | 1    | supergrzałki     |
| rke  | 3   | 1    | 1    | 1    | marksizmu        |
| rkf  | 3   | 0    | 0    | 1    | cerkwi           |
| rkj  | 3   | 1    | 1    | 1    | parkiecie        |
| rkl  | 3   | 0    | 0    | 1    | merkla           |

|      |   |   |   |   |                  |
|------|---|---|---|---|------------------|
| rkm  | 3 | 0 | 0 | 1 | turkmenistanu    |
| rkn  | 3 | 1 | 1 | 1 | warknał          |
| rkn  | 3 | 0 | 1 | 1 | warknie          |
| rkr  | 3 | 0 | 0 | 1 | superkrótkiej    |
| rks  | 3 | 0 | 0 | 1 | marksa           |
| rkt  | 3 | 0 | 0 | 1 | antarktydy       |
| rkw  | 3 | 0 | 0 | 1 | uświerkłem       |
| rks  | 3 | 0 | 0 | 1 | durkheima        |
| rlb  | 3 | 0 | 0 | 1 | marlboro         |
| rlj  | 3 | 0 | 0 | 1 | berlioza         |
| rls  | 3 | 0 | 0 | 1 | girlsy           |
| rmj  | 3 | 1 | 1 | 1 | firmie           |
| rmk  | 3 | 0 | 0 | 1 | giermków         |
| rml  | 3 | 0 | 0 | 1 | supermleczne     |
| rmm  | 3 | 0 | 1 | 0 | nakarmmyż        |
| rmn  | 3 | 0 | 0 | 1 | karmniki         |
| rmtē | 3 | 0 | 1 | 0 | nakarmcież       |
| rmz  | 3 | 0 | 1 | 0 | karmże           |
| rns  | 3 | 0 | 0 | 1 | burnsa           |
| rnts | 3 | 0 | 0 | 1 | garncarz         |
| rnv  | 3 | 0 | 0 | 1 | kornwalii        |
| rnx  | 3 | 0 | 0 | 1 | kornhauser       |
| rng  | 3 | 0 | 0 | 1 | horngacher       |
| rnk  | 3 | 1 | 1 | 1 | garnki           |
| rpe  | 3 | 0 | 0 | 1 | superpsiarnia    |
| rpj  | 3 | 1 | 1 | 1 | sierpień         |
| rpk  | 3 | 1 | 1 | 1 | uszczerbku       |
| rpl  | 3 | 1 | 1 | 1 | cierpliwość      |
| rpm  | 3 | 0 | 1 | 0 | cierpmyż         |
| rpn  | 3 | 1 | 1 | 1 | zaczerpnął       |
| rpn  | 3 | 1 | 1 | 1 | sierpnia         |
| rpr  | 3 | 0 | 0 | 1 | interpretacji    |
| rps  | 3 | 0 | 0 | 1 | cyberprzestrzeni |
| rpt  | 3 | 0 | 0 | 1 | ekscerpty        |
| rpts | 3 | 0 | 1 | 1 | skarbcza         |
| rptē | 3 | 0 | 1 | 0 | cierpcie         |
| rptš | 3 | 0 | 0 | 1 | skarbczyk        |
| rpw  | 3 | 0 | 0 | 1 | ścierpła         |
| rpz  | 3 | 0 | 1 | 0 | czerpże          |
| rrj  | 3 | 0 | 0 | 1 | barrie           |

| cl    | len | dict | para | corp | example       |
|-------|-----|------|------|------|---------------|
| rsf   | 3   | 0    | 0    | 1    | perswazji     |
| rsj   | 3   | 1    | 1    | 1    | wersji        |
| rsk   | 3   | 1    | 1    | 1    | maklerskich   |
| rsn   | 3   | 0    | 0    | 1    | supersnajper  |
| rsp   | 3   | 1    | 1    | 1    | perspektywy   |
| rst   | 3   | 1    | 1    | 1    | horsta        |
| rsts  | 3   | 0    | 1    | 1    | holenderscy   |
| rsw   | 3   | 0    | 0    | 1    | supersłużby   |
| rşf   | 3   | 0    | 0    | 1    | hirszfelda    |
| rşk   | 3   | 0    | 0    | 1    | konsjerżki    |
| rşl   | 3   | 0    | 0    | 1    | durszlak      |
| rşm   | 3   | 0    | 1    | 1    | forszmak      |
| rşn   | 3   | 0    | 0    | 1    | charsznica    |
| rşr   | 3   | 0    | 0    | 1    | marszruty     |
| rşt   | 3   | 1    | 1    | 1    | warsztaty     |
| rşts  | 3   | 0    | 0    | 1    | samodzierżą   |
| rşte  | 3   | 0    | 1    | 0    | zgorszcie     |
| rştş  | 3   | 1    | 1    | 1    | zmarszczek    |
| rşz   | 3   | 0    | 1    | 0    | zgorsze       |
| rte   | 3   | 0    | 0    | 1    | najtwardsi    |
| rtf   | 3   | 1    | 1    | 1    | portfela      |
| rtj   | 3   | 1    | 1    | 1    | partia        |
| rtk   | 3   | 1    | 1    | 1    | kartki        |
| rtl   | 3   | 0    | 0    | 1    | portland      |
| rtm   | 3   | 1    | 1    | 1    | dortmund      |
| rtn   | 3   | 0    | 0    | 1    | partnerów     |
| rtñ   | 3   | 0    | 0    | 1    | bartnik       |
| rtr   | 3   | 1    | 1    | 1    | portret       |
| rtş   | 3   | 0    | 0    | 1    | twardsze      |
| rtts  | 3   | 0    | 1    | 1    | kartce        |
| rttş  | 3   | 0    | 0    | 1    | bartczak      |
| rtw   | 3   | 0    | 0    | 1    | bartłomiej    |
| rtşj  | 3   | 1    | 1    | 1    | konsorcjum    |
| rtşk  | 3   | 0    | 0    | 1    | frankfurckich |
| rtşm  | 3   | 0    | 0    | 1    | harc mistrz   |
| rtşp  | 3   | 0    | 0    | 1    | marcpol       |
| rtşt  | 3   | 0    | 0    | 1    | arctowskiego  |
| rtştş | 3   | 0    | 0    | 1    | stargardzcy   |
| rtşv  | 3   | 0    | 0    | 1    | szwarcwaldzie |

|                                   |   |   |   |   |                        |
|-----------------------------------|---|---|---|---|------------------------|
| rt <sup>~</sup> sx                | 3 | 0 | 0 | 1 | szwarzcharakter        |
| rt <sup>~</sup> ed                | 3 | 0 | 0 | 1 | ćwierćdolarówki        |
| rt <sup>~</sup> ef                | 3 | 0 | 0 | 1 | ćwierćfinale           |
| rt <sup>~</sup> ek                | 3 | 0 | 0 | 1 | ćwierćkontaktów        |
| rt <sup>~</sup> el                | 3 | 0 | 0 | 1 | ćwierćlitrową          |
| rt <sup>~</sup> em                | 3 | 0 | 1 | 1 | ćwierćmilionowego      |
| rt <sup>~</sup> ep                | 3 | 0 | 0 | 1 | ćwierćpolityków        |
| rt <sup>~</sup> er                | 3 | 0 | 0 | 1 | ćwierćromanse          |
| rt <sup>~</sup> et                | 3 | 0 | 0 | 1 | ćwierćtonowy           |
| rt <sup>~</sup> e <sup>~</sup> te | 3 | 0 | 1 | 0 | stwierdżcie            |
| rt <sup>~</sup> e <sup>~</sup> z  | 3 | 0 | 1 | 0 | wierćze                |
| rt <sup>~</sup> sk                | 3 | 0 | 0 | 1 | utarczki               |
| rt <sup>~</sup> sl                | 3 | 0 | 0 | 1 | kurczliwość            |
| rt <sup>~</sup> sm                | 3 | 0 | 1 | 1 | karczma                |
| rt <sup>~</sup> sn                | 3 | 0 | 0 | 1 | jarmarczne             |
| rt <sup>~</sup> sp                | 3 | 0 | 0 | 1 | jarmarcznie            |
| rt <sup>~</sup> sts               | 3 | 0 | 0 | 1 | utarczce               |
| rt <sup>~</sup> ste               | 3 | 0 | 1 | 0 | dostarczcież           |
| rt <sup>~</sup> sz                | 3 | 0 | 1 | 0 | skurczże               |
| rvj                               | 3 | 1 | 1 | 1 | przerwie               |
| rvn                               | 3 | 1 | 1 | 1 | barwne                 |
| rvp                               | 3 | 0 | 1 | 1 | barwnie                |
| rvr                               | 3 | 0 | 0 | 1 | superwroga             |
| rvz                               | 3 | 0 | 0 | 1 | superwziernik          |
| rwb                               | 3 | 0 | 1 | 1 | poparłby               |
| rwş                               | 3 | 0 | 1 | 1 | dotarłszy              |
| rx <sup>~</sup> f                 | 3 | 0 | 1 | 1 | marchwi                |
| rxj                               | 3 | 0 | 0 | 1 | hierarchia             |
| rxl                               | 3 | 0 | 0 | 1 | purchła                |
| rxn                               | 3 | 0 | 0 | 1 | szturchnął             |
| rxp                               | 3 | 0 | 0 | 1 | szturchnięcie          |
| rzeb                              | 3 | 0 | 0 | 1 | petersburgu            |
| rzl                               | 3 | 0 | 1 | 1 | zmarzliny              |
| rnz                               | 3 | 1 | 1 | 1 | marzną                 |
| rzp                               | 3 | 0 | 1 | 1 | zamarznięte            |
| rzw                               | 3 | 0 | 1 | 1 | zamarzło               |
| rz <sup>~</sup> m                 | 3 | 0 | 1 | 1 | skarżmy                |
| rz <sup>~</sup> n                 | 3 | 0 | 0 | 1 | wyrz <sup>~</sup> nęli |
| rz <sup>~</sup> p                 | 3 | 0 | 0 | 1 | katorżnicy             |
| rz <sup>~</sup> z                 | 3 | 0 | 1 | 0 | skarżże                |

| cl   | len | dict | para | corp | example        |
|------|-----|------|------|------|----------------|
| sef  | 3   | 0    | 0    | 1    | rozświtu       |
| seɲ  | 3   | 0    | 0    | 1    | bezsnieżne     |
| ser  | 3   | 0    | 0    | 1    | rozśrodkowania |
| sete | 3   | 0    | 0    | 1    | bezściółkową   |
| ske  | 3   | 0    | 0    | 1    | bezksiężycową  |
| skdź | 3   | 0    | 0    | 1    | dyskdźokeja    |
| skf  | 3   | 1    | 1    | 1    | moskwy         |
| skj  | 3   | 0    | 1    | 1    | wnioskiem      |
| skl  | 3   | 1    | 1    | 1    | troskliwie     |
| skm  | 3   | 0    | 0    | 1    | discman        |
| skr  | 3   | 1    | 1    | 1    | dyskryminacji  |
| skś  | 3   | 0    | 1    | 1    | oskrzeli       |
| skw  | 3   | 1    | 1    | 1    | rozkład        |
| smj  | 3   | 0    | 1    | 1    | pasmie         |
| smr  | 3   | 0    | 0    | 1    | zasmrodzenie   |
| sɲk  | 3   | 1    | 1    | 1    | czosnku        |
| spj  | 3   | 1    | 1    | 1    | bezpieczeństwa |
| spl  | 3   | 1    | 1    | 1    | biznesplan     |
| spɲ  | 3   | 0    | 0    | 1    | wespnie        |
| spr  | 3   | 1    | 1    | 1    | rozprawy       |
| spś  | 3   | 0    | 0    | 1    | wesprzeć       |
| spw  | 3   | 1    | 1    | 1    | bezpłatnie     |
| ssk  | 3   | 0    | 0    | 1    | bezsukutecznie |
| ssm  | 3   | 0    | 0    | 1    | rozsmakowywać  |
| ssn  | 3   | 0    | 0    | 1    | rozsnuć        |
| ssp  | 3   | 0    | 0    | 1    | bezsborne      |
| sst  | 3   | 1    | 1    | 1    | rozstanie      |
| ssw  | 3   | 0    | 0    | 1    | rozślawił      |
| sśf  | 3   | 0    | 0    | 1    | bezszywowych   |
| sśk  | 3   | 0    | 0    | 1    | bezszkodowy    |
| sśm  | 3   | 0    | 0    | 1    | bezszerowo     |
| sśɲ  | 3   | 0    | 0    | 1    | bezsnurowowych |
| sśtś | 3   | 0    | 0    | 1    | rozszczepienia |
| stb  | 3   | 0    | 0    | 1    | postbank       |
| std  | 3   | 0    | 0    | 1    | postdata       |
| stf  | 3   | 1    | 1    | 1    | towarzystwa    |
| stj  | 3   | 1    | 1    | 1    | kwestia        |
| stk  | 3   | 1    | 1    | 1    | wszystkich     |
| stl  | 3   | 0    | 0    | 1    | whistler       |

|      |   |   |   |   |                    |
|------|---|---|---|---|--------------------|
| stm  | 3 | 0 | 0 | 1 | investment         |
| stn  | 3 | 1 | 1 | 1 | korzystne          |
| stñ  | 3 | 1 | 1 | 1 | istnieje           |
| stp  | 3 | 0 | 0 | 1 | postpenitencjarnej |
| str  | 3 | 1 | 1 | 1 | ministra           |
| sts  | 3 | 0 | 0 | 1 | bestsellerów       |
| stś  | 3 | 1 | 1 | 1 | mistrzostw         |
| stt  | 3 | 0 | 0 | 1 | posttotalitarna    |
| stts | 3 | 0 | 1 | 1 | jednostce          |
| stw  | 3 | 0 | 0 | 1 | beztłuszczową      |
| stx  | 3 | 0 | 0 | 1 | posthumanistycznej |
| stz  | 3 | 0 | 0 | 1 | postzimnowojennej  |
| stż  | 3 | 0 | 0 | 1 | postrzegany        |
| stsf | 3 | 0 | 0 | 1 | biznescwaniacy     |
| stsw | 3 | 0 | 0 | 1 | bezcłowy           |
| stśw | 3 | 0 | 0 | 1 | rozcłonkowania     |
| swb  | 3 | 0 | 1 | 1 | wyniosłby          |
| swś  | 3 | 0 | 1 | 1 | podrósłszy         |
| sxf  | 3 | 0 | 0 | 1 | rozchwytywane      |
| sxl  | 3 | 0 | 1 | 1 | oschle             |
| sxm  | 3 | 0 | 0 | 1 | bezhmurne          |
| sxn  | 3 | 1 | 1 | 1 | wyschną            |
| sxp  | 3 | 0 | 1 | 1 | wyschnie           |
| sxr  | 3 | 0 | 0 | 1 | pseudoschronisk    |
| sxw  | 3 | 0 | 1 | 1 | wyszło             |
| śdźb | 3 | 0 | 0 | 1 | wieszczby          |
| śfj  | 3 | 0 | 1 | 1 | podeszwie          |
| śkf  | 3 | 0 | 0 | 1 | paszkwil           |
| śkj  | 3 | 0 | 1 | 1 | iwaszkiewicz       |
| śkl  | 3 | 0 | 0 | 1 | przeszkłone        |
| śkp  | 3 | 0 | 0 | 1 | zgorzkniały        |
| ślp  | 3 | 0 | 0 | 1 | kaszlnięć          |
| śmj  | 3 | 0 | 0 | 1 | oszmiańskiego      |
| śpj  | 3 | 0 | 0 | 1 | przeszpiegi        |
| śpl  | 3 | 0 | 0 | 1 | raszplą            |
| śpr  | 3 | 0 | 0 | 1 | naszprycowanego    |
| śtk  | 3 | 1 | 1 | 1 | resztki            |
| śtr  | 3 | 0 | 0 | 1 | musztry            |
| śtś  | 3 | 0 | 0 | 1 | musztrze           |
| śtts | 3 | 0 | 1 | 0 | resztce            |

| cl    | len | dict | para | corp | example       |
|-------|-----|------|------|------|---------------|
| stsk  | 3   | 1    | 1    | 1    | przejażdżki   |
| stsm  | 3   | 0    | 1    | 1    | niszcmy       |
| ststs | 3   | 0    | 1    | 1    | przejażdżce   |
| stste | 3   | 0    | 1    | 1    | niszczcie     |
| stsz  | 3   | 0    | 1    | 0    | pogłaszczce   |
| sxl   | 3   | 0    | 0    | 1    | rozpierzchli  |
| sxn   | 3   | 0    | 0    | 1    | zwierzchności |
| sxp   | 3   | 1    | 1    | 1    | powierzchni   |
| sxw   | 3   | 0    | 0    | 1    | zamierzchłych |
| tef   | 3   | 0    | 0    | 1    | przedświt     |
| tel   | 3   | 0    | 0    | 1    | przedślubnej  |
| ten   | 3   | 0    | 0    | 1    | odśnieżania   |
| ter   | 3   | 0    | 0    | 1    | odśrodkowych  |
| tete  | 3   | 0    | 0    | 1    | podścienie    |
| tfj   | 3   | 1    | 1    | 1    | potwierdził   |
| tfm   | 3   | 0    | 1    | 1    | ułatwmy       |
| tfr   | 3   | 0    | 0    | 1    | gotfryd       |
| tfs   | 3   | 0    | 0    | 1    | odwszeniu     |
| tfte  | 3   | 0    | 1    | 0    | załatwcie     |
| tftś  | 3   | 1    | 1    | 1    | przedwczesne  |
| tfz   | 3   | 0    | 1    | 0    | ułatwze       |
| tkf   | 3   | 1    | 1    | 1    | utkwila       |
| tkj   | 3   | 1    | 1    | 1    | podatkiem     |
| tkl   | 3   | 0    | 0    | 1    | dotkliwie     |
| tkn   | 3   | 1    | 1    | 1    | dotknęła      |
| tkp   | 3   | 0    | 1    | 1    | dotkniętych   |
| tkp   | 3   | 0    | 0    | 1    | podkpiwał     |
| tkr   | 3   | 1    | 1    | 1    | podkreślił    |
| tkś   | 3   | 0    | 0    | 1    | odkrzyknął    |
| tkw   | 3   | 1    | 1    | 1    | przedkłada    |
| tlm   | 3   | 0    | 1    | 0    | wyświetlmy    |
| tln   | 3   | 1    | 1    | 1    | świetlnych    |
| tlp   | 3   | 0    | 1    | 0    | świetlni      |
| tlte  | 3   | 0    | 1    | 1    | prześwietlcie |
| tlz   | 3   | 0    | 1    | 0    | oświetlże     |
| tmj   | 3   | 0    | 1    | 1    | rytmie        |
| tpj   | 3   | 1    | 1    | 1    | odpiera       |
| tpl   | 3   | 0    | 0    | 1    | ponadplanowy  |
| tpr   | 3   | 1    | 1    | 1    | odprawy       |

|       |   |   |   |   |                     |
|-------|---|---|---|---|---------------------|
| tps   | 3 | 0 | 0 | 1 | nadpsuty            |
| tpş   | 3 | 1 | 1 | 1 | odprzedaży          |
| tpw   | 3 | 1 | 1 | 1 | nieodpłatnie        |
| trf   | 3 | 1 | 1 | 1 | potrwa              |
| trj   | 3 | 1 | 1 | 1 | repatriacji         |
| trk   | 3 | 0 | 0 | 1 | piotrków            |
| tsf   | 3 | 0 | 0 | 1 | botswana            |
| tsk   | 3 | 0 | 0 | 1 | piłsudski           |
| tsm   | 3 | 0 | 0 | 1 | przedsmak           |
| tsp   | 3 | 0 | 0 | 1 | nadspodziewanie     |
| tst   | 3 | 1 | 1 | 1 | podstawie           |
| tsts  | 3 | 0 | 0 | 1 | podscenia           |
| tsw   | 3 | 1 | 1 | 1 | odśłania            |
| tşk   | 3 | 1 | 1 | 1 | odszkodowania       |
| tşl   | 3 | 0 | 0 | 1 | podszlifować        |
| tşm   | 3 | 0 | 1 | 1 | popatrzmy           |
| tşn   | 3 | 1 | 1 | 1 | powietrznych        |
| tşp   | 3 | 0 | 1 | 1 | nieopatrnie         |
| tşp   | 3 | 0 | 0 | 1 | przedszpitalnych    |
| tşt   | 3 | 0 | 0 | 1 | nadsztygar          |
| tşte  | 3 | 0 | 1 | 1 | patrzcie            |
| tstş  | 3 | 0 | 0 | 1 | podszczecińskich    |
| tşz   | 3 | 0 | 1 | 0 | popatrzże           |
| ttf   | 3 | 1 | 1 | 1 | odtworzyć           |
| ttj   | 3 | 0 | 0 | 1 | donizettiego        |
| ttl   | 3 | 0 | 0 | 1 | podtlenek           |
| ttr   | 3 | 0 | 0 | 1 | odtracony           |
| ttş   | 3 | 1 | 1 | 1 | podtrzymał          |
| ttw   | 3 | 0 | 0 | 1 | odtłuszczonego      |
| ttşm  | 3 | 0 | 1 | 0 | świadczmyż          |
| ttşt  | 3 | 0 | 0 | 1 | ponadczterogodzinne |
| ttşte | 3 | 0 | 1 | 0 | świadczcie          |
| ttşw  | 3 | 0 | 0 | 1 | nadczołwieka        |
| ttşz  | 3 | 0 | 1 | 0 | świadczże           |
| twb   | 3 | 0 | 1 | 1 | wymiótlby           |
| twş   | 3 | 0 | 1 | 0 | zamiótlszy          |
| txf   | 3 | 0 | 0 | 1 | podchwycili         |
| txl   | 3 | 0 | 0 | 1 | nadchloroetylenu    |
| txn   | 3 | 1 | 1 | 1 | odetchnąć           |
| txp   | 3 | 1 | 1 | 1 | wytchnienia         |

| cl   | len | dict | para | corp | example               |
|------|-----|------|------|------|-----------------------|
| txr  | 3   | 0    | 0    | 1    | podhrubieszowskiej    |
| txş  | 3   | 0    | 0    | 1    | przedchrześcijańskich |
| txw  | 3   | 0    | 0    | 1    | otchłań               |
| tzw  | 3   | 0    | 0    | 1    | kilkusetzłotową       |
| tşgr | 3   | 0    | 0    | 1    | specgrupa             |
| tşkj | 3   | 0    | 1    | 1    | niemieckiej           |
| tşkn | 3   | 0    | 0    | 1    | ocknął                |
| tşkp | 3   | 0    | 0    | 1    | ocknie                |
| tşpj | 3   | 0    | 0    | 1    | hucpiarskim           |
| tşpr | 3   | 0    | 0    | 1    | kacpra                |
| tşpş | 3   | 0    | 0    | 1    | kacprzak              |
| tşsw | 3   | 0    | 0    | 1    | specsłużb             |
| tşşp | 3   | 0    | 0    | 1    | specszpitali          |
| tştf | 3   | 1    | 1    | 1    | województwa           |
| tştr | 3   | 0    | 0    | 1    | spectrum              |
| tşzn | 3   | 0    | 0    | 1    | specznak              |
| tefj | 3   | 0    | 0    | 1    | poćwiartowano         |
| tekj | 3   | 0    | 0    | 1    | maćkiem               |
| temj | 3   | 0    | 0    | 1    | zaciemnienie          |
| tşfn | 3   | 0    | 0    | 1    | wieczfnia             |
| tşkj | 3   | 0    | 1    | 1    | miasteczkiem          |
| tşmj | 3   | 0    | 0    | 1    | kucznie               |
| tştm | 3   | 0    | 0    | 1    | pocztmistrza          |
| vbj  | 3   | 0    | 0    | 1    | przeciwbiegun         |
| vbw  | 3   | 0    | 0    | 1    | przeciwbłotne         |
| vdj  | 3   | 0    | 0    | 1    | awdiejew              |
| vdr  | 3   | 0    | 0    | 1    | niewdrożenia          |
| vđzm | 3   | 0    | 1    | 1    | sprawdźmy             |
| vđzz | 3   | 0    | 1    | 0    | sprawdźże             |
| vgn  | 3   | 0    | 0    | 1    | powgniatani           |
| vgr  | 3   | 0    | 0    | 1    | przeciwgruźlicze      |
| vgz  | 3   | 0    | 0    | 1    | przeciwgrzybiczymi    |
| vmj  | 3   | 0    | 0    | 1    | przeciwmiażdżycowe    |
| vvj  | 3   | 0    | 0    | 1    | przeciwwiatrową       |
| vwv  | 3   | 0    | 0    | 1    | przeciwwłamaniowymi   |
| vvz  | 3   | 0    | 0    | 1    | przeciwwrzodowy       |
| vzb  | 3   | 0    | 0    | 1    | niewzbożacy           |
| vzd  | 3   | 0    | 0    | 1    | powzdychać            |
| vzg  | 3   | 0    | 0    | 1    | płaskowzgórze         |

|      |   |   |   |   |                       |
|------|---|---|---|---|-----------------------|
| vzm  | 3 | 0 | 0 | 1 | przeciwzmarszczkowego |
| vzn  | 3 | 0 | 0 | 1 | niewznawianej         |
| vzn  | 3 | 0 | 1 | 1 | uwzniożenia           |
| vzr  | 3 | 0 | 0 | 1 | krótkowzroczność      |
| vzd  | 3 | 0 | 0 | 1 | zawždy                |
| wbj  | 3 | 0 | 0 | 1 | kołbiel               |
| wbr  | 3 | 0 | 0 | 1 | współbraci            |
| wbz  | 3 | 0 | 0 | 1 | laubzega              |
| wbz  | 3 | 0 | 0 | 1 | wałbrzych             |
| wep  | 3 | 0 | 0 | 1 | półśnie               |
| wer  | 3 | 0 | 0 | 1 | półśredniej           |
| wete | 3 | 0 | 0 | 1 | holokauście           |
| wdj  | 3 | 0 | 0 | 1 | audio                 |
| wdp  | 3 | 0 | 0 | 1 | półdniową             |
| wdr  | 3 | 1 | 1 | 1 | kołdry                |
| wdw  | 3 | 0 | 0 | 1 | współdłużników        |
| wdz  | 3 | 0 | 1 | 1 | kołdrze               |
| wfe  | 3 | 0 | 0 | 1 | półwsi                |
| wfm  | 3 | 0 | 0 | 1 | kaufmana              |
| wfr  | 3 | 0 | 0 | 1 | półfrak               |
| wfs  | 3 | 0 | 0 | 1 | półwsad               |
| wgj  | 3 | 0 | 1 | 1 | czołgiem              |
| wgr  | 3 | 0 | 0 | 1 | współgra              |
| wgw  | 3 | 1 | 1 | 1 | półgłosem             |
| wke  | 3 | 0 | 0 | 1 | półksięzycza          |
| wkj  | 3 | 0 | 1 | 1 | całkiem               |
| wkl  | 3 | 0 | 0 | 1 | zgiełkliwe            |
| wkn  | 3 | 1 | 1 | 1 | przełknąć             |
| wkp  | 3 | 0 | 1 | 1 | przełknięcia          |
| wkr  | 3 | 0 | 0 | 1 | dwuipółkrotnie        |
| wkx  | 3 | 0 | 0 | 1 | półkrzyżem            |
| wkts | 3 | 0 | 0 | 1 | aukcyjny              |
| wkw  | 3 | 0 | 0 | 1 | pożółkłych            |
| wlj  | 3 | 0 | 0 | 1 | pauliańska            |
| wmg  | 3 | 0 | 0 | 1 | baumgart              |
| wmj  | 3 | 0 | 1 | 1 | chełmie               |
| wmk  | 3 | 0 | 0 | 1 | chełmka               |
| wmn  | 3 | 0 | 0 | 1 | chełmna               |
| wmp  | 3 | 0 | 0 | 1 | chełmnie              |
| wmr  | 3 | 0 | 0 | 1 | półmroku              |

| cl    | len | dict | para | corp | example           |
|-------|-----|------|------|------|-------------------|
| wmz   | 3   | 0    | 0    | 1    | chełmżyńska       |
| wpm   | 3   | 0    | 1    | 0    | pełny             |
| wpte  | 3   | 0    | 1    | 0    | spełnicieź        |
| wpz   | 3   | 0    | 1    | 0    | spełnze           |
| wpj   | 3   | 0    | 1    | 1    | małpie            |
| wpk   | 3   | 0    | 0    | 1    | małpki            |
| wpl   | 3   | 0    | 0    | 1    | współplemieńcy    |
| wpr   | 3   | 1    | 1    | 1    | współpracy        |
| wps   | 3   | 0    | 0    | 1    | półprzewodników   |
| wpts  | 3   | 0    | 0    | 1    | małpce            |
| wpw   | 3   | 0    | 0    | 1    | współpłatności    |
| wsk   | 3   | 0    | 0    | 1    | sosnowski         |
| wsn   | 3   | 0    | 0    | 1    | półsnu            |
| wsp   | 3   | 0    | 0    | 1    | auspicjami        |
| wst   | 3   | 0    | 0    | 1    | holokaustu        |
| wsts  | 3   | 0    | 0    | 1    | współscenarzystą  |
| wsw   | 3   | 0    | 0    | 1    | półsłówkami       |
| wśf   | 3   | 0    | 0    | 1    | półszwadronem     |
| wśl   | 3   | 0    | 0    | 1    | półszlachetnych   |
| wst   | 3   | 0    | 0    | 1    | półsztywna        |
| wsts  | 3   | 0    | 0    | 1    | bełzcu            |
| wtf   | 3   | 0    | 0    | 1    | współtwórca       |
| wtk   | 3   | 1    | 1    | 1    | kształtki         |
| wtn   | 3   | 0    | 0    | 1    | człekokształtnych |
| wtp   | 3   | 0    | 0    | 1    | przekształtnik    |
| wtr   | 3   | 0    | 0    | 1    | półtrykotach      |
| wts   | 3   | 0    | 0    | 1    | autsajder         |
| wtw   | 3   | 0    | 0    | 1    | półtłustego       |
| wtsj  | 3   | 1    | 1    | 1    | kaucji            |
| wtsk  | 3   | 0    | 0    | 1    | ełcki             |
| wtsts | 3   | 0    | 0    | 1    | ełccy             |
| wtem  | 3   | 0    | 1    | 0    | kształćmy         |
| wtete | 3   | 0    | 1    | 0    | wykształćcieź     |
| wtez  | 3   | 0    | 1    | 0    | wykształćće       |
| wtsf  | 3   | 0    | 0    | 1    | półczwarta        |
| wtsw  | 3   | 0    | 0    | 1    | półczłowiekiem    |
| wvj   | 3   | 1    | 1    | 1    | półwiecza         |
| wvw   | 3   | 0    | 0    | 1    | współwłasności    |
| wwb   | 3   | 0    | 1    | 0    | mełłbyś           |

|       |   |   |   |   |                     |
|-------|---|---|---|---|---------------------|
| wwę   | 3 | 0 | 1 | 0 | zmełszy             |
| wxl   | 3 | 0 | 0 | 1 | półchleba           |
| wxw   | 3 | 0 | 0 | 1 | półchłopek          |
| wzl   | 3 | 0 | 1 | 1 | pełziśmy            |
| wzn   | 3 | 1 | 1 | 1 | okielznać           |
| wzn   | 3 | 0 | 1 | 1 | pełźnie             |
| wzw   | 3 | 0 | 1 | 1 | spelzły             |
| wzl   | 3 | 0 | 1 | 0 | pełzliby            |
| wzn   | 3 | 0 | 1 | 0 | pełzniecie          |
| ~wfş  | 3 | 0 | 1 | 1 | począwszy           |
| ~wsk  | 3 | 1 | 1 | 1 | związku             |
| ~wsn  | 3 | 1 | 1 | 1 | mięsnych            |
| ~wst  | 3 | 1 | 1 | 1 | często              |
| ~wsts | 3 | 0 | 1 | 1 | zwycięzca           |
| ~wsw  | 3 | 1 | 1 | 1 | przesła             |
| ~wşe  | 3 | 0 | 0 | 1 | najciężsi           |
| ~wşk  | 3 | 1 | 1 | 1 | książki             |
| ~wşp  | 3 | 0 | 0 | 1 | zamażpójścia        |
| ~wşş  | 3 | 0 | 0 | 1 | najcięższych        |
| ~wştş | 3 | 0 | 1 | 1 | książce             |
| ~wşte | 3 | 0 | 1 | 1 | rozwiązcie          |
| ~wştş | 3 | 1 | 1 | 1 | mężczyzn            |
| ~wvz  | 3 | 0 | 0 | 1 | przedsięwzięcia     |
| ~wxl  | 3 | 0 | 0 | 1 | stęchlizny          |
| ~wxw  | 3 | 0 | 0 | 1 | zatechły            |
| ~wzn  | 3 | 0 | 0 | 1 | grzęzną             |
| ~wzv  | 3 | 0 | 0 | 1 | gązwa               |
| ~wzw  | 3 | 1 | 1 | 1 | węzła               |
| ~wzl  | 3 | 0 | 0 | 1 | uciążliwe           |
| ~wzm  | 3 | 0 | 1 | 1 | księżmi             |
| ~wzn  | 3 | 1 | 1 | 1 | pieniężnej          |
| ~wzn  | 3 | 1 | 1 | 1 | księżniczka         |
| ~wzz  | 3 | 0 | 1 | 0 | rozwiąże            |
| xfj   | 3 | 1 | 1 | 1 | zachwiać            |
| xfw   | 3 | 0 | 0 | 1 | wszechwładnego      |
| xmj   | 3 | 1 | 1 | 1 | natychmiast         |
| xps   | 3 | 0 | 0 | 1 | wszechprzenikającym |
| xsk   | 3 | 0 | 0 | 1 | olbrychski          |
| xst   | 3 | 0 | 0 | 1 | kazachstan          |
| xsts  | 3 | 0 | 0 | 1 | kazachscy           |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b>       |
|-----------|------------|-------------|-------------|-------------|----------------------|
| xsw       | 3          | 0           | 0           | 1           | wszechsłowaniańskich |
| xst       | 3          | 0           | 0           | 1           | hochsztapler         |
| xste      | 3          | 0           | 0           | 1           | ochrzcił             |
| xstę      | 3          | 0           | 0           | 1           | ochrzczony           |
| xtj       | 3          | 0           | 0           | 1           | ichtiolog            |
| xtr       | 3          | 0           | 0           | 1           | blichtru             |
| xtsk      | 3          | 0           | 0           | 1           | utrechckim           |
| xtsn      | 3          | 0           | 0           | 1           | psychcznej           |
| xwb       | 3          | 0           | 1           | 1           | wybuchłby            |
| xwę       | 3          | 0           | 1           | 0           | zdechłszy            |
| zbj       | 3          | 1           | 1           | 1           | izbie                |
| zbr       | 3          | 1           | 1           | 1           | uzbrojenia           |
| zbw       | 3          | 0           | 0           | 1           | bezbłędnie           |
| zbz       | 3          | 0           | 0           | 1           | bezbrzeżna           |
| zdb       | 3          | 0           | 0           | 1           | rostbefu             |
| zdz       | 3          | 0           | 0           | 1           | minizdjęciami        |
| zdm       | 3          | 0           | 0           | 1           | rozdmuchany          |
| zdn       | 3          | 1           | 1           | 1           | przyjezdnych         |
| zdn       | 3          | 1           | 1           | 1           | jezdni               |
| zdr       | 3          | 1           | 1           | 1           | zazdrości            |
| zdv       | 3          | 0           | 0           | 1           | rozdwojenie          |
| zdz       | 3          | 0           | 0           | 1           | nozdrza              |
| zdzv      | 3          | 0           | 0           | 1           | rozdzwoniły          |
| zgd       | 3          | 0           | 0           | 1           | rozgdakały           |
| zgj       | 3          | 0           | 1           | 1           | mózgiem              |
| zgl       | 3          | 1           | 1           | 1           | rozgląda             |
| zgn       | 3          | 1           | 1           | 1           | poślizgnął           |
| zgn       | 3          | 1           | 1           | 1           | rozgniewany          |
| zgr       | 3          | 1           | 1           | 1           | rozgrywek            |
| zgv       | 3          | 0           | 0           | 1           | rozgwaru             |
| zgw       | 3          | 0           | 0           | 1           | rozgłośni            |
| zgz       | 3          | 1           | 1           | 1           | termozgrzewalna      |
| zmj       | 3          | 1           | 1           | 1           | rozmiarów            |
| zml       | 3          | 0           | 0           | 1           | bezmlecznych         |
| zmn       | 3          | 1           | 1           | 1           | rozmnażanie          |
| zmp       | 3          | 0           | 0           | 1           | niezmniejszającej    |
| zmr       | 3          | 0           | 0           | 1           | rozmrażanie          |
| zvj       | 3          | 1           | 1           | 1           | rozwiązania          |
| zvl       | 3          | 0           | 0           | 1           | rozwlekle            |

|       |   |   |   |   |                   |
|-------|---|---|---|---|-------------------|
| zvr   | 3 | 0 | 0 | 1 | niezwrotne        |
| zvw   | 3 | 0 | 0 | 1 | niezwłocznie      |
| zvz   | 3 | 0 | 0 | 1 | rozwrzeszczanych  |
| zwb   | 3 | 0 | 1 | 1 | znalazłby         |
| zwş   | 3 | 0 | 1 | 1 | znalazłszy        |
| zww   | 3 | 0 | 0 | 1 | rozżłościł        |
| zby   | 3 | 1 | 1 | 1 | rzeźbiarz         |
| zby   | 3 | 0 | 1 | 0 | wyrzeźbmyż        |
| zbz   | 3 | 0 | 1 | 0 | rzeźbże           |
| zdzdz | 3 | 0 | 0 | 1 | sześdziesiątych   |
| zdzm  | 3 | 0 | 1 | 1 | gwoździ           |
| zdzż  | 3 | 0 | 1 | 0 | jeźdźże           |
| zmj   | 3 | 0 | 1 | 1 | weźmie            |
| zpm   | 3 | 0 | 1 | 0 | zaprzyjaźmyż      |
| zpte  | 3 | 0 | 1 | 0 | zaprzyjaźnicie    |
| zpz   | 3 | 0 | 1 | 0 | zaprzyjaźniże     |
| zvj   | 3 | 0 | 1 | 1 | wytrzeźwień       |
| zby   | 3 | 0 | 1 | 1 | służbie           |
| zbn   | 3 | 0 | 0 | 1 | złowróbne         |
| zbn   | 3 | 0 | 0 | 1 | wierzbnie         |
| zdzm  | 3 | 0 | 1 | 0 | gwiździ           |
| zdzż  | 3 | 0 | 1 | 0 | gwiździże         |
| zgn   | 3 | 0 | 0 | 1 | przedzierzgnął    |
| zgn   | 3 | 0 | 0 | 1 | zadzierzgnięta    |
| zmj   | 3 | 0 | 0 | 1 | ujarzmienia       |
| zvj   | 3 | 0 | 1 | 1 | łyżwiarstwo       |
| brsk  | 4 | 0 | 0 | 1 | dobrski           |
| brvj  | 4 | 0 | 0 | 1 | międzybrwiowe     |
| bwkj  | 4 | 0 | 1 | 1 | jabłkiem          |
| bzdr  | 4 | 0 | 0 | 1 | klabzdra          |
| bzmj  | 4 | 1 | 1 | 1 | zabrzmiało        |
| ddzvj | 4 | 0 | 0 | 1 | oddźwięk          |
| dgvj  | 4 | 0 | 0 | 1 | przedgwiadzkowych |
| drvj  | 4 | 0 | 0 | 1 | wydrwione         |
| dvzg  | 4 | 0 | 0 | 1 | podwzgórza        |
| dvzr  | 4 | 0 | 0 | 1 | nadwzrocność      |
| dzbj  | 4 | 0 | 0 | 1 | podzbiory         |
| dzvj  | 4 | 0 | 0 | 1 | odzwierciedla     |
| dzvr  | 4 | 0 | 0 | 1 | podzwrotnikowych  |
| dzvj  | 4 | 0 | 0 | 1 | czterodrzwiowy    |

| cl    | len | dict | para | corp | example            |
|-------|-----|------|------|------|--------------------|
| ffsk  | 4   | 0    | 0    | 1    | przeciwwskazań     |
| fskj  | 4   | 0    | 1    | 1    | warszawskiej       |
| fspj  | 4   | 0    | 0    | 1    | niewspierania      |
| fstf  | 4   | 1    | 1    | 1    | rybołówstwa        |
| fstr  | 4   | 0    | 0    | 1    | wodowstręt         |
| fstş  | 4   | 0    | 0    | 1    | powstrzymać        |
| fşkj  | 4   | 0    | 0    | 1    | wawrzkievicz       |
| gzbj  | 4   | 0    | 0    | 1    | białogrzbiety      |
| gzmj  | 4   | 1    | 1    | 1    | zagrzmiał          |
| jbgv  | 4   | 0    | 0    | 1    | lejbguardia        |
| jefj  | 4   | 0    | 0    | 1    | najświętszej       |
| jemj  | 4   | 0    | 0    | 1    | najśmielsze        |
| jfsp  | 4   | 0    | 0    | 1    | najwspanialszych   |
| jfst  | 4   | 0    | 0    | 1    | najwstydlivszy     |
| jkrf  | 4   | 0    | 0    | 1    | najkrwawszych      |
| jmkj  | 4   | 0    | 0    | 1    | dejmkiem           |
| jpfr  | 4   | 0    | 0    | 1    | grejpfruty         |
| jskf  | 4   | 0    | 0    | 1    | najskwapliwiej     |
| jskj  | 4   | 0    | 1    | 1    | europeskiej        |
| jskr  | 4   | 0    | 1    | 1    | najskromniejszą    |
| jspr  | 4   | 0    | 1    | 1    | najsprawniej       |
| jstf  | 4   | 1    | 1    | 1    | zabójstwo          |
| jstr  | 4   | 0    | 1    | 1    | trójstronnej       |
| jstş  | 4   | 0    | 1    | 1    | majstrze           |
| jsxl  | 4   | 0    | 0    | 1    | najschludniejszych |
| jtkej | 4   | 0    | 0    | 1    | wojtkiewicza       |
| jtkl  | 4   | 0    | 1    | 1    | najtkliwszą        |
| jtrf  | 4   | 0    | 1    | 1    | najtrwalsze        |
| jtskj | 4   | 0    | 0    | 1    | kuwejskie          |
| jvzn  | 4   | 0    | 0    | 1    | najwznościejsza    |
| jzbr  | 4   | 0    | 0    | 1    | rajzbretu          |
| jzdr  | 4   | 0    | 1    | 1    | najzdrowszy        |
| jzgr  | 4   | 0    | 1    | 1    | najzgrabniejsze    |
| jskj  | 4   | 0    | 1    | 1    | amerykańskiej      |
| jskf  | 4   | 1    | 1    | 1    | państwa            |
| krfj  | 4   | 0    | 0    | 1    | niedokrwienniej    |
| krsk  | 4   | 0    | 0    | 1    | mokrsko            |
| ksbr  | 4   | 0    | 0    | 1    | eksbramkarz        |
| kske  | 4   | 0    | 0    | 1    | eksksięży          |

|       |   |   |   |   |                 |
|-------|---|---|---|---|-----------------|
| kskl  | 4 | 0 | 0 | 1 | ekskluzywnych   |
| kskr  | 4 | 0 | 0 | 1 | ekskrementy     |
| kspj  | 4 | 0 | 0 | 1 | ekspiacji       |
| kspł  | 4 | 0 | 0 | 1 | eksploatacji    |
| kspr  | 4 | 1 | 1 | 1 | express         |
| kspę  | 4 | 0 | 0 | 1 | eksprzyjaciel   |
| kssp  | 4 | 0 | 0 | 1 | ekssportowiec   |
| kstr  | 4 | 0 | 0 | 1 | ekstradycji     |
| ksvj  | 4 | 0 | 0 | 1 | ekswięźniów     |
| ksxw  | 4 | 0 | 0 | 1 | ekschłopaka     |
| lbmj  | 4 | 0 | 0 | 1 | skalbmierska    |
| lgzd  | 4 | 0 | 0 | 1 | folksdojcz      |
| lkwb  | 4 | 0 | 1 | 0 | zamilkłby       |
| lkwę  | 4 | 0 | 1 | 0 | zamilkłszy      |
| lmsk  | 4 | 0 | 0 | 1 | sztokholmski    |
| lskj  | 4 | 0 | 1 | 1 | polskiej        |
| lstf  | 4 | 1 | 1 | 1 | obywatelstwa    |
| lstr  | 4 | 0 | 0 | 1 | alstromeria     |
| ltkr  | 4 | 0 | 0 | 1 | szyldkretu      |
| ltst  | 4 | 0 | 0 | 1 | goldstar        |
| łtskj | 4 | 0 | 0 | 1 | grunwaldzkiej   |
| łtškj | 4 | 0 | 0 | 1 | milczkiem       |
| lzdr  | 4 | 0 | 0 | 1 | socjalzdrajcami |
| lzbj  | 4 | 0 | 0 | 1 | elżbieta        |
| mblj  | 4 | 0 | 0 | 1 | lamblia         |
| mbrj  | 4 | 0 | 0 | 1 | embrionów       |
| mbrn  | 4 | 0 | 0 | 1 | krnąbrnych      |
| mbrņ  | 4 | 0 | 0 | 1 | krnąbrni        |
| mpkj  | 4 | 0 | 1 | 1 | występkiem      |
| mpsk  | 4 | 0 | 0 | 1 | dębski          |
| mpsts | 4 | 0 | 0 | 1 | porębscy        |
| mpstj | 4 | 0 | 0 | 1 | konsumpcji      |
| mskj  | 4 | 0 | 1 | 1 | żeromskiego     |
| mskn  | 4 | 0 | 0 | 1 | wymsknęło       |
| mskn  | 4 | 0 | 0 | 1 | wymsknie        |
| mstf  | 4 | 1 | 1 | 1 | kłamstwa        |
| ndgr  | 4 | 0 | 0 | 1 | lindgren        |
| ndrj  | 4 | 0 | 0 | 1 | andriej         |
| ndrk  | 4 | 0 | 0 | 1 | mędrkowaniem    |
| ndrn  | 4 | 0 | 0 | 1 | jędrne          |

| <b>cl</b>       | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b>        |
|-----------------|------------|-------------|-------------|-------------|-----------------------|
| ndr $\bar{n}$   | 4          | 0           | 0           | 1           | ujędnia               |
| ndrts           | 4          | 0           | 0           | 1           | mędrców               |
| ndvw            | 4          | 0           | 0           | 1           | blondwłosą            |
| ndwb            | 4          | 0           | 1           | 0           | wiadłby               |
| ndw $\bar{s}$   | 4          | 0           | 1           | 0           | zwiadłszy             |
| nsgr            | 4          | 0           | 0           | 1           | transgranicznego      |
| nskj            | 4          | 0           | 0           | 1           | komenskigo            |
| nskr            | 4          | 0           | 0           | 1           | inskrypcji            |
| nsml            | 4          | 0           | 0           | 1           | transmlecz            |
| nspl            | 4          | 0           | 0           | 1           | transplantacji        |
| nspr            | 4          | 0           | 0           | 1           | transprojekt          |
| nstr            | 4          | 1           | 1           | 1           | konstrukcji           |
| nstw            | 4          | 0           | 0           | 1           | transtłuszczowe       |
| ntfj            | 4          | 0           | 0           | 1           | odrętwienia           |
| ntkj            | 4          | 0           | 1           | 1           | wyjątkiem             |
| ntpj            | 4          | 0           | 1           | 1           | wątpienia             |
| ntpl            | 4          | 1           | 1           | 1           | wątpliwości           |
| ntpm            | 4          | 0           | 1           | 0           | zwątpmy               |
| ntp $\bar{t}$ e | 4          | 0           | 1           | 0           | zwątpcie              |
| ntp $\bar{z}$   | 4          | 0           | 1           | 0           | zwątpże               |
| ntrd            | 4          | 0           | 0           | 1           | kontrdemonstracją     |
| ntrk            | 4          | 0           | 0           | 1           | kontrkandydata        |
| ntrl            | 4          | 0           | 0           | 1           | kontrlist             |
| ntrm            | 4          | 0           | 0           | 1           | kontrmanifestację     |
| ntrn            | 4          | 0           | 0           | 1           | kontrnatarcia         |
| ntrp            | 4          | 0           | 0           | 1           | kontrpartnerów        |
| ntrr            | 4          | 0           | 0           | 1           | kontrrewolucji        |
| ntrs            | 4          | 0           | 0           | 1           | kontrsojusze          |
| nrt $\bar{s}$   | 4          | 0           | 0           | 1           | kontrykliczna         |
| ntrv            | 4          | 0           | 0           | 1           | kontrwywiadu          |
| ntrw            | 4          | 0           | 0           | 1           | kontrłatami           |
| ntrz            | 4          | 0           | 0           | 1           | kontrzarzuty          |
| nt $\bar{s}$ b  | 4          | 0           | 0           | 1           | wewnątrzbankowym      |
| nt $\bar{s}$ dz | 4          | 0           | 0           | 1           | wewnątrzdzielnicowe   |
| nt $\bar{s}$ f  | 4          | 0           | 0           | 1           | wewnątrzfirmowych     |
| nt $\bar{s}$ g  | 4          | 0           | 0           | 1           | wewnątrzgałkowe       |
| nt $\bar{s}$ k  | 4          | 0           | 0           | 1           | wewnątrz korporacyjne |
| nt $\bar{s}$ l  | 4          | 0           | 0           | 1           | wewnątrzludzkie       |
| nt $\bar{s}$ m  | 4          | 0           | 0           | 1           | wewnątrzmaciczne      |

|         |   |   |   |   |                      |
|---------|---|---|---|---|----------------------|
| nt̄sn   | 4 | 1 | 1 | 1 | wewnętrznych         |
| nt̄sp   | 4 | 0 | 1 | 1 | wewnętrznie          |
| nt̄sr   | 4 | 0 | 0 | 1 | wewnątrzpartyjnych   |
| nt̄ss   | 4 | 0 | 0 | 1 | wewnątrzresortowa    |
| nt̄st   | 4 | 0 | 0 | 1 | wewnątrzsystemowej   |
| nt̄sts  | 4 | 0 | 0 | 1 | wewnątrztekstowe     |
| nt̄st̄s | 4 | 0 | 0 | 1 | wewnątrzcerkiewnych  |
| nt̄sx   | 4 | 0 | 0 | 1 | wewnątrzczaszkowe    |
| nt̄sz   | 4 | 0 | 0 | 1 | wewnątrzhiszpańskich |
| nt̄sz̄  | 4 | 0 | 0 | 1 | wewnątrzzakładowych  |
| nt̄skj  | 4 | 0 | 1 | 1 | wewnątrzżydowską     |
| nt̄skr  | 4 | 0 | 0 | 1 | prezydenckie         |
| nt̄stf  | 4 | 0 | 0 | 1 | tysiącrotnie         |
| nt̄szw  | 4 | 0 | 0 | 1 | malkontenctwo        |
| nt̄skj  | 4 | 0 | 1 | 1 | tysiącłotową         |
| nt̄smj  | 4 | 1 | 1 | 1 | pączkiem             |
| nd̄zvj  | 4 | 0 | 0 | 1 | jęczmień             |
| ηglj    | 4 | 0 | 0 | 1 | łędźwiowego          |
| ηgrj    | 4 | 0 | 0 | 1 | anglia               |
| ηgvj    | 4 | 0 | 0 | 1 | sangrię              |
| ηgwb    | 4 | 0 | 0 | 1 | chorągwie            |
| ηkkj    | 4 | 0 | 1 | 1 | przysiagłby          |
| ηkpl    | 4 | 0 | 0 | 1 | miękkie              |
| ηksj    | 4 | 0 | 0 | 1 | longplaye            |
| ηkst    | 4 | 0 | 0 | 1 | banksja              |
| ηksm    | 4 | 0 | 1 | 0 | gangsterów           |
| ηk̄sts  | 4 | 0 | 0 | 1 | powiększmy           |
| ηk̄ste  | 4 | 0 | 1 | 0 | węgrzce              |
| ηk̄sz̄  | 4 | 0 | 1 | 0 | zwiększcież          |
| ηk̄tsj  | 4 | 1 | 1 | 1 | powiększe            |
| ηkwb    | 4 | 0 | 1 | 0 | funkcji              |
| ηkw̄s   | 4 | 0 | 1 | 1 | pekłby               |
| pemj    | 4 | 0 | 0 | 1 | złakłszy             |
| pskj    | 4 | 0 | 1 | 1 | obsmiewać            |
| pskr    | 4 | 0 | 0 | 1 | arabskie             |
| pstf    | 4 | 1 | 1 | 1 | subskrypcji          |
| pstr    | 4 | 1 | 1 | 1 | przekupstwa          |
| pst̄s   | 4 | 0 | 0 | 1 | abstrakcyjne         |
| ptrj    | 4 | 0 | 0 | 1 | upstrzone            |
|         |   |   |   |   | dioptria             |

| cl     | len | dict | para | corp | example          |
|--------|-----|------|------|------|------------------|
| refj   | 4   | 0    | 0    | 1    | superświadek     |
| rdzy   | 4   | 0    | 0    | 1    | superdrzwi       |
| rdzvj  | 4   | 0    | 0    | 1    | hiperdźwiękowego |
| rfsk   | 4   | 0    | 0    | 1    | superwskaźnik    |
| rgvj   | 4   | 0    | 0    | 1    | supergwiazdą     |
| rkfj   | 4   | 0    | 0    | 1    | cerkwie          |
| rmkj   | 4   | 0    | 0    | 1    | giermkciem       |
| rmsk   | 4   | 0    | 0    | 1    | permskim         |
| rkj    | 4   | 0    | 1    | 1    | ziarnkiem        |
| rpkj   | 4   | 0    | 1    | 1    | uszczerbkiem     |
| rpsk   | 4   | 0    | 0    | 1    | serbskich        |
| rpsts  | 4   | 0    | 0    | 1    | serbscy          |
| rptsj  | 4   | 0    | 0    | 1    | absorpcji        |
| rskj   | 4   | 0    | 1    | 1    | maklerskie       |
| rskl   | 4   | 0    | 0    | 1    | hipersklepu      |
| rskn   | 4   | 0    | 0    | 1    | parsknął         |
| rskp   | 4   | 0    | 0    | 1    | parsknięciem     |
| rskr   | 4   | 0    | 0    | 1    | superskrót       |
| rspr   | 4   | 0    | 0    | 1    | supersprinter    |
| rstf   | 4   | 1    | 1    | 1    | przedsiębiorstwa |
| rstk   | 4   | 1    | 1    | 1    | garstka          |
| rstr   | 4   | 0    | 0    | 1    | superstrona      |
| rstş   | 4   | 0    | 0    | 1    | superstrzelca    |
| rstts  | 4   | 0    | 1    | 1    | garstce          |
| rşpj   | 4   | 0    | 0    | 1    | superszpieg      |
| rştşk  | 4   | 1    | 1    | 1    | zmarszczki       |
| rştşm  | 4   | 0    | 1    | 0    | zmarszczmyż      |
| rştşts | 4   | 0    | 1    | 0    | zmarszczce       |
| rştşte | 4   | 0    | 1    | 0    | zmarszczcież     |
| rştşz  | 4   | 0    | 1    | 0    | zmarszczce       |
| rtfj   | 4   | 0    | 1    | 1    | martwią          |
| rtfm   | 4   | 0    | 1    | 1    | martwmy          |
| rtfte  | 4   | 0    | 1    | 1    | martwcie         |
| rtfz   | 4   | 0    | 1    | 0    | martwze          |
| rtkj   | 4   | 0    | 1    | 1    | czwartkiem       |
| rtsm   | 4   | 0    | 0    | 1    | sportsmenkę      |
| rtskj  | 4   | 0    | 0    | 1    | frankfurckiej    |
| rtstf  | 4   | 0    | 0    | 1    | bękarctwa        |
| rtepr  | 4   | 0    | 0    | 1    | ćwierćprawdach   |

|       |   |   |   |   |                    |
|-------|---|---|---|---|--------------------|
| rtvj  | 4 | 0 | 0 | 1 | ćwierćwiecza       |
| rtmj  | 4 | 0 | 0 | 1 | karczmie           |
| rxj   | 4 | 0 | 1 | 1 | marchwią           |
| rzwb  | 4 | 0 | 1 | 0 | marzłby            |
| rzws  | 4 | 0 | 1 | 0 | zmarzłszy          |
| sefj  | 4 | 0 | 0 | 1 | rozświetla         |
| semj  | 4 | 0 | 0 | 1 | rozśmiesza         |
| sepj  | 4 | 0 | 0 | 1 | rozśpiewana        |
| sfete | 4 | 0 | 0 | 1 | rozwścieczony      |
| sfkw  | 4 | 0 | 0 | 1 | bezwkładową        |
| sfst  | 4 | 0 | 0 | 1 | bezwstyd           |
| skfj  | 4 | 0 | 0 | 1 | moskwie            |
| skrf  | 4 | 0 | 0 | 1 | bezkrwawą          |
| skst  | 4 | 0 | 0 | 1 | bezkształną        |
| snkj  | 4 | 0 | 1 | 1 | czoskiem           |
| spwte | 4 | 0 | 0 | 1 | bezpłciowe         |
| ssks  | 4 | 0 | 0 | 1 | bezskrzydłego      |
| sskw  | 4 | 0 | 0 | 1 | bezskładkowo       |
| sspš  | 4 | 0 | 0 | 1 | bezsprzecznie      |
| sstr  | 4 | 1 | 1 | 1 | bezzstronności     |
| sstš  | 4 | 1 | 1 | 1 | rozstrzygnięcia    |
| sškj  | 4 | 0 | 0 | 1 | bezszkieletowych   |
| stfj  | 4 | 0 | 1 | 1 | towarzystwie       |
| stfr  | 4 | 0 | 0 | 1 | postfreudowskich   |
| stkj  | 4 | 0 | 1 | 1 | wszystkie          |
| stkl  | 4 | 0 | 0 | 1 | roztkliwiać        |
| stmj  | 4 | 0 | 0 | 1 | astmie             |
| stpj  | 4 | 0 | 0 | 1 | postpieriestrojką  |
| stpr  | 4 | 0 | 0 | 1 | postprodukcja      |
| stps  | 4 | 0 | 0 | 1 | postprzemysłowej   |
| strf  | 4 | 0 | 0 | 1 | roztrwonila        |
| strj  | 4 | 0 | 0 | 1 | maestria           |
| stst  | 4 | 0 | 0 | 1 | poststalinowskim   |
| stsl  | 4 | 0 | 0 | 1 | postszlacheckiej   |
| stsm  | 4 | 0 | 1 | 0 | ostrzmy            |
| stste | 4 | 0 | 1 | 1 | ostrzcie           |
| stsz  | 4 | 0 | 1 | 0 | ostrzże            |
| sttr  | 4 | 0 | 0 | 1 | posttransakcyjnych |
| stšw  | 4 | 0 | 0 | 1 | postczłowiek       |
| stxn  | 4 | 1 | 1 | 1 | westchnął          |

| cl    | len | dict | para | corp | example              |
|-------|-----|------|------|------|----------------------|
| stxɲ  | 4   | 0    | 1    | 1    | westchnień           |
| stxɕ  | 4   | 0    | 0    | 1    | postchrześcijańskiej |
| stɛfj | 4   | 0    | 0    | 1    | rozéwiartowaniu      |
| sxfj  | 4   | 0    | 0    | 1    | rozehwianie          |
| sxwɕ  | 4   | 0    | 1    | 0    | uschłszy             |
| strj  | 4   | 0    | 0    | 1    | zaratusztrianin      |
| stɕkj | 4   | 0    | 0    | 1    | pyszczkiem           |
| stɕkn | 4   | 0    | 0    | 1    | uszczańć             |
| stɕkɲ | 4   | 0    | 0    | 1    | uszczańnie           |
| tɛfj  | 4   | 0    | 0    | 1    | przedświątecznym     |
| temj  | 4   | 0    | 0    | 1    | podśmiewają          |
| tepj  | 4   | 0    | 0    | 1    | odśpiewał            |
| tfsk  | 4   | 0    | 0    | 1    | podwskaźniki         |
| tfst  | 4   | 0    | 0    | 1    | przedwstępną         |
| tkfj  | 4   | 0    | 1    | 1    | utkwionymi           |
| tkɕt  | 4   | 0    | 0    | 1    | odkształcenia        |
| tskj  | 4   | 0    | 0    | 1    | piłsudskiego         |
| tskr  | 4   | 0    | 0    | 1    | nadskrupulatne       |
| tspr  | 4   | 0    | 0    | 1    | nadsprawiedliwie     |
| tspɕ  | 4   | 1    | 1    | 1    | odsprzedaży          |
| tstr  | 4   | 0    | 0    | 1    | odstrasza            |
| tstɕ  | 4   | 0    | 0    | 1    | odstrzał             |
| tɕkl  | 4   | 0    | 0    | 1    | podszkliwe           |
| txmj  | 4   | 0    | 0    | 1    | podchmielony         |
| tstfj | 4   | 0    | 1    | 1    | województwie         |
| tɛstf | 4   | 0    | 0    | 1    | uchodźstwa           |
| vdrɔ  | 4   | 0    | 0    | 1    | przeciwdrɔgawkowe    |
| vmɔj  | 4   | 0    | 0    | 1    | przeciwɔmgielne      |
| vmɔw  | 4   | 0    | 0    | 1    | przeciwɔmgłowych     |
| vzgl  | 4   | 1    | 1    | 1    | uwzględnieniem       |
| wbɕm  | 4   | 0    | 0    | 1    | współbrɕmi           |
| wɛfj  | 4   | 0    | 0    | 1    | półświatka           |
| wɔɕvj | 4   | 0    | 0    | 1    | półdźwięki           |
| wɔɕb  | 4   | 0    | 0    | 1    | augsburskiego        |
| wkrf  | 4   | 0    | 0    | 1    | półkrwi              |
| wkɕt  | 4   | 0    | 0    | 1    | współkształtują      |
| wktsj | 4   | 0    | 0    | 1    | aukcji               |
| wmsk  | 4   | 0    | 0    | 1    | chełmska             |
| wmsts | 4   | 0    | 0    | 1    | chełmscy             |

|         |   |   |   |   |                    |
|---------|---|---|---|---|--------------------|
| wmst̃s  | 4 | 0 | 0 | 1 | chełmszczyźnie     |
| wnts̃j  | 4 | 0 | 0 | 1 | szesnastuncjowych  |
| wskj    | 4 | 0 | 0 | 1 | sosnowskiego       |
| wspr    | 4 | 0 | 0 | 1 | współsprawców      |
| wstr    | 4 | 0 | 0 | 1 | australia          |
| wtkj    | 4 | 0 | 1 | 1 | zółtkiem           |
| wtskj   | 4 | 0 | 0 | 1 | ełckiego           |
| wzwb    | 4 | 0 | 1 | 0 | pełzłby            |
| ũskj    | 4 | 0 | 1 | 1 | śląskiego          |
| ũskn    | 4 | 1 | 1 | 1 | tęsknota           |
| ũskn    | 4 | 1 | 1 | 1 | tęskni             |
| ũstf    | 4 | 1 | 1 | 1 | zwycięstwo         |
| ũstk    | 4 | 1 | 1 | 1 | częstki            |
| ũstn    | 4 | 0 | 0 | 1 | chrześnej          |
| ũstn    | 4 | 0 | 0 | 1 | gęstnieje          |
| ũst̃s   | 4 | 0 | 0 | 1 | częstsze           |
| ũstts   | 4 | 0 | 1 | 1 | częstce            |
| ũswb    | 4 | 0 | 1 | 0 | trząsłby           |
| ũsw̃s   | 4 | 0 | 1 | 0 | zatrząsłszy        |
| ũskj    | 4 | 0 | 1 | 1 | ciężkie            |
| ũxfj    | 4 | 0 | 0 | 1 | warzących          |
| xcfj    | 4 | 1 | 1 | 1 | wszechświata       |
| xfst    | 4 | 0 | 0 | 1 | zmartwychwstanie   |
| xskj    | 4 | 0 | 0 | 1 | olbrychskiego      |
| xstr    | 4 | 0 | 0 | 1 | wszechstronnie     |
| xtkl    | 4 | 0 | 0 | 1 | jachtklubach       |
| xzvj    | 4 | 0 | 0 | 1 | wszechzwiązkowego  |
| zbz̃m   | 4 | 0 | 0 | 1 | rozbrzmi           |
| zd̃zṽj | 4 | 0 | 0 | 1 | rozdźwięk          |
| zgvj    | 4 | 0 | 0 | 1 | rozwieżdzone       |
| zzbj    | 4 | 0 | 0 | 1 | bezzbiornikowe     |
| zzvj    | 4 | 0 | 0 | 1 | bezzwiązkowa       |
| zzvr    | 4 | 0 | 0 | 1 | bezzwrotnej        |
| zzvw    | 4 | 0 | 0 | 1 | bezzwłocznie       |
| brskj   | 5 | 0 | 0 | 1 | dobrskiego         |
| ddzvj   | 5 | 0 | 0 | 1 | naddrzwiowych      |
| ffst̃s  | 5 | 0 | 0 | 1 | przeciwstrząsowych |
| fstfj   | 5 | 0 | 1 | 1 | ustawodawstwie     |
| jfstr   | 5 | 0 | 1 | 1 | najwstrętniejszej  |
| jjskj   | 5 | 0 | 0 | 1 | bahrajńskiej       |

| cl     | len | dict | para | corp | example                  |
|--------|-----|------|------|------|--------------------------|
| jstfj  | 5   | 0    | 1    | 1    | zabójstwo                |
| juzgl  | 5   | 0    | 0    | 1    | najwzględniejszej        |
| ̂stfj  | 5   | 0    | 1    | 1    | państwie                 |
| krskj  | 5   | 0    | 0    | 1    | mokrskiego               |
| lfstr  | 5   | 0    | 0    | 1    | golfsztromie             |
| lmskj  | 5   | 0    | 0    | 1    | sztokholmskiej           |
| lstfj  | 5   | 0    | 1    | 1    | obywatelstwo             |
| mpskj  | 5   | 0    | 0    | 1    | dębskiego                |
| mpstf  | 5   | 1    | 1    | 1    | przestępstwa             |
| mstfj  | 5   | 0    | 1    | 1    | kłamstwo                 |
| ntrpr  | 5   | 0    | 0    | 1    | kontrpropozycję          |
| ntrps  | 5   | 0    | 0    | 1    | kontrprzykład            |
| ntrvn  | 5   | 0    | 0    | 1    | kontrwniosków            |
| ntskj  | 5   | 0    | 0    | 1    | kąskiego                 |
| nt̂bz  | 5   | 0    | 0    | 1    | wewnątrzbrzusznym        |
| nt̂er  | 5   | 0    | 0    | 1    | wewnątrzśrodowiskowej    |
| nt̂gr  | 5   | 0    | 0    | 1    | wewnątrzgrupowych        |
| nt̂kl  | 5   | 0    | 0    | 1    | wewnątrzklubowym         |
| nt̂kr  | 5   | 0    | 0    | 1    | wewnątrzkrajowym         |
| nt̂mj  | 5   | 0    | 0    | 1    | wewnątrzmiejską          |
| nt̂pl  | 5   | 0    | 0    | 1    | wewnątrzplemienne        |
| nt̂pr  | 5   | 0    | 0    | 1    | wewnątrzproduktowy       |
| nt̂pw  | 5   | 0    | 0    | 1    | wewnątrzpłodowej         |
| nt̂sp  | 5   | 0    | 0    | 1    | wewnątrzspółdzielczego   |
| nt̂st  | 5   | 0    | 0    | 1    | wewnątrzstanowym         |
| nt̂sk  | 5   | 0    | 0    | 1    | wewnątrzszkolnym         |
| nt̂sp  | 5   | 0    | 0    | 1    | wewnątrzszpitalnych      |
| nt̂vw  | 5   | 0    | 0    | 1    | wewnątrzwłaścicielskich  |
| nt̂xs  | 5   | 0    | 0    | 1    | wewnątrzchrześcijańskich |
| ntstfj | 5   | 0    | 0    | 1    | niemowlęctwie            |
| pstfj  | 5   | 0    | 1    | 1    | hrabstwo                 |
| rmskj  | 5   | 0    | 0    | 1    | permskiego               |
| rmstr  | 5   | 0    | 0    | 1    | armstrong                |
| rpskj  | 5   | 0    | 0    | 1    | serbskiej                |
| rstfj  | 5   | 0    | 1    | 1    | ministerstwo             |
| rstkj  | 5   | 0    | 1    | 1    | szorstkie                |
| stzvj  | 5   | 0    | 0    | 1    | postzwiązkowego          |
| testfj | 5   | 0    | 0    | 1    | uchodźstwo               |
| wbẑmj | 5   | 0    | 0    | 1    | współbrzmienie           |

|         |   |   |   |   |                      |
|---------|---|---|---|---|----------------------|
| wmskj   | 5 | 0 | 0 | 1 | chełmskiej           |
| wstrj   | 5 | 0 | 0 | 1 | austria              |
| ŵstfj   | 5 | 0 | 1 | 1 | zwycięstwie          |
| zbz̄mj  | 5 | 0 | 0 | 1 | rozbrzmiewa          |
| zvzgl   | 5 | 1 | 1 | 1 | bezwzględnie         |
| mpstfj  | 6 | 0 | 1 | 1 | przestępstwie        |
| nt̄sfsp | 6 | 0 | 0 | 1 | wewnątrzspółnotowego |
| nt̄szvj | 6 | 0 | 0 | 1 | wewnątrzwiązkową     |

**Final clusters**

| cl  | len | dict | para | corp | example    |
|-----|-----|------|------|------|------------|
| bl  | 2   | 0    | 0    | 1    | notabl     |
| br  | 2   | 1    | 1    | 1    | dóbr       |
| bw  | 2   | 0    | 0    | 1    | osłabł     |
| el  | 2   | 0    | 1    | 1    | myśl       |
| em  | 2   | 0    | 1    | 1    | taśm       |
| en  | 2   | 1    | 1    | 1    | pieśń      |
| ep  | 2   | 0    | 1    | 1    | rzeźb      |
| ete | 2   | 1    | 1    | 1    | wartość    |
| dl  | 2   | 0    | 0    | 1    | módl       |
| dm  | 2   | 0    | 0    | 1    | wydm       |
| dr  | 2   | 0    | 1    | 1    | kadr       |
| dw  | 2   | 0    | 1    | 1    | spadł      |
| dzm | 2   | 0    | 0    | 1    | wiedźm     |
| fl  | 2   | 0    | 0    | 1    | trefl      |
| fr  | 2   | 0    | 1    | 1    | cyfr       |
| ft  | 2   | 1    | 1    | 1    | luft       |
| fts | 2   | 0    | 0    | 1    | szewc      |
| fte | 2   | 0    | 1    | 1    | sprawdź    |
| gm  | 2   | 0    | 0    | 1    | diafragm   |
| gw  | 2   | 0    | 1    | 1    | mógł       |
| je  | 2   | 0    | 1    | 1    | jakiejs    |
| jf  | 2   | 0    | 0    | 1    | sejf       |
| jk  | 2   | 1    | 1    | 1    | strajk     |
| jl  | 2   | 0    | 0    | 1    | koktajl    |
| jm  | 2   | 1    | 1    | 1    | sejm       |
| jn  | 2   | 0    | 0    | 1    | sejn       |
| jp  | 2   | 0    | 0    | 1    | knajp      |
| js  | 2   | 0    | 0    | 1    | rejs       |
| js  | 2   | 0    | 1    | 1    | spójrz     |
| jt  | 2   | 1    | 1    | 1    | wójt       |
| jts | 2   | 0    | 0    | 1    | bajc       |
| jte | 2   | 0    | 1    | 1    | znajdź     |
| jtš | 2   | 0    | 0    | 1    | folksdojcz |
| je  | 2   | 1    | 1    | 1    | jakąś      |
| kf  | 2   | 0    | 0    | 1    | sakw       |
| kl  | 2   | 1    | 1    | 1    | spektakl   |
| kr  | 2   | 0    | 0    | 1    | akr        |

|     |   |   |   |   |           |
|-----|---|---|---|---|-----------|
| ks  | 2 | 1 | 1 | 1 | indeks    |
| kş  | 2 | 0 | 0 | 1 | riksz     |
| kt  | 2 | 1 | 1 | 1 | projekt   |
| kw  | 2 | 0 | 1 | 1 | orzekł    |
| lf  | 2 | 0 | 0 | 1 | rudolf    |
| lk  | 2 | 1 | 1 | 1 | ulg       |
| lm  | 2 | 1 | 1 | 1 | film      |
| ln  | 2 | 0 | 1 | 1 | kopalń    |
| lp  | 2 | 0 | 0 | 1 | alp       |
| ls  | 2 | 0 | 0 | 1 | puls      |
| lş  | 2 | 1 | 1 | 1 | olsz      |
| lt  | 2 | 1 | 1 | 1 | witold    |
| łts | 2 | 1 | 1 | 1 | kielc     |
| łtş | 2 | 0 | 1 | 1 | jelez     |
| lx  | 2 | 0 | 0 | 1 | olch      |
| mę  | 2 | 0 | 1 | 1 | czymś     |
| mf  | 2 | 1 | 1 | 1 | triumf    |
| ml  | 2 | 0 | 0 | 1 | kreml     |
| mn  | 2 | 1 | 1 | 1 | hymn      |
| mp  | 2 | 1 | 1 | 1 | dostęp    |
| ms  | 2 | 0 | 0 | 1 | ryms      |
| mş  | 2 | 1 | 1 | 1 | zamsz     |
| mte | 2 | 0 | 0 | 1 | imć       |
| mtş | 2 | 0 | 0 | 1 | niemcz    |
| nf  | 2 | 0 | 0 | 1 | tynf      |
| nn  | 2 | 0 | 1 | 1 | ann       |
| ns  | 2 | 1 | 1 | 1 | szans     |
| nş  | 2 | 1 | 1 | 1 | czynsz    |
| nt  | 2 | 1 | 1 | 1 | sąd       |
| nts | 2 | 1 | 1 | 1 | więc      |
| ntş | 2 | 1 | 1 | 1 | wręcz     |
| nx  | 2 | 0 | 0 | 1 | konch     |
| nts | 2 | 0 | 1 | 1 | słońc     |
| nte | 2 | 1 | 1 | 1 | pięc      |
| ntş | 2 | 0 | 1 | 1 | pomarańcz |
| nk  | 2 | 1 | 1 | 1 | bank      |
| pn  | 2 | 0 | 0 | 1 | wapń      |
| pr  | 2 | 0 | 0 | 1 | cypr      |
| ps  | 2 | 1 | 1 | 1 | gips      |
| pş  | 2 | 1 | 1 | 1 | wieprz    |

| cl  | len | dict | para | corp | example   |
|-----|-----|------|------|------|-----------|
| pt  | 2   | 0    | 1    | 1    | egipt     |
| pte | 2   | 0    | 0    | 1    | babé      |
| ptş | 2   | 0    | 1    | 0    | podepcz   |
| pw  | 2   | 0    | 0    | 1    | okrzepl   |
| re  | 2   | 1    | 1    | 1    | piers     |
| rf  | 2   | 1    | 1    | 1    | najpierw  |
| rk  | 2   | 1    | 1    | 1    | przetarg  |
| rl  | 2   | 0    | 0    | 1    | karl      |
| rm  | 2   | 1    | 1    | 1    | firm      |
| rn  | 2   | 0    | 1    | 1    | koncern   |
| rn  | 2   | 0    | 1    | 1    | księgarń  |
| rp  | 2   | 1    | 1    | 1    | skarb     |
| rs  | 2   | 1    | 1    | 1    | kurs      |
| rş  | 2   | 1    | 1    | 1    | marsz     |
| rt  | 2   | 1    | 1    | 1    | robert    |
| rts | 2   | 0    | 1    | 1    | serc      |
| rte | 2   | 1    | 1    | 1    | śmierć    |
| rtş | 2   | 0    | 1    | 1    | kurcz     |
| rw  | 2   | 0    | 1    | 1    | zmarł     |
| rx  | 2   | 0    | 0    | 1    | wierch    |
| sf  | 2   | 0    | 1    | 1    | nazw      |
| sk  | 2   | 1    | 1    | 1    | zysk      |
| sm  | 2   | 0    | 1    | 1    | pism      |
| sp  | 2   | 0    | 1    | 1    | izb       |
| st  | 2   | 1    | 1    | 1    | jest      |
| sw  | 2   | 1    | 1    | 1    | pomysł    |
| şf  | 2   | 0    | 1    | 0    | łyżw      |
| şp  | 2   | 0    | 1    | 1    | służb     |
| şş  | 2   | 0    | 0    | 1    | wywyższ   |
| şt  | 2   | 1    | 1    | 1    | koszt     |
| ştş | 2   | 1    | 1    | 1    | bydgoszcz |
| şx  | 2   | 1    | 1    | 1    | zmierzch  |
| tf  | 2   | 0    | 1    | 1    | gonitw    |
| tl  | 2   | 0    | 1    | 1    | pejotl    |
| tm  | 2   | 1    | 1    | 1    | rytm      |
| tr  | 2   | 1    | 1    | 1    | piotr     |
| tş  | 2   | 0    | 1    | 1    | patrz     |
| ttş | 2   | 0    | 1    | 1    | świadcz   |
| tw  | 2   | 0    | 1    | 1    | kwitł     |

|      |   |   |   |   |                  |
|------|---|---|---|---|------------------|
| t̄sp | 2 | 0 | 1 | 1 | liczb            |
| t̄st | 2 | 0 | 1 | 1 | poczt            |
| vr   | 2 | 0 | 0 | 1 | manewr           |
| wf   | 2 | 1 | 1 | 1 | zółw             |
| wk   | 2 | 1 | 1 | 1 | czołg            |
| wl   | 2 | 0 | 0 | 1 | kraul            |
| wm   | 2 | 1 | 1 | 1 | chełm            |
| wn   | 2 | 0 | 0 | 1 | braun            |
| w̄p  | 2 | 0 | 1 | 0 | spełń            |
| wp   | 2 | 0 | 1 | 1 | małp             |
| wr   | 2 | 0 | 0 | 1 | laur             |
| ws   | 2 | 0 | 0 | 1 | klaus            |
| wś   | 2 | 0 | 0 | 1 | falsz            |
| wt   | 2 | 1 | 1 | 1 | aut              |
| wte  | 2 | 0 | 1 | 1 | zólc             |
| wt̄s | 2 | 0 | 0 | 1 | wałcz            |
| w̄s  | 2 | 1 | 1 | 1 | mięś             |
| w̄ś  | 2 | 1 | 1 | 1 | wciąż            |
| w̄x  | 2 | 1 | 1 | 1 | węch             |
| xe   | 2 | 0 | 1 | 1 | jakichś          |
| xm   | 2 | 0 | 0 | 1 | drachm           |
| xr   | 2 | 0 | 0 | 1 | ochr             |
| xś   | 2 | 0 | 0 | 1 | spichrz          |
| xt   | 2 | 0 | 1 | 1 | jacht            |
| xw   | 2 | 0 | 1 | 1 | wybuchł          |
| zm   | 2 | 1 | 1 | 1 | mechanizm        |
| zn   | 2 | 0 | 1 | 1 | mężczyzn         |
| zw   | 2 | 0 | 1 | 1 | znalazł          |
| z̄p  | 2 | 0 | 1 | 1 | przyjaźń         |
| fsk  | 3 | 0 | 0 | 1 | dniepropietrowsk |
| jete | 3 | 1 | 1 | 1 | dojść            |
| jsk  | 3 | 0 | 1 | 1 | wojsk            |
| jsts | 3 | 0 | 1 | 1 | miejsc           |
| ĵete | 3 | 1 | 1 | 1 | część            |
| ĵsk  | 3 | 0 | 0 | 1 | gdańsk           |
| kst  | 3 | 1 | 1 | 1 | tekst            |
| lkw  | 3 | 0 | 1 | 1 | zamilkł          |
| lsk  | 3 | 0 | 0 | 1 | bielsk           |
| ltr  | 3 | 1 | 1 | 1 | filtr            |
| mbl  | 3 | 0 | 0 | 1 | ansambl          |

| cl   | len | dict | para | corp | example     |
|------|-----|------|------|------|-------------|
| mbr  | 3   | 0    | 0    | 1    | tembr       |
| mpt  | 3   | 0    | 0    | 1    | asumpt      |
| msk  | 3   | 0    | 0    | 1    | tomsk       |
| mst  | 3   | 0    | 1    | 1    | zemst       |
| ndw  | 3   | 0    | 1    | 1    | zwiądł      |
| nst  | 3   | 0    | 0    | 1    | kunszt      |
| nte  | 3   | 0    | 0    | 1    | skądś       |
| ntf  | 3   | 0    | 0    | 1    | mątw        |
| ntn  | 3   | 0    | 1    | 0    | tętn        |
| ntp  | 3   | 0    | 1    | 1    | wątp        |
| ntr  | 3   | 0    | 1    | 1    | kontr       |
| ntş  | 3   | 1    | 1    | 1    | zewnątrz    |
| ngw  | 3   | 0    | 0    | 1    | poprzysiągł |
| ņks  | 3   | 0    | 0    | 1    | sfinks      |
| ņkş  | 3   | 0    | 1    | 1    | zwiększ     |
| ņkt  | 3   | 1    | 1    | 1    | punkt       |
| ņkw  | 3   | 0    | 1    | 1    | pękł        |
| psk  | 3   | 0    | 0    | 1    | ślupsk      |
| rete | 3   | 1    | 1    | 1    | garść       |
| rks  | 3   | 0    | 0    | 1    | marks       |
| rpw  | 3   | 0    | 0    | 1    | ścierpł     |
| rsk  | 3   | 0    | 0    | 1    | kursk       |
| rst  | 3   | 0    | 0    | 1    | horst       |
| rş   | 3   | 0    | 0    | 1    | herszt      |
| rşş  | 3   | 1    | 1    | 1    | barszcz     |
| rtf  | 3   | 0    | 1    | 1    | martw       |
| rzw  | 3   | 0    | 1    | 1    | zamarzł     |
| stf  | 3   | 0    | 1    | 1    | mistrzostw  |
| stm  | 3   | 0    | 0    | 1    | astm        |
| str  | 3   | 1    | 1    | 1    | rejestr     |
| stş  | 3   | 1    | 1    | 1    | mistrz      |
| tstf | 3   | 0    | 1    | 1    | województw  |
| wkw  | 3   | 0    | 0    | 1    | pożółkł     |
| wst  | 3   | 0    | 0    | 1    | holokaust   |
| wzw  | 3   | 0    | 1    | 1    | spełzł      |
| ŵsk  | 3   | 0    | 1    | 1    | śląsk       |
| ŵst  | 3   | 0    | 0    | 1    | pochrzęst   |
| ŵsw  | 3   | 0    | 1    | 1    | zatrząsł    |
| ŵşş  | 3   | 0    | 0    | 1    | miąsz       |

|       |   |   |   |   |                   |
|-------|---|---|---|---|-------------------|
| ŵst̃  | 3 | 0 | 0 | 1 | gąszcz            |
| ŵzw   | 3 | 0 | 0 | 1 | ugrzął            |
| xtr   | 3 | 0 | 0 | 1 | blichtr           |
| fstf  | 4 | 0 | 1 | 1 | ustawodawstw      |
| jstf  | 4 | 0 | 1 | 1 | zabójstw          |
| ĵstf  | 4 | 0 | 1 | 1 | państw            |
| lstf  | 4 | 0 | 1 | 1 | przedstawicielstw |
| mstf  | 4 | 0 | 1 | 1 | kłamstw           |
| ntstf | 4 | 0 | 0 | 1 | natręctw          |
| pstf  | 4 | 0 | 1 | 1 | głupstw           |
| rstf  | 4 | 0 | 1 | 1 | przedsiębiorstw   |
| ŵstf  | 4 | 0 | 1 | 1 | zwycięstw         |
| mpstf | 5 | 0 | 1 | 1 | przestępstw       |



## Appendix D. Distribution of English consonant clusters: raw data

### Dictionary

#### Cluster Types

| len   | i  | m   | f  | total |
|-------|----|-----|----|-------|
| 2     | 48 | 309 | 59 | 416   |
| 3     | 9  | 327 | 21 | 357   |
| 4     | 0  | 64  | 1  | 65    |
| 5     | 0  | 3   | 0  | 3     |
| 6     | 0  | 0   | 0  | 0     |
| total | 57 | 703 | 81 | 841   |

#### Word Types

| len   | i    | m     | f    | total |
|-------|------|-------|------|-------|
| 2     | 5178 | 17775 | 3432 | 26385 |
| 3     | 332  | 3167  | 74   | 3573  |
| 4     | 0    | 309   | 2    | 311   |
| 5     | 0    | 3     | 0    | 3     |
| 6     | 0    | 0     | 0    | 0     |
| total | 5510 | 21254 | 3508 | 30272 |

### Paradigms

#### Cluster Types

| len   | i  | m   | f   | total |
|-------|----|-----|-----|-------|
| 2     | 48 | 310 | 61  | 419   |
| 3     | 9  | 329 | 50  | 388   |
| 4     | 0  | 64  | 10  | 74    |
| 5     | 0  | 3   | 0   | 3     |
| 6     | 0  | 0   | 0   | 0     |
| total | 57 | 706 | 121 | 884   |

## Word Types

| len   | i     | m     | f     | total |
|-------|-------|-------|-------|-------|
| 2     | 10405 | 29140 | 14615 | 54160 |
| 3     | 731   | 5053  | 2024  | 7808  |
| 4     | 0     | 527   | 20    | 547   |
| 5     | 0     | 5     | 0     | 5     |
| 6     | 0     | 0     | 0     | 0     |
| total | 11136 | 34725 | 16659 | 62520 |

**Corpus**

## Cluster Types

| len   | i  | m   | f   | total |
|-------|----|-----|-----|-------|
| 2     | 46 | 306 | 61  | 413   |
| 3     | 9  | 317 | 50  | 376   |
| 4     | 0  | 62  | 8   | 70    |
| 5     | 0  | 2   | 0   | 2     |
| 6     | 0  | 0   | 0   | 0     |
| total | 55 | 687 | 119 | 861   |

## Word Types

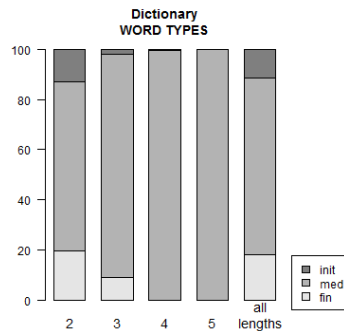
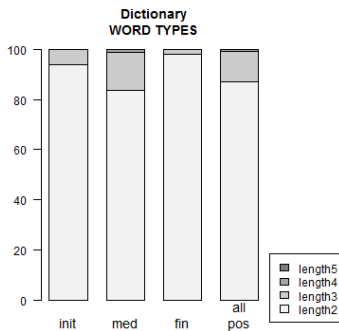
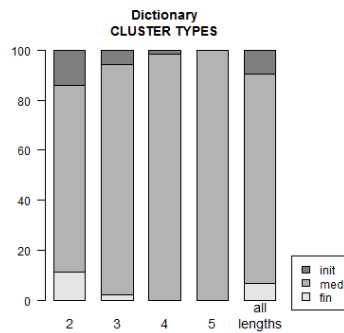
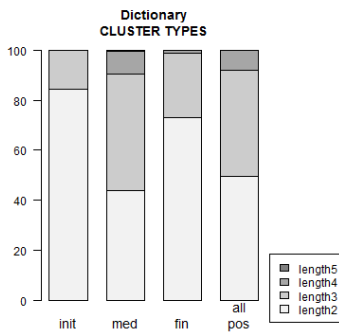
| len   | i     | m     | f     | total |
|-------|-------|-------|-------|-------|
| 2     | 9451  | 25919 | 12924 | 48294 |
| 3     | 651   | 4440  | 1821  | 6912  |
| 4     | 0     | 496   | 18    | 514   |
| 5     | 0     | 3     | 0     | 3     |
| 6     | 0     | 0     | 0     | 0     |
| total | 10102 | 30858 | 14763 | 55723 |

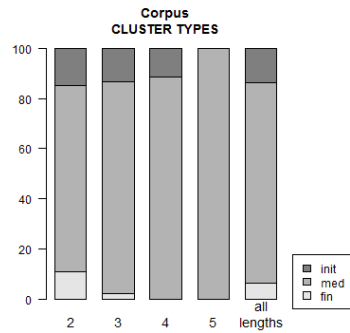
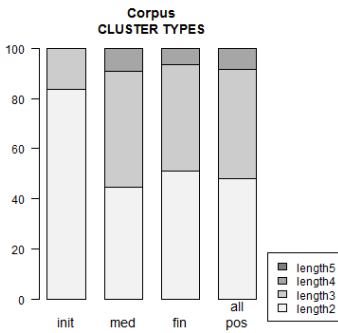
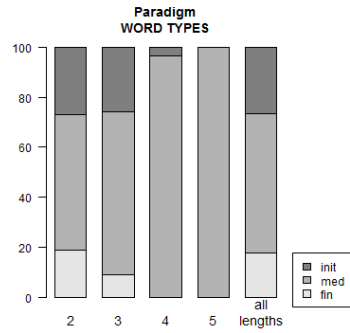
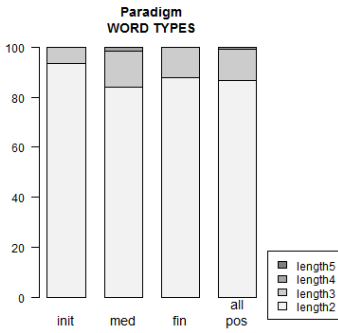
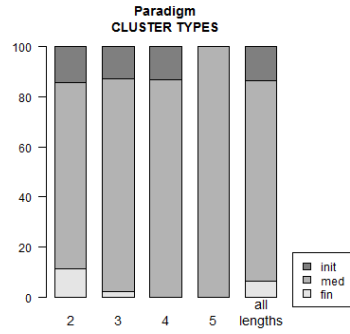
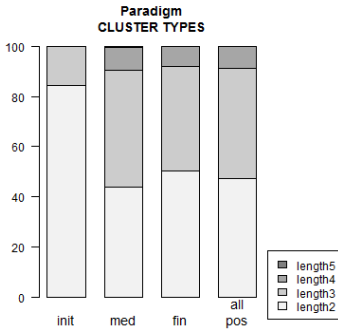
## Word Tokens

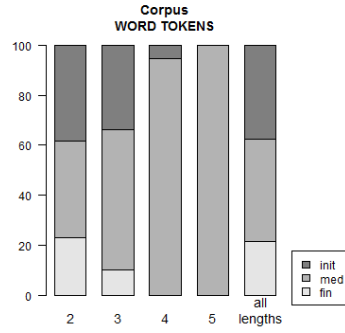
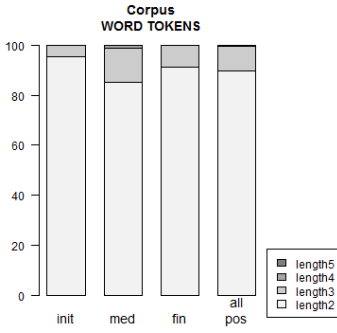
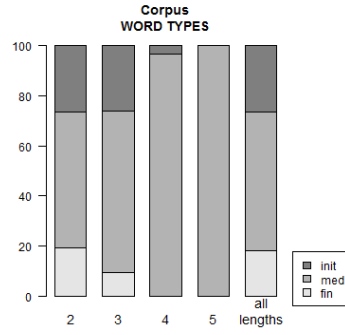
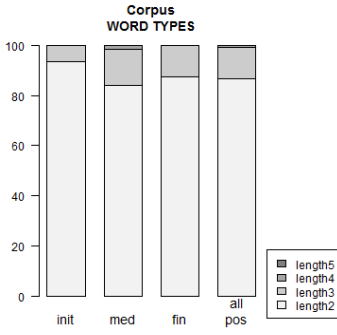
| len   | i        | m        | f        | total     |
|-------|----------|----------|----------|-----------|
| 2     | 31247588 | 52387660 | 51977317 | 135612565 |
| 3     | 1575598  | 8332815  | 5069983  | 14978396  |
| 4     | 0        | 805836   | 47802    | 853638    |
| 5     | 0        | 642      | 0        | 642       |
| 6     | 0        | 0        | 0        | 0         |
| total | 32823186 | 61526953 | 57095102 | 151445241 |

# Appendix E. Cluster distribution in English

Interpretation of the graphs: *cluster types* are unique clusters, *word types* are unique words in which clusters with specific characteristics were recorded, *word tokens* are all occurrences of words in which clusters with specific characteristics were recorded. Column 2 in graph 1 below shows that among medial clusters recorded in the dictionary data: 44% are length 2, 46% are length 3, 9% are length 4, and less than 1% are length 5.









## Appendix F. English cluster statistics

The following tables present English clusters in all positions and of all sizes with word type frequency statistics for the three resources studied and word token statistics for the corpus resource.

### Legend

pos = position  
len = length (size)  
cl = cluster  
dict = dictionary  
para = paradigms  
corp = corpus  
wtyp = word types  
wtok = words tokens

A value in the *dictwtyp* (dictionary word type) column indicates the number of word types (unique words excluding inflected forms) featuring a particular cluster in one of the three positions in the dictionary resource. Values in the *parawtyp* (paradigms word type) and *corpwtyp* (corpus word types) columns indicate similar statistics for the paradigms and corpus resources. A value in the *corpwtok* (corpus word token) column indicates the number of word tokens (all, possibly repeated, occurrences of words including inflected forms) featuring a particular cluster in one of the three positions in the corpus resource. A zero in a specific column means that no words with a given cluster in a particular position were recorded in the resource represented by the column.

**Initial clusters**

| pos | len | cl | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|----|----------|----------|----------|----------|
| i   | 2   | bj | 19       | 26       | 25       | 87148    |
| i   | 2   | bl | 180      | 400      | 363      | 721214   |
| i   | 2   | br | 241      | 492      | 464      | 1046417  |
| i   | 2   | dj | 46       | 75       | 63       | 129049   |
| i   | 2   | dr | 160      | 321      | 294      | 936624   |
| i   | 2   | dw | 7        | 16       | 16       | 12993    |
| i   | 2   | fj | 38       | 70       | 64       | 610803   |
| i   | 2   | fl | 228      | 473      | 428      | 696651   |
| i   | 2   | fr | 214      | 393      | 346      | 2690596  |
| i   | 2   | gl | 103      | 209      | 196      | 273360   |
| i   | 2   | gr | 248      | 508      | 463      | 1580126  |
| i   | 2   | gw | 2        | 3        | 3        | 403      |
| i   | 2   | hj | 35       | 59       | 55       | 218371   |
| i   | 2   | kj | 39       | 59       | 58       | 83340    |
| i   | 2   | kl | 221      | 445      | 410      | 1227721  |
| i   | 2   | kr | 285      | 592      | 548      | 1069832  |
| i   | 2   | 1  | 1        | 0        | 0        |          |
| i   | 2   | kw | 164      | 329      | 288      | 698651   |
| i   | 2   | lj | 5        | 8        | 7        | 12578    |
| i   | 2   | mj | 41       | 68       | 66       | 207472   |
| i   | 2   | nj | 81       | 126      | 117      | 1840996  |
| i   | 2   | pj | 43       | 69       | 63       | 53017    |
| i   | 2   | pl | 196      | 388      | 344      | 1481162  |
| i   | 2   | pr | 663      | 1290     | 1156     | 4157189  |
| i   | 2   | pw | 3        | 4        | 4        | 990      |
| i   | 2   | sf | 7        | 10       | 10       | 11793    |
| i   | 2   | sj | 7        | 12       | 10       | 24295    |
| i   | 2   | sk | 230      | 501      | 446      | 979040   |
| i   | 2   | sl | 169      | 367      | 329      | 462570   |
| i   | 2   | sm | 56       | 137      | 123      | 472426   |
| i   | 2   | sn | 93       | 231      | 204      | 153470   |
| i   | 2   | sp | 248      | 492      | 449      | 1185997  |
| i   | 2   | st | 420      | 851      | 782      | 3602097  |
| i   | 2   | sw | 97       | 229      | 208      | 248852   |
| i   | 2   | fm | 4        | 4        | 3        | 217      |
| i   | 2   | fn | 2        | 3        | 3        | 284      |
| i   | 2   | fp | 1        | 2        | 2        | 326      |

|   |   |     |     |     |     |         |
|---|---|-----|-----|-----|-----|---------|
| i | 2 | jr  | 31  | 66  | 59  | 52141   |
| i | 2 | jw  | 2   | 3   | 0   | 0       |
| i | 2 | tj  | 35  | 52  | 49  | 53250   |
| i | 2 | tr  | 396 | 779 | 716 | 2188713 |
| i | 2 | tw  | 35  | 89  | 78  | 118373  |
| i | 2 | θj  | 1   | 3   | 1   | 5       |
| i | 2 | θr  | 67  | 122 | 110 | 1635783 |
| i | 2 | θw  | 2   | 8   | 8   | 2511    |
| i | 2 | vj  | 6   | 12  | 12  | 217770  |
| i | 2 | vw  | 5   | 6   | 6   | 889     |
| i | 2 | zl  | 1   | 2   | 2   | 83      |
| i | 3 | skj | 6   | 10  | 10  | 2790    |
| i | 3 | skl | 1   | 1   | 1   | 866     |
| i | 3 | skr | 64  | 165 | 139 | 151160  |
| i | 3 | skw | 47  | 111 | 103 | 103645  |
| i | 3 | spj | 7   | 10  | 10  | 2582    |
| i | 3 | spl | 24  | 52  | 43  | 32012   |
| i | 3 | spr | 38  | 87  | 73  | 145559  |
| i | 3 | stj | 18  | 30  | 30  | 322950  |
| i | 3 | str | 127 | 265 | 242 | 814034  |

**Medial clusters**

| pos | len | cl  | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-----|----------|----------|----------|----------|
| m   | 2   | bd  | 10       | 23       | 22       | 9666     |
| m   | 2   | bdz | 25       | 48       | 39       | 157300   |
| m   | 2   | bf  | 5        | 9        | 7        | 294      |
| m   | 2   | bg  | 1        | 2        | 2        | 28       |
| m   | 2   | bh  | 5        | 11       | 11       | 3561     |
| m   | 2   | bj  | 57       | 106      | 97       | 196539   |
| m   | 2   | bk  | 9        | 17       | 15       | 5663     |
| m   | 2   | bl  | 439      | 517      | 466      | 1146808  |
| m   | 2   | bm  | 13       | 21       | 20       | 17522    |
| m   | 2   | bn  | 14       | 21       | 20       | 5369     |
| m   | 2   | br  | 125      | 219      | 191      | 159799   |
| m   | 2   | bs  | 51       | 76       | 73       | 128668   |
| m   | 2   | bt  | 10       | 19       | 16       | 28493    |
| m   | 2   | bθ  | 0        | 2        | 0        | 0        |
| m   | 2   | bv  | 9        | 18       | 17       | 58463    |
| m   | 2   | bw  | 3        | 5        | 5        | 9533     |
| m   | 2   | bz  | 16       | 29       | 27       | 84819    |
| m   | 2   | db  | 26       | 39       | 37       | 38892    |
| m   | 2   | dd  | 5        | 6        | 5        | 4542     |
| m   | 2   | df  | 22       | 28       | 21       | 6901     |
| m   | 2   | dg  | 4        | 6        | 6        | 925      |
| m   | 2   | dh  | 15       | 26       | 25       | 11510    |
| m   | 2   | dj  | 72       | 139      | 120      | 444739   |
| m   | 2   | dk  | 18       | 32       | 27       | 23665    |
| m   | 2   | dl  | 347      | 395      | 338      | 377779   |
| m   | 2   | dm  | 50       | 80       | 72       | 186609   |
| m   | 2   | dn  | 64       | 71       | 62       | 53454    |
| m   | 2   | dp  | 12       | 21       | 19       | 2478     |
| m   | 2   | dr  | 96       | 172      | 154      | 127288   |
| m   | 2   | ds  | 15       | 26       | 23       | 35272    |
| m   | 2   | dʃ  | 14       | 21       | 20       | 7840     |
| m   | 2   | dt  | 4        | 7        | 6        | 2048     |
| m   | 2   | dθ  | 2        | 2        | 2        | 338      |
| m   | 2   | dʒ  | 2        | 2        | 2        | 74       |
| m   | 2   | dv  | 47       | 87       | 70       | 214199   |
| m   | 2   | dw  | 40       | 57       | 47       | 43384    |
| m   | 2   | dz  | 1        | 2        | 2        | 106      |

|   |   |     |     |     |     |         |
|---|---|-----|-----|-----|-----|---------|
| m | 2 | ðh  | 2   | 4   | 4   | 3993    |
| m | 2 | ðl  | 4   | 4   | 4   | 8022    |
| m | 2 | ðm  | 4   | 6   | 6   | 2640    |
| m | 2 | ðn  | 1   | 1   | 1   | 437     |
| m | 2 | ðr  | 1   | 1   | 1   | 1079    |
| m | 2 | ðs  | 2   | 2   | 2   | 282     |
| m | 2 | ɖʒh | 4   | 10  | 6   | 755     |
| m | 2 | ɖʒk | 1   | 2   | 2   | 380     |
| m | 2 | ɖʒl | 12  | 13  | 11  | 51464   |
| m | 2 | ɖʒm | 15  | 26  | 22  | 63525   |
| m | 2 | ɖʒn | 2   | 2   | 2   | 81      |
| m | 2 | ɖʒp | 2   | 4   | 1   | 23      |
| m | 2 | ɖʒr | 2   | 3   | 3   | 445     |
| m | 2 | ɖʒw | 8   | 10  | 6   | 2016    |
| m | 2 | fb  | 10  | 17  | 14  | 2375    |
| m | 2 | fd  | 2   | 3   | 2   | 182     |
| m | 2 | fg  | 3   | 7   | 7   | 4339    |
| m | 2 | fh  | 9   | 10  | 7   | 1445    |
| m | 2 | fj  | 39  | 72  | 67  | 78481   |
| m | 2 | fk  | 5   | 9   | 5   | 2989    |
| m | 2 | fl  | 110 | 170 | 153 | 182649  |
| m | 2 | fm  | 1   | 3   | 3   | 25      |
| m | 2 | fn  | 11  | 12  | 11  | 6044    |
| m | 2 | fp  | 4   | 5   | 2   | 332     |
| m | 2 | fr  | 70  | 137 | 120 | 350854  |
| m | 2 | fs  | 6   | 11  | 11  | 5915    |
| m | 2 | fʃ  | 4   | 5   | 5   | 8893    |
| m | 2 | ft  | 82  | 166 | 143 | 1245943 |
| m | 2 | fθ  | 8   | 10  | 9   | 509     |
| m | 2 | fw  | 3   | 3   | 3   | 14959   |
| m | 2 | gb  | 9   | 14  | 12  | 2373    |
| m | 2 | gf  | 2   | 4   | 1   | 65      |
| m | 2 | gh  | 4   | 8   | 6   | 601     |
| m | 2 | gj  | 44  | 77  | 72  | 216471  |
| m | 2 | gl  | 58  | 91  | 75  | 48769   |
| m | 2 | gm  | 42  | 75  | 64  | 44009   |
| m | 2 | gn  | 138 | 214 | 190 | 364606  |
| m | 2 | gp  | 5   | 10  | 7   | 960     |
| m | 2 | gr  | 229 | 410 | 366 | 880693  |
| m | 2 | gs  | 3   | 5   | 3   | 1279    |

| pos | len | cl  | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-----|----------|----------|----------|----------|
| m   | 2   | gf  | 2        | 4        | 4        | 1669     |
| m   | 2   | gt  | 5        | 7        | 7        | 824      |
| m   | 2   | gw  | 9        | 16       | 13       | 1724     |
| m   | 2   | gz  | 70       | 149      | 143      | 461465   |
| m   | 2   | gż  | 2        | 2        | 2        | 1999     |
| m   | 2   | hj  | 4        | 8        | 5        | 909      |
| m   | 2   | kb  | 29       | 60       | 46       | 11975    |
| m   | 2   | kd  | 16       | 27       | 25       | 13884    |
| m   | 2   | kdź | 3        | 4        | 2        | 100      |
| m   | 2   | kf  | 9        | 16       | 14       | 18339    |
| m   | 2   | kg  | 1        | 1        | 1        | 123      |
| m   | 2   | kh  | 12       | 22       | 16       | 2114     |
| m   | 2   | kj  | 203      | 358      | 325      | 654061   |
| m   | 2   | kk  | 5        | 10       | 8        | 2059     |
| m   | 2   | kl  | 171      | 280      | 244      | 654324   |
| m   | 2   | km  | 20       | 35       | 34       | 5597     |
| m   | 2   | kn  | 48       | 81       | 75       | 213159   |
| m   | 2   | kp  | 10       | 22       | 19       | 8973     |
| m   | 2   | kr  | 171      | 292      | 258      | 493383   |
| m   | 2   | ks  | 275      | 555      | 500      | 838122   |
| m   | 2   | kś  | 189      | 302      | 287      | 894672   |
| m   | 2   | kt  | 401      | 744      | 687      | 1928479  |
| m   | 2   | kθ  | 1        | 2        | 1        | 26       |
| m   | 2   | kʃ  | 61       | 100      | 97       | 494726   |
| m   | 2   | kw  | 204      | 334      | 293      | 543915   |
| m   | 2   | lb  | 55       | 94       | 82       | 51205    |
| m   | 2   | ld  | 119      | 257      | 221      | 494781   |
| m   | 2   | ldź | 35       | 52       | 46       | 70985    |
| m   | 2   | lf  | 109      | 147      | 126      | 96580    |
| m   | 2   | lg  | 22       | 40       | 35       | 6603     |
| m   | 2   | lh  | 23       | 35       | 28       | 3538     |
| m   | 2   | lj  | 81       | 143      | 126      | 610299   |
| m   | 2   | lk  | 53       | 101      | 92       | 93553    |
| m   | 2   | ll  | 38       | 38       | 32       | 23388    |
| m   | 2   | lm  | 96       | 144      | 118      | 211946   |
| m   | 2   | ln  | 74       | 77       | 67       | 49822    |
| m   | 2   | lp  | 37       | 75       | 55       | 41335    |
| m   | 2   | lr  | 38       | 57       | 52       | 37233    |
| m   | 2   | ls  | 69       | 108      | 99       | 1010891  |

|   |   |                |     |      |      |         |
|---|---|----------------|-----|------|------|---------|
| m | 2 | lf             | 17  | 30   | 20   | 8214    |
| m | 2 | lt             | 182 | 352  | 309  | 526218  |
| m | 2 | lθ             | 15  | 27   | 27   | 82728   |
| m | 2 | lf             | 29  | 50   | 44   | 220395  |
| m | 2 | lv             | 48  | 88   | 78   | 93824   |
| m | 2 | lw             | 30  | 41   | 33   | 9359    |
| m | 2 | lz             | 5   | 9    | 6    | 1483    |
| m | 2 | mb             | 250 | 511  | 452  | 855780  |
| m | 2 | md             | 7   | 10   | 5    | 12056   |
| m | 2 | mdz            | 2   | 3    | 2    | 106     |
| m | 2 | mf             | 52  | 81   | 75   | 64204   |
| m | 2 | mh             | 8   | 11   | 11   | 49547   |
| m | 2 | mj             | 88  | 161  | 143  | 300189  |
| m | 2 | mk             | 9   | 19   | 15   | 1803    |
| m | 2 | ml             | 115 | 143  | 131  | 141404  |
| m | 2 | mm             | 2   | 3    | 1    | 599     |
| m | 2 | mn             | 70  | 106  | 83   | 34944   |
| m | 2 | mp             | 392 | 731  | 675  | 1487899 |
| m | 2 | mr             | 9   | 16   | 16   | 5132    |
| m | 2 | ms             | 11  | 19   | 15   | 92349   |
| m | 2 | m <sup>l</sup> | 7   | 16   | 10   | 1797    |
| m | 2 | mt             | 5   | 8    | 7    | 96574   |
| m | 2 | mθ             | 1   | 1    | 1    | 267681  |
| m | 2 | m <sup>l</sup> | 2   | 3    | 3    | 3044    |
| m | 2 | mv             | 5   | 11   | 6    | 1307    |
| m | 2 | mw             | 20  | 26   | 25   | 196684  |
| m | 2 | mz             | 20  | 34   | 28   | 10723   |
| m | 2 | nb             | 60  | 105  | 93   | 24889   |
| m | 2 | nd             | 927 | 1709 | 1553 | 3293511 |
| m | 2 | nð             | 4   | 5    | 3    | 163     |
| m | 2 | ndz            | 82  | 144  | 140  | 246665  |
| m | 2 | nf             | 204 | 338  | 312  | 518685  |
| m | 2 | ng             | 17  | 29   | 27   | 57620   |
| m | 2 | nh             | 76  | 139  | 129  | 96274   |
| m | 2 | nj             | 162 | 266  | 231  | 610686  |
| m | 2 | nk             | 200 | 312  | 282  | 243525  |
| m | 2 | nl             | 219 | 265  | 240  | 1513780 |
| m | 2 | nm             | 82  | 113  | 106  | 398097  |
| m | 2 | nn             | 36  | 41   | 38   | 30767   |
| m | 2 | np             | 36  | 58   | 52   | 28927   |

| <b>pos</b> | <b>len</b> | <b>cl</b> | <b>dictwtyp</b> | <b>parawtyp</b> | <b>corpwtyp</b> | <b>corpwtok</b> |
|------------|------------|-----------|-----------------|-----------------|-----------------|-----------------|
| m          | 2          | nr        | 82              | 118             | 98              | 51133           |
| m          | 2          | ns        | 751             | 1345            | 1223            | 2419503         |
| m          | 2          | nʃ        | 217             | 292             | 279             | 824776          |
| m          | 2          | nt        | 1036            | 1938            | 1740            | 5168950         |
| m          | 2          | nθ        | 27              | 39              | 37              | 16863           |
| m          | 2          | nʃ̣       | 75              | 202             | 179             | 244887          |
| m          | 2          | nv        | 153             | 302             | 277             | 713020          |
| m          | 2          | nw        | 69              | 94              | 87              | 79272           |
| m          | 2          | nz        | 46              | 73              | 67              | 56465           |
| m          | 2          | ŋb        | 9               | 16              | 16              | 2672            |
| m          | 2          | ŋd        | 9               | 12              | 12              | 13821           |
| m          | 2          | ŋf        | 5               | 6               | 6               | 9810            |
| m          | 2          | ŋg        | 145             | 286             | 244             | 513578          |
| m          | 2          | ŋh        | 9               | 13              | 9               | 7305            |
| m          | 2          | ŋk        | 142             | 317             | 273             | 273442          |
| m          | 2          | ŋl        | 382             | 386             | 332             | 206849          |
| m          | 2          | ŋm        | 4               | 5               | 5               | 466             |
| m          | 2          | ŋn        | 4               | 5               | 5               | 8454            |
| m          | 2          | ŋp        | 1               | 2               | 2               | 424             |
| m          | 2          | ŋr        | 5               | 8               | 4               | 2551            |
| m          | 2          | ŋs        | 5               | 7               | 6               | 8346            |
| m          | 2          | ŋʃ        | 3               | 3               | 3               | 551             |
| m          | 2          | ŋt        | 7               | 10              | 9               | 36992           |
| m          | 2          | ŋw        | 5               | 6               | 5               | 492             |
| m          | 2          | ŋz        | 4               | 5               | 4               | 15406           |
| m          | 2          | pb        | 7               | 11              | 10              | 2036            |
| m          | 2          | pd        | 5               | 10              | 10              | 12463           |
| m          | 2          | pdʒ       | 1               | 2               | 2               | 68              |
| m          | 2          | pf        | 10              | 15              | 11              | 14718           |
| m          | 2          | pg        | 4               | 7               | 7               | 1245            |
| m          | 2          | ph        | 14              | 25              | 23              | 13291           |
| m          | 2          | pj        | 56              | 101             | 85              | 183670          |
| m          | 2          | pk        | 10              | 18              | 15              | 7378            |
| m          | 2          | pl        | 173             | 317             | 280             | 447910          |
| m          | 2          | pm        | 20              | 30              | 28              | 132658          |
| m          | 2          | pn        | 20              | 29              | 24              | 5233            |
| m          | 2          | pp        | 2               | 4               | 2               | 269             |
| m          | 2          | pr        | 206             | 414             | 376             | 678373          |
| m          | 2          | ps        | 52              | 91              | 80              | 37163           |

|   |   |      |     |      |      |         |
|---|---|------|-----|------|------|---------|
| m | 2 | pf   | 50  | 82   | 76   | 175301  |
| m | 2 | pt   | 135 | 214  | 199  | 233010  |
| m | 2 | pf̣  | 26  | 41   | 40   | 53734   |
| m | 2 | pv   | 1   | 2    | 2    | 703     |
| m | 2 | pw   | 20  | 31   | 17   | 8177    |
| m | 2 | sb   | 38  | 74   | 58   | 43549   |
| m | 2 | sd   | 13  | 28   | 18   | 7805    |
| m | 2 | sdz  | 6   | 9    | 9    | 967     |
| m | 2 | sf   | 84  | 135  | 112  | 198760  |
| m | 2 | sg   | 10  | 24   | 19   | 13969   |
| m | 2 | sh   | 17  | 36   | 34   | 25266   |
| m | 2 | sj   | 21  | 37   | 34   | 96597   |
| m | 2 | sk   | 199 | 422  | 357  | 430368  |
| m | 2 | sl   | 750 | 789  | 685  | 507384  |
| m | 2 | sm   | 82  | 140  | 124  | 162181  |
| m | 2 | sn   | 138 | 148  | 120  | 27223   |
| m | 2 | sp   | 212 | 383  | 353  | 839026  |
| m | 2 | sr   | 26  | 47   | 43   | 50381   |
| m | 2 | ss   | 5   | 10   | 9    | 1271    |
| m | 2 | ṣf  | 6   | 9    | 9    | 4447    |
| m | 2 | st   | 924 | 1703 | 1546 | 2821462 |
| m | 2 | sθ   | 15  | 31   | 26   | 11694   |
| m | 2 | sṭf | 50  | 92   | 82   | 317948  |
| m | 2 | sw   | 43  | 67   | 59   | 25315   |
| m | 2 | ʃb   | 8   | 17   | 17   | 3624    |
| m | 2 | ʃd   | 4   | 9    | 3    | 58      |
| m | 2 | ʃf   | 6   | 7    | 4    | 1133    |
| m | 2 | ʃg   | 1   | 2    | 1    | 4       |
| m | 2 | ʃh   | 1   | 2    | 2    | 5750    |
| m | 2 | ʃk   | 5   | 9    | 6    | 200     |
| m | 2 | ʃl   | 107 | 113  | 91   | 23567   |
| m | 2 | ʃm   | 24  | 35   | 29   | 45649   |
| m | 2 | ʃn   | 32  | 32   | 24   | 2996    |
| m | 2 | ʃp   | 1   | 2    | 1    | 28      |
| m | 2 | ʃr   | 2   | 6    | 6    | 7620    |
| m | 2 | ʃt   | 1   | 2    | 2    | 101     |
| m | 2 | ʃṭf | 2   | 3    | 2    | 28      |
| m | 2 | ʃw   | 6   | 8    | 8    | 3590    |
| m | 2 | tb   | 30  | 59   | 51   | 68576   |
| m | 2 | td   | 11  | 19   | 17   | 25496   |

| pos | len | cl  | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-----|----------|----------|----------|----------|
| m   | 2   | tdz | 3        | 5        | 4        | 277      |
| m   | 2   | tf  | 72       | 103      | 96       | 71507    |
| m   | 2   | tg  | 7        | 11       | 9        | 7685     |
| m   | 2   | th  | 40       | 70       | 54       | 18112    |
| m   | 2   | tj  | 126      | 190      | 169      | 288309   |
| m   | 2   | tk  | 17       | 31       | 28       | 35484    |
| m   | 2   | tl  | 318      | 361      | 329      | 802568   |
| m   | 2   | tm  | 69       | 105      | 88       | 277064   |
| m   | 2   | tn  | 65       | 83       | 78       | 143269   |
| m   | 2   | tp  | 14       | 26       | 21       | 11964    |
| m   | 2   | tr  | 317      | 536      | 448      | 410878   |
| m   | 2   | ts  | 61       | 94       | 81       | 201196   |
| m   | 2   | tʃ  | 2        | 4        | 4        | 331      |
| m   | 2   | tt  | 12       | 17       | 15       | 12760    |
| m   | 2   | tθ  | 1        | 2        | 0        | 0        |
| m   | 2   | tʃ  | 1        | 4        | 4        | 515      |
| m   | 2   | tv  | 1        | 4        | 2        | 88       |
| m   | 2   | tw  | 48       | 81       | 71       | 362050   |
| m   | 2   | θb  | 2        | 4        | 4        | 687      |
| m   | 2   | θd  | 2        | 3        | 3        | 28009    |
| m   | 2   | θf  | 19       | 20       | 20       | 13278    |
| m   | 2   | θk  | 2        | 2        | 2        | 3038     |
| m   | 2   | θl  | 30       | 31       | 28       | 74600    |
| m   | 2   | θm  | 5        | 7        | 5        | 1635     |
| m   | 2   | θn  | 10       | 13       | 10       | 28318    |
| m   | 2   | θp  | 5        | 9        | 9        | 2940     |
| m   | 2   | θr  | 20       | 32       | 30       | 27322    |
| m   | 2   | θs  | 3        | 4        | 4        | 280      |
| m   | 2   | θt  | 2        | 3        | 3        | 3242     |
| m   | 2   | θw  | 15       | 23       | 19       | 33344    |
| m   | 2   | ʃb  | 5        | 10       | 7        | 1933     |
| m   | 2   | ʃd  | 6        | 11       | 8        | 9911     |
| m   | 2   | ʃf  | 6        | 7        | 7        | 1676     |
| m   | 2   | ʃg  | 3        | 5        | 3        | 389      |
| m   | 2   | ʃh  | 4        | 12       | 10       | 1358     |
| m   | 2   | ʃj  | 1        | 2        | 2        | 328      |
| m   | 2   | ʃk  | 4        | 9        | 4        | 342      |
| m   | 2   | ʃl  | 12       | 15       | 13       | 4463     |
| m   | 2   | ʃm  | 17       | 24       | 22       | 15024    |

|   |   |     |     |     |     |        |
|---|---|-----|-----|-----|-----|--------|
| m | 2 | fn  | 2   | 2   | 2   | 1880   |
| m | 2 | fj  | 2   | 3   | 0   | 0      |
| m | 2 | fj  | 1   | 4   | 2   | 76     |
| m | 2 | ft  | 3   | 7   | 5   | 410    |
| m | 2 | fjw | 11  | 16  | 15  | 2545   |
| m | 2 | vb  | 3   | 9   | 9   | 209    |
| m | 2 | vd  | 3   | 3   | 1   | 170    |
| m | 2 | vð  | 3   | 3   | 3   | 756    |
| m | 2 | vf  | 2   | 2   | 2   | 749    |
| m | 2 | vh  | 2   | 2   | 0   | 0      |
| m | 2 | vj  | 30  | 47  | 45  | 197287 |
| m | 2 | vk  | 2   | 3   | 3   | 47     |
| m | 2 | vl  | 277 | 286 | 267 | 258239 |
| m | 2 | vm  | 10  | 16  | 15  | 104626 |
| m | 2 | vn  | 42  | 44  | 40  | 22371  |
| m | 2 | vp  | 2   | 3   | 3   | 361    |
| m | 2 | vr  | 16  | 19  | 19  | 856573 |
| m | 2 | vs  | 1   | 1   | 1   | 134    |
| m | 2 | vt  | 2   | 5   | 5   | 535    |
| m | 2 | vw  | 11  | 13  | 13  | 18131  |
| m | 2 | zb  | 17  | 30  | 27  | 74108  |
| m | 2 | zd  | 3   | 7   | 5   | 11707  |
| m | 2 | zđ  | 1   | 1   | 1   | 18     |
| m | 2 | zf  | 1   | 1   | 1   | 4      |
| m | 2 | zg  | 2   | 3   | 3   | 100    |
| m | 2 | zj  | 10  | 20  | 16  | 20381  |
| m | 2 | zk  | 3   | 8   | 7   | 2939   |
| m | 2 | zl  | 21  | 25  | 24  | 15156  |
| m | 2 | zm  | 58  | 77  | 70  | 39352  |
| m | 2 | zn  | 4   | 5   | 5   | 169551 |
| m | 2 | zp  | 5   | 7   | 5   | 39054  |
| m | 2 | zr  | 2   | 4   | 4   | 1844   |
| m | 2 | zř  | 1   | 2   | 1   | 5      |
| m | 2 | zv  | 1   | 2   | 1   | 10     |
| m | 2 | zw  | 11  | 12  | 9   | 2070   |
| m | 2 | zf  | 3   | 3   | 2   | 487    |
| m | 2 | zl  | 2   | 2   | 2   | 6160   |
| m | 2 | zm  | 8   | 12  | 11  | 16385  |
| m | 2 | zn  | 1   | 1   | 1   | 565    |
| m | 2 | zw  | 3   | 3   | 3   | 5087   |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| m   | 3   | bdj  | 5        | 8        | 8        | 2835     |
| m   | 3   | bgr  | 1        | 2        | 2        | 1922     |
| m   | 3   | bhj  | 1        | 1        | 1        | 136      |
| m   | 3   | bkj  | 1        | 1        | 1        | 292      |
| m   | 3   | bkr  | 1        | 4        | 1        | 5        |
| m   | 3   | bsj  | 1        | 4        | 4        | 751      |
| m   | 3   | bsk  | 1        | 4        | 4        | 154      |
| m   | 3   | bsl  | 3        | 5        | 2        | 264      |
| m   | 3   | bst  | 34       | 57       | 50       | 73235    |
| m   | 3   | btj  | 3        | 3        | 3        | 288      |
| m   | 3   | btr  | 9        | 18       | 17       | 3302     |
| m   | 3   | bzm  | 2        | 2        | 2        | 491      |
| m   | 3   | dbl  | 1        | 2        | 2        | 1211     |
| m   | 3   | dbr  | 1        | 3        | 2        | 395      |
| m   | 3   | ddr  | 1        | 2        | 2        | 826      |
| m   | 3   | dfl  | 2        | 3        | 3        | 331      |
| m   | 3   | dfr  | 1        | 2        | 1        | 197      |
| m   | 3   | dgr  | 1        | 2        | 2        | 226      |
| m   | 3   | dhj  | 1        | 1        | 1        | 4        |
| m   | 3   | dkl  | 2        | 2        | 2        | 277      |
| m   | 3   | dkr  | 2        | 4        | 3        | 504      |
| m   | 3   | dkw  | 0        | 1        | 1        | 11541    |
| m   | 3   | dpl  | 2        | 2        | 2        | 120      |
| m   | 3   | dsk  | 1        | 2        | 2        | 403      |
| m   | 3   | dsl  | 1        | 2        | 2        | 41       |
| m   | 3   | dsp  | 2        | 3        | 3        | 1059     |
| m   | 3   | dst  | 20       | 36       | 28       | 4623     |
| m   | 3   | dtr  | 1        | 4        | 4        | 326      |
| m   | 3   | dθw  | 2        | 2        | 0        | 0        |
| m   | 3   | dvj  | 2        | 2        | 2        | 162      |
| m   | 3   | dzm  | 20       | 21       | 18       | 2234     |
| m   | 3   | dzp  | 1        | 1        | 1        | 45       |
| m   | 3   | ðdr  | 6        | 9        | 9        | 26126    |
| m   | 3   | ðst  | 3        | 5        | 5        | 5408     |
| m   | 3   | ðzh  | 1        | 2        | 2        | 36       |
| m   | 3   | ðzl  | 1        | 2        | 2        | 609      |
| m   | 3   | dźkr | 1        | 1        | 1        | 73       |
| m   | 3   | dźsk | 1        | 1        | 1        | 3897     |
| m   | 3   | fbl  | 1        | 1        | 1        | 391      |

|   |   |     |    |     |     |        |
|---|---|-----|----|-----|-----|--------|
| m | 3 | fbr | 2  | 4   | 4   | 544    |
| m | 3 | fdr | 2  | 5   | 0   | 0      |
| m | 3 | fhj | 1  | 1   | 1   | 280    |
| m | 3 | fkr | 1  | 1   | 1   | 24     |
| m | 3 | fpr | 2  | 3   | 2   | 154    |
| m | 3 | fst | 2  | 3   | 3   | 725    |
| m | 3 | ftb | 2  | 2   | 2   | 1767   |
| m | 3 | ftl | 12 | 12  | 7   | 23051  |
| m | 3 | ftm | 2  | 2   | 0   | 0      |
| m | 3 | ftn | 3  | 3   | 3   | 1096   |
| m | 3 | ftt | 2  | 5   | 4   | 254    |
| m | 3 | ftw | 2  | 3   | 3   | 1243   |
| m | 3 | gfl | 1  | 1   | 1   | 81     |
| m | 3 | gpl | 1  | 2   | 2   | 1890   |
| m | 3 | gsk | 1  | 1   | 1   | 116    |
| m | 3 | gst | 6  | 10  | 9   | 3975   |
| m | 3 | gsw | 1  | 1   | 0   | 0      |
| m | 3 | gtr | 1  | 1   | 1   | 18     |
| m | 3 | gzb | 1  | 2   | 1   | 13     |
| m | 3 | gzh | 1  | 2   | 2   | 50     |
| m | 3 | gzj | 8  | 14  | 14  | 3616   |
| m | 3 | kbr | 3  | 6   | 4   | 768    |
| m | 3 | kdr | 1  | 2   | 2   | 3053   |
| m | 3 | kfr | 3  | 4   | 1   | 17     |
| m | 3 | kgr | 1  | 2   | 2   | 28175  |
| m | 3 | kkk | 2  | 3   | 3   | 189    |
| m | 3 | kkr | 1  | 1   | 1   | 5      |
| m | 3 | knj | 1  | 2   | 1   | 5      |
| m | 3 | ksf | 3  | 5   | 4   | 278    |
| m | 3 | ksh | 7  | 17  | 11  | 2948   |
| m | 3 | ksk | 16 | 39  | 32  | 7631   |
| m | 3 | ksl | 6  | 7   | 5   | 367    |
| m | 3 | ksm | 9  | 11  | 9   | 15175  |
| m | 3 | ksp | 89 | 175 | 156 | 572926 |
| m | 3 | ksj | 1  | 2   | 2   | 79     |
| m | 3 | kst | 66 | 123 | 104 | 142424 |
| m | 3 | ksf | 14 | 23  | 20  | 82907  |
| m | 3 | ksw | 2  | 3   | 3   | 315    |
| m | 3 | ktf | 6  | 6   | 4   | 1098   |
| m | 3 | ktl | 21 | 21  | 21  | 92686  |

| <b>pos</b> | <b>len</b> | <b>cl</b> | <b>dictwtyp</b> | <b>parawtyp</b> | <b>corpwtyp</b> | <b>corpwtok</b> |
|------------|------------|-----------|-----------------|-----------------|-----------------|-----------------|
| m          | 3          | ktm       | 1               | 2               | 2               | 1100            |
| m          | 3          | ktn       | 6               | 6               | 6               | 1721            |
| m          | 3          | ktr       | 44              | 75              | 67              | 127537          |
| m          | 3          | kθr       | 2               | 3               | 3               | 8097            |
| m          | 3          | l bj      | 1               | 1               | 1               | 38              |
| m          | 3          | l bl      | 3               | 5               | 3               | 1175            |
| m          | 3          | l br      | 5               | 6               | 5               | 508             |
| m          | 3          | l db      | 1               | 1               | 1               | 1288            |
| m          | 3          | l df      | 2               | 2               | 0               | 0               |
| m          | 3          | l dh      | 3               | 4               | 3               | 16968           |
| m          | 3          | l dj      | 2               | 5               | 5               | 1086            |
| m          | 3          | l dl      | 12              | 12              | 12              | 7547            |
| m          | 3          | l dn      | 3               | 3               | 3               | 1182            |
| m          | 3          | l dr      | 18              | 23              | 21              | 239272          |
| m          | 3          | l dʃ      | 1               | 2               | 1               | 9               |
| m          | 3          | l dw      | 2               | 2               | 2               | 1586            |
| m          | 3          | l fd      | 3               | 3               | 3               | 2112            |
| m          | 3          | l fh      | 2               | 3               | 3               | 1627            |
| m          | 3          | l fj      | 1               | 2               | 1               | 31              |
| m          | 3          | l fk      | 8               | 8               | 8               | 6743            |
| m          | 3          | l fl      | 4               | 5               | 5               | 1243            |
| m          | 3          | l fm      | 1               | 1               | 1               | 487             |
| m          | 3          | l fp      | 3               | 3               | 3               | 792             |
| m          | 3          | l fr      | 14              | 19              | 17              | 14225           |
| m          | 3          | l fs      | 7               | 10              | 8               | 2691            |
| m          | 3          | l fw      | 2               | 3               | 2               | 28              |
| m          | 3          | l gr      | 9               | 13              | 11              | 6425            |
| m          | 3          | l kh      | 1               | 2               | 2               | 727             |
| m          | 3          | l kj      | 10              | 20              | 19              | 21531           |
| m          | 3          | l kl      | 4               | 5               | 5               | 6764            |
| m          | 3          | l km      | 3               | 4               | 4               | 295             |
| m          | 3          | l kr      | 8               | 11              | 9               | 507             |
| m          | 3          | l ks      | 1               | 2               | 1               | 4               |
| m          | 3          | l kt      | 0               | 2               | 0               | 0               |
| m          | 3          | l kw      | 3               | 5               | 5               | 382             |
| m          | 3          | l nj      | 1               | 1               | 1               | 835             |
| m          | 3          | l pf      | 5               | 5               | 5               | 11373           |
| m          | 3          | l pl      | 3               | 4               | 2               | 4308            |
| m          | 3          | l pm      | 2               | 4               | 4               | 102             |

|   |   |      |     |     |     |        |
|---|---|------|-----|-----|-----|--------|
| m | 3 | lpr  | 3   | 4   | 4   | 4204   |
| m | 3 | lpt  | 2   | 5   | 5   | 5398   |
| m | 3 | lptf | 3   | 6   | 6   | 16479  |
| m | 3 | lsh  | 1   | 2   | 2   | 543    |
| m | 3 | lsk  | 5   | 7   | 6   | 1478   |
| m | 3 | lsl  | 4   | 4   | 4   | 2650   |
| m | 3 | lsn  | 2   | 3   | 3   | 957    |
| m | 3 | lsp  | 8   | 14  | 10  | 17534  |
| m | 3 | lst  | 25  | 46  | 41  | 17415  |
| m | 3 | lsw  | 2   | 2   | 2   | 31444  |
| m | 3 | ltf  | 1   | 1   | 1   | 13     |
| m | 3 | lth  | 1   | 1   | 1   | 2477   |
| m | 3 | ltl  | 3   | 3   | 3   | 201    |
| m | 3 | ltp  | 2   | 3   | 0   | 0      |
| m | 3 | ltr  | 29  | 47  | 32  | 14848  |
| m | 3 | lts  | 4   | 4   | 2   | 118    |
| m | 3 | ltw  | 2   | 2   | 2   | 867    |
| m | 3 | l0f  | 1   | 1   | 1   | 1097   |
| m | 3 | lvj  | 2   | 3   | 2   | 54     |
| m | 3 | lvm  | 2   | 4   | 3   | 16372  |
| m | 3 | lzm  | 5   | 5   | 3   | 4511   |
| m | 3 | lzw  | 2   | 2   | 2   | 374    |
| m | 3 | mbj  | 8   | 18  | 15  | 5583   |
| m | 3 | mbl  | 27  | 57  | 53  | 42844  |
| m | 3 | mbr  | 27  | 49  | 47  | 35626  |
| m | 3 | mdr  | 1   | 1   | 1   | 198    |
| m | 3 | mfl  | 1   | 2   | 2   | 20     |
| m | 3 | mfr  | 2   | 2   | 2   | 214    |
| m | 3 | mft  | 4   | 4   | 4   | 34099  |
| m | 3 | mkr  | 1   | 1   | 1   | 9      |
| m | 3 | mpb  | 1   | 2   | 2   | 335    |
| m | 3 | mpj  | 20  | 44  | 43  | 116248 |
| m | 3 | mpl  | 102 | 186 | 178 | 644644 |
| m | 3 | mpm  | 1   | 2   | 2   | 666    |
| m | 3 | mpn  | 2   | 2   | 2   | 371    |
| m | 3 | mpr  | 69  | 117 | 112 | 178502 |
| m | 3 | mps  | 1   | 4   | 4   | 1521   |
| m | 3 | mpf  | 2   | 3   | 3   | 10852  |
| m | 3 | mpt  | 24  | 47  | 46  | 108487 |
| m | 3 | msk  | 3   | 5   | 3   | 38     |

| <b>pos</b> | <b>len</b> | <b>cl</b> | <b>dictwtyp</b> | <b>parawtyp</b> | <b>corpwtyp</b> | <b>corpwtok</b> |
|------------|------------|-----------|-----------------|-----------------|-----------------|-----------------|
| m          | 3          | m sl      | 1               | 2               | 0               | 0               |
| m          | 3          | m sp      | 4               | 4               | 4               | 910             |
| m          | 3          | m st      | 17              | 29              | 24              | 28416           |
| m          | 3          | m zh      | 1               | 2               | 2               | 118             |
| m          | 3          | m zm      | 1               | 1               | 1               | 136             |
| m          | 3          | n bl      | 5               | 9               | 6               | 113             |
| m          | 3          | n br      | 7               | 9               | 7               | 2434            |
| m          | 3          | n dh      | 3               | 5               | 5               | 518             |
| m          | 3          | n dj      | 25              | 47              | 38              | 30337           |
| m          | 3          | n dl      | 31              | 42              | 36              | 46136           |
| m          | 3          | n dm      | 1               | 2               | 2               | 1118            |
| m          | 3          | n dn      | 2               | 2               | 2               | 1017            |
| m          | 3          | n dr      | 51              | 78              | 67              | 96559           |
| m          | 3          | n dw      | 7               | 9               | 7               | 1348            |
| m          | 3          | n dzl     | 2               | 3               | 3               | 590             |
| m          | 3          | n fj      | 8               | 22              | 17              | 29871           |
| m          | 3          | n fl      | 45              | 73              | 69              | 131587          |
| m          | 3          | n fr      | 18              | 37              | 31              | 33872           |
| m          | 3          | n gl      | 8               | 11              | 11              | 5137            |
| m          | 3          | n gr      | 29              | 59              | 48              | 41791           |
| m          | 3          | n hj      | 4               | 5               | 4               | 1283            |
| m          | 3          | n kj      | 4               | 5               | 5               | 656             |
| m          | 3          | n kl      | 30              | 55              | 45              | 333129          |
| m          | 3          | n kr      | 19              | 37              | 33              | 109543          |
| m          | 3          | n kw      | 24              | 35              | 33              | 20178           |
| m          | 3          | n pl      | 11              | 22              | 14              | 7301            |
| m          | 3          | n pr      | 23              | 25              | 23              | 13916           |
| m          | 3          | n pw      | 2               | 2               | 0               | 0               |
| m          | 3          | n sd      | 1               | 2               | 2               | 26              |
| m          | 3          | n sf      | 16              | 32              | 32              | 75992           |
| m          | 3          | n sj      | 25              | 39              | 35              | 68674           |
| m          | 3          | n sk      | 6               | 12              | 10              | 2182            |
| m          | 3          | n sl      | 32              | 53              | 41              | 42684           |
| m          | 3          | n sm      | 10              | 19              | 19              | 16868           |
| m          | 3          | n sn      | 6               | 9               | 9               | 535             |
| m          | 3          | n sp      | 50              | 84              | 76              | 106021          |
| m          | 3          | n sf      | 2               | 6               | 5               | 148             |
| m          | 3          | n st      | 80              | 141             | 129             | 359240          |
| m          | 3          | n sw      | 7               | 8               | 8               | 758             |

|   |   |      |     |     |     |         |
|---|---|------|-----|-----|-----|---------|
| m | 3 | nr   | 2   | 8   | 8   | 833     |
| m | 3 | ntb  | 1   | 2   | 1   | 15      |
| m | 3 | ntd  | 3   | 4   | 4   | 4155    |
| m | 3 | ntf  | 5   | 5   | 5   | 1575    |
| m | 3 | nth  | 3   | 4   | 4   | 1343    |
| m | 3 | ntj  | 18  | 30  | 21  | 6632    |
| m | 3 | ntl  | 345 | 354 | 325 | 624138  |
| m | 3 | ntm  | 12  | 20  | 19  | 23715   |
| m | 3 | ntn  | 9   | 11  | 11  | 969     |
| m | 3 | ntr  | 197 | 325 | 290 | 1176769 |
| m | 3 | nts  | 7   | 11  | 5   | 1095    |
| m | 3 | ntw  | 2   | 5   | 5   | 683     |
| m | 3 | n0j  | 6   | 10  | 10  | 15893   |
| m | 3 | n0l  | 4   | 5   | 3   | 18462   |
| m | 3 | n0r  | 19  | 30  | 27  | 15314   |
| m | 3 | ntfb | 2   | 4   | 4   | 331     |
| m | 3 | ntfl | 2   | 2   | 2   | 722     |
| m | 3 | ntfm | 4   | 6   | 6   | 791     |
| m | 3 | ntfn | 1   | 1   | 1   | 4       |
| m | 3 | ntfp | 2   | 4   | 3   | 396     |
| m | 3 | ntft | 1   | 2   | 2   | 1194    |
| m | 3 | nzf  | 2   | 2   | 2   | 255     |
| m | 3 | nzk  | 1   | 1   | 1   | 322     |
| m | 3 | nzm  | 16  | 25  | 24  | 20136   |
| m | 3 | nzp  | 2   | 2   | 1   | 809     |
| m | 3 | nzt  | 2   | 2   | 2   | 1562    |
| m | 3 | nzv  | 4   | 5   | 5   | 803     |
| m | 3 | nzw  | 2   | 2   | 1   | 23      |
| m | 3 | ndr  | 1   | 1   | 1   | 24      |
| m | 3 | ngj  | 9   | 12  | 12  | 8019    |
| m | 3 | ngl  | 15  | 21  | 16  | 6153    |
| m | 3 | ngr  | 22  | 38  | 35  | 157954  |
| m | 3 | ngw  | 21  | 39  | 39  | 92753   |
| m | 3 | nkf  | 5   | 7   | 6   | 3745    |
| m | 3 | nkj  | 7   | 14  | 12  | 2120    |
| m | 3 | nkf  | 22  | 32  | 29  | 35942   |
| m | 3 | nkj  | 4   | 5   | 3   | 916     |
| m | 3 | nkj  | 3   | 4   | 4   | 409     |
| m | 3 | nkr  | 21  | 35  | 32  | 98805   |
| m | 3 | nks  | 2   | 9   | 6   | 97      |

| pos | len | cl  | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-----|----------|----------|----------|----------|
| m   | 3   | ηkj | 4        | 6        | 6        | 3737     |
| m   | 3   | ηkt | 17       | 23       | 23       | 12349    |
| m   | 3   | ηkθ | 5        | 14       | 14       | 20162    |
| m   | 3   | ηkj | 5        | 8        | 7        | 3901     |
| m   | 3   | ηkw | 17       | 36       | 33       | 12873    |
| m   | 3   | ηpl | 2        | 3        | 3        | 231      |
| m   | 3   | ηsp | 1        | 2        | 2        | 295      |
| m   | 3   | ηst | 3        | 5        | 5        | 278      |
| m   | 3   | pbr | 4        | 9        | 7        | 1677     |
| m   | 3   | pdr | 4        | 12       | 6        | 210      |
| m   | 3   | pfl | 1        | 1        | 1        | 252      |
| m   | 3   | pfr | 4        | 10       | 8        | 1636     |
| m   | 3   | pgr | 2        | 6        | 5        | 6464     |
| m   | 3   | pkl | 2        | 2        | 2        | 12       |
| m   | 3   | psj | 1        | 2        | 2        | 2649     |
| m   | 3   | psk | 3        | 5        | 4        | 439      |
| m   | 3   | pst | 13       | 24       | 21       | 13053    |
| m   | 3   | ptf | 2        | 2        | 0        | 0        |
| m   | 3   | ptj | 1        | 2        | 2        | 90       |
| m   | 3   | ptl | 8        | 8        | 8        | 12358    |
| m   | 3   | ptn | 3        | 3        | 3        | 139      |
| m   | 3   | ptr | 2        | 3        | 3        | 167      |
| m   | 3   | pts | 2        | 3        | 3        | 13545    |
| m   | 3   | sbr | 3        | 7        | 7        | 345      |
| m   | 3   | sfl | 5        | 7        | 7        | 220      |
| m   | 3   | sfr | 2        | 5        | 3        | 109      |
| m   | 3   | sgr | 7        | 10       | 8        | 3513     |
| m   | 3   | skj | 18       | 27       | 24       | 40254    |
| m   | 3   | skl | 15       | 24       | 23       | 16158    |
| m   | 3   | skm | 1        | 2        | 2        | 122      |
| m   | 3   | skn | 3        | 3        | 3        | 53       |
| m   | 3   | skr | 51       | 92       | 79       | 156761   |
| m   | 3   | skw | 12       | 27       | 21       | 2485     |
| m   | 3   | spj | 8        | 13       | 12       | 17554    |
| m   | 3   | spl | 17       | 31       | 29       | 63279    |
| m   | 3   | spn | 1        | 1        | 1        | 142      |
| m   | 3   | spr | 20       | 39       | 33       | 10162    |
| m   | 3   | spw | 1        | 1        | 1        | 16       |
| m   | 3   | ssk | 1        | 4        | 2        | 152      |

|   |   |     |     |     |     |        |
|---|---|-----|-----|-----|-----|--------|
| m | 3 | ssp | 4   | 9   | 7   | 530    |
| m | 3 | sst | 4   | 9   | 7   | 641    |
| m | 3 | stf | 9   | 9   | 9   | 1548   |
| m | 3 | stg | 1   | 2   | 1   | 59     |
| m | 3 | sth | 5   | 7   | 7   | 2032   |
| m | 3 | stj | 14  | 23  | 22  | 11341  |
| m | 3 | stl | 35  | 41  | 36  | 34597  |
| m | 3 | stm | 2   | 3   | 3   | 462    |
| m | 3 | stn | 3   | 3   | 3   | 667    |
| m | 3 | str | 217 | 360 | 312 | 886577 |
| m | 3 | stw | 6   | 9   | 9   | 2084   |
| m | 3 | tkl | 2   | 4   | 4   | 532    |
| m | 3 | jpl | 1   | 2   | 1   | 5      |
| m | 3 | jst | 1   | 2   | 2   | 68     |
| m | 3 | jtr | 2   | 3   | 3   | 2060   |
| m | 3 | tbj | 1   | 1   | 1   | 338    |
| m | 3 | tbl | 2   | 2   | 2   | 315    |
| m | 3 | tbr | 13  | 19  | 13  | 7840   |
| m | 3 | tdr | 1   | 2   | 2   | 131    |
| m | 3 | tfl | 5   | 9   | 8   | 1400   |
| m | 3 | tfr | 2   | 2   | 2   | 150    |
| m | 3 | tgl | 1   | 1   | 1   | 168    |
| m | 3 | tgr | 6   | 9   | 9   | 2115   |
| m | 3 | tkl | 2   | 6   | 6   | 2271   |
| m | 3 | tkr | 5   | 13  | 10  | 2931   |
| m | 3 | tpl | 5   | 11  | 9   | 6269   |
| m | 3 | tpr | 4   | 5   | 5   | 3192   |
| m | 3 | tsk | 2   | 3   | 3   | 2136   |
| m | 3 | tsl | 1   | 7   | 1   | 225    |
| m | 3 | tsm | 14  | 17  | 17  | 3699   |
| m | 3 | tsp | 6   | 10  | 3   | 2062   |
| m | 3 | tst | 8   | 15  | 11  | 11430  |
| m | 3 | tsw | 1   | 2   | 0   | 0      |
| m | 3 | ttr | 2   | 3   | 0   | 0      |
| m | 3 | θbl | 2   | 3   | 2   | 50     |
| m | 3 | θbr | 1   | 2   | 2   | 1112   |
| m | 3 | θkw | 1   | 2   | 2   | 7153   |
| m | 3 | θpl | 1   | 2   | 2   | 1171   |
| m | 3 | θpr | 1   | 4   | 1   | 5      |
| m | 3 | θbl | 2   | 2   | 2   | 212    |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| m   | 3   | fđj  | 1        | 2        | 2        | 203      |
| m   | 3   | fkr  | 1        | 1        | 1        | 1150     |
| m   | 3   | fst  | 1        | 2        | 2        | 700      |
| m   | 3   | vst  | 2        | 3        | 3        | 4096     |
| m   | 3   | zbl  | 1        | 2        | 2        | 201      |
| m   | 3   | zfl  | 1        | 2        | 1        | 12       |
| m   | 3   | zkl  | 1        | 1        | 1        | 252      |
| m   | 3   | zkr  | 1        | 4        | 4        | 402      |
| m   | 3   | zpr  | 1        | 1        | 1        | 575      |
| m   | 3   | zst  | 1        | 2        | 2        | 869      |
| m   | 4   | bskj | 6        | 11       | 10       | 12027    |
| m   | 4   | bskr | 3        | 11       | 11       | 7513     |
| m   | 4   | bstr | 19       | 30       | 26       | 16434    |
| m   | 4   | dspr | 2        | 3        | 3        | 10283    |
| m   | 4   | dstr | 2        | 2        | 2        | 469      |
| m   | 4   | fspr | 1        | 1        | 1        | 3207     |
| m   | 4   | fstr | 1        | 1        | 1        | 34       |
| m   | 4   | ftsm | 9        | 9        | 7        | 2306     |
| m   | 4   | ksfr | 2        | 2        | 2        | 1848     |
| m   | 4   | ksgl | 1        | 2        | 2        | 144      |
| m   | 4   | kshj | 2        | 6        | 6        | 533      |
| m   | 4   | kskj | 4        | 7        | 6        | 17285    |
| m   | 4   | kskl | 7        | 14       | 14       | 31982    |
| m   | 4   | kskr | 10       | 16       | 14       | 2919     |
| m   | 4   | kskw | 3        | 3        | 3        | 3014     |
| m   | 4   | kspl | 24       | 45       | 45       | 208161   |
| m   | 4   | kspr | 12       | 21       | 21       | 66232    |
| m   | 4   | ksst | 1        | 1        | 1        | 29       |
| m   | 4   | kstj | 1        | 1        | 1        | 1478     |
| m   | 4   | kstr | 41       | 67       | 64       | 131985   |
| m   | 4   | ldbl | 1        | 1        | 1        | 592      |
| m   | 4   | ldst | 2        | 3        | 3        | 108      |
| m   | 4   | ldzm | 2        | 2        | 0        | 0        |
| m   | 4   | lfpr | 1        | 1        | 1        | 376      |
| m   | 4   | lfst | 2        | 3        | 3        | 443      |
| m   | 4   | lmzm | 2        | 2        | 2        | 178      |
| m   | 4   | lptr | 1        | 2        | 1        | 20       |
| m   | 4   | lstr | 1        | 2        | 2        | 394      |
| m   | 4   | ltst | 2        | 3        | 0        | 0        |

|   |   |       |    |     |     |        |
|---|---|-------|----|-----|-----|--------|
| m | 4 | mpbl  | 1  | 1   | 1   | 18     |
| m | 4 | mptj  | 4  | 5   | 5   | 2761   |
| m | 4 | mptl  | 2  | 2   | 2   | 6080   |
| m | 4 | mptn  | 1  | 1   | 1   | 59     |
| m | 4 | mptr  | 1  | 2   | 2   | 107    |
| m | 4 | mskr  | 4  | 9   | 9   | 886    |
| m | 4 | mstr  | 3  | 6   | 6   | 2094   |
| m | 4 | ndzm  | 6  | 6   | 3   | 21     |
| m | 4 | nsfj  | 2  | 6   | 5   | 1085   |
| m | 4 | nskr  | 14 | 32  | 31  | 17984  |
| m | 4 | nspl  | 2  | 6   | 6   | 7037   |
| m | 4 | nspr  | 1  | 2   | 2   | 36     |
| m | 4 | nstr  | 57 | 104 | 103 | 215044 |
| m | 4 | ntbl  | 1  | 1   | 1   | 413    |
| m | 4 | ntbr  | 1  | 2   | 2   | 492    |
| m | 4 | ntfr  | 2  | 2   | 2   | 242    |
| m | 4 | ntpr  | 1  | 1   | 1   | 53     |
| m | 4 | ntsm  | 4  | 4   | 2   | 306    |
| m | 4 | ntst  | 1  | 1   | 1   | 159    |
| m | 4 | ntfdr | 1  | 1   | 1   | 61     |
| m | 4 | nzgr  | 3  | 8   | 8   | 1515   |
| m | 4 | nzmj  | 2  | 6   | 6   | 407    |
| m | 4 | ηksg  | 1  | 2   | 2   | 6258   |
| m | 4 | ηkst  | 6  | 9   | 8   | 9897   |
| m | 4 | ηktl  | 6  | 6   | 6   | 5840   |
| m | 4 | ηktn  | 3  | 3   | 3   | 155    |
| m | 4 | ηkθw  | 2  | 2   | 2   | 1133   |
| m | 4 | ηspr  | 1  | 2   | 1   | 34     |
| m | 4 | ηstr  | 1  | 2   | 1   | 93     |
| m | 4 | pskr  | 1  | 2   | 2   | 134    |
| m | 4 | pskw  | 1  | 2   | 2   | 86     |
| m | 4 | pstr  | 2  | 3   | 2   | 1929   |
| m | 4 | tspr  | 1  | 1   | 1   | 80     |
| m | 4 | tstr  | 4  | 7   | 7   | 2202   |
| m | 4 | vzdr  | 1  | 6   | 6   | 1141   |
| m | 5 | mpstr | 1  | 2   | 0   | 0      |
| m | 5 | stskr | 1  | 2   | 2   | 392    |
| m | 5 | ststr | 1  | 1   | 1   | 250    |

**Final clusters**

| pos | len | cl  | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|-----|----------|----------|----------|----------|
| f   | 2   | bd  | 3        | 66       | 60       | 92158    |
| f   | 2   | bz  | 2        | 127      | 119      | 142994   |
| f   | 2   | dθ  | 3        | 3        | 3        | 4974     |
| f   | 2   | dz  | 13       | 632      | 554      | 1156210  |
| f   | 2   | ðd  | 6        | 22       | 21       | 11224    |
| f   | 2   | ðz  | 3        | 34       | 29       | 49638    |
| f   | 2   | dʒd | 14       | 106      | 99       | 190266   |
| f   | 2   | fs  | 1        | 143      | 115      | 87412    |
| f   | 2   | ft  | 62       | 113      | 105      | 384444   |
| f   | 2   | fθ  | 1        | 1        | 1        | 15357    |
| f   | 2   | gd  | 5        | 66       | 57       | 45165    |
| f   | 2   | gz  | 0        | 138      | 118      | 191172   |
| f   | 2   | ks  | 135      | 789      | 697      | 1679023  |
| f   | 2   | kt  | 145      | 355      | 337      | 1652907  |
| f   | 2   | lb  | 3        | 3        | 3        | 2837     |
| f   | 2   | ld  | 130      | 700      | 637      | 2137481  |
| f   | 2   | ldʒ | 5        | 5        | 5        | 3637     |
| f   | 2   | lf  | 18       | 18       | 16       | 302964   |
| f   | 2   | lk  | 13       | 13       | 13       | 38877    |
| f   | 2   | lm  | 6        | 6        | 6        | 64082    |
| f   | 2   | ln  | 2        | 2        | 2        | 572      |
| f   | 2   | lp  | 9        | 9        | 9        | 195224   |
| f   | 2   | ls  | 7        | 7        | 7        | 115425   |
| f   | 2   | lf  | 1        | 1        | 1        | 1456     |
| f   | 2   | lt  | 72       | 72       | 67       | 430738   |
| f   | 2   | lθ  | 6        | 6        | 6        | 147741   |
| f   | 2   | lf  | 6        | 6        | 6        | 2454     |
| f   | 2   | lv  | 14       | 14       | 13       | 51124    |
| f   | 2   | lz  | 8        | 1299     | 1128     | 1815294  |
| f   | 2   | md  | 26       | 179      | 169      | 397098   |
| f   | 2   | mf  | 6        | 6        | 6        | 6765     |
| f   | 2   | mp  | 58       | 58       | 55       | 115815   |
| f   | 2   | mt  | 1        | 1        | 1        | 592      |
| f   | 2   | mθ  | 1        | 1        | 1        | 5608     |
| f   | 2   | mz  | 9        | 488      | 428      | 1437128  |
| f   | 2   | nd  | 308      | 712      | 660      | 15151219 |
| f   | 2   | ndʒ | 7        | 7        | 7        | 187012   |

|   |   |      |     |      |      |         |
|---|---|------|-----|------|------|---------|
| f | 2 | ns   | 407 | 409  | 394  | 2000132 |
| f | 2 | nt   | 920 | 920  | 884  | 4903475 |
| f | 2 | nθ   | 23  | 23   | 22   | 97612   |
| f | 2 | ntʃ  | 42  | 42   | 41   | 116547  |
| f | 2 | nz   | 16  | 2034 | 1770 | 3029495 |
| f | 2 | ŋd   | 1   | 20   | 18   | 15810   |
| f | 2 | ŋk   | 87  | 87   | 83   | 925769  |
| f | 2 | ŋz   | 9   | 283  | 256  | 643061  |
| f | 2 | ps   | 23  | 399  | 334  | 685168  |
| f | 2 | pt   | 48  | 205  | 191  | 556637  |
| f | 2 | pθ   | 1   | 1    | 1    | 9971    |
| f | 2 | sk   | 37  | 37   | 36   | 242402  |
| f | 2 | sp   | 11  | 11   | 11   | 15988   |
| f | 2 | st   | 573 | 1405 | 1106 | 5436034 |
| f | 2 | ʃt   | 10  | 102  | 97   | 234380  |
| f | 2 | ts   | 16  | 1580 | 1342 | 2846858 |
| f | 2 | tθ   | 1   | 1    | 1    | 6455    |
| f | 2 | θs   | 2   | 64   | 49   | 26555   |
| f | 2 | θt   | 3   | 7    | 5    | 1248    |
| f | 2 | ʈt   | 8   | 84   | 80   | 187498  |
| f | 2 | vd   | 14  | 103  | 95   | 491189  |
| f | 2 | vz   | 15  | 211  | 189  | 470892  |
| f | 2 | zd   | 19  | 337  | 317  | 600571  |
| f | 2 | zd   | 3   | 31   | 31   | 25030   |
| f | 3 | dst  | 4   | 4    | 4    | 7629    |
| f | 3 | dθs  | 0   | 3    | 3    | 399     |
| f | 3 | fts  | 0   | 33   | 29   | 31761   |
| f | 3 | fθs  | 0   | 1    | 1    | 119     |
| f | 3 | kst  | 8   | 36   | 34   | 315961  |
| f | 3 | ksθ  | 1   | 1    | 1    | 8882    |
| f | 3 | kts  | 0   | 119  | 108  | 345448  |
| f | 3 | lbz  | 0   | 3    | 2    | 2918    |
| f | 3 | ldz  | 0   | 53   | 45   | 80807   |
| f | 3 | ldʒd | 0   | 4    | 4    | 1237    |
| f | 3 | lfs  | 0   | 4    | 4    | 265     |
| f | 3 | lft  | 1   | 3    | 3    | 1254    |
| f | 3 | lfθ  | 1   | 1    | 1    | 1043    |
| f | 3 | lks  | 0   | 10   | 9    | 977     |
| f | 3 | lkt  | 1   | 6    | 5    | 671     |
| f | 3 | lmd  | 0   | 3    | 3    | 6614    |

| pos | len | cl   | dictwtyp | parawtyp | corpwtyp | corpwtok |
|-----|-----|------|----------|----------|----------|----------|
| f   | 3   | lmz  | 0        | 6        | 6        | 20403    |
| f   | 3   | lnz  | 0        | 2        | 1        | 199      |
| f   | 3   | lps  | 0        | 7        | 7        | 23844    |
| f   | 3   | lpt  | 1        | 7        | 7        | 45693    |
| f   | 3   | lst  | 1        | 5        | 5        | 1727     |
| f   | 3   | lſt  | 0        | 1        | 1        | 12       |
| f   | 3   | lts  | 2        | 44       | 39       | 118035   |
| f   | 3   | lſt  | 0        | 4        | 4        | 492      |
| f   | 3   | lvd  | 2        | 13       | 13       | 86217    |
| f   | 3   | lvz  | 4        | 21       | 20       | 139690   |
| f   | 3   | mfs  | 0        | 3        | 3        | 1336     |
| f   | 3   | mft  | 0        | 2        | 2        | 496      |
| f   | 3   | mps  | 2        | 56       | 51       | 43805    |
| f   | 3   | mpt  | 6        | 45       | 41       | 71509    |
| f   | 3   | ndz  | 3        | 198      | 180      | 670435   |
| f   | 3   | ndzd | 1        | 7        | 7        | 63334    |
| f   | 3   | nst  | 6        | 60       | 59       | 377826   |
| f   | 3   | nts  | 3        | 517      | 468      | 1427232  |
| f   | 3   | nθs  | 0        | 20       | 11       | 106194   |
| f   | 3   | nſt  | 0        | 27       | 24       | 28592    |
| f   | 3   | nzd  | 0        | 3        | 2        | 663      |
| f   | 3   | ŋks  | 10       | 76       | 66       | 155078   |
| f   | 3   | ŋkt  | 9        | 45       | 40       | 54026    |
| f   | 3   | ŋkθ  | 4        | 4        | 4        | 46713    |
| f   | 3   | pst  | 0        | 7        | 7        | 8742     |
| f   | 3   | pts  | 0        | 27       | 25       | 31093    |
| f   | 3   | pθs  | 0        | 1        | 1        | 3609     |
| f   | 3   | sks  | 0        | 27       | 27       | 56950    |
| f   | 3   | skt  | 1        | 11       | 10       | 157180   |
| f   | 3   | sps  | 0        | 10       | 10       | 2934     |
| f   | 3   | spt  | 0        | 6        | 6        | 5700     |
| f   | 3   | sts  | 0        | 450      | 390      | 490310   |
| f   | 3   | tst  | 0        | 1        | 1        | 130      |
| f   | 3   | tθs  | 0        | 1        | 1        | 73       |
| f   | 4   | ksts | 0        | 3        | 3        | 13355    |
| f   | 4   | ksθs | 0        | 1        | 1        | 17       |
| f   | 4   | lſθs | 0        | 1        | 0        | 0        |
| f   | 4   | lkts | 0        | 1        | 0        | 0        |
| f   | 4   | lpts | 0        | 1        | 1        | 70       |

|   |   |      |   |   |   |       |
|---|---|------|---|---|---|-------|
| f | 4 | mpst | 0 | 1 | 1 | 1185  |
| f | 4 | mpts | 0 | 4 | 4 | 16100 |
| f | 4 | ηkst | 2 | 2 | 2 | 3324  |
| f | 4 | ηkts | 0 | 3 | 3 | 3992  |
| f | 4 | ηkθs | 0 | 3 | 3 | 9759  |



## Appendix G. English cluster examples

The following tables present representative examples of words featuring English initial, medial and final clusters in all positions. Wherever possible, a high frequency example word from the COCA corpus resource is shown. If a cluster was not present in the corpus resource (which is indicated by a '0' in the *corp* column), a random example from one of the other resources is given. The tables can also be used as a quick reference resource showing in which of the three resources words with particular clusters in a specific position were present.

### Initial clusters

| cl | len | dict | para | corp  | example   |
|----|-----|------|------|-------|-----------|
| bj | 2   | 1    | 1    | 1     | beautiful |
| bl | 2   | 1    | 1    | 1     | black     |
| br | 2   | 1    | 1    | 1     | bring     |
| dj | 2   | 1    | 1    | 1     | due       |
| dr | 2   | 1    | 1    | 1     | drug      |
| dw | 2   | 1    | 1    | 1     | dwarf     |
| fj | 2   | 1    | 1    | 1     | few       |
| fl | 2   | 1    | 1    | 1     | floor     |
| fr | 2   | 1    | 1    | 1     | from      |
| gl | 2   | 1    | 1    | 1     | glass     |
| gr | 2   | 1    | 1    | 1     | great     |
| gw | 2   | 1    | 1    | 1     | guava     |
| hj | 2   | 1    | 1    | 1     | human     |
| kj | 2   | 1    | 1    | 1     | curious   |
| kl | 2   | 1    | 1    | 1     | close     |
| kr | 2   | 1    | 1    | 1     | create    |
| 2  | 1   | 1    | 0    | kvass |           |
| kw | 2   | 1    | 1    | 1     | question  |
| lj | 2   | 1    | 1    | 1     | lure      |
| mj | 2   | 1    | 1    | 1     | music     |
| nj | 2   | 1    | 1    | 1     | new       |
| pj | 2   | 1    | 1    | 1     | pure      |
| pl | 2   | 1    | 1    | 1     | place     |
| pr | 2   | 1    | 1    | 1     | president |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| pw        | 2          | 1           | 1           | 1           | pueblo         |
| sf        | 2          | 1           | 1           | 1           | sphere         |
| sj        | 2          | 1           | 1           | 1           | sue            |
| sk        | 2          | 1           | 1           | 1           | school         |
| sl        | 2          | 1           | 1           | 1           | sleep          |
| sm        | 2          | 1           | 1           | 1           | small          |
| sn        | 2          | 1           | 1           | 1           | snow           |
| sp        | 2          | 1           | 1           | 1           | special        |
| st        | 2          | 1           | 1           | 1           | still          |
| sw        | 2          | 1           | 1           | 1           | sweet          |
| ʃm        | 2          | 1           | 1           | 1           | schmaltz       |
| ʃn        | 2          | 1           | 1           | 1           | schnapps       |
| ʃp        | 2          | 1           | 1           | 1           | spiel          |
| ʃr        | 2          | 1           | 1           | 1           | shrugged       |
| ʃw        | 2          | 1           | 1           | 0           | schwa          |
| tj        | 2          | 1           | 1           | 1           | tube           |
| tr        | 2          | 1           | 1           | 1           | trying         |
| tw        | 2          | 1           | 1           | 1           | twice          |
| θj        | 2          | 1           | 1           | 1           | thurable       |
| θr        | 2          | 1           | 1           | 1           | through        |
| θw        | 2          | 1           | 1           | 1           | thwart         |
| vj        | 2          | 1           | 1           | 1           | view           |
| vw        | 2          | 1           | 1           | 1           | voyeurism      |
| zl        | 2          | 1           | 1           | 1           | zloty          |
| skj       | 3          | 1           | 1           | 1           | skewers        |
| skl       | 3          | 1           | 1           | 1           | sclerosis      |
| skr       | 3          | 1           | 1           | 1           | screen         |
| skw       | 3          | 1           | 1           | 1           | square         |
| spj       | 3          | 1           | 1           | 1           | spewing        |
| spl       | 3          | 1           | 1           | 1           | split          |
| spr       | 3          | 1           | 1           | 1           | spring         |
| stj       | 3          | 1           | 1           | 1           | students       |
| str       | 3          | 1           | 1           | 1           | strong         |

**Medial clusters**

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| bd        | 2          | 1           | 1           | 1           | abdomen        |
| bdʒ       | 2          | 1           | 1           | 1           | subject        |
| bf        | 2          | 1           | 1           | 1           | clubfoot       |
| bg        | 2          | 1           | 1           | 1           | hobgoblin      |
| bh        | 2          | 1           | 1           | 1           | clubhouse      |
| bj        | 2          | 1           | 1           | 1           | abuse          |
| bk        | 2          | 1           | 1           | 1           | subcommittee   |
| bl        | 2          | 1           | 1           | 1           | public         |
| bm        | 2          | 1           | 1           | 1           | submitted      |
| bn        | 2          | 1           | 1           | 1           | abnormal       |
| br        | 2          | 1           | 1           | 1           | library        |
| bs        | 2          | 1           | 1           | 1           | absolutely     |
| bt        | 2          | 1           | 1           | 1           | obtained       |
| bθ        | 2          | 0           | 1           | 0           | tub-thumper    |
| bv        | 2          | 1           | 1           | 1           | obviously      |
| bw        | 2          | 1           | 1           | 1           | subway         |
| bz        | 2          | 1           | 1           | 1           | observed       |
| db        | 2          | 1           | 1           | 1           | feedback       |
| dd        | 2          | 1           | 1           | 1           | midday         |
| df        | 2          | 1           | 1           | 1           | dreadful       |
| dg        | 2          | 1           | 1           | 1           | headgear       |
| dh        | 2          | 1           | 1           | 1           | adherence      |
| dj        | 2          | 1           | 1           | 1           | education      |
| dk        | 2          | 1           | 1           | 1           | broadcast      |
| dl        | 2          | 1           | 1           | 1           | hardly         |
| dm        | 2          | 1           | 1           | 1           | administration |
| dn        | 2          | 1           | 1           | 1           | midnight       |
| dp        | 2          | 1           | 1           | 1           | woodpecker     |
| dr        | 2          | 1           | 1           | 1           | address        |
| ds        | 2          | 1           | 1           | 1           | medicine       |
| df        | 2          | 1           | 1           | 1           | hardship       |
| dt        | 2          | 1           | 1           | 1           | bedtime        |
| dθ        | 2          | 1           | 1           | 1           | bloodthirsty   |
| df        | 2          | 1           | 1           | 1           | godchild       |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b>  |
|-----------|------------|-------------|-------------|-------------|-----------------|
| dv        | 2          | 1           | 1           | 1           | advantage       |
| dw        | 2          | 1           | 1           | 1           | hardware        |
| dz        | 2          | 1           | 1           | 1           | odds-on         |
| ðh        | 2          | 1           | 1           | 1           | withheld        |
| ðl        | 2          | 1           | 1           | 1           | smoothly        |
| ðm        | 2          | 1           | 1           | 1           | rhythmic        |
| ðn        | 2          | 1           | 1           | 1           | smoothness      |
| ðr        | 2          | 1           | 1           | 1           | brethren        |
| ðs        | 2          | 1           | 1           | 1           | loathsome       |
| dʒh       | 2          | 1           | 1           | 1           | sledgehammer    |
| dʒk       | 2          | 1           | 1           | 1           | stagecoach      |
| dʒl       | 2          | 1           | 1           | 1           | largely         |
| dʒm       | 2          | 1           | 1           | 1           | management      |
| dʒn       | 2          | 1           | 1           | 1           | largeness       |
| dʒp       | 2          | 1           | 1           | 1           | ridgepole       |
| dʒr       | 2          | 1           | 1           | 1           | hedgerow        |
| dʒw       | 2          | 1           | 1           | 1           | passageway      |
| fb        | 2          | 1           | 1           | 1           | offbeat         |
| fd        | 2          | 1           | 1           | 1           | serfdom         |
| fg        | 2          | 1           | 1           | 1           | safeguard       |
| fh        | 2          | 1           | 1           | 1           | offhand         |
| fj        | 2          | 1           | 1           | 1           | refused         |
| fk        | 2          | 1           | 1           | 1           | briefcase       |
| fl        | 2          | 1           | 1           | 1           | reflect         |
| fm        | 2          | 1           | 1           | 1           | mischief-making |
| fn        | 2          | 1           | 1           | 1           | stiffness       |
| fp        | 2          | 1           | 1           | 1           | off-putting     |
| fr        | 2          | 1           | 1           | 1           | different       |
| fs        | 2          | 1           | 1           | 1           | offset          |
| ff        | 2          | 1           | 1           | 1           | offshore        |
| ft        | 2          | 1           | 1           | 1           | after           |
| fθ        | 2          | 1           | 1           | 1           | diphtheria      |
| fw        | 2          | 1           | 1           | 1           | halfway         |
| gb        | 2          | 1           | 1           | 1           | rugby           |
| gf        | 2          | 1           | 1           | 1           | dogfish         |
| gh        | 2          | 1           | 1           | 1           | doghouse        |

|     |   |   |   |   |              |
|-----|---|---|---|---|--------------|
| gj  | 2 | 1 | 1 | 1 | regular      |
| gl  | 2 | 1 | 1 | 1 | ugly         |
| gm  | 2 | 1 | 1 | 1 | segment      |
| gn  | 2 | 1 | 1 | 1 | significant  |
| gp  | 2 | 1 | 1 | 1 | flagpole     |
| gr  | 2 | 1 | 1 | 1 | program      |
| gs  | 2 | 1 | 1 | 1 | jigsaw       |
| gf  | 2 | 1 | 1 | 1 | flagship     |
| gt  | 2 | 1 | 1 | 1 | ragtime      |
| gw  | 2 | 1 | 1 | 1 | dogwood      |
| gz  | 2 | 1 | 1 | 1 | example      |
| gʒ  | 2 | 1 | 1 | 1 | luxurious    |
| hj  | 2 | 1 | 1 | 1 | superhuman   |
| kb  | 2 | 1 | 1 | 1 | blockbuster  |
| kd  | 2 | 1 | 1 | 1 | breakdown    |
| kɔʒ | 2 | 1 | 1 | 1 | lockjaw      |
| kf  | 2 | 1 | 1 | 1 | breakfast    |
| kg  | 2 | 1 | 1 | 1 | backgammon   |
| kh  | 2 | 1 | 1 | 1 | stockholders |
| kj  | 2 | 1 | 1 | 1 | security     |
| kk  | 2 | 1 | 1 | 1 | bookcase     |
| kl  | 2 | 1 | 1 | 1 | likely       |
| km  | 2 | 1 | 1 | 1 | blackmail    |
| kn  | 2 | 1 | 1 | 1 | technology   |
| kp  | 2 | 1 | 1 | 1 | cockpit      |
| kr  | 2 | 1 | 1 | 1 | across       |
| ks  | 2 | 1 | 1 | 1 | success      |
| kʃ  | 2 | 1 | 1 | 1 | action       |
| kt  | 2 | 1 | 1 | 1 | director     |
| kθ  | 2 | 1 | 1 | 1 | blackthorn   |
| kʧ  | 2 | 1 | 1 | 1 | actually     |
| kw  | 2 | 1 | 1 | 1 | required     |
| lb  | 2 | 1 | 1 | 1 | album        |
| ld  | 2 | 1 | 1 | 1 | building     |
| ldʒ | 2 | 1 | 1 | 1 | soldiers     |
| lf  | 2 | 1 | 1 | 1 | welfare      |
| lg  | 2 | 1 | 1 | 1 | vulgar       |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| lh        | 2          | 1           | 1           | 1           | schoolhouse    |
| lj        | 2          | 1           | 1           | 1           | million        |
| lk        | 2          | 1           | 1           | 1           | welcome        |
| ll        | 2          | 1           | 1           | 1           | solely         |
| lm        | 2          | 1           | 1           | 1           | almost         |
| ln        | 2          | 1           | 1           | 1           | illness        |
| lp        | 2          | 1           | 1           | 1           | helping        |
| lr        | 2          | 1           | 1           | 1           | jewelry        |
| ls        | 2          | 1           | 1           | 1           | also           |
| lf        | 2          | 1           | 1           | 1           | bullshit       |
| lt        | 2          | 1           | 1           | 1           | alternative    |
| lθ        | 2          | 1           | 1           | 1           | healthy        |
| lf        | 2          | 1           | 1           | 1           | culture        |
| lv        | 2          | 1           | 1           | 1           | silver         |
| lw        | 2          | 1           | 1           | 1           | railway        |
| lz        | 2          | 1           | 1           | 1           | palsy          |
| mb        | 2          | 1           | 1           | 1           | members        |
| md        | 2          | 1           | 1           | 1           | someday        |
| mdʒ       | 2          | 1           | 1           | 1           | ramjet         |
| mf        | 2          | 1           | 1           | 1           | emphasis       |
| mh        | 2          | 1           | 1           | 1           | somehow        |
| mj        | 2          | 1           | 1           | 1           | community      |
| mk        | 2          | 1           | 1           | 1           | homecoming     |
| ml        | 2          | 1           | 1           | 1           | extremely      |
| mm        | 2          | 1           | 1           | 1           | custom-made    |
| mn        | 2          | 1           | 1           | 1           | columnist      |
| mp        | 2          | 1           | 1           | 1           | important      |
| mr        | 2          | 1           | 1           | 1           | comrades       |
| ms        | 2          | 1           | 1           | 1           | themselves     |
| mʃ        | 2          | 1           | 1           | 1           | bombshell      |
| mt        | 2          | 1           | 1           | 1           | sometimes      |
| mθ        | 2          | 1           | 1           | 1           | something      |
| mʃ        | 2          | 1           | 1           | 1           | armchair       |
| mv        | 2          | 1           | 1           | 1           | circumvent     |
| mw        | 2          | 1           | 1           | 1           | someone        |
| mz        | 2          | 1           | 1           | 1           | crimson        |

|     |   |   |   |   |                  |
|-----|---|---|---|---|------------------|
| nb  | 2 | 1 | 1 | 1 | unbelievable     |
| nd  | 2 | 1 | 1 | 1 | under            |
| nð  | 2 | 1 | 1 | 1 | dyed-in-the-wool |
| ndʒ | 2 | 1 | 1 | 1 | changes          |
| nf  | 2 | 1 | 1 | 1 | information      |
| ng  | 2 | 1 | 1 | 1 | engaged          |
| nh  | 2 | 1 | 1 | 1 | manhattan        |
| nj  | 2 | 1 | 1 | 1 | continue         |
| nk  | 2 | 1 | 1 | 1 | income           |
| nl  | 2 | 1 | 1 | 1 | only             |
| nm  | 2 | 1 | 1 | 1 | government       |
| nn  | 2 | 1 | 1 | 1 | unknown          |
| np  | 2 | 1 | 1 | 1 | input            |
| nr  | 2 | 1 | 1 | 1 | genre            |
| ns  | 2 | 1 | 1 | 1 | inside           |
| nf  | 2 | 1 | 1 | 1 | attention        |
| nt  | 2 | 1 | 1 | 1 | into             |
| nθ  | 2 | 1 | 1 | 1 | synthetic        |
| nf  | 2 | 1 | 1 | 1 | century          |
| nv  | 2 | 1 | 1 | 1 | involved         |
| nw  | 2 | 1 | 1 | 1 | meanwhile        |
| nz  | 2 | 1 | 1 | 1 | transition       |
| nb  | 2 | 1 | 1 | 1 | springboard      |
| nd  | 2 | 1 | 1 | 1 | kingdom          |
| nf  | 2 | 1 | 1 | 1 | meaningful       |
| ng  | 2 | 1 | 1 | 1 | longer           |
| nh  | 2 | 1 | 1 | 1 | shanghai         |
| nk  | 2 | 1 | 1 | 1 | thinking         |
| nl  | 2 | 1 | 1 | 1 | increasingly     |
| nm  | 2 | 1 | 1 | 1 | hangman          |
| nn  | 2 | 1 | 1 | 1 | willingness      |
| np  | 2 | 1 | 1 | 1 | kingpin          |
| nr  | 2 | 1 | 1 | 1 | long-range       |
| ns  | 2 | 1 | 1 | 1 | alongside        |
| nf  | 2 | 1 | 1 | 1 | kingship         |
| nt  | 2 | 1 | 1 | 1 | long-term        |
| nw  | 2 | 1 | 1 | 1 | gangway          |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| ηz        | 2          | 1           | 1           | 1           | anxiety        |
| pb        | 2          | 1           | 1           | 1           | scrapbook      |
| pd        | 2          | 1           | 1           | 1           | update         |
| pdʒ       | 2          | 1           | 1           | 1           | flapjacks      |
| pf        | 2          | 1           | 1           | 1           | hopefully      |
| pg        | 2          | 1           | 1           | 1           | scapegoat      |
| ph        | 2          | 1           | 1           | 1           | uphill         |
| pj        | 2          | 1           | 1           | 1           | population     |
| pk        | 2          | 1           | 1           | 1           | popcorn        |
| pl        | 2          | 1           | 1           | 1           | supply         |
| pm        | 2          | 1           | 1           | 1           | development    |
| pn        | 2          | 1           | 1           | 1           | hypnosis       |
| pp        | 2          | 1           | 1           | 1           | stepparents    |
| pr        | 2          | 1           | 1           | 1           | approach       |
| ps        | 2          | 1           | 1           | 1           | upset          |
| pf        | 2          | 1           | 1           | 1           | options        |
| pt        | 2          | 1           | 1           | 1           | accepted       |
| ptʃ       | 2          | 1           | 1           | 1           | captured       |
| pʋ        | 2          | 1           | 1           | 1           | grapevine      |
| pw        | 2          | 1           | 1           | 1           | upward         |
| sb        | 2          | 1           | 1           | 1           | baseball       |
| sd        | 2          | 1           | 1           | 1           | jurisdiction   |
| sdʒ       | 2          | 1           | 1           | 1           | disjointed     |
| sf        | 2          | 1           | 1           | 1           | successful     |
| sg        | 2          | 1           | 1           | 1           | disgust        |
| sh        | 2          | 1           | 1           | 1           | household      |
| sj        | 2          | 1           | 1           | 1           | assume         |
| sk        | 2          | 1           | 1           | 1           | asking         |
| sl        | 2          | 1           | 1           | 1           | obviously      |
| sm        | 2          | 1           | 1           | 1           | assessment     |
| sn        | 2          | 1           | 1           | 1           | consciousness  |
| sp        | 2          | 1           | 1           | 1           | especially     |
| sr        | 2          | 1           | 1           | 1           | classroom      |
| ss        | 2          | 1           | 1           | 1           | cross-section  |
| sf        | 2          | 1           | 1           | 1           | spaceship      |
| st        | 2          | 1           | 1           | 1           | system         |

|     |   |   |   |   |               |
|-----|---|---|---|---|---------------|
| sθ  | 2 | 1 | 1 | 1 | aesthetic     |
| stʃ | 2 | 1 | 1 | 1 | question      |
| sw  | 2 | 1 | 1 | 1 | persuade      |
| ʃb  | 2 | 1 | 1 | 1 | dashboard     |
| ʃd  | 2 | 1 | 1 | 1 | splashdown    |
| ʃf  | 2 | 1 | 1 | 1 | wishful       |
| ʃg  | 2 | 1 | 1 | 1 | flashgun      |
| ʃh  | 2 | 1 | 1 | 1 | threshold     |
| ʃk  | 2 | 1 | 1 | 1 | pushcart      |
| ʃl  | 2 | 1 | 1 | 1 | freshly       |
| ʃm  | 2 | 1 | 1 | 1 | establishment |
| ʃn  | 2 | 1 | 1 | 1 | freshness     |
| ʃp  | 2 | 1 | 1 | 1 | fleshpots     |
| ʃr  | 2 | 1 | 1 | 1 | mushrooms     |
| ʃt  | 2 | 1 | 1 | 1 | washtub       |
| ʃtʃ | 2 | 1 | 1 | 1 | pushchair     |
| ʃw  | 2 | 1 | 1 | 1 | freshwater    |
| tb  | 2 | 1 | 1 | 1 | football      |
| td  | 2 | 1 | 1 | 1 | outdoor       |
| tdʒ | 2 | 1 | 1 | 1 | straitjacket  |
| tf  | 2 | 1 | 1 | 1 | platform      |
| tg  | 2 | 1 | 1 | 1 | shotgun       |
| th  | 2 | 1 | 1 | 1 | apartheid     |
| tj  | 2 | 1 | 1 | 1 | opportunity   |
| tk  | 2 | 1 | 1 | 1 | outcome       |
| tl  | 2 | 1 | 1 | 1 | absolutely    |
| tm  | 2 | 1 | 1 | 1 | treatment     |
| tn  | 2 | 1 | 1 | 1 | partner       |
| tp  | 2 | 1 | 1 | 1 | output        |
| tr  | 2 | 1 | 1 | 1 | literature    |
| ts  | 2 | 1 | 1 | 1 | outside       |
| tʃ  | 2 | 1 | 1 | 1 | outshine      |
| tt  | 2 | 1 | 1 | 1 | short-term    |
| tθ  | 2 | 1 | 1 | 0 | whitethorn    |
| tʃʃ | 2 | 1 | 1 | 1 | short-changed |
| tv  | 2 | 1 | 1 | 1 | outvoted      |
| tw  | 2 | 1 | 1 | 1 | between       |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| θb        | 2          | 1           | 1           | 1           | deathbed       |
| θd        | 2          | 1           | 1           | 1           | birthday       |
| θf        | 2          | 1           | 1           | 1           | faithful       |
| θk        | 2          | 1           | 1           | 1           | forthcoming    |
| θl        | 2          | 1           | 1           | 1           | catholic       |
| θm        | 2          | 1           | 1           | 1           | arithmetic     |
| θn        | 2          | 1           | 1           | 1           | ethnic         |
| θp        | 2          | 1           | 1           | 1           | toothpaste     |
| θr        | 2          | 1           | 1           | 1           | bathroom       |
| θs        | 2          | 1           | 1           | 1           | soothsayer     |
| θt        | 2          | 1           | 1           | 1           | bath tub       |
| θw        | 2          | 1           | 1           | 1           | northwest      |
| ʃb        | 2          | 1           | 1           | 1           | archbishop     |
| ʃd        | 2          | 1           | 1           | 1           | touchdown      |
| ʃf        | 2          | 1           | 1           | 1           | watchful       |
| ʃg        | 2          | 1           | 1           | 1           | churchgoers    |
| ʃh        | 2          | 1           | 1           | 1           | hitchhiking    |
| ʃj        | 2          | 1           | 1           | 1           | churchyard     |
| ʃk        | 2          | 1           | 1           | 1           | latchkey       |
| ʃl        | 2          | 1           | 1           | 1           | richly         |
| ʃm        | 2          | 1           | 1           | 1           | impeachment    |
| ʃn        | 2          | 1           | 1           | 1           | richness       |
| ʃp        | 2          | 1           | 1           | 0           | catchpenny     |
| ʃs        | 2          | 1           | 1           | 1           | vouchsafed     |
| ʃt        | 2          | 1           | 1           | 1           | watchtower     |
| ʃw        | 2          | 1           | 1           | 1           | patchwork      |
| vb        | 2          | 1           | 1           | 1           | lovebirds      |
| vd        | 2          | 1           | 1           | 1           | out-of-date    |
| vð        | 2          | 1           | 1           | 1           | out-of-the-way |
| vf        | 2          | 1           | 1           | 1           | matter-of-fact |
| vh        | 2          | 1           | 1           | 0           | yo-heave-ho    |
| vj        | 2          | 1           | 1           | 1           | interview      |
| vk        | 2          | 1           | 1           | 1           | dovecote       |
| vl        | 2          | 1           | 1           | 1           | relatively     |
| vm        | 2          | 1           | 1           | 1           | movement       |
| vn        | 2          | 1           | 1           | 1           | effectiveness  |

|     |   |   |   |   |                 |
|-----|---|---|---|---|-----------------|
| vp  | 2 | 1 | 1 | 1 | mother-of-pearl |
| vr  | 2 | 1 | 1 | 1 | every           |
| vs  | 2 | 1 | 1 | 1 | lovesick        |
| vt  | 2 | 1 | 1 | 1 | dovetail        |
| vw  | 2 | 1 | 1 | 1 | driveway        |
| zb  | 2 | 1 | 1 | 1 | husband         |
| zd  | 2 | 1 | 1 | 1 | wisdom          |
| zdʒ | 2 | 1 | 1 | 1 | phosgene        |
| zf  | 2 | 1 | 1 | 1 | reposeful       |
| zg  | 2 | 1 | 1 | 1 | nosegay         |
| zj  | 2 | 1 | 1 | 1 | presumably      |
| zk  | 2 | 1 | 1 | 1 | newscast        |
| zl  | 2 | 1 | 1 | 1 | newsletter      |
| zm  | 2 | 1 | 1 | 1 | cosmic          |
| zn  | 2 | 1 | 1 | 1 | business        |
| zp  | 2 | 1 | 1 | 1 | newspaper       |
| zr  | 2 | 1 | 1 | 1 | newsroom        |
| zʃ  | 2 | 1 | 1 | 1 | newsheets       |
| zv  | 2 | 1 | 1 | 1 | newsvendor      |
| zw  | 2 | 1 | 1 | 1 | newsworthy      |
| ʒf  | 2 | 1 | 1 | 1 | vengeful        |
| ʒl  | 2 | 1 | 1 | 1 | strangely       |
| ʒm  | 2 | 1 | 1 | 1 | arrangements    |
| ʒn  | 2 | 1 | 1 | 1 | strangeness     |
| ʒw  | 2 | 1 | 1 | 1 | bourgeois       |
| bdj | 3 | 1 | 1 | 1 | subdued         |
| bgr | 3 | 1 | 1 | 1 | subgroups       |
| bhj | 3 | 1 | 1 | 1 | subhuman        |
| bkj | 3 | 1 | 1 | 1 | subcutaneous    |
| bkr | 3 | 1 | 1 | 1 | pub-crawling    |
| bsj | 3 | 1 | 1 | 1 | subsumed        |
| bsk | 3 | 1 | 1 | 1 | absconded       |
| bsl | 3 | 1 | 1 | 1 | bobsled         |
| bst | 3 | 1 | 1 | 1 | substantial     |
| btj | 3 | 1 | 1 | 1 | obtuse          |
| btr | 3 | 1 | 1 | 1 | subtract        |
| bzm | 3 | 1 | 1 | 1 | tribesmen       |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b>  |
|-----------|------------|-------------|-------------|-------------|-----------------|
| dbl       | 3          | 1           | 1           | 1           | roadblocks      |
| dbr       | 3          | 1           | 1           | 1           | record-breaking |
| ddr       | 3          | 1           | 1           | 1           | headdress       |
| dfi       | 3          | 1           | 1           | 1           | gadfly          |
| dfr       | 3          | 1           | 1           | 1           | breadfruit      |
| dgr       | 3          | 1           | 1           | 1           | bridegroom      |
| dhj       | 3          | 1           | 1           | 1           | good-humoured   |
| dki       | 3          | 1           | 1           | 1           | bedclothes      |
| dkr       | 3          | 1           | 1           | 1           | breadcrumbs     |
| dkw       | 3          | 0           | 1           | 1           | headquarters    |
| dpl       | 3          | 1           | 1           | 1           | swordplay       |
| dsk       | 3          | 1           | 1           | 1           | redskins        |
| dsl       | 3          | 1           | 1           | 1           | mudslingers     |
| dsp       | 3          | 1           | 1           | 1           | loudspeaker     |
| dst       | 3          | 1           | 1           | 1           | foodstuffs      |
| dtr       | 3          | 1           | 1           | 1           | sidetracked     |
| dθw       | 3          | 1           | 1           | 0           | breadthways     |
| dvj       | 3          | 1           | 1           | 1           | side-view       |
| dzm       | 3          | 1           | 1           | 1           | ombudsman       |
| dzp       | 3          | 1           | 1           | 1           | tradespeople    |
| ðdr       | 3          | 1           | 1           | 1           | withdrawal      |
| ðst       | 3          | 1           | 1           | 1           | notwithstanding |
| ðzh       | 3          | 1           | 1           | 1           | clotheshorse    |
| ðzl       | 3          | 1           | 1           | 1           | clothesline     |
| dʒkr      | 3          | 1           | 1           | 1           | stagecraft      |
| dʒsk      | 3          | 1           | 1           | 1           | large-scale     |
| fbl       | 3          | 1           | 1           | 1           | lifeblood       |
| fbr       | 3          | 1           | 1           | 1           | half-brother    |
| fdr       | 3          | 1           | 1           | 0           | rough-dry       |
| fhj       | 3          | 1           | 1           | 1           | rough-hewn      |
| fkr       | 3          | 1           | 1           | 1           | half-crazed     |
| fpr       | 3          | 1           | 1           | 1           | half-price      |
| fst       | 3          | 1           | 1           | 1           | offstage        |
| fth       | 3          | 1           | 1           | 1           | left-handed     |
| ftl       | 3          | 1           | 1           | 1           | softly          |
| ftm       | 3          | 1           | 1           | 0           | aircraftman     |

|      |   |   |   |   |               |
|------|---|---|---|---|---------------|
| ftn  | 3 | 1 | 1 | 1 | softness      |
| fr   | 3 | 1 | 1 | 1 | half-truths   |
| ftw  | 3 | 1 | 1 | 1 | left-wing     |
| gfl  | 3 | 1 | 1 | 1 | stagflation   |
| gpl  | 3 | 1 | 1 | 1 | eggplant      |
| gsk  | 3 | 1 | 1 | 1 | pigskin       |
| gst  | 3 | 1 | 1 | 1 | drugstore     |
| gsw  | 3 | 1 | 1 | 0 | pigswill      |
| gtr  | 3 | 1 | 1 | 1 | dogtrot       |
| gzb  | 3 | 1 | 1 | 1 | dogsbody      |
| gzh  | 3 | 1 | 1 | 1 | hogshead      |
| gzj  | 3 | 1 | 1 | 1 | exuberant     |
| kbr  | 3 | 1 | 1 | 1 | stockbroker   |
| kdr  | 3 | 1 | 1 | 1 | backdrop      |
| kfr  | 3 | 1 | 1 | 1 | quick-frozen  |
| kgr  | 3 | 1 | 1 | 1 | background    |
| kk   | 3 | 1 | 1 | 1 | rock-climbing |
| kk   | 3 | 1 | 1 | 1 | rock-crystal  |
| knj  | 3 | 1 | 1 | 1 | ichneumon     |
| ksf  | 3 | 1 | 1 | 1 | sixfold       |
| ksh  | 3 | 1 | 1 | 1 | exhale        |
| ksk  | 3 | 1 | 1 | 1 | excavation    |
| ksl  | 3 | 1 | 1 | 1 | backsliding   |
| ksm  | 3 | 1 | 1 | 1 | spokesman     |
| ksp  | 3 | 1 | 1 | 1 | experience    |
| ksf  | 3 | 1 | 1 | 1 | six-shooter   |
| kst  | 3 | 1 | 1 | 1 | extent        |
| kstf | 3 | 1 | 1 | 1 | exchange      |
| ksw  | 3 | 1 | 1 | 1 | boxwood       |
| ktf  | 3 | 1 | 1 | 1 | disrespectful |
| ktl  | 3 | 1 | 1 | 1 | directly      |
| ktm  | 3 | 1 | 1 | 1 | enactment     |
| ktn  | 3 | 1 | 1 | 1 | correctness   |
| ktr  | 3 | 1 | 1 | 1 | electronic    |
| kθr  | 3 | 1 | 1 | 1 | breakthrough  |
| lby  | 3 | 1 | 1 | 1 | albumen       |
| lbl  | 3 | 1 | 1 | 1 | full-blown    |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b>    |
|-----------|------------|-------------|-------------|-------------|-------------------|
| lbr       | 3          | 1           | 1           | 1           | double-breasted   |
| ldb       | 3          | 1           | 1           | 1           | childbirth        |
| ldf       | 3          | 1           | 1           | 0           | uncalled-for      |
| ldh       | 3          | 1           | 1           | 1           | childhood         |
| ldj       | 3          | 1           | 1           | 1           | mildew            |
| ldl       | 3          | 1           | 1           | 1           | boldly            |
| ldn       | 3          | 1           | 1           | 1           | boldness          |
| ldr       | 3          | 1           | 1           | 1           | children          |
| ldf       | 3          | 1           | 1           | 1           | cold-shouldered   |
| ldw       | 3          | 1           | 1           | 1           | fieldwork         |
| lfd       | 3          | 1           | 1           | 1           | self-etermination |
| lfh       | 3          | 1           | 1           | 1           | self-help         |
| lfj       | 3          | 1           | 1           | 1           | sulphuric         |
| lfk       | 3          | 1           | 1           | 1           | self-confidence   |
| lfl       | 3          | 1           | 1           | 1           | full-fledged      |
| lfm       | 3          | 1           | 1           | 1           | self-made         |
| lfp       | 3          | 1           | 1           | 1           | self-pity         |
| lfr       | 3          | 1           | 1           | 1           | girlfriend        |
| lfs       | 3          | 1           | 1           | 1           | self-sufficient   |
| lfw       | 3          | 1           | 1           | 1           | self-willed       |
| lgr       | 3          | 1           | 1           | 1           | pilgrimage        |
| lkh       | 3          | 1           | 1           | 1           | bulkhead          |
| lkj       | 3          | 1           | 1           | 1           | calculated        |
| lkl       | 3          | 1           | 1           | 1           | middle-class      |
| lkm       | 3          | 1           | 1           | 1           | milkman           |
| lkr       | 3          | 1           | 1           | 1           | fulcrum           |
| lks       | 3          | 1           | 1           | 1           | milksop           |
| lkt       | 3          | 0           | 1           | 0           | mulcted           |
| lkw       | 3          | 1           | 1           | 1           | milkweed          |
| lnj       | 3          | 1           | 1           | 1           | malnutrition      |
| lpf       | 3          | 1           | 1           | 1           | helpful           |
| lpl       | 3          | 1           | 1           | 1           | helpless          |
| lpm       | 3          | 1           | 1           | 1           | helpmate          |
| lpr       | 3          | 1           | 1           | 1           | malpractice       |
| lpt       | 3          | 1           | 1           | 1           | sculptor          |
| lptf      | 3          | 1           | 1           | 1           | sculpture         |

|     |   |   |   |   |             |
|-----|---|---|---|---|-------------|
| lsh | 3 | 1 | 1 | 1 | falsehood   |
| lsk | 3 | 1 | 1 | 1 | full-scale  |
| lsl | 3 | 1 | 1 | 1 | falsely     |
| lsn | 3 | 1 | 1 | 1 | rattlesnake |
| lsp | 3 | 1 | 1 | 1 | tablespoons |
| lst | 3 | 1 | 1 | 1 | bolster     |
| lsw | 3 | 1 | 1 | 1 | elsewhere   |
| ltf | 3 | 1 | 1 | 1 | volte-face  |
| lth | 3 | 1 | 1 | 1 | adulthood   |
| ltl | 3 | 1 | 1 | 1 | faultless   |
| ltp | 3 | 1 | 1 | 0 | saltpan     |
| ltr | 3 | 1 | 1 | 1 | poultry     |
| lts | 3 | 1 | 1 | 1 | schmaltzy   |
| ltw | 3 | 1 | 1 | 1 | saltwater   |
| lθf | 3 | 1 | 1 | 1 | healthful   |
| lvj | 3 | 1 | 1 | 1 | valvular    |
| lvm | 3 | 1 | 1 | 1 | involvement |
| lzm | 3 | 1 | 1 | 1 | salesman    |
| lzw | 3 | 1 | 1 | 1 | saleswoman  |
| mbj | 3 | 1 | 1 | 1 | ambulance   |
| mbl | 3 | 1 | 1 | 1 | assembly    |
| mbr | 3 | 1 | 1 | 1 | embrace     |
| mdr | 3 | 1 | 1 | 1 | humdrum     |
| mfl | 3 | 1 | 1 | 1 | circumflex  |
| mfr | 3 | 1 | 1 | 1 | comfrey     |
| mft | 3 | 1 | 1 | 1 | comfortable |
| mkr | 3 | 1 | 1 | 1 | gimcrack    |
| mpb | 3 | 1 | 1 | 1 | humpback    |
| mpj | 3 | 1 | 1 | 1 | computer    |
| mpl | 3 | 1 | 1 | 1 | simply      |
| mpm | 3 | 1 | 1 | 1 | encampment  |
| mpn | 3 | 1 | 1 | 1 | dampness    |
| mpr | 3 | 1 | 1 | 1 | improve     |
| mps | 3 | 1 | 1 | 1 | glimpses    |
| mpf | 3 | 1 | 1 | 1 | consumption |
| mpt | 3 | 1 | 1 | 1 | empty       |
| msk | 3 | 1 | 1 | 1 | lambskin    |

| <b>cl</b>        | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|------------------|------------|-------------|-------------|-------------|----------------|
| m <sub>sl</sub>  | 3          | 1           | 1           | 0           | gymslip        |
| m <sub>sp</sub>  | 3          | 1           | 1           | 1           | homespun       |
| m <sub>st</sub>  | 3          | 1           | 1           | 1           | circumstances  |
| m <sub>zh</sub>  | 3          | 1           | 1           | 1           | alms-house     |
| m <sub>zm</sub>  | 3          | 1           | 1           | 1           | gamesmanship   |
| n <sub>bl</sub>  | 3          | 1           | 1           | 1           | unblocked      |
| n <sub>br</sub>  | 3          | 1           | 1           | 1           | unbroken       |
| n <sub>dh</sub>  | 3          | 1           | 1           | 1           | roundhouse     |
| n <sub>dj</sub>  | 3          | 1           | 1           | 1           | enduring       |
| n <sub>dl</sub>  | 3          | 1           | 1           | 1           | handling       |
| n <sub>dm</sub>  | 3          | 1           | 1           | 1           | commandments   |
| n <sub>dn</sub>  | 3          | 1           | 1           | 1           | fondness       |
| n <sub>dr</sub>  | 3          | 1           | 1           | 1           | hundred        |
| n <sub>dw</sub>  | 3          | 1           | 1           | 1           | bandwagon      |
| n <sub>dʒl</sub> | 3          | 1           | 1           | 1           | changeling     |
| n <sub>fj</sub>  | 3          | 1           | 1           | 1           | confused       |
| n <sub>fl</sub>  | 3          | 1           | 1           | 1           | influence      |
| n <sub>fr</sub>  | 3          | 1           | 1           | 1           | infrastructure |
| n <sub>gl</sub>  | 3          | 1           | 1           | 1           | sunglasses     |
| n <sub>gr</sub>  | 3          | 1           | 1           | 1           | congressional  |
| n <sub>hj</sub>  | 3          | 1           | 1           | 1           | inhuman        |
| n <sub>kj</sub>  | 3          | 1           | 1           | 1           | incurable      |
| n <sub>kl</sub>  | 3          | 1           | 1           | 1           | including      |
| n <sub>kr</sub>  | 3          | 1           | 1           | 1           | increased      |
| n <sub>kw</sub>  | 3          | 1           | 1           | 1           | inquiry        |
| n <sub>pl</sub>  | 3          | 1           | 1           | 1           | unpleasant     |
| n <sub>pr</sub>  | 3          | 1           | 1           | 1           | unprecedented  |
| n <sub>pw</sub>  | 3          | 1           | 1           | 0           | embonpoint     |
| n <sub>sd</sub>  | 3          | 1           | 1           | 1           | princedom      |
| n <sub>sf</sub>  | 3          | 1           | 1           | 1           | transfer       |
| n <sub>sj</sub>  | 3          | 1           | 1           | 1           | consumer       |
| n <sub>sk</sub>  | 3          | 1           | 1           | 1           | unskilled      |
| n <sub>sl</sub>  | 3          | 1           | 1           | 1           | translated     |
| n <sub>sm</sub>  | 3          | 1           | 1           | 1           | announcement   |
| n <sub>sn</sub>  | 3          | 1           | 1           | 1           | ensnared       |
| n <sub>sp</sub>  | 3          | 1           | 1           | 1           | transportation |

|      |   |   |   |   |                  |
|------|---|---|---|---|------------------|
| nsf  | 3 | 1 | 1 | 1 | transshipment    |
| nst  | 3 | 1 | 1 | 1 | instead          |
| nsw  | 3 | 1 | 1 | 1 | unsweetened      |
| nfr  | 3 | 1 | 1 | 1 | enshrined        |
| ntb  | 3 | 1 | 1 | 1 | paintbox         |
| ntd  | 3 | 1 | 1 | 1 | present-day      |
| ntf  | 3 | 1 | 1 | 1 | resentful        |
| nth  | 3 | 1 | 1 | 1 | penthouse        |
| ntj  | 3 | 1 | 1 | 1 | intuitive        |
| ntl  | 3 | 1 | 1 | 1 | recently         |
| ntm  | 3 | 1 | 1 | 1 | appointment      |
| ntn  | 3 | 1 | 1 | 1 | unpleasantness   |
| ntr  | 3 | 1 | 1 | 1 | country          |
| nts  | 3 | 1 | 1 | 1 | intelligentsia   |
| ntw  | 3 | 1 | 1 | 1 | entwined         |
| nθj  | 3 | 1 | 1 | 1 | enthusiasm       |
| nθl  | 3 | 1 | 1 | 1 | monthly          |
| nθr  | 3 | 1 | 1 | 1 | anthropology     |
| ntfb | 3 | 1 | 1 | 1 | hunchback        |
| ntfl | 3 | 1 | 1 | 1 | staunchly        |
| ntfm | 3 | 1 | 1 | 1 | retrenchment     |
| ntfn | 3 | 1 | 1 | 1 | staunchness      |
| ntfp | 3 | 1 | 1 | 1 | linchpin         |
| ntft | 3 | 1 | 1 | 1 | lunchtime        |
| nzf  | 3 | 1 | 1 | 1 | townsfolk        |
| nzk  | 3 | 1 | 1 | 1 | transcontinental |
| nzm  | 3 | 1 | 1 | 1 | transmission     |
| nzp  | 3 | 1 | 1 | 1 | townspeople      |
| nzt  | 3 | 1 | 1 | 1 | painstakingly    |
| nzv  | 3 | 1 | 1 | 1 | transverse       |
| nzw  | 3 | 1 | 1 | 1 | kinswoman        |
| ndr  | 3 | 1 | 1 | 1 | long-drawn-out   |
| ngj  | 3 | 1 | 1 | 1 | singular         |
| ngl  | 3 | 1 | 1 | 1 | anglers          |
| ngr  | 3 | 1 | 1 | 1 | congress         |
| ngw  | 3 | 1 | 1 | 1 | language         |
| nkf  | 3 | 1 | 1 | 1 | thankful         |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b>   |
|-----------|------------|-------------|-------------|-------------|------------------|
| ηkj       | 3          | 1           | 1           | 1           | incubator        |
| ηkl       | 3          | 1           | 1           | 1           | frankly          |
| ηkm       | 3          | 1           | 1           | 1           | embankment       |
| ηkn       | 3          | 1           | 1           | 1           | frankness        |
| ηkr       | 3          | 1           | 1           | 1           | increase         |
| ηks       | 3          | 1           | 1           | 1           | sphinxes         |
| ηkj       | 3          | 1           | 1           | 1           | conjunction      |
| ηkt       | 3          | 1           | 1           | 1           | distinctive      |
| ηkθ       | 3          | 1           | 1           | 1           | strengthen       |
| ηktf      | 3          | 1           | 1           | 1           | sanctuary        |
| ηkw       | 3          | 1           | 1           | 1           | conquest         |
| ηpl       | 3          | 1           | 1           | 1           | gangplank        |
| ηsp       | 3          | 1           | 1           | 1           | wingspan         |
| ηst       | 3          | 1           | 1           | 1           | stepping-stone   |
| pbr       | 3          | 1           | 1           | 1           | upbringing       |
| pdr       | 3          | 1           | 1           | 1           | snapdragons      |
| pfl       | 3          | 1           | 1           | 1           | top-flight       |
| pfr       | 3          | 1           | 1           | 1           | grapefruit       |
| pgr       | 3          | 1           | 1           | 1           | upgrade          |
| pkl       | 3          | 1           | 1           | 1           | pipeclay         |
| psj       | 3          | 1           | 1           | 1           | capsule          |
| psk       | 3          | 1           | 1           | 1           | sheepskin        |
| pst       | 3          | 1           | 1           | 1           | upstairs         |
| ptf       | 3          | 1           | 1           | 0           | unhoped-for      |
| ptj       | 3          | 1           | 1           | 1           | septuagenarian   |
| ptl       | 3          | 1           | 1           | 1           | abruptly         |
| ptn       | 3          | 1           | 1           | 1           | abruptness       |
| ptr       | 3          | 1           | 1           | 1           | scriptwriter     |
| pts       | 3          | 1           | 1           | 1           | bankruptcy       |
| sbr       | 3          | 1           | 1           | 1           | icebreaker       |
| sfl       | 3          | 1           | 1           | 1           | horseflesh       |
| sfr       | 3          | 1           | 1           | 1           | disfranchisement |
| sgr       | 3          | 1           | 1           | 1           | disgrace         |
| skj       | 3          | 1           | 1           | 1           | rescue           |
| skl       | 3          | 1           | 1           | 1           | disclosure       |
| skm       | 3          | 1           | 1           | 1           | taskmaster       |

|     |   |   |   |   |               |
|-----|---|---|---|---|---------------|
| skn | 3 | 1 | 1 | 1 | brusqueness   |
| skr | 3 | 1 | 1 | 1 | described     |
| skw | 3 | 1 | 1 | 1 | disqualified  |
| spj | 3 | 1 | 1 | 1 | dispute       |
| spl | 3 | 1 | 1 | 1 | display       |
| spn | 3 | 1 | 1 | 1 | crispness     |
| spr | 3 | 1 | 1 | 1 | aspirin       |
| spw | 3 | 1 | 1 | 1 | wasp-waisted  |
| ssk | 3 | 1 | 1 | 1 | ice-skating   |
| ssp | 3 | 1 | 1 | 1 | misspelled    |
| sst | 3 | 1 | 1 | 1 | misstated     |
| stf | 3 | 1 | 1 | 1 | wasteful      |
| stg | 3 | 1 | 1 | 1 | coastguard    |
| sth | 3 | 1 | 1 | 1 | priesthood    |
| stj | 3 | 1 | 1 | 1 | costume       |
| stl | 3 | 1 | 1 | 1 | costly        |
| stm | 3 | 1 | 1 | 1 | maladjustment |
| stn | 3 | 1 | 1 | 1 | earnestness   |
| str | 3 | 1 | 1 | 1 | history       |
| stw | 3 | 1 | 1 | 1 | westward      |
| jkl | 3 | 1 | 1 | 1 | washcloth     |
| jpl | 3 | 1 | 1 | 1 | fishplates    |
| jst | 3 | 1 | 1 | 1 | washstand     |
| jtr | 3 | 1 | 1 | 1 | ashtray       |
| tbj | 3 | 1 | 1 | 1 | politburo     |
| tbl | 3 | 1 | 1 | 1 | jet-black     |
| tbr | 3 | 1 | 1 | 1 | outbreak      |
| tdr | 3 | 1 | 1 | 1 | nightdress    |
| tfl | 3 | 1 | 1 | 1 | outflow       |
| tfr | 3 | 1 | 1 | 1 | scot-free     |
| tgl | 3 | 1 | 1 | 1 | plate-glass   |
| tgr | 3 | 1 | 1 | 1 | outgrowth     |
| tkl | 3 | 1 | 1 | 1 | nightclub     |
| tkr | 3 | 1 | 1 | 1 | outcry        |
| tpl | 3 | 1 | 1 | 1 | marketplace   |
| tpr | 3 | 1 | 1 | 1 | footprints    |
| tsk | 3 | 1 | 1 | 1 | outskirts     |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| tsl       | 3          | 1           | 1           | 1           | shirtsleeves   |
| tsm       | 3          | 1           | 1           | 1           | statesman      |
| tsp       | 3          | 1           | 1           | 1           | outspoken      |
| tst       | 3          | 1           | 1           | 1           | outstanding    |
| tsw       | 3          | 1           | 1           | 0           | carpet-sweeper |
| ttr       | 3          | 1           | 1           | 0           | out-tray       |
| θbl       | 3          | 1           | 1           | 1           | deathblow      |
| θbr       | 3          | 1           | 1           | 1           | toothbrush     |
| θkw       | 3          | 1           | 1           | 1           | earthquake     |
| θpl       | 3          | 1           | 1           | 1           | birthplace     |
| θpr       | 3          | 1           | 1           | 1           | mothproofing   |
| ʃbl       | 3          | 1           | 1           | 1           | pitch-black    |
| ʃdj       | 3          | 1           | 1           | 1           | archduke       |
| ʃkr       | 3          | 1           | 1           | 1           | witchcraft     |
| ʃst       | 3          | 1           | 1           | 1           | touchstone     |
| vst       | 3          | 1           | 1           | 1           | livestock      |
| zbl       | 3          | 1           | 1           | 1           | nosebleed      |
| zfl       | 3          | 1           | 1           | 1           | newsflash      |
| zkl       | 3          | 1           | 1           | 1           | cheesecloth    |
| zkr       | 3          | 1           | 1           | 1           | wisecracking   |
| zpr       | 3          | 1           | 1           | 1           | newsprint      |
| zst       | 3          | 1           | 1           | 1           | newsstand      |
| bskj      | 4          | 1           | 1           | 1           | obscure        |
| bskr      | 4          | 1           | 1           | 1           | subscribers    |
| bstr      | 4          | 1           | 1           | 1           | abstract       |
| dspr      | 4          | 1           | 1           | 1           | widespread     |
| dstr      | 4          | 1           | 1           | 1           | headstrong     |
| fspr      | 4          | 1           | 1           | 1           | offspring      |
| fstr      | 4          | 1           | 1           | 1           | off-street     |
| ftsm      | 4          | 1           | 1           | 1           | craftsmen      |
| ksfr      | 4          | 1           | 1           | 1           | tax-free       |
| ksgl      | 4          | 1           | 1           | 1           | foxglove       |
| kshj      | 4          | 1           | 1           | 1           | exhumed        |
| kskj      | 4          | 1           | 1           | 1           | excuse         |
| kskl      | 4          | 1           | 1           | 1           | exclusive      |
| kskr      | 4          | 1           | 1           | 1           | excruciating   |

|       |   |   |   |   |                   |
|-------|---|---|---|---|-------------------|
| kswk  | 4 | 1 | 1 | 1 | exquisite         |
| kspl  | 4 | 1 | 1 | 1 | explain           |
| kspr  | 4 | 1 | 1 | 1 | expression        |
| ksst  | 4 | 1 | 1 | 1 | sex-starved       |
| kstj  | 4 | 1 | 1 | 1 | contextual        |
| kstr  | 4 | 1 | 1 | 1 | extra             |
| ldbl  | 4 | 1 | 1 | 1 | cold-blooded      |
| ldst  | 4 | 1 | 1 | 1 | oldsters          |
| ldzm  | 4 | 1 | 1 | 0 | fieldsman         |
| lfpr  | 4 | 1 | 1 | 1 | self-preservation |
| lfst  | 4 | 1 | 1 | 1 | self-styled       |
| lmzm  | 4 | 1 | 1 | 1 | helmsman          |
| lptr  | 4 | 1 | 1 | 1 | sculptress        |
| lstr  | 4 | 1 | 1 | 1 | maelstrom         |
| ltst  | 4 | 1 | 1 | 0 | maltster          |
| mpbl  | 4 | 1 | 1 | 1 | lamp-black        |
| mptj  | 4 | 1 | 1 | 1 | impromptu         |
| mptl  | 4 | 1 | 1 | 1 | promptly          |
| mptn  | 4 | 1 | 1 | 1 | promptness        |
| mptr  | 4 | 1 | 1 | 1 | temptress         |
| mskr  | 4 | 1 | 1 | 1 | circumscribed     |
| mstr  | 4 | 1 | 1 | 1 | hamstring         |
| ndzm  | 4 | 1 | 1 | 1 | bandsmen          |
| nsfj  | 4 | 1 | 1 | 1 | transfusion       |
| nskr  | 4 | 1 | 1 | 1 | transcribed       |
| nspl  | 4 | 1 | 1 | 1 | transplant        |
| nspr  | 4 | 1 | 1 | 1 | mainspring        |
| nstr  | 4 | 1 | 1 | 1 | construction      |
| ntbl  | 4 | 1 | 1 | 1 | point-blank       |
| ntbr  | 4 | 1 | 1 | 1 | paintbrush        |
| ntfr  | 4 | 1 | 1 | 1 | rent-free         |
| ntpr  | 4 | 1 | 1 | 1 | accident-prone    |
| ntsm  | 4 | 1 | 1 | 1 | hunter            |
| ntst  | 4 | 1 | 1 | 1 | flintstone        |
| ntfdr | 4 | 1 | 1 | 1 | punch-drunk       |
| nzgr  | 4 | 1 | 1 | 1 | transgression     |
| nzmj  | 4 | 1 | 1 | 1 | transmuted        |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| ηksg      | 4          | 1           | 1           | 1           | thanksgiving   |
| ηkst      | 4          | 1           | 1           | 1           | youngsters     |
| ηktl      | 4          | 1           | 1           | 1           | distinctly     |
| ηktn      | 4          | 1           | 1           | 1           | distinctness   |
| ηkθw      | 4          | 1           | 1           | 1           | lengthwise     |
| ηspr      | 4          | 1           | 1           | 1           | wingspread     |
| ηstr      | 4          | 1           | 1           | 1           | songstress     |
| pskr      | 4          | 1           | 1           | 1           | typescript     |
| pskw      | 4          | 1           | 1           | 1           | pipsqueak      |
| pstr      | 4          | 1           | 1           | 1           | upstream       |
| tspr      | 4          | 1           | 1           | 1           | outspread      |
| tstr      | 4          | 1           | 1           | 1           | outstretched   |
| vzdr      | 4          | 1           | 1           | 1           | eavesdropping  |
| mpstr     | 5          | 1           | 1           | 0           | sempstress     |
| stskr     | 5          | 1           | 1           | 1           | postscript     |
| ststr     | 5          | 1           | 1           | 1           | breaststroke   |

**Final clusters**

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| bd        | 2          | 1           | 1           | 1           | described      |
| bz        | 2          | 1           | 1           | 1           | jobs           |
| dθ        | 2          | 1           | 1           | 1           | width          |
| dz        | 2          | 1           | 1           | 1           | kids           |
| ðd        | 2          | 1           | 1           | 1           | breathed       |
| ðz        | 2          | 1           | 1           | 1           | clothes        |
| dʒd       | 2          | 1           | 1           | 1           | managed        |
| fs        | 2          | 1           | 1           | 1           | photographs    |
| ft        | 2          | 1           | 1           | 1           | left           |
| fθ        | 2          | 1           | 1           | 1           | fifth          |
| gd        | 2          | 1           | 1           | 1           | shrugged       |
| gz        | 2          | 0           | 1           | 1           | drugs          |
| ks        | 2          | 1           | 1           | 1           | makes          |
| kt        | 2          | 1           | 1           | 1           | fact           |
| lb        | 2          | 1           | 1           | 1           | bulb           |
| ld        | 2          | 1           | 1           | 1           | world          |
| ldʒ       | 2          | 1           | 1           | 1           | indulge        |
| lf        | 2          | 1           | 1           | 1           | itself         |
| lk        | 2          | 1           | 1           | 1           | milk           |
| lm        | 2          | 1           | 1           | 1           | film           |
| ln        | 2          | 1           | 1           | 1           | kiln           |
| lp        | 2          | 1           | 1           | 1           | help           |
| ls        | 2          | 1           | 1           | 1           | else           |
| lj        | 2          | 1           | 1           | 1           | welsh          |
| lt        | 2          | 1           | 1           | 1           | felt           |
| lθ        | 2          | 1           | 1           | 1           | health         |
| ljʃ       | 2          | 1           | 1           | 1           | mulch          |
| lv        | 2          | 1           | 1           | 1           | solve          |
| lz        | 2          | 1           | 1           | 1           | schools        |
| md        | 2          | 1           | 1           | 1           | seemed         |
| mf        | 2          | 1           | 1           | 1           | triumph        |
| mp        | 2          | 1           | 1           | 1           | camp           |
| mt        | 2          | 1           | 1           | 1           | dreamt         |
| mθ        | 2          | 1           | 1           | 1           | warmth         |
| mz        | 2          | 1           | 1           | 1           | times          |
| nd        | 2          | 1           | 1           | 1           | and            |
| ndʒ       | 2          | 1           | 1           | 1           | change         |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| ns        | 2          | 1           | 1           | 1           | since          |
| nt        | 2          | 1           | 1           | 1           | want           |
| nθ        | 2          | 1           | 1           | 1           | month          |
| ntf       | 2          | 1           | 1           | 1           | lunch          |
| nz        | 2          | 1           | 1           | 1           | means          |
| nd        | 2          | 1           | 1           | 1           | belonged       |
| nk        | 2          | 1           | 1           | 1           | think          |
| nz        | 2          | 1           | 1           | 1           | things         |
| ps        | 2          | 1           | 1           | 1           | perhaps        |
| pt        | 2          | 1           | 1           | 1           | kept           |
| pθ        | 2          | 1           | 1           | 1           | depth          |
| sk        | 2          | 1           | 1           | 1           | ask            |
| sp        | 2          | 1           | 1           | 1           | grasp          |
| st        | 2          | 1           | 1           | 1           | just           |
| ft        | 2          | 1           | 1           | 1           | published      |
| ts        | 2          | 1           | 1           | 1           | its            |
| tθ        | 2          | 1           | 1           | 1           | eighth         |
| θs        | 2          | 1           | 1           | 1           | deaths         |
| θt        | 2          | 1           | 1           | 1           | unearthed      |
| ft        | 2          | 1           | 1           | 1           | reached        |
| vd        | 2          | 1           | 1           | 1           | moved          |
| vz        | 2          | 1           | 1           | 1           | lives          |
| zd        | 2          | 1           | 1           | 1           | raised         |
| zd        | 2          | 1           | 1           | 1           | arranged       |
| dst       | 3          | 1           | 1           | 1           | midst          |
| dθs       | 3          | 0           | 1           | 1           | widths         |
| fts       | 3          | 0           | 1           | 1           | gifts          |
| fθs       | 3          | 0           | 1           | 1           | fifths         |
| kst       | 3          | 1           | 1           | 1           | next           |
| ksθ       | 3          | 1           | 1           | 1           | sixth          |
| kts       | 3          | 0           | 1           | 1           | effects        |
| lbz       | 3          | 0           | 1           | 1           | bulbs          |
| ldz       | 3          | 0           | 1           | 1           | fields         |
| ldzd      | 3          | 0           | 1           | 1           | indulged       |
| lfs       | 3          | 0           | 1           | 1           | engulfs        |
| lft       | 3          | 1           | 1           | 1           | engulfed       |
| lfθ       | 3          | 1           | 1           | 1           | twelfth        |
| lks       | 3          | 0           | 1           | 1           | silks          |
| lkt       | 3          | 1           | 1           | 1           | milked         |

|      |   |   |   |   |             |
|------|---|---|---|---|-------------|
| lmd  | 3 | 0 | 1 | 1 | overwhelmed |
| lmz  | 3 | 0 | 1 | 1 | films       |
| lnz  | 3 | 0 | 1 | 1 | kilns       |
| lps  | 3 | 0 | 1 | 1 | helps       |
| lpt  | 3 | 1 | 1 | 1 | helped      |
| lst  | 3 | 1 | 1 | 1 | pulsed      |
| lft  | 3 | 0 | 1 | 1 | welshed     |
| lts  | 3 | 1 | 1 | 1 | results     |
| lft  | 3 | 0 | 1 | 1 | belched     |
| lvd  | 3 | 1 | 1 | 1 | involved    |
| lvz  | 3 | 1 | 1 | 1 | themselves  |
| mfs  | 3 | 0 | 1 | 1 | triumphs    |
| mft  | 3 | 0 | 1 | 1 | triumphed   |
| mps  | 3 | 1 | 1 | 1 | camps       |
| mpt  | 3 | 1 | 1 | 1 | attempt     |
| ndz  | 3 | 1 | 1 | 1 | hands       |
| ndzd | 3 | 1 | 1 | 1 | changed     |
| nst  | 3 | 1 | 1 | 1 | against     |
| nts  | 3 | 1 | 1 | 1 | students    |
| nθs  | 3 | 0 | 1 | 1 | months      |
| nft  | 3 | 0 | 1 | 1 | launched    |
| nzd  | 3 | 0 | 1 | 1 | cleansed    |
| ηks  | 3 | 1 | 1 | 1 | thanks      |
| ηkt  | 3 | 1 | 1 | 1 | linked      |
| ηkθ  | 3 | 1 | 1 | 1 | strength    |
| pst  | 3 | 0 | 1 | 1 | collapsed   |
| pts  | 3 | 0 | 1 | 1 | concepts    |
| pθs  | 3 | 0 | 1 | 1 | depths      |
| sks  | 3 | 0 | 1 | 1 | asks        |
| skt  | 3 | 1 | 1 | 1 | asked       |
| sps  | 3 | 0 | 1 | 1 | gasps       |
| spt  | 3 | 0 | 1 | 1 | gaspd       |
| sts  | 3 | 0 | 1 | 1 | costs       |
| tst  | 3 | 0 | 1 | 1 | blitzed     |
| tθs  | 3 | 0 | 1 | 1 | eighths     |
| ksts | 4 | 0 | 1 | 1 | texts       |
| ksθs | 4 | 0 | 1 | 1 | sixths      |
| lfθs | 4 | 0 | 1 | 0 | twelfths    |
| lkts | 4 | 0 | 1 | 0 | mulcts      |
| lpts | 4 | 0 | 1 | 1 | sculpts     |

| <b>cl</b> | <b>len</b> | <b>dict</b> | <b>para</b> | <b>corp</b> | <b>example</b> |
|-----------|------------|-------------|-------------|-------------|----------------|
| mpst      | 4          | 0           | 1           | 1           | glimpsed       |
| mpts      | 4          | 0           | 1           | 1           | attempts       |
| ŋkst      | 4          | 1           | 1           | 1           | amongst        |
| ŋkts      | 4          | 0           | 1           | 1           | instincts      |
| ŋkθs      | 4          | 0           | 1           | 1           | strengths      |

## References

- Bańko, M. (ed.). 2000. *Inny słownik języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Bargielówna, M. 1950. "Grupy fonemów spółgłoskowych współczesnej polszczyzny kulturalnej", *Biuletyn Polskiego Towarzystwa Językoznawczego* 10: 1-25.
- Baroni, A. 2014. "On the importance of being noticed: The role of acoustic salience in phonotactics and casual speech", *Language Sciences* 46A: 18-36.
- Bartnicka-Dąbkowska, B. 1968. "Fonetyka", in: W. Doroszewski and B. Wieczorkiewicz (eds.), *Gramatyka opisowa języka polskiego z ćwiczeniami*. Warszawa: Państwowe Zakłady Wydawnictw Szkolnych, 3-118.
- Bartnicka, B. and R. Sinielnikoff. 1999. *Słownik podstawowy języka polskiego dla cudzoziemców*. Kielce: Wydawnictwo Takt.
- Baudouin de Courtenay, J. 1895. *Versuch einer Theorie phonetischer Alternationen*. Strassburg: Trübner.
- Benni, T. 1926. "Polska wymowa sceniczna", *Język polski* 11(6): 161-168.
- Benni, T. 1924. *Ortofonia polska*. Warszawa: Książnica Polska.
- Benni, T. 1959. *Fonetyka opisowa języka polskiego*. Wrocław: Zakład Narodowy im. Ossolińskich.
- Biedrzycki, L. 1963. "Fonologiczna interpretacja polskich głosek nosowych", *Biuletyn Polskiego Towarzystwa Językoznawczego* 22: 25-45.
- Biedrzycki, L. 1978. *Fonologia angielskich i polskich rezonantów. Porównanie samogłosek oraz spółgłosek rezonantowych*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Clements, N. 1990. "The role of sonority in core syllabification", in: J. Kingston and M.E. Beckman (eds.), *Papers in laboratory phonology: Between the grammar and physics of speech*. Cambridge: Cambridge University Press, 283-333.
- Crystal, D. 1992. *An encyclopedic dictionary of language and languages*. Middlesex: Blackwell.
- Cyran, E. (ed.). 1998. *Structure and interpretation. Studies in phonology*. Lublin: Folium.
- Cyran, E. and E. Gussmann. 1999. "Consonantal clusters and governing relations: Polish initial consonant sequences", in: H. van der Hulst and A.N. Ritter (eds.), *The syllable. Views and facts*. Berlin: Mouton de Gruyter, 219-249.
- Davies, M. 2011. Word frequency data from the Corpus of Contemporary American English (COCA). Downloaded from <http://www.wordfrequency>. (date of access: 24 Jan. 2011).
- Dobrogowska, K. 1984. "Śródgłosowe grupy spółgłosek w polskich tekstach popularnonaukowych", *Polonica* 10: 15-34.
- Dobrogowska, K. 1990. "Word internal consonant clusters in Polish artistic prose", *Studia Phonetica Posnaniensia* 2: 43-67.
- Dobrogowska, K. 1992. "Word initial and word final consonant clusters in Polish popular science text and in artistic prose", *Studia Phonetica Posnaniensia* 3: 47-121.

- Doroszewski, W. and B. Wieczorkiewicz. 1947. *Zasady poprawnej wymowy polskiej (ze słowniczkiem)*. Warszawa: Państwowe Zakłady Wydawnictw Szkolnych.
- Dressler, W.U. 1985. *Morphonology: the Dynamics of Derivation*. Ann Arbor: Karoma.
- Dressler, W.U. 1996. "A functionalist semiotic model of morphonology", in: S. Rajendra (ed.), *Trubetzkoy's Orphan*. Amsterdam: Benjamins, 67-83.
- Dressler, W.U. and K. Dziubalska-Kołaczyk. 2006. "Proposing morphonotactics", *Rivista di Linguistica*, 18(2): 249-266.
- Dressler, W.U., K. Dziubalska-Kołaczyk and R. Spina. 2001. "Sources of markedness in language structures", *Folia Linguistica Historica* 22(1-2): 103-135.
- Dubisz, S. (ed.). 2003. *Uniwersalny słownik języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Dukiewicz, L. 1958. "Analiza mowy nagranej wstecz", *Język Polski* 38(3): 197-204.
- Dukiewicz, L. 1967. *Polskie głoski nosowe*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Dukiewicz, L. 1980. "Rodzaje i częstość występowania nagłosowych grup fonemów typu CC(V)-, CCC(V)- i CCCC(V)- w polskich tekstach prasowych", *Lětopis Instytutu za Serbski Ludospyt* 27(2): 142-157.
- Dukiewicz, L. 1985. "Nagłosowe grupy spółgłosek w polskich tekstach popularnonaukowych i prasowych", *Studia Gramatyczne* 6: 17-34.
- Dukiewicz, L. and I. Sawicka. 1995. *Fonetyka i fonologia (Gramatyka współczesnego języka polskiego)*. Kraków: Wydawnictwo Instytutu Języka Polskiego PAN.
- Dunaj, B. 1985. *Grupy spółgłoskowe współczesnej polszczyzny mówionej*. Warszawa: Polska Akademia Nauk.
- Dunaj, B. 1986. "Wygłosowe grupy spółgłoskowe współczesnej polszczyzny mówionej", *Zeszyty Naukowe Uniwersytetu Jagiellońskiego: Prace Językoznawcze* 82: 103-117.
- Dunaj, B. 1991. "Dwa dyskusyjne problemy polskiej fonologii", *Prace Językoznawcze* 19: 40-46.
- Dunaj, B. 1994. "Zmiany fonetyczne (i fonologiczne) w polszczyźnie po 1945 roku", in: S. Gajda and Z. Adamiszyn (eds.), *Przemiany współczesnej polszczyzny*. Opole: WSP, 41-52.
- Dunaj, B. 2006. "Zasady poprawnej wymowy polskiej", *Język Polski* 86(3): 161-172.
- Durand, P. and R. Gubrynowicz. 1999. "On the reduction of consonant clusters in Polish according to emotional situation", in: *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 2113-2115.
- Dziubalska-Kołaczyk, K. (ed.). 2001. *Constraints and preferences*. Berlin: Mouton de Gruyter.
- Dziubalska-Kołaczyk, K. 2002. *Beats-and-Binding Phonology*. Frankfurt am Main: Peter Lang.
- Dziubalska-Kołaczyk, K. 2009. "NP extension: B&B phonotactics", *Poznań Studies in Contemporary Linguistics* 45(1): 55-71.
- Dziubalska-Kołaczyk, K. 2014. "Explaining phonotactics using NAD", *Language Sciences* 46A: 6-17.

- Dziubalska-Kołodziej, K. and G. Krynicki. 2007. Predicting difficulty in second language acquisition. (A paper presented at New Sounds 2007, 25-28 November 2007) (<http://ifa.amu.edu.pl/~dkasia/calculator/>) (date of access: 6 Oct. 2008)
- Dziubalska-Kołodziej, K. and D. Zielińska. 2010. "Predicting phonotactic difficulty in second language acquisition", in: A. Rauber and M. Watkins (eds.), *The acquisition of second language speech: Studies in honor of Professor Barbara O. Baptista*. Florianópolis: Graduate Program in English/Federal University of Santa Catarina, 281-306.
- Dziubalska-Kołodziej, K. and D. Zielińska. 2011. "Universal phonotactic and morphonotactic preferences in second language acquisition", in: K. Dziubalska-Kołodziej, M. Wrembel and M. Kul (eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010*. Frankfurt am Main: Peter Lang, 53-63.
- Dziubalska-Kołodziej, K., P. Wierzchoń, M. Jankowski, P. Orzechowska, P. Zydorowicz and D. Pietrala. 2012. Phonotactics and morphonotactics of Polish and English: description, tools and applications. [Unpublished research project report, Adam Mickiewicz University, Poznań].
- Dziubalska-Kołodziej, K., P. Zydorowicz and M. Jankowski. 2013. "English morphonotactics: A corpus study". *The Phonetician. A Publication of ISPhS International Society of Phonetic Sciences* 107-108: 53-67.
- Greenberg, J. 1978. "Some generalisations concerning initial and final consonant clusters", in: J. Greenberg (ed.) *Universals of human language. Vol. 2: Phonology*. Stanford, California: Stanford University Press, 243-279.
- Grzegorzczakowa, R., R. Laskowski and H. Wróbel (eds.) 1998. *Gramatyka współczesnego języka polskiego. Morfologia*. Warszawa: Wydawnictwo Naukowe PWN.
- Gussmann, E. and E. Cyran. 1998. "Polish consonantal sequences: A phonological testing ground", in: E. Cyran (ed.), *Structure and interpretation. Studies in phonology*. Folium: Lublin, 127-138.
- Hill, A.A. 1958. *Introduction to linguistic structures: From sound to sentence in English*. New York: Harcourt, Brace & Co.
- Hooper, J.B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.
- Hornby, A.S. 1974. *Oxford Advanced Learner's Dictionary of Current English*. London: Oxford University Press.
- Jadacka, H. (ed.). 2001a. *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego*. Vol. 2. Kraków: Universitas.
- Jadacka H. 2001b. *System słowotwórczy polszczyzny (1945-2000)*. Warszawa: Wydawnictwo Naukowe PWN.
- Jankowski, M. 1994. "Practical automatic phonemic transcription systems", *Studia Anglica Posnaniensia* 28: 143-150. <http://ifa.amu.edu.pl/sap/files/28/09Jankowski.pdf>
- Jassem, W. 1958. "A phonologic and acoustic classification of Polish vowels", *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft* 11(4): 299-319.
- Jassem, W. 1973. *Podstawy fonetyki akustycznej*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Jassem, W. 2003. "Polish" *Journal of International Phonetic Association* 33(1): 103-108.

- Karpowicz, T. 2009. *Kultura języka polskiego. Wymowa, ortografia, interpunkcja*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Kiparsky, Paul. 1979. "Metrical structure assignment is cyclic", *Linguistic Inquiry*, 10(3): 521-541.
- Klemensiewicz, Z. 1930. *Prawidła poprawnej wymowy polskiej*. Kraków: Gebethner & Wolff.
- Koneczna, H. 1934. "Studium eksperymentalne artykulacji głosek polskich", *Prace Filologiczne* 16: 33-174.
- Koneczna, H. 1965. *Charakterystyka fonetyczna języka polskiego na tle innych języków słowiańskich*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Kurcz, I., A. Lewicki, J. Sambor, K. Szafran and J. Woronczak. 1990. *Słownik frekwencyjny polszczyzny współczesnej*. Kraków: Instytut Języka Polskiego PAN.
- Kuryłowicz, J. 1952. "Uwagi o polskich grupach spółgłoskowych", *Biuletyn Polskiego Towarzystwa Językoznawczego* 11: 54-69.
- Ladefoged, P. 2006. *A course in phonetics*. Boston: Heinle & Heinle.
- Łukaszewicz, B. 2007. "Reduction in syllable onsets in the acquisition of Polish: deletion, coalescence, metathesis and gemination", *Journal of child language* 34: 53-82.
- Maddieson, I. 1999. "In search of universals", in: J. Ohala, Y. Hasegawa, M. Ohala, D. Granville and A.C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences 1999*. Vol. 3: 2521-2528.
- Maddieson, I. 2008. "Syllable structure", in: M. Haspelmath, M.S. Dryer, D. Gil, and B. Comrie (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. (<http://wals.info/feature/39>).
- Madejowa, M. 1989. "Zasady wymowy polskiej", *Biuletyn Audiofonologii* 1(2-4): 69-83.
- Madejowa, M. 1990. "Modern Polish linguistic norms with reference to the pronunciation of consonant clusters", *Studia Phonetica Posnaniensia* 2: 69-105.
- Madejowa, M. 1992. "Zasady współczesnej wymowy polskiej (w zakresie samogłosek nosowych i grup spółgłoskowych) oraz ich przydatność w praktyce szkolnej", *Język Polski* 2-3: 187-198.
- Madejowa, M. 1993. "Normative rules of modern Polish pronunciation", *Studia Phonetica Posnaniensia* 4: 19-30.
- Madelska, L. and M. Witaszek-Gambarska. 2015. *Zapis fonetyczny*. Poznań: Wydawnictwo Naukowe UAM.
- Milewski, S. 2005. "Grupy spółgłoskowe w języku mówionym dzieci przedszkolnych", *Logopeda* 1: 5-32.
- Mitton, R. 1992. "A description of a computer-usable dictionary file based on the Oxford Advanced Learner's Dictionary of Current English." (<http://coral.lili.uni-bielefeld.de/Classes/Winter98/ExPhon/SynthDBNotes/synthdbnotes/node18.html>) (date of access: 19 May 2014).
- Nowakowski, P. 1997. *Wariantywność współczesnej polskiej wymowy scenicznej*. Poznań: Sarus.
- Ohala, J. 1990a. "Alternatives to the sonority hierarchy for explaining segmental sequential constraints", *Papers from the Parasession on the Syllable in Phonetics and Phonology*. Chicago Linguistic Society 26,2: 319-338.

- Ohala, J. 1990b. "The phonetics and phonology of aspects of assimilation", in: J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I*. Cambridge: Cambridge University Press, 258-275.
- Ohala, J. and H. Kawasaki. 1984. "Prosodic phonology and phonetics", *Phonology Yearbook* 11: 113-129.
- Orzechowska, P. 2009. Morphonotactics in English and Polish. A dictionary- and corpus-based study of word-final consonant clusters. [Unpublished Ph.D. dissertation, Adam Mickiewicz University, Poznań].
- Orzechowska, P. 2012. "Effects of morphology on the syllables structure", in: T. Stolz, N. Nau and C. Stroh (eds.), *Monosyllables: From Phonology to Typology*. Berlin: Akademie Verlag, 93-113.
- Osowicka-Kondratowicz, M. 2005. Assimilative palatalization within consonantal clusters in Polish. *Studia Phonetica Posnaniensia* 7: 5-22.
- Osowicka-Kondratowicz, M. and A. Serowik. 2007. "Palatalność asymilacyjna w języku polskim", in: I. Sawicka (ed.), *Komparacja systemów i funkcjonowania współczesnych języków słowiańskich. Fonetyka/Fonologia*. Opole: Uniwersytet Opolski, Instytut Filologii Polskiej, 61-102.
- Pietrała, D. 2014. The application of linguistic principles in programming: Phonemic transcription and phonotactic constraints. [Unpublished Ph.D. dissertation, Adam Mickiewicz University, Poznań].
- Ritt, N. 2010. "Middle English (mor-)phonotactics: system, usage, change", *SELIM: Journal of the Spanish society of medieval English language and literature* 17: 43-68.
- Ritt, N. and K. Kaźmierski. in press. "How rarities like gold come to exist: on co-evolutionary interactions between morphology and lexical phonotactics", *English Language and Linguistics*.
- Rochoń, M. 2000. *Optimality in Complexity: The Case of Polish Consonant Clusters (Studia Grammatica 48)*. Berlin: Akademie Verlag.
- Roślawski, B. 1986. *Zarys fonologii, fonetyki, fonotaktyki i fonostatystyki współczesnego języka polskiego*. Gdańsk: Wydawnictwo Uczelniane UG.
- Sawicka, I. 1988. *Fonologia konfrontatywna polsko-serbsko-chorwacka*. Wrocław: Zakład Narodowy im. Ossolińskich.
- Sawicka, I. and S. Grzybowski. 1999. *Studia z palatalności w językach słowiańskich*. Toruń: Wydawnictwo Naukowe Uniwersytetu im. Mikołaja Kopernika.
- Scheer, T. 2007. "On the status of word-initial clusters in Slavic (and elsewhere)", in: R. Compton, M. Goledzinowska and U. Savchenko (eds.), *Annual workshop on formal approaches to Slavic linguistics. The Toronto Meeting 2006*. Ann Arbor: Michigan Slavic Publications, 346-364.
- Scheer, T. 2008. "Syllabic and trapped consonants in (Western) Slavic: Different but still the same", in: G. Zybatow, L. Szucsich, U. Junghanns and R. Meyer (eds.), *Formal Description of Slavic Languages: The Fifth Conference, Leipzig 2003*. Frankfurt am Main: Peter Lang, 149-167.
- Selkirk, E. O. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge: Massachusetts Institute of Technology Press.

- Skarżyński, M., M. Berendt, M. Bondkowska, I. Burkacka, H. Jadacka, M. Olejniczak and T. Vogelegang. 2004. *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego*. Vol. 3. Kraków: Towarzystwo Wydawnicze "Historia Iagellonica" z siedzibą w Instytucie Historii UJ.
- Sobkowiak, W. and W. Ferlacka. 2011. "PDI as a tool of phonetic enhancements to graded e-readers", in: J. Arabski and A. Wojtaszek (eds.), *The acquisition of L2 phonology*. Bristol: Multilingual Matters, 138-158.
- Sobkowiak, W. 2006. "PDI revisited: lexical cooccurrence of phonetic difficulty codes", in: W. Sobkowiak and E. Waniek-Klimczak (eds.), *Dydaktyka fonetyki języka obcego. Neofilologia VIII. Zeszyty naukowe Państwowej Wyższej Szkoły Zawodowej w Płocku*. Płock: Wydawnictwo Państwowej Wyższej Szkoły Zawodowej. *Proceedings of the Fifth Phonetics in FLT Conference*, Soczewka, 25-27.04.2005, 225-238.
- Steffen-Batogowa, M. 1975. *Automatyzacja transkrypcji fonemacyjnej tekstów polskich*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Steffen-Batóg, M. 1996. "On the pronunciation of some Polish consonant clusters containing sonorants", *Studia Phonetica Posnaniensia* 5: 61-85.
- Steffen-Batóg, M. and P. Nowakowski. 1993. "An Algorithm for Phonetic Transcription of Orthographic Texts in Polish", *Studia Phonetica Posnaniensia* 3: 135-184.
- Steriade, D. 1982. Greek Prosodies and the Nature of Syllabification. [Unpublished Ph.D. dissertation, Massachusetts Institute of Technology].
- Stieber, Z. 1966. *Historyczna i współczesna fonologia języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Szymczak, M. (ed.). 1981. *Słownik języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Śledziński, D. 2005. Indeks zbitek spółgłoskowych języka polskiego z przykładami. [Unpublished index, Adam Mickiewicz University, Poznań: Instytut Językoznawstwa].
- Tokarski, J. 1993. *Schematyczny indeks a tergo polskich form wyrazowych*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Vennemann, T. 1988. *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Berlin: Mouton.
- Wierzchowska, B. 1967. *Opis fonetyczny języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Wierzchowska, B. 1971. *Wymowa polska*. Warszawa: Państwowe Zakłady Wydawnictw Szkolnych.
- Wierzchowska, B. 1980. *Fonetyka i fonologia języka polskiego*. Wrocław: Zakład Narodowy im. Ossolińskich.
- Zec, D. 1995. "Sonority Constraints on Syllable Structure", *Phonology* 12: 85-129.
- Zwoliński, P. 1949. "Przejście ł > u w języku polskim", *Biuletyn Polskiego Towarzystwa Językoznawczego* 9: 81-96.
- Zydorowicz, P. 2009. English and Polish morphonotactics in first language acquisition. [Unpublished Ph.D. dissertation, Adam Mickiewicz University, Poznań].

- Zydorowicz, P. 2010. "Consonant clusters across morpheme boundaries: Polish morphotactic inventory and its acquisition". *Poznań Studies in Contemporary Linguistics* 46(4): 565-588.
- Zydorowicz, P., K. Dziubalska-Kołaczyk and M. Jankowski. 2015. "English word-medial morphotactics: A corpus study", in: *The Scottish Consortium for ICPhS 2015 (ed.) Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow, online.



## Streszczenie

# Fonotaktyka i morfonotaktyka języka polskiego i angielskiego. Teoria, opis, narzędzia i zastosowania

Fonotaktyka określa fonologiczne warunki i ograniczenia występowania i współwystępowania dźwięków (spółgłosek i samogłosek) w danym języku. Morfonotaktyka odnosi się do interakcji między fonotaktyką a morfotaktyką. Główne podejście do zbitek spółgłoskowych oparte jest na podziale zbitek na leksykalne oraz morfonotaktyczne. Zbitki leksykalne pojawiają się wewnątrz morfemu. Zbitki morfonotaktyczne powstają za sprawą operacji morfologicznych takich jak konkatencja czy przegłos. Zbitki morfonotaktyczne często mają charakter fonotaktycznie nacechowany oraz bywają zbitkami, które nie występują wewnątrzmorfemowo. Przedmiotem badania są zarówno zbitki leksykalne (wewnątrzmorfemowe) jak i te o cechach morfonotaktycznych. Przewodnym motywem niniejszej pracy jest zweryfikowanie hipotezy dotyczącej stopnia nacechowania zbitek w zależności od ich fono- i morfonotaktycznego charakteru. Uzyskane dane słownikowe i korpusowe o występowaniu oraz statystycznym zachowaniu zbitek spółgłoskowych zostały zbadane pod kątem stopnia nacechowania opartego na uniwersalnych preferencjach fonotaktycznych. Jako narzędzie pomiaru stworzony został kalkulator fonotaktyczny, który posłużył do oceny preferencyjności fonotaktycznych oraz morfonotaktycznych zbitek spółgłoskowych.

Fonotaktyka to klasa teorii, w których rozpatruje się fonologiczne warunki i ograniczenia występowania i współwystępowania dźwięków (spółgłosek i samogłosek) w danym języku, tj. określa się specyficzne dla języka fonologiczne warunki i ograniczenia występowania i współwystępowania spółgłosek i samogłosek. Morfonotaktyka z kolei odnosi się do korelacji między fonotaktyką a morfotaktyką. Morfonotaktyka jest terminem zaproponowanym przez Wolfganga Dresslera i Katarzynę Dziubalską-Kołączyk (2006) w odniesieniu do interakcji między fonotaktyką a morfotaktyką. Morfonotaktykę można uznać za komponent morfonologii, która z kolei obejmuje relacje pomiędzy fonologią a morfologią (por. Dressler

1985, 1996). Dzięki badaniom morfonotaktycznym, podjętym w projekcie ze szczególną uwagą i na stosunkowo obszernym materiale empirycznym, ukazano wpływ morfologii danego języka, jego struktury morfemowej itp. na kształtowanie się własności fonotaktycznych tego języka, tj. na kształtowanie się dystrybucji pewnych ciągów spółgłoskowych. To podejście do analizy struktury wyrazu pozwala wyjaśnić te aspekty pojawiania się, występowania i funkcjonowania zbitek, których nie można wyjaśnić na gruncie teorii wyłącznie fonotaktycznej.

Kluczową rolę w badaniu morfonotaktycznym pełni parametr różnicujący zbitki na leksykalne oraz morfonotaktyczne. Jest to parametr związany z koniecznością rozpoznania semantycznego (leksykalnego oraz gramatycznego) analizowanej formy wyrazowej. Zbitki leksykalne pojawiają się wewnątrz morfemu. Zbitki morfonotaktyczne powstają wskutek operacji morfologicznych, takich jak konkatenacja (taktonifikacja) czy przegłos. Zbitki morfonotaktyczne często mają charakter fonotaktycznie nacechowany oraz bywają zbitkami, które nie występują wewnątrz morfemów. Przedmiotem badań w ramach projektu stały się zarówno zbitki leksykalne (wewnątrzmorfemowe), jak i te o cechach morfonotaktycznych. Dla tak określonych parametrów badań oraz dla scharakteryzowanej poniżej bazy empirycznej sformułowano motyw przewodni projektu, jakim było zweryfikowanie hipotezy dotyczącej stopnia nacechowania zbitek w zależności od ich fono- i morfonotaktycznego charakteru. Przewiduje się, że prototypowe zbitki morfonotaktyczne będą zawsze nacechowane pod względem fonologicznie preferowanych cech zbitek fonotaktycznych. Reasumując, w projekcie przebadano własności fonotaktyczne i morfonotaktyczne polskiego i angielskiego wyrazu w izolacji (od spacji do spacji) z perspektywy synchronicznej w świetle tak sformułowanej hipotezy.

Wspomniane badanie rzeczywistości konsonantycznej języka polskiego oraz angielskiego odbywało się w pewnym kompleksie operacji. Kompleks ten wymusił określenie rozmiaru zasobów, jakimi dysponował zespół dla wykonania przewidzianego badania. Takie parametry organizacji badań zdeterminowały (i) poszukiwania komponentu empirycznego, służącego za podstawę do przeprowadzenia założonych badań, (ii) liczbę przebadanych zjawisk (np. w zakresie wykrycia korelacji) językowych. Dla języka polskiego przyjęto 3 rodzaje danych, tj. (i) zbiór tekstów dziennika „Rzeczpospolita” z dwu lat; wybrano dwa pełne roczniki, na które składały się teksty w liczbie ok. 190 plików na jedno wydanie dzienne (4,5 tys. tekstów miesięcznie), co dało łącznie ok. 108 tys. plików z 48,6 milionami słów (to-

kens), z czego uzyskano ok. 630.000 unikalnych słów (types), (ii) Słownika podstawowego języka polskiego dla cudzoziemców autorstwa Bartnickiej-Dąbkowskiej, Sinielnikoff, co umożliwiło uzyskanie zwalidowanego zbioru słów w ich formach podstawowych w wymiarze ok. 8000 jednostek (przy zapowiedzi 5000 we Wstępie Słownika), (iii) zbioru paradygmatów fundowanych formami hasłowymi słownika podstawowego oraz kategoriami fleksyjnymi wykorzystanymi z projektu morfologik.org. Każdy z tych zasobów pozwalał na uzyskanie innych informacji. Dane prasowe informowały o częstości użyciu wyrazów, dane słownikowe pozwalały uniknąć w analizie wadliwych (nieistniejących w polszczyźnie) zbitek konsonantycznych, tj. wynikłych np. z błędów literowych lub zbitek właściwych wyłącznie dla nazw obcojęzycznych (np. obcych nazwisk), natomiast postać paradygmatyczna danych umożliwiała (i) analizę różnych zbitek w formach zależnych (typu: len – lnu), (ii) monitorowanie tych danych ze względu na jawne etykiety gramatyczne. W ramach prac podjęto się także opracowania i przetestowania różnych metod określania typu zbitki konsonantycznej, tj. typu leksykalnego (wt- we wtorek) oraz morfologicznego (-ln- w stolnica). W ramach prac wykonano analizy zarówno manualne (oparte na introspekcyjnej wiedzy użytkowników języka, udokumentowanej serią dyrektyw operacyjnych dotyczących konkretnych rozstrzygnięć), jak i opartej na pewnych zbiorach danych typu np. słowniki derywacyjne czy fleksyjne. Ostatecznie dla dychotomicznego scharakteryzowania pięciu najczęstszych zbitek języka polskiego opracowano metodę quasi-konkordancyjną, polegającą na obserwacji zbioru słów z daną zbitką w uszeregowaniu ze względu na tę zbitkę (sort a medio, inaczej: klaster w kontekście lewo- i prawostronnym).

Podobne rozważania zasobu podjęto dla języka angielskiego. W celu stworzenia możliwie reprezentatywnej listy słów dla wykonania założonej analizy zbitek spółgłoskowych wykorzystano listę zbudowaną na podstawie korpusu COCA. Korpus ten, stworzony przez Marka Daviesa z Brigham Young University w latach 1990-2012, zawiera ponad 450 milionów słów i uwzględnia treści mówione, literaturę piękną, magazyny popularne, gazety oraz teksty akademickie. Lista słów niepowtarzalnych zawierała ok. 500.000 jednostek wraz z ich charakterystyką w obrębie części mowy. Dobór opracowany został na bazie korpusu ze stycznia 2011 (w tym czasie korpus zawierał 410 milionów wyrazów).

Model analizy zbitek spółgłoskowych zastosowany w projekcie przewidywał dla języka polskiego repertuar 31 fonemów spółgłoskowych:

p, b, t, d, k, g, ts, dz, tʂ, dʂ, te, dz, f, v, s, z, ʂ, zʂ, ɛ, z, x, m, n, ɲ, ŋ, l, r, w, j, ɰ, ʃ, a dla języka angielskiego repertuar 24 fonemów spółgłoskowych: b, d, ð, dʒ, ʃ, f, g, h, j, k, l, m, n, ŋ, p, r, s, t, θ, ʧ, v, w, z, ʒ.

Wszystkie listy słów przeznaczone do analizy polskich i angielskich zbitek były transkrybowane specjalnie na potrzeby projektu przy użyciu oprogramowania i zasobów stworzonych przez członków zespołu. System transkrypcji dla j. polskiego korzysta się z dwu zestawów po 360–380 reguł kontekstowych uwzględniających wyjątki. System dla j. angielskiego korzysta z leksykonu ok. 70.000 słów wraz z ich transkrypcjami brytyjskimi.

Tak przygotowany materiał pozwolił na zweryfikowanie 3 szczegółowych hipotez. Hipoteza pierwsza dotyczyła relacji między długością zbitki i złożonością morfologiczną. Hipoteza ta zakłada, iż, długość zbitki spółgłoskowej koreluje z jej strukturą morfologiczną. Hipoteza 2 przewiduje, iż stopień preferencyjności zbitki koreluje ze złożonością morfologiczną – zbitki morfonotaktyczne będą miały nacechowany charakter pod względem NAD, podczas gdy zbitki fonotaktyczne będą fonologicznie preferowane. Hipoteza 3 przewiduje, iż stopień preferencyjności koreluje z częstotliwością zbitki: preferowane zbitki będą częstsze w języku. W celu uzyskania potwierdzenia hipotez zbadano zbitki początkowe oraz końcowe w języku polskim oraz zbitki końcowe w języku angielskim.

Wykonano także obliczenia dotyczące prawdopodobieństwa występowania zbitek spółgłoskowych w języku polskim. Zasobem wykorzystywanym do szacowania prawdopodobieństwa była lista ok. 185 tys. form fleksyjnych. Brano pod uwagę następujące możliwe sytuacje: co najmniej jedna zbitka w słowie, dokładnie jedna, dwie itd., w danej pozycji – początkowej, środkowej i końcowej, konkretną liczbę zbitek (maksymalnie 5), oraz prawdopodobieństwo występowania konkretnej zbitki – na dowolnej pozycji i na jednej z trzech pozycji. Uzyskane zasoby wykorzystane zostały do budowy prototypu kalkulatora probabilistycznego dla zbitek spółgłoskowych w j. polskim.

Reasumując, w projekcie uzyskano dane słownikowe i korpusowe o występowaniu oraz statystycznej charakterystyce zbitek spółgłoskowych. Zbitki zostały zbadane pod kątem stopnia nacechowania opartego na uniwersalnych preferencjach fonotaktycznych. Jako narzędzie pomiaru stworzono kalkulator fonotaktyczny, który służy do oceny preferencyjności fonotaktycznych oraz morfonotaktycznych zbitek spółgłoskowych.

“The book presents a coherent, methodologically sound, and theoretically determined attempt to make sense of the types and the frequencies of consonant clusters attested in two contemporary languages, namely English and Polish. First, the book demonstrates in an exemplary manner how solid research in the language sciences needs to be conducted. Secondly, the book is an excellent representation of the work that is being done in (what I think deserves to be called) the Poznań School of Natural Phonology, and will clearly increase its already very high international visibility.

The authors’ theoretical grounding is in the framework of Natural Linguistics, as represented in Europe most prominently by Wolfgang Dressler, and by one of the co-authors of the present volume, namely, Katarzyna Dziubalska-Kołaczyk. The volume takes up an extremely interesting issue about the interactions of morphological and phonological structure, which is attracting increasing attention in the linguistic community. It contributes to the study of frequency effects, whose importance has only recently begun to be fully realized. The book represents a model case of good empirical research, and it is at the same time an excellent representation of the high quality work being done at the Faculty of English of Adam Mickiewicz University.”

Nikolaus Ritt (University of Vienna)



ISBN 978-83-232-3109-7

ISSN 0554-8144