WHICH DICTIONARY FOR WHOM?

RECEPTIVE USE OF BILINGUAL, MONOLINGUAL AND SEMI-BILINGUAL DICTIONARIES BY POLISH LEARNERS OF ENGLISH

Robert Lew

WHICH DICTIONARY FOR WHOM?

RECEPTIVE USE OF BILINGUAL, MONOLINGUAL AND SEMI-BILINGUAL DICTIONARIES BY POLISH LEARNERS OF ENGLISH

motivex
Poznań 2004

Recenzent: prof. dr hab. Włodzimierz Sobkowiak

Książka ukazuje się dzięki wsparciu finansowemu
Uniwersytetu im. Adama Mickiewicza w Poznaniu

Published by
motivex

Printed in Poland

ISBN 83-87314-42-0

Acknowledgements

It is a pleasure to acknowledge the help which I have received from various quarters during my work on this book.

My words of gratitude go first of all to the Head of the School of English, Adam Mickiewicz University, Professor Jacek Fisiak, and the Head of the Department of Lexicology and Lexicography, Professor Arleta Adamska-Sałaciak, who have both offered their continuing support and encouragement during my work on the project, including the difficult time when unexpected health problems brought my work on this book to a sudden halt.

I am indebted to Arleta Adamska-Sałaciak, Anna Cieślicka, Don McCreary, Anna Dziemianko, Philippe Humblé, Katarzyna Lew, Kristen Mackintosh, Hilary Nesi, Renata Szczepaniak, Włodzimierz Sobkowiak and Bogusława Whyatt for kindly presenting me with copies of their work or sharing their copies of other materials.

Special thanks are due to the following students and teachers who helped with the piloting and administering of the questionnaires and tests: Adam Kałużny, Agnieszka Politańska, Agnieszka Szymańska, Anna Rux, Anna Wiszniewska, Beata Kouhan, Blanka Ślotała, Ewa Zywert, Iwona Kikolska, Izabela Niklas, Justyna Tokarska, Leszek Ratajczak, Marcin Jakubowski, Marcin Jędrzejewski, Mariusz Głowacki, Natalia Kroma, Sabina Siemaszko, Szymon Nowak, Wyszomira Głowacka. Ewa Zywert deserves my special gratitude for acting as independent evaluator for a selection of questionnaire and test forms.

Jakub Saternus was very kind to provide invaluable expert advice on some highly technical aspects of relational database structuring and querying.

Anna Cieślicka, Hilary Nesi and Włodzimierz Sobkowiak generously volunteered their time to read a complete draft version of the book, providing insightful comments and helpful suggestions, for which I am very grateful. I have also benefitted from comments on parts of the book from Agnieszka Rzepa and Renata Szczepaniak.

Finally, but perhaps most importantly, I would like to thank the 712 anonymous learners of English who participated in the study as subjects, and thus contributed the data that this study is based on.

Robert Lew

Contents

0. Introduction

0.1 Context and aims of this study

This is a study in dictionary use. It focuses on several aspects of receptive dictionary use by Polish learners of English, and on the effectiveness of various dictionary types in providing receptive lexical support.

Systematic research into dictionary use has a relatively short tradition. The first significant study of dictionary use which directly involved actual dictionary users was published in 1979 by Jerzy Tomaszczyk. An even earlier study by Barnhart (1962) was also an important landmark, yet it was based on second-hand opinion by teachers on what they believed their students did as dictionary users, rather than more directly on dictionary users' reports, their performance while using dictionaries, or their look-up behaviour.

Barnhart's method based on surveying teachers of dictionary users did not attract many followers in later studies of dictionary use by learners, probably because the relationship between teachers' reports and actual dictionary use is far from direct and may have been seen as too distant and tenuous to yield reliable results. Though well-founded in general, the above reservation may not apply to certain aspects surrounding the dictionary-using act about which the teacher may actually be able to offer a more accurate report than the dictionary-using learners themselves. To take an example, if dictionaries are used in class and provided by the school, teachers are actually in a far better position to identify and report faithfully the particular dictionary titles employed by their learners in class than are the learners themselves.

There are, though, other aspects of the dictionary-using act which may not be so readily available to the teacher. In fact, the teacher may not be present in the context of dictionary use at all. While the scope of this study is restricted to dictionary use by Polish learners of English, dictionary use is not in general restricted to learners or students, or to learning contexts involving the teacher. Whatever the role of teachers may be, surveying dictionary users directly with a written questionnaire is the most popular technique of collecting data from dictionary users (Tomaszczyk 1979). The use of questionnaires in dictionary use research has come under criticism (Hatherall 1984), but remains an important and useful methodology (Lew 2002a), especially when complemented with other methods.

Systematic observation of instances of actual dictionary use, still fairly rare, focuses on the dictionary consultation act and the way the dictionary and the dictionary user interact, and may be attempted with a variety of recording techniques (videotaping, think-aloud protocols, recording sheets filled by the user or by a human monitor, computer logging), all with their own methodological chal-

lenges. Observation-based studies are difficult at the data-analysis stage, and are thus typically limited in their scope to small samples.

Performance of dictionary users can be measured by evaluating the products of their work with dictionaries. Such performance tests can be based on more or less naturalistic instances of dictionary use at a variety of tasks, and can sometimes take the form of experimental or quasi-experimental set-ups.

Between Tomaszczyk's (1979) pioneering study and now, research on dictionary use has been gaining steadily in importance, and a substantial body of research has accumulated, even though some scholars are sometimes critical of its academic quality (McCreary and Dolezal 1998) or of the general direction in which dictionary use research appears to be moving (Humblé 2001). Doubts of this kind, as well as grounds for them, are probably part and parcel of any new area of research.

In this study, largely exploratory in nature, a broad sample of English language learners as dictionary users were examined in the hope of revealing patterns that might throw some light on the little-researched area of receptive dictionary use by language learners. On top of this general goal, this study was designed with a fairly wide range of more specific research questions in mind (they are set out in section 3.1 below), and it employs a number of techniques to this end. A questionnaire (henceforth, Teachers' Questionnaire) was used to collect information from teachers on the type and level of educational institutions and learner groups from which subjects for the study had been recruited. These learner subjects were in turn surveyed with another questionnaire (henceforth, Learners' Questionnaire), yielding some basic demographic data, information on the duration and intensity of instruction in EFL (English as a foreign language), and self-assessment ratings of proficiency level. Learners provided reports on how frequently they consulted dictionaries of different types, and how often they accessed the various information categories typically offered by dictionaries. They also gave details on their dictionary preferences, and were asked to identify and rate the dictionaries they used.

The same learners who completed the Learners' Questionnaire also acted as subjects in an experiment whose principal aim was to test the effectiveness of various dictionary entries in a series of lexical comprehension tasks with varying amount of textual context (the Dictionary Effectiveness Test). The dictionary entries used in the Dictionary Effectiveness Test were specially designed to facilitate an objective comparison of the effectiveness of different dictionary types for receptive dictionary use, focusing on entry structure (*microstructure*). Receptive use is here understood (as in Scholfield 1999) as the use of dictionaries for immediate lexical support during comprehension-related tasks, without including any long-term retention or learning effects. Data from the three sources (i.e. Teachers' Questionnaire, Learners' Questionnaire, Dictionary Effectiveness Test) are combined to yield a more complete picture of receptive dictionary use by Polish learners of English.

Details of the questionnaire surveys and the Dictionary Effectiveness Test are set out in Chapter 3, and the results are presented and discussed in Chapter 4. In Chapter 1 I review the findings of previous dictionary use studies in those selected areas which are pertinent to and addressed in the present study. I raise and discuss some methodological issues relevant for this study in Chapter 2. Conclusions are presented in Chapter 5. Testing instruments with their English translations are given in the final Appendices.

0.2 Are dictionaries useful in foreign language learning?

Dictionaries are often seen as a basic tool in the process of foreign language learning. It seems that the conviction of the usefulness of dictionaries is common among lexicographers, as well as language learners themselves. Language teachers, on the other hand, appear to be more divided on the issue of dictionaries: some believe that dictionaries offer substantial lexical benefits, others fear that the consultation process is distracting and might upset the learning process (Hosenfeld 1977). Those who hold an enthusiastic view of the dictionary in foreign language learning, do so on the (often implicit) assumption that dictionaries can be helpful to the foreign language learner: after all, both by design and by actual practice, the main use of dictionaries is for lexical information, and lexical knowledge, in turn, is uncontroversially of the utmost importance in foreign language learning (Anderson and Freebody 1981). This view is also reflected in dictionaries being commonly classified among language learning aids. And yet this seemingly uncontroversial assumption has failed to receive confirmation from some well-designed empirical studies (as we shall see in 1.5.1). Not without good reason, Tomaszczyk (1987: 145) said that "dictionaries are not nearly as important to the average learner as some lexicographers and most teachers consider them to be."

However, despite the dynamic growth of the research into dictionary use since its inception in the 1960's, we are still far from getting definitive answers to many important questions regarding dictionary use, including dictionary use by language learners. Perhaps the most fundamental question of all is whether dictionaries are at all helpful to learners. Though the question may appear to be an obvious one to those unfamiliar with research into the effectiveness of dictionaries, the available empirical evidence suggests that the question is actually not at all trivial. The relevant evidence will be reviewed in section 1.5.1 below.

Further, having established – or assumed – that dictionaries can indeed offer help to learners, the next question of interest to lexicographers, language learners, and (where applicable) teachers, is which dictionary types and what dictionary features are of greatest benefit to learners. In the following section, I review some arguments for and against bilingual and monolingual dictionaries in the context of foreign language learning. I also include the semi-bilingual dictionary, a relatively new addition to the lexicographic landscape, which combines the features of the two more traditional dictionary types.

0.3 Bilingual, monolingual, and semi-bilingual dictionaries

0.3.1 Bilingual versus monolingual dictionaries and the language learner

As Piotrowski (1989: 72) and Hartmann (1994: 207) point out, it is the bilingual dictionary that has been the traditional lexical resource of the language learner. In contrast, the monolingual dictionary for EFL, and, more generally, foreign language learning, is a relatively new development (for historical details see Cowie 1999). In this section, arguments and issues related to the choice of dictionary type will be presented. The discussion is dominated by bilingual and monolingual dictionaries, as the more recent semi-bilingual dictionaries have not (yet?) reached anything like the level of popularity of the two more traditional types.

Wingate (2002) offers a good overview of the theoretical issues behind the choice between monolingual and bilingual dictionaries, while Piotrowski (1994: 71-73) looks at how this choice is affected by the assumptions typical of the various approaches to foreign language teaching. Piotrowski (1989; also 1994: 64-70) presents an insightful account of the fundamental differences between bilingual and monolingual dictionaries. Piotrowski's discussion of the provision of meaning in the two dictionary types is especially relevant in this context and will be discussed more extensively in section 0.3.3 below. Apart from the meaning-as-concept approach, Piotrowski (1989: 80-81) considers meaning provision in dictionaries from the alternative meaning-as-use point of view. Under this rubric, Piotrowski mainly raises issues concerning access to semantic information held in dictionaries.

Drawing partly on Thompson (1987), Wingate (2002: 23) lists the following as arguments voiced against monolingual dictionaries:

1. If learners use [the monolingual dictionary] for production purposes, they cannot find words they are looking for, because they do not know them. If, however, they use it for comprehension, much of the information provided, such as the grammatical behaviour of words, is not necessary.
2. The definitions may be too difficult for learners to understand.
3. Circularity can never be completely avoided.
4. Even if the dictionary has a restricted defining vocabulary, the grammatical structures can be complex.
5. Learners will not benefit from the exposure to the target language in the dictionary, because for the definitions a lexicographic metalanguage is used. This language represents a special register which does not follow the rules of the normal language (...).
6. Learners often pass over important information such as semantic restrictions of words, because they are not able to understand the basic content of words.

I will take up the above arguments in turn below.

Charge 1 above addresses the issue of access to information sought by the dictionary user. In an onomasiological dictionary consultation act, the dictionary user wants to get to the linguistic form that best coveys the ideas he or she intends to express. Since in a conventional monolingual dictionary the arrangement of entries is alphabetical, locating an unknown headword form is a practical impossibility if the orthographic form of the lexical item is not known. Further, as the access structure (Hausmann and Wiegand 1989) of a monolingual dictionary relies on the source language of the dictionary, which in our case is the learner's foreign language, the learner will not as a rule be able to locate the desired entry, because the lexical goal of the search is in principle required as the search term in the access structure, thus resulting in a vicious circle.

Piotrowski (1989: 80-81) also points out the inherent difficulty in locating the information in a monolingual dictionary in an encoding look-up situation (an onomasiological, or meaning-to-word consultation act). Piotrowski notes that monolingual dictionary users find themselves facing the paradoxical situation where they need to know the L2 item in order to look it up, but that L2 item is precisely what they do not know and are trying to find out. An adequate bilingual dictionary does not have this problem, because it uses the L1 lexical system as the framework for access structure. What is more, the L1 lexicon provides an efficient indexing system for meaning because it appears to be the native speaker's best available mnemonic for conceptualizations (Piotrowski 1994: 78).

Some types of monolingual reference works, however, such as traditional thesauri and dictionaries of synonyms, exhibit access structures with some degree of semantic organization, which can be a more helpful access facilitator than a mere alphabetic list. Perhaps thesauri can be included in the broad definition of dictionaries in a lexicographic typology, though some scholars seem unwilling to allow such a possibility (e.g. Kipfer 1987). Nevertheless, thesauri generally only offer list forms without providing semantic information beyond that embedded in the classification itself. Electronic monolingual dictionaries may be less affected by the access problem than paper-format dictionaries if they offer sophisticated search facilities that make onomasiological consultation easier, either indirectly such as through definition text searches, or more explicitly through some kind of semantic-relationship tagging (Nesi 2000a; de Schryver 2003).

There is yet another type of monolingual reference work that combines a semantically-based access structure with more detailed lexical information, such as the British *Longman Lexicon of Contemporary English* (LLCE), the American *Random House Word Menu* (RHWM), and perhaps most of all the *Longman Language Activator* (LLA). Although the idea of a semantically-based access structure is theoretically attractive, in practice creating an intuitive and efficient semantic taxonomy (or an *ontology*) of a natural language is a daunting task. But creating such a taxonomy is one thing, using it is another. Any such system would require extensive learning and sophisticated metalinguistic competence from the user, because naming the nodes of the hierarchy would necessitate the

use of some metalanguage, which in a monolingual reference work would probably have to be based on the natural language described. It seems reasonable that the natural lexical system of the user's native language would hold an advantage over such an artificial taxonomy. I suspect the average user would find it easier to think of a native language term and locate it on an alphabetically arranged list than to navigate through a complex, unfamiliar semantic network encoded in a foreign metalanguage.

On the other hand, the charge under point 1 above does not do full justice to monolingual dictionaries, because not all dictionary consultations in a production task must take the form of onomasiological queries. It may well be that the user has already decided which item to use but he or she may be unsure about the syntactic pattern or collocation to use with this word. This is a type of question that a monolingual learner's dictionary can help answer in principle, and often in practice. A bilingual dictionary can also help answer this question in principle, but not always in practice, because there are relatively few bilingual dictionaries that actually do provide this type of information.[1] Because of these practical limitations, users will sometimes decide to use two dictionaries in turn during a single search: a bilingual dictionary first, to find a foreign language item to use, and then a learner's monolingual dictionary, to get guidance on usage or seek reassurance (Varantola 1998: 184).

Issues of lexicographic access related to the use of L2-based metalanguage in the explanation of L2 items, as discussed above, are directly relevant to productive (active) dictionary use, which is not our direct concern in the context of the present work. It is worth remembering that the present study focuses on receptive dictionary use, where the monolingual dictionary is on a more equal footing with bilinguals in terms of access path, which is usually through the foreign language lemma spelling form.

The charge of the superfluity of certain information types, made in the latter part of the statement under point 1 above, is a telling example of the widespread confusion between dictionary types and dictionary products (see 2.6.2 below for discussion), since the statement refers to lexicographic information that is essentially independent of dictionary type, and thus invoking it does not seem appropriate as a general argument against monolingual dictionaries.

Charge 2 above, in turn, receives support from both questionnaire-based and experimental studies, which confirm that users often find it difficult to understand definitions or words in the definitions (Nesi and Haill 2002; Neubach and Cohen 1988; Wingate 2002: 95,115). The introduction of controlled vocabulary into learners' dictionaries does not necessarily improve the comprehension of definitions, "since these more frequent words are also the most polysemous and

[1] Those that do, include, on the Polish market, *Podręczny słownik angielsko-polski, polsko-angielski* (LongPodr) or *Longman słownik współczesny angielsko-polski, polsko-angielski* (LSW).

idiomatic in the target language" (Hartmann 1989a: 184; see also Jansen, Mergeai and Vanadroye 1987). Restricted vocabulary definitions may also lack sufficient precision (Cowie 1999: 111-112; Kirkpatrick 1985; Zöfgen 1994), and make it necessary to use complex and unnatural syntax (Carter 1987: 127; Herbst 1996; Kühn 1996; Zöfgen 1994). Neubach and Cohen (1988) quote the following comments from students to illustrate the problem with understanding dictionary definitions; it should be stressed that Neubach and Cohen's subjects were using the *Longman Active Study Dictionary* (LASD), whose definitions *do* use a controlled vocabulary:

> I don't understand this definition. What should I do – look up meanings of words in the definitions? Where does it stop?
>
> Actually the dictionary hardly ever helps me. I don't understand the definition and I feel that it hinders me more than it helps me. (Neubach and Cohen 1988: 8)

The circularity problem (point 3 above) of monolingual dictionary definitions (Calzolari 1977; Wierzbicka 1985; 1993) is theoretically interesting, but its practical consequences for the foreign-language speaking dictionary user have been overrated, in my opinion. The extent to which the comprehension of definitions is likely to be impeded by circularity is probably negligible when compared to more general comprehension problems resulting from the unfamiliarity with the foreign language lexical system, unless the level of circularity of definitions is very high due to editorial incompetence; but that does not really happen with modern learners' dictionaries. It is true, though, that bilingual dictionaries are in principle unaffected by this problem.

Point 4, the complexity of grammatical structures in lexicographic definitions, has already been partially addressed above, in the discussion of the comprehensibility of dictionary definitions. Thompson (1987: 284) puts the problem this way: "even if the defining vocabulary is restricted, the grammatical structures used are not – for example, very frequent use is made of participial clauses, with and without conjunctions, which are structures normally handled at an advanced level." In fact, it is unavoidable that the simplicity of vocabulary used must to some extent be paid for by the increased complexity of syntax, if approximately the same meaning is to be conveyed.

Point 5 appears to be a response to the claims that by using monolingual dictionaries learners may benefit from additional exposure to the foreign language in the definitions. As argued repeatedly (Hanks 1987; Piotrowski 1989; Rundell 1988), the metalanguage of definitions differs from natural language in several important respects: not just in terms of register, as mentioned in the quote above, but also in terms of lexis, syntax, collocation, and various more or less cryptic abbreviatory conventions typical of lexicographic description.

The problem described under point 6 above is again related to point 2. It is interesting to note that such comprehension problems are not restricted to foreign learners, but also occur with the native-speaking users of the monolingual dic-

tionary, as evidenced by the kidrule phenomenon[2] (Miller and Gildea 1985; 1987; Mitchell 1983b), which has also been observed with foreign learners, including advanced ones (Nesi and Meara 1994; Szczepaniak 2003; 2004; Wingate 2002).

Tomaszczyk (1983) presented arguments for the use of bilingual dictionaries by foreign language learners under four headings, and he did so in the context of foreign language learners' needs. He first pointed out that since the vocabulary of a given language is largely culture-specific, the scope of a monolingual dictionary cannot cover the needs of a speaker of another language from another culture. Bogaards (1991) expressed a similar view, pointing out that learners' monolingual dictionaries, being written with no particular first language in mind, do not address the problem of false friends. Since then, some learners' dictionaries, such as the *Cambridge International Dictionary of English* (CIDE), have tried to include information on false friends for the so called *major languages*. Such information is, however, highly redundant to a learner who only needs it for his or her native language. And, of course, the issue of false friends is just one small, albeit quite fashionable, aspect of language specificity.

The second aspect discussed by Tomaszczyk was that of interference between L1 and L2. Tomaszczyk's suggestion was that "whether one likes it or not, language learners do rely on their mother tongue to quite a considerable extent. If this cannot be avoided, why not capitalize on it?" (1983: 44). Bogaards (1991) seconded Tomaszczyk's view, advocating the comparison of words and concepts in the foreign language and the mother tongue in a dictionary.

The third of Tomaszczyk's (1983) arguments concerned the coverage of interlingual contrasts, which, Tomaszczyk pointed out, is achievable in a bilingual dictionary, but not in a monolingual one, which is not made with speakers of a particular language in mind. Tomaszczyk expressed the opinion that making the lexicographic treatment target-language-sensitive in this way would be beneficial.

The fourth and final point raised by Tomaszczyk (1983) in support of bilingual dictionaries was the marked preference of the great majority of dictionary users for bilingual dictionaries as suggested by results of questionnaire studies. This argument was based on the assumption that if the users themselves elected to use bilingual dictionaries, they must find some real value in them.

On the other side of the argument, bilingual dictionaries have been accused of a number of deficiencies. The most frequent charges can be itemized as fol-

[2] The kidrule phenomenon consists in the extraction of a readily known substring from an item's definition and treating that substring as equivalent in meaning to the item defined. For example, from a definition of the word tenet as 'opinion, belief, principle, or doctrine held as true', only the final word 'true' was extracted and used to produce the ill-formed *The news was very tenet* (Miller and Gildea 1987: 88).

lows (the list draws partially on: Nakamoto 1995; Thompson 1987: 282; and Wingate 2002: 24):

1. bilingual dictionaries reinforce learners' tendency to translate from the native language (Baxter 1980);
2. they discourage learners from thinking directly in the foreign language;
3. they reinforce the belief in one-to-one lexical equivalents between the two languages (Atkins 1985: 19; Béjoint and Moulin 1987: 100-101; Snell-Hornby 1987: 159-160, 165; Stein 1990);
4. learners do not develop their paraphrasing or defining skills (Baxter 1980: 329-330);
5. bilingual dictionaries give little information about semantically related words (synonyms, antonyms, hyperonyms), word formation, and syntactic behaviour of words (Thompson 1987; Wingate 2002);
6. there may be no equivalent in L2 to provide (Kromann, Riiber and Rosbach 1991: 2718; Snell-Hornby 1987: 165; Tomaszczyk 1983: 48);
7. the equivalent may differ from the L1 lexical item in terms of denotation, style, and other dimensions of meaning (Kromann, Riiber and Rosbach 1991: 2718).

I will address the above points in turn.

The tendency-to-translate charge presented under point 1 above may well be correct, though systematic evidence for it has yet to be presented; at this time, it appears to be based mostly on speculation and presumption. Further, one may wonder whether foreign language learners have any viable alternative to relying on their L1 knowledge, that is "does the L2 definition not merely send the learner 'back' to an item in the L1 that most closely corresponds to the referent described?" (Béjoint and Moulin 1987: 103). And if Béjoint is right in his characterization of what happens in the learner's mind, is this scenario necessarily a disadvantage? In fact, this charge is related to the next one.

With respect to item 2 above, Piotrowski (1989: 72) points out that "monolingual dictionaries seem to be indispensable within the framework of all 'direct' methodologies, which equate foreign language acquisition with thinking in the foreign language." Now, "thinking in the foreign language" may sound nice as a marketing slogan for a language school, but is it anything more than just that? "Thinking" suggests a manipulation of concepts, and concepts – according to the most popular view of the L2 learner's lexical memory as a hierarchical (layered) construct – are shared by the L1 and L2, at what is often referred to as the conceptual level of representation. The view of the shared conceptual level finds broad support from experimental evidence involving lexical decision tasks, word-recall, and semantic priming experiments (Chen and Ng 1989; Gerard and Scarborough 1989; de Groot and Nas 1991; Jin 1990; Tzelgov and Eben-Ezra 1992). If this view is accepted, "thinking in the foreign language" is a dubious concept from the psycholinguistic point of view. The asymmetry between L1 and L2 may, however, be more relevant for issues of lexical access, because these issues

are directly dependent on the nature of the relationship between concepts (shared by the two languages) and lexical forms (separate for the two languages). Here, there is the intriguing question of whether (and to what extent) L2 lexical forms are linked to L1 forms, or to concepts directly. Current work on the structure of language learners' lexicon suggests that the dependence of the L2 mental lexicon on the L1 lexicon is mediated by several factors such as foreign language proficiency, word type, and word frequency, but the degree of this dependence is substantial (Altarriba and Mathis 1997; Chen 1990; Ellis and Beaton 1995; de Groot and Keijzer 2000; Jiang 2000; Kroll 1993; Kroll and Sholl 1992; Kroll and Stewart 1994; Kroll and Tokowicz 2001; MacWhinney 1997). In light of this evidence, Béjoint's (1987: 103) suspicions about the L2 definition "merely sending the learner back to an item in the L1" appear to be well-founded.

As for point 3, the meaning, or even the (narrower) denotation of two words in two different languages is virtually never identical, except perhaps for certain technical terms in restricted specialist usage. This implies a criticism of the mode of meaning provision that is typically used in bilingual dictionaries. The important thing to realize here, though, is that monolingual dictionaries do not really have any better option to offer in that regard, as discussed earlier in this section. Furthermore, most bilingual dictionaries, except very small ones, usually offer a larger number of target language equivalents than just one per headword. When a list of L1 (target language[3]) equivalents is presented to the user under a single L2 (source language) headword, the charge that such a microstructure reinforces the picture of one-to-one lexical equivalents between the two languages appears to be misdirected. What is more, bilingual dictionaries often have sense indicators, thus further emphasizing the message that a single equivalent is not sufficient. And, because of anisomorphism between the two lexical systems (cf. Zgusta et al. 1971), a source language headword item treated with the same depth in the two types of dictionaries will typically have finer sense and subsense splits in a bilingual dictionary than in a monolingual dictionary, especially if the former attempts to provide translation equivalents, rather than cognitive equivalents[4].

---

[3] The terms *source language* (SL) and *target language* (TL) as used here and much throughout the lexicographic literature should not be confused with identically-sounding terms as used in translation studies or language acquisition. Unlike in the latter disciplines, the terms SL and TL refer to the language of the dictionary entry, quite apart from the language or languages spoken by the human user (in this book, L1 and L2 will be used for the user's native and second/foreign language, respectively). SL refers to the language of the left-hand side in a lexicographic pair of equivalent expressions, such as the entry headword, a phrase or an idiom whose meaning is being explained, while TL refers to the right-hand side, basically the semantic explanation. See Al-Kasimi (1984) and Piotrowski (1994: 21) for a fuller explanation of the terms L1, L2, SL and TL as often used in lexicography.

[4] On translation and cognitive equivalents in bilingual dictionaries see Piotrowski (1994: chapter 5).

Baxter's (1980) charge listed under point 4 above is not based on empirical evidence. The rationale presented by Baxter is that exposure to the defining language of the monolingual dictionaries would train users in their paraphrasing skills in the foreign language. One problem with this assumption is that the defining language used in monolingual dictionaries is actually a special metalanguage that passes off as the target language itself but, for most dictionaries, is actually markedly different, as already discussed above. Baxter does not seem to be aware of this important factor. Another unknown is the amount of transfer from the passive reading of dictionary definitions to the active use of language for paraphrasing. Nevertheless, a recent small-scale experimental study (Kroma 2001) suggests that indeed exposure to L2 definitions may be beneficial in the development of language learners' defining skills. Kroma exposed three groups of Polish learners of English to three types of glosses accompanying their reading texts: English definitions, Polish equivalents, and a combination of definitions and equivalents. After four weeks of such training, subjects from the three groups were asked to write their own definitions of new words. These definitions were then presented to another group of Polish learners of English at a similar proficiency level, who were asked to provide Polish equivalents corresponding to each definition, so successful definitions were operationalized as those that allowed other Polish students to accurately guess the meaning and express it in Polish. Polish equivalents provided by the second group of students were then compared with the original English words that the subjects were asked to define. Kroma found that the most successful definitions were written by the definition-only and combined definition-and-equivalent groups, with the former group achieving slightly but not significantly higher scores than the latter group. Both these groups performed significantly better than the Polish equivalent glosses group. Kroma's results are very interesting but need to be confirmed by further study, especially as Kroma did not use random assignment of subjects or evaluators (partially for practical reasons). It must also be remembered that Kroma's study concerned written defining skills, which is not quite the same thing as conversational paraphrase skills apparently meant by Baxter (1980).

Point 5 above is certainly valid, but is really a charge against the weaknesses of specific dictionary titles rather than bilingual dictionaries as a type (see 2.6.2 below for a discussion of the methodological issues involved). There is no principled reason why syntactic information in bilingual dictionaries should be less comprehensive than in monolingual dictionaries. There is, however, a possible commercial reason. Monolingual EFL dictionaries are potentially marketable in all countries where there are learners of English, irrespective of the native languages spoken there. Bilingual dictionaries, in contrast, are only useful to speakers or learners of a given pair of languages, which generally tends to be a much smaller market. When translated into corresponding sale volumes, development cost of monolingual dictionaries is spread over a larger number of units sold, thus a popular monolingual dictionary publisher can, on average, afford better lexi-

cographers, better tools (such as large corpora), and more frequent updates, which statistically speaking tends to produce better dictionary content.

The existence of isolated problematic items for which no suitable L2 equivalents exist to provide in a bilingual dictionary (point 6) is a well-known issue. However, in such cases, a bilingual dictionary can supply a gloss (definition), which is essentially the standard treatment found in the monolingual dictionary, except that in a bilingual dictionary when used for comprehension the definition is presented in the user's native language. In fact, a bilingual dictionary also has an option of offering both an approximate equivalent and a gloss.

The criticism under point 7 is right, of course, given that no two lexical items from two different lexical systems can be fully equivalent on all dimensions of meaning, except perhaps in specialist terminology; after all, different languages have their own unique ways of naming and organizing reality. Equally obviously, though, there is no solution to this problem, short of explaining an item with the exact same item, which is of course perfectly circular and completely useless. Specifically, there is certainly no perfect semantic equivalence in any case between a foreign language lexical item to be defined and its definition in a learner's dictionary, however ingenious the definition happens to be; so again, although this is a valid charge, it does not represent a disadvantage of the bilingual dictionary versus any other dictionary type (although pictorial and possibly other multimedia dictionaries might have something to offer in the restricted area of concrete noun items mainly). Finally, even those authors that are sceptical of the bilingual dictionary do concede that the problem of anisomorphism in decoding may be largely overrated, if only because one looks at individual senses rather than the totality of conceptual meaning of an item (Béjoint 1988: 145; Mackintosh 1995: 27; Snell-Hornby 1987: 167).

As seen from the above discussion, there is no shortage of arguments for and against both types of dictionaries. It is relatively easy to theorize about the hypothetical advantages or disadvantages of a specific dictionary type. The proof of the pudding, though, is in the eating, and, as will be shown in 1.2 below, questionnaire studies suggest dictionary users are not particularly avid consumers of monolingual dictionaries, unless coerced to use them.

0.3.2 Semi-bilingual dictionaries

A recent third alternative to monolingual and bilingual dictionaries is the *semi-bilingual* dictionary, sometimes referred to as the *hybrid* or *bilingualized* dictionary. The three terms are often used interchangeably. *Hybrid*, historically the earliest term, has also been used to denote a variety of other mixed-type reference works, such as encyclopaedic dictionaries. In the present context, *hybrid* dictionaries are understood as combining the source and the target language in meaning explanation, which is also what the name *semi-bilingual* suggests. The term *bilingualized*, while also referring to the use of two languages for semantic explanation, carries the extra suggestion that the dictionary has been produced as an

adaptation of a monolingual work. The emergence of the bilingualized dictionary was so described by Hartmann (1994: 207):

> the recent dogma of the English learner's dictionary as a monolingual pedagogical tool (...) becomes almost an aberration, particularly, as Henry Sweet pointed out over a hundred years ago, learners regularly seek the psycholinguistic assurance of translation equivalents by consulting bilingual dictionaries. For all these reasons, the notion of a bilingualized learner's dictionary is no longer an oddity.

The first modern bilingualized learner's dictionary, according to Laufer (1995) and Hartmann (1994), was the English-English-Hebrew *Oxford Student's Dictionary for Hebrew Speakers* published in 1986 by Kernerman & Kahn. Soon after, several other similar titles from the same publisher followed, including the *English Dictionary for Speakers of Polish* in 1990 (in cooperation with the Polish publishing house SAWW), as well as from some other publishers.

The advantage of semi-bilingual dictionaries for the foreign language learner is seen in the combination of foreign language definitions and native language equivalents. Here is how the leading publisher of semi-bilingual dictionaries describes the attractiveness of the concept:

> By providing a brief equivalent in the language of the learner, the dictionary incorporates the useful features of both monolingual and bilingual dictionaries, while avoiding their drawbacks. The learner can immerse in the English language, with active support from the mother tongue. In this way the dictionary user is encouraged to read the information in English, whereas the translation serves to provide psychological reassurance, to reinforce understanding, and to correct misunderstanding when it arises. (http://kdictionaries.com/concept.html)

Since semi-bilingual dictionaries have only been in use for a relatively short time, the circumstances and aspects of their use have not yet been studied as much as has been the case for the more traditional dictionary types. What little evidence is available, will be summarized in 1.5 below.

Having reviewed the highly divergent opinions on bilingual, monolingual and bilingualized dictionaries, it is difficult not to agree with Wingate (2002: 26) when she says that "[i]t is surprising that so far recommendations about dictionaries are based on pedagogical intuitions rather than on empirical evidence." Further, Wingate (2002: 230) mentions the teachers' acceptance of the "orthodoxy" of the superiority of the monolingual dictionary, as well as their being swayed by the label word *learner's* found in many titles of pedagogical monolingual dictionaries. No doubt, marketing plays an important role in selling the positive image of the monolingual dictionary, as well as the dictionaries themselves. The large publishing houses that specialize in international learners' dictionaries can afford the type of aggressive marketing campaigns (targeting learners, teachers, and educational institutions) that smaller, national bilingual dictionary publishers simply cannot compete with.

It is important, then, to produce objective evidence of the effectiveness of the different dictionary types in different situations and for different learners, and this is one of the chief aims of the present work. The findings of previous empirical studies of this aspect of dictionary use will be reported and discussed in 1.5 below. The combination of definition and equivalent appears to be an intuitively appealing innovation, and this is one aspect that will be subjected to scrutiny in the present study.

0.3.3 Explanation of meaning

The provision of semantic information in dictionary entries is the major focus of this study, and so the issue of how meaning is explained in monolingual and bilingual dictionaries deserves a special discussion in the present section.

Modes of provision of semantic information in dictionaries may be roughly placed into four categories: definition, equivalent, example, picture. Let us leave example and picture aside and focus on definition and equivalent. Definition is the classical lexicographic device[5], and it is the canonical method of meaning explanation in monolingual dictionaries targeted at both native speakers and language learners. In contrast, equivalent is the meaning-provision method of choice in bilingual dictionaries. A bilingual dictionary explains the lexical items of the source language (SL) using the target language (TL)[6], and it is routinely the TL equivalent that is employed. The idea of a semi-bilingual dictionary is to combine definition and equivalent.

The above account characterizes the typical situation when it comes to the utilization of definition and equivalent in monolingual and bilingual dictionaries. However, other combinations are also possible and are sometimes employed. Definitions in the L1 of the target user may in principle be – and in practice are – employed in bilingual dictionaries[7]. Such definitions are used systematically in national adaptations of learners' dictionaries, where original definitions are translated into the target language. As an illustration of such an adaptation consider the *Cobuild Bridge Bilingual English-Portuguese Dictionary* (CobuildBBEP), where the monolingual *Cobuild Student's Dictionary* definitions were translated into Portuguese, as in the following example:

> **begrudge**, se você **begrudge someone** algo, você sente que essa pessoa não merece isso e sente inveja dela por tê-lo.

Definitions in addition to equivalents are sometimes encountered in special-purpose dictionaries, such as the *Dictionnaire de la comptabilité* (DC; see Mack-

[5] The literature on lexicographic definition is extensive (e.g. Benson, Benson and Ilson 1986; Hanks 1987; Ilson 1984; Landau 2001; MacFarquhar and Richards 1983; McCawley 1993; Wierzbicka 1985; 1993; Zgusta et al. 1971).

[6] See footnote 3 on page 10 above.

[7] Definitions in the user's L1 corresponding to the TL of a bilingual dictionary were proposed as an explanation of meaning by Lev Shcherba in the 1940's (1995) and by Iannucci (1957).

intosh 1995 for further examples), a dictionary of accounting in which both the definitions and equivalents are present. One argument for the inclusion of definitions in specialized bilingual dictionaries is that they are likely to contain terms whose native language designations may be unknown (Rey-Debove, 1991: 2860). Another could be that common word-forms may be used with unusual, technical or legal designations.

Definitions in the TL are also occasionally used (as so-called 'glosses') for those entries or senses for which no acceptable equivalent can be found in the target language, or when the lexicographers could not think of one. A definition in place of the usual equivalent may or may not be distinguished typographically from the usual equivalent. Below is an example from a Polish-English dictionary (*Wielki słownik polsko-angielski*, STAG) of a definition showing in place of an equivalent, rendered in the same typography as the usual equivalent:

> **bigos** *sm G.* **~u** 1. *kulin.* dish of hashed sausage, pork and beef stewed in sauerkraut; **~ hultajski** the same dish with a variety of meat

On the other hand, some monolingual dictionaries, especially the small portable ones, utilize meaning explanation by providing synonyms rather than definitions (in the narrower sense). Synonyms in the the same language as the SL of the dictionary may be seen as *intralingual* equivalents. Here is an example from the *Collins Gem English Dictionary* (CGED):

> **nim'ble** *a.* agile, active, quick, dexterous

The above shows, then, that although there are preferences and traditions, there is in fact no simple one-to-one relationship between the use of definitions versus equivalents and the status of the dictionary as monolingual or bilingual.

Piotrowski (1989; also Piotrowski 1994: 64-70) offers an interesting view on the provision of meaning in bilingual and monolingual dictionaries, and he does so from two perspectives: as an abstract concept, and as a functional matter of practical language use. Piotrowski recounts Lev Shcherba's views (see Shcherba 1995) on the inherent inadequacy of L2 equivalents in conveying the senses of L1 headwords resulting from anisomorphism of two lexical systems (cf. Zgusta et al. 1971), but he points out that monolingual dictionaries do not fare any better in their definition-based systems of meaning provision. Piotrowski's arguments are valid and revealing. While the L1 headword and its L2 dictionary equivalent come from two different lexical systems, they nevertheless occupy corresponding positions within their respective systems, and this may give them an advantage over meaning provision by definition (see also Bogaards 1991). In contrast, definition, the standard semantic explanation device in monolingual dictionaries, requires the dictionary user to work on two levels of abstraction, the linguistic and the metalinguistic. The metalanguage of EFL dictionaries is based on English. When Polish learners of English use EFL dictionaries for comprehension, they must possess a number of fairly complex skills to successfully complete the steps involved in extracting the meaning of the problematic item. First, they need to have an appropriate command of English that will allow them

to interpret the English-based metalanguage of the definitions; second, they would have to learn the conventions of the metalanguage itself; and third, they would have to assemble the meaning of the defined word or expression from the distributed analytical components of the definitions. In the words of Piotrowski:

> Meaning [in monolingual EFL dictionaries] is described (...) in an analytical way, and the users have to synthesize the various bits into a meaningful whole: a lexical unit. It certainly is not easy and depends on the abilities, intelligence, and overall linguistic competence of the users. (Piotrowski 1989: 78)

To the above, one might add *metalinguistic* competence needed to efficiently interpret the definitional metalanguage. Of course, this just accounts for but one element in the dictionary look-up act (as described in Bogaards 1993; Hartmann 1989b: 105; 2001: 90-91; Müllich 1990; Scholfield 1982; 1999: 13-14), namely that of interpreting the semantic information for a given headword sense, which corresponds to step 6 in Scholfield (1982: 190): "Understand the definition and integrate it into the context where the unknown was met," stage 6 in Hartmann (1989b: 105) and Hartmann (2001: 91): "extract relevant data."

Piotrowski (1989: 80-81) also refers to Lyons' (1977: 438-452) distinction between first-order, second-order, and third-order words, and points out that the semantic explanation in the bilingual dictionary is based on first-order words, while EFL monolingual dictionaries rely in their meaning explanation on second-order and third-order relational words. According to this distinction, first-order words are the concrete words that most directly relate to the external world and tend to be semantic centres in texts, while second-order and third-order words provide more of a structure to first-order words. This distinction would presumably give another point of advantage to the mode of meaning provision typical of bilingual dictionaries over that used in monolingual dictionaries.

Steiner (1989) argues that the bilingual dictionary with its native language equivalents provides the most efficient path for decoding because

> [t]o understand and use these equivalents, the translator uses an internalized monolingual dictionary. (This internalization is the result of learning experiences.) If there are gaps in the retrieval of material because of the inability to recall a certain item, the translator can refer for correction or emendation to a written monolingual dictionary, to an informant, to a thesaurus, or to an encyclopedia. (Steiner 1989: 255)

Although Steiner's arguments are made in the context of L2→L1 translation, they appear to be almost as valid for other L2 decoding tasks, except that the dictionary user may not have a need to *use* the L1 equivalent, it may be sufficient for him or her to understand it.

1. Studies of receptive dictionary use: findings

Several excellent, up-to-date overviews of research in the field of dictionary use are already available (Hulstijn and Atkins 1998; Nesi 2000b; Tono 2001; Wingate 2002). It would thus be wastefully repetitive to present here a comprehensive overview of all the published literature on dictionary use. In what follows, I selectively concentrate on the issues particularly relevant in the context of the present study.

1.1 Frequency of dictionary consultation

Answering the seemingly innocuous question of *How often do users consult their dictionaries?* is much harder than it might at first appear. Tomaszczyk (1987: 140), when discussing translation errors in written texts, identified a general "unwillingness to consult reference books at all." Indeed, the short and sweet answer to the question posed at the beginning of this paragraph appears to be *not enough*. User questionnaires have addressed the issue of the frequency of dictionary consultation by users, but findings obtained in this way have to be viewed with caution, since their factual correctness requires reliable long-term recollection by respondents of statistical facts involving a rather unremarkable activity. In fact, this could be one area where Hatherall's (1984) otherwise rather one-sided objections to the questionnaire technique may be warranted, especially as the data on consultation frequency returned from questionnaires tend to be of the relative type (such as 'often', 'rarely', etc.), and do not straightforwardly translate into unequivocal statements of the absolute kind.

The first major study of the dictionary consultation habits of non-native dictionary users was Tomaszczyk (1979). Amongst the 449 subjects taking part in this study were foreign language learners (284 subjects), teachers and professional translators. As many as 16 foreign languages were represented in the study (the highest numbers being for English, Russian, Polish, French, and German). The majority of the learner subjects (167 out of 284) were Polish learners of foreign languages. Writing later about this study, Tomaszczyk (1987: footnote 5) reported: "When developing the questionnaire for the 1979 study I spoke to many successful language learners, from various language backgrounds, who insisted they had never used any dictionaries."

Available results on the relationship between the frequency of dictionary consultation and learner level are contradictory. Some studies (Hatherall 1984; Knight 1994; Wingate 2002) found the frequency of dictionary use to increase with level, as recorded during dictionary use tests. Other studies (Atkins and Varantola 1998a; Tomaszczyk 1979) revealed a reverse tendency, with lower-proficiency users tending to use their dictionaries more frequently. Atkins and Varantola (1997; 1998b) monitored dictionary use in translation by a group of 71

ESL speakers from fifteen different language backgrounds, and found no consistent pattern across the range of L2 skills in their sample. Intermediate users registered the highest rates of dictionary use, while beginners appeared to have consulted their dictionaries the least. Advanced users ranked between the intermediate group and the beginners. The differences between the group frequencies were not large; no statistical significance measures were reported.

Jakubowski (2001) investigated the use of bilingual and monolingual dictionaries by Polish high school learners and found that learners used dictionaries with an average frequency of a few times a week. He also found the frequency of use to be higher for students of higher proficiency level.

Questions about the frequency of dictionary consultation are often asked in the context of reasons for dictionary look-up, and here Hartmann (2001) further distinguishes between the activity task that is accompanied by dictionary look-up to meet specific reference needs – all part of the context of the dictionary consultation act – and the information category that the user looks up in the dictionary, which is part of the dictionary text. The distinction between the dictionary text and the context of the consultation act, as well as the study of user needs as a whole, owes much to Wiegand's (1977b: 81) initial call for a study of the social context of dictionary use, or what is sometimes referred to as the sociology of dictionary use and of dictionary user (Hartmann 1989b).

Most early studies of dictionary user needs were done with native-speaking dictionary users (Greenbaum, Meyer and Taylor 1984; Kipfer 1987; Quirk 1974; Summers 1988), or – indirectly – with their teachers (Barnhart 1962), and it was general-purpose monolingual dictionaries that were targeted. The most striking similarity as regards the findings of those studies was that of the preference for semantic information in dictionaries (see 1.4 below).

1.2 Dictionary preference

One of the principal findings of Tomaszczyk (1979: 104) was that "[a]lmost all subjects, no matter how sophisticated they are, use bilingual dictionaries." Monolingual dictionaries were used much less than bilingual dictionaries, even though most of Tomaszczyk's subjects were either advanced students or language professionals: instructors and translators. The preference for bilingual dictionaries was also evident in the responses of those 228 subjects in Tomaszczyk's sample who possessed both types of dictionaries.

Similarly, in a survey of 342 Japanese learners of English, Baxter (1980) found that a great majority of learners, even at the university level, preferred bilingual dictionaries. However, Baxter's questions referred to dictionary ownership rather than use.

Not all studies, though, have found bilingual dictionaries to be preferred to monolinguals. In Béjoint's (1981) study only 17% of the subjects claimed to prefer bilingual to monolingual dictionaries. Béjoint's questionnaire, however, was

exclusively concerned with the use of monolingual dictionaries, which may have shaped the responses of the subjects (French students majoring in English) to a considerable extent. In addition, as many as 85% of the students in the study had chosen their dictionary following the recommendation of their teacher, so one could say that there was something of a school policy involved here. Likewise, high-proficiency students investigated by Neubach and Cohen (1988) preferred the monolingual dictionary, claiming "more precise meanings of words" to be the reason for their preference.

In their study of dictionary use in translation, Atkins and Varantola (1997) again found a clear preference for bilingual dictionaries over monolingual dictionaries, with 71% of the look-ups recorded being performed in bilingual dictionaries, versus 28% look-ups in monolingual dictionaries (a small number of look-ups did not have the dictionary type specified). The dominance of bilingual dictionaries was greater for L1→L2 translation than for L2→L1 translation. This means that monolingual dictionaries were used somewhat more in translation into the subjects' native language than in translation from the subjects' native language, while bilingual dictionary remained the reference tool of choice in both types of translation. Translation, the complex task that it is, involves a combination of decoding (of the source text) and encoding (of the target text). However, unless the L2 proficiency of the translator is very high (which was not the case in the study, at least not for the Oxford participants, whose results are being discussed here), it is the L2-related activity that presents the greater challenge, so it can be assumed that the difficult part of L1→L2 translation primarily requiring lexicographic support is in the encoding, while for L2→L1 translation most of the difficulty is in the decoding, though for technical texts encoding might also pose problems. The results obtained by Atkins and Varantola (1997) generally confirm the preference for bilingual dictionaries. When monolingual dictionaries are consulted at all, it is for decoding rather than for encoding. The likely reasons for this have been discussed in section 0.3 above.

In the large-scale EURALEX/AILA Research Project on Dictionary Use (Atkins and Varantola 1998a), a general preference for bilingual dictionaries was found, with well over half of the subjects (learners of English as a foreign language) electing to use a bilingual dictionary, and only a minority opting for a monolingual dictionary (see also 1.2.1 below for other details).

Looking at the types of semantic information preferred by dictionary users, Laufer and Kimmel (1997) reported that Hebrew EFL learners, when given the choice, consulted the Hebrew equivalent more often than they did the English definition. Laufer (2000) gave her subjects, advanced learners of English as a foreign language in Israel, a choice of information to be displayed on screen during a reading task as any combination of L1 translation, English definition and example. Only three combinations out of the seven possibilities were actually used, and over 90 percent of all selections were for L1 translation only. A combination of translation and definition was selected in seven percent of the cases,

with a three-way entry of translation plus definition plus example accounting for the remaining less than three percent of the selections.

Overall, results of published studies indicate that users exhibit a marked degree of preference for bilingual dictionaries over monolingual dictionaries. In the following section, I briefly examine the question of whether this relationship is affected by user proficiency level.

1.2.1 Dictionary preference and proficiency level

A number of studies (Al-Ajmi 1992; Atkins and Varantola 1998a; Baxter 1980; Jakubowski 2001; Tomaszczyk 1979) noted a tendency for higher-proficiency learners to use monolingual dictionaries relatively more often compared to bilingual dictionaries, although bilingual dictionaries usually remained the dictionary type of choice at any proficiency level. There is also the question of the different situations and tasks for which various dictionary types are employed. For example, Atkins and Varantola (1998a) reported that the principal use of monolingual dictionaries by the higher-proficiency students in her sample was to aid in the comprehension of L2 expressions and to help with using a known item in production. However, subjects in the study were not free to independently select dictionary types for the different tasks. Overall, the proportion of learners preferring bilingual dictionaries decreased with proficiency, from 88% for Grade D students (lowest proficiency level in the study), to 60% for Grade A (highest proficiency) students. The intermediate grades C and B registered rates of 81% and 79%, respectively. Atkins and Varantola (1998a) do not test for statistical significance of these differences, but the figures point to a tendency for the use of monolingual dictionaries to increase only moderately at lower and intermediate levels of proficiency, with a sharper rise for the highest levels.

Jakubowski (2001) looked at the use of bilingual and monolingual dictionaries by Polish secondary school learners at two levels, and found, at both levels, a strong preference for the bilingual dictionary in L2→L1 and L1→L2 translation and writing, but less of a preference in reading and listening tasks, and least of all in speaking. It should be stressed in this connection, though, that exercising oral skills is definitely not a salient context for dictionary use in general.

What little evidence we have suggests that while there exists amongst language learners a general preference for bilingual dictionaries, the preference is less marked for higher-proficiency learners.

1.3 Dictionary evaluation

Monolingual dictionaries have been found to be more highly regarded by users than bilingual dictionaries (Baxter 1980; Béjoint 1981; Kharma 1985; Tomaszczyk 1979). This effect could be due to the users' inability to perceive the failings in a dictionary whose semantic explanations are not fully understood. That this is a likely factor is confirmed by the finding (Wingate 2002) that users evaluate bilingual dictionaries involving two languages other than their native

language less critically than they do bilingual dictionaries including their native language as either the source or target language. Wingate's study involved Chinese as the native language, English as a second language, and German or French as foreign languages. Another reason could be ideological, as Nesi (2000b) suggested while commenting on Tomaszczyk's (1979) study:

> Tomaszczyk refers to the school of thought that has condemned the use of the first language in the second language classroom, and has disapproved of bilingual dictionaries. It may be that Tomaszczyk's subjects had been influenced by this attitude, feeling that in some way the monolingual dictionary was superior, despite the difficulties they encountered when consulting it. (Nesi 2000b: 6)

Cowie (1999: 184) rightly points out that part of the learners' enthusiasm for the monolingual dictionary mimics that of their teachers, whose recommendations are valued highly and followed unquestioningly by large proportions of learners (Bareggi 1989: 79%; Béjoint 1981: 85%). As a result, "a wide gap often exists between a student's perception of the dictionary's value and its actual usefulness as an aid to learning" (Cowie 1999: 184).

Within the bilingual category itself, L2→L1 dictionaries have been reported as earning slightly higher user ratings than L1→L2 dictionaries (Tomaszczyk 1979).

1.4 Information categories consulted

Word meaning is the information category which has consistently (though not exclusively: Harvey and Yuill 1997) been identified as the one most frequently sought by learners (and other users) when consulting dictionaries (Battenburg 1991; Béjoint 1981; Bishop 1998; Galisson 1983; Hartmann 1982; 1999a; Jakubowski 2001; Li 1998; Marello 1989; Nuccorini 1992; Snell-Hornby 1987; Summers 1988; Wingate 2002). These consistent findings point to "the dictionary user's overwhelming preoccupation with meaning" (Cowie 1999: 181). Of the above studies, Battenburg (1991) is the only survey where data are crosstabulated according to proficiency level (Battenburg 1991: 94-95; see also Tono 2001: 48 for some statistical testing of Battenburg's figures). The percentages of native speaker users who declared meaning as the primary reason for dictionary use typically ranges from 40 to 75 percent, although the measures from the different studies cannot be compared directly due to substantial differences in data collection protocols.

Apart from the unquestionable dominance of semantic information, it is hard to rank the remaining types of information, because collecting such data is difficult and complex, and the results of the different studies are hard to compare directly (Bogaards 1988: 137). The problem exists even in studies that are meant as replications (see Lew 2002a for a discussion of these issues). To give an indication of how widely the declared usage figures can vary, synonyms and antonyms received 72 percent in one survey (Quirk 1974), but only 14 percent in another

similar survey (Greenbaum, Meyer and Taylor 1984) designed specifically to replicate the former study. In most cases, comparing results of different studies is even more problematic, as questionnaire items differ in the number of options available, in the phrasing of the different options, and in whether multiple responses or only single responses are accepted. For example, as far as semantic information is concerned, questionnaires may or may not distinguish between equivalents and definitions, both serving to supply lexical semantic information. Clearly, the breakdown of responses will depend on the categorization and clustering of the available options. Also, there may be a significant ordering effect at play here. For example, the recording sheet in Harvey and Yuill (1997) included an eight-item table listing the suggested reasons for looking up an item, the top of which looked as follows (Harvey and Yuill 1997: 276):

01	To find out if the item exists.
02	To check how the item is spelt.
03	To check the meaning I want.

It is with the above order of presentation in mind that one should evaluate the results of Harvey and Yuill's survey (see below). At the same time, it needs to be pointed out that the relative need for specific information types is likely to depend on the particular task that the dictionary is being used for. Harvey and Yuill (1997) investigated the use of the Collins COBUILD English Language Dictionary by learners of English engaged in composition writing. Their subjects reported that during this task they most often looked for spelling information (24.4% of all look-ups), with meaning only in the second position (18.3%). Finding out whether the item existed was reported as the third most important reason (12.8%). Clearly, one might suspect the order effect to have been partially responsible for the somewhat unexpected ranking of responses, by giving a positive boost to spelling information relative to meaning, and in promoting the response listed as number one, which is an option rarely seen in other similar studies. Synonyms and grammar practically tied in the fourth place (10.6% and 10.5% of look-ups, respectively). Further choices were, in turn: register (9.3%), collocation (8.2%), and inflection (5.9%). Quite interestingly, synonyms and antonyms were the information types that differed most sharply in the ranking of information types for monolingual and bilingual dictionaries in Bogaards (1988). Admittedly, it may be quite difficult to determine to what degree the different results are artefacts of different methodologies, techniques, and materials, and to what extent they reflect actual differences.

The above difficulties notwithstanding, spelling and grammatical information usually rank fairly high among the categories of information sought in dictionaries, though spelling tends to be wanted more by native speaking dictionary users than by language learners (e.g. Béjoint 1981), whereas the reverse appears to be the case for grammatical information. As far as spelling is concerned, its status vis-à-vis other types of lexical information is unique, because spelling information is automatically given as long as word forms are spelled out for dic-

tionary entries. This may not be forever into the future, but for now the orthographic representation as the basic citation form seems to rule supreme, although it does sometimes differ from conventional spelling encountered in normal texts, such as through the addition of special typographics to indicate vowel pronunciation, stress placement, or hyphenation. So, while dictionaries without semantic information abound, such as pronouncing dictionaries or spelling dictionaries (the latter type being very popular with Polish dictionary users), dictionaries deprived of spelling information are clearly not a viable option for the time being. In the modern literate society, spelling seems to be indispensable and carries a fundamental indexical-representational function.

Sobkowiak (1999: 115-121) presents a detailed overview of the reported interest in pronunciation information in dictionaries. The general pattern that emerges is that dictionary users' interest in pronunciation is low. Sobkowiak's own questionnaire did reveal a rather high level of interest in pronunciation, but his subjects were all English majors, more advanced than in most other studies.

Somewhat surprisingly, Bareggi (1989) found pronunciation to be the most sought-after information category among Italian learners of English at their first year of university using monolingual learners' dictionaries. However, it is apparently bilingual dictionaries that the students primarily used for meaning information, so these results do not indicate that pronunciation information was in general wanted more than meaning.

The results of those few studies that controlled or recorded the dictionary users' L2 proficiency level (Atkins and Varantola 1997; 1998b; Battenburg 1991) suggest that the users' interest in the different types of information tends to vary with their L2 proficiency level, with perhaps the more proficient users exhibiting more interest in the (on average) less common information types. In a study of written translation (Atkins and Varantola 1998b), only 1% of look-ups conducted by beginners were aimed at locating collocational information, compared to 15% of look-ups by intermediate users and 10% by advanced users. Further, Atkins and Varantola (1998b: Table 4) report that no look-ups were recorded for "other types of information" (a catch-all category for the remaining non-central types of information) in the beginners group, while "other types of information" were sought in 4% of look-ups by intermediate proficiency users, and 5% of look-ups by advanced users. However, these figures may be misleading, because subjects in the study had free choice as to the direction of the translation tasks: the majority elected to translate into L2, while a minority chose to translate into their own language. As shown in one of the tables (Atkins and Varantola 1998b: 96), all of the "other types of information" look-ups, and a disproportionate majority of the collocation look-ups, were conducted during L1→L2 translation. Further, one might expect that less proficient subjects were relatively less likely to select L1→L2 translation and relatively more likely to prefer L2→L1 translation instead, since the latter is (rightly) seen as an easier task. In fact, the authors do note themselves that "most of the Beginners chose to do a translation into their

own language rather than into a foreign language, and that fact must also be reflected in these figures" (Atkins and Varantola 1998b: 97). The difference in the frequencies of look-ups that appear to be explained in terms of user proficiency level may in fact be due to the different task being selected. It is a pity that the data were not cross-tabulated by these two interacting variables.

Finally, etymology is one information type that has consistently registered very low interest figures in user questionnaires (Battenburg 1991; Béjoint 1981; Greenbaum, Meyer and Taylor 1984; Quirk 1974; Tomaszczyk 1979). Also, surveys indicate (e.g. Atkins and Varantola 1998a) that introductory matter is read only very infrequently, but introductory matter is not part of the entry structure, so its status in a dictionary is distinctly different; nor is introductory matter lexical information, but rather metalexicographic.

1.5 Dictionary effectiveness

1.5.1 Is there any use in the dictionary at all?

The views on the usefulness of dictionaries are mixed (Bensoussan and Laufer 1984; Laufer 1993). Hilary Nesi describes the situation well:

> Implicit in the design of the questionnaires (...) was the assumption that a dictionary was a useful tool for students. (...) It remained to be proved, however, whether dictionary use brought these students actual benefit, or was merely the result of habit, supported by their teachers, and the publishers who promote dictionary purchase. (Nesi 2000b: 55)

Studies of receptive dictionary use for lexical support have most commonly focused on one or more of the following aspects: immediate vocabulary comprehension, short-term recall, and long-term recall, although more original set-ups have also been used. Vocabulary comprehension has typically been measured in reading comprehension studies, and vocabulary recall has been the measure of choice in studies of vocabulary learning with the use of dictionaries.

The present section will review the results of studies of dictionary effectiveness in receptive use under two headings that appear to have received most of the attention of dictionary use researchers: vocabulary comprehension (see 1.5.1.1) and vocabulary learning (see 1.5.1.2). Following these two subsections, I will report on the results of studies comparing the effectiveness of several dictionary products or types (see 1.5.2 below).

The effectiveness of dictionaries in encoding use (production) has been investigated in a number of studies (Ard 1982; Atkins and Varantola 1997; Harvey and Yuill 1997; Hatherall 1984; Laufer and Hadar 1997; Meara and English 1988; Nesi 1994; Nesi and Meara 1994; Nuccorini 1994), but this lies outside the focus of the present work and will not be pursued here.

1.5.1.1 Vocabulary comprehension

A seminal paper by Bensoussan, Sim and Weiss (1984) reported no effect of dictionary use on comprehension test scores. The investigation used a very large sample (around 1500 subjects overall) and a careful methodology, with three independent studies at two institutions. In one of the three studies subjects had control over their choice of dictionary, and in this case it might be claimed that the self-selected sample of dictionary non-users could differ systematically from dictionary users in that it could include an unfair proportion of the higher-ability and more confident students, whose lack of dictionaries during the test might have been compensated by their better language skills. However, in the other two studies there was random assignment of dictionaries, so such a hypothetical effect could not be used to explain away the lack of difference in those cases. Learners in the study used institution-assigned dictionaries: perhaps the results would have been different, had they been using their own dictionaries with which they were more familiar. The finding that dictionaries did not help the test scores was described as surprising by the researchers themselves. Later replications (Nesi 2000b, Chapter 2; Nesi and Meara 1991) generally confirmed the results, and pointed to three sources of reasons for the failure of dictionary access to have positively affected the test scores: first, the comprehension test questions did not in their majority rely on the comprehension of individual words; second, the dictionary used in the test did not always have the information needed to answer the comprehension questions; and third, the users did not look up the relevant items. The effectiveness of dictionaries for word comprehension was also the topic of a study by Neubach and Cohen (1988), who found that dictionaries – either monolingual or bilingual – did not appear to be of any help to users on reading comprehension tasks. Another study by Tono (1989; 2001: Chapter 5) found a positive correlation between advanced learners' dictionary using skills and their success with reading *without*, but not *with*, a dictionary. This somewhat surprising finding is interpreted by Tono as evidence that dictionary use might have a long-term effect on overall reading comprehension ability. An earlier study by Padrón and Waxman (1988) found a negative correlation between the amount of dictionary use reported by ESL learners and improvement in reading proficiency.

In contrast, a test of dictionary effectiveness conducted by Longman (Black 1986; Summers 1988) found that dictionary entries were helpful for word comprehension by adult ESL learners, with blank entries used as controls. Three types of entries were used: definitions only, examples only, and definitions followed by examples; though no difference was found between the three types, all three improved subjects' performance in comprehension (as well as in production).

Knight (1994) used on-screen dictionary entries to assess the effectiveness of dictionaries in vocabulary learning by Spanish ESL learners. She found that access to dictionary entries significantly improved subjects' performance on immediate supply-definition and select-definition tasks, which may be taken as a

measure of vocabulary comprehension. However, when learners were stratified into two language proficiency groups, it was only the lower-proficiency subjects who benefited from dictionary access, and no significant effect was found for high-proficiency learners. This last finding is consistent with Bensoussan et al.'s (1984) results, where all subjects were at a relatively high proficiency level. Knight's (1994) main focus was on vocabulary learning, for which see 1.5.1.2 below.

McCreary and Dolezal (1999) tested the usefulness of *The American Heritage Dictionary* (AHD2) for foreign learners of English at an American college during a vocabulary test. The performance of three groups of learners was compared: Group 1 had access to the dictionary; Group 2 did not have dictionary access, but were given a short story to read that included the tested vocabulary; Group 3 had both the story and dictionary access. The three groups scored 69%, 68%, and 79%, respectively, on a multiple choice vocabulary test. Group 3 performed better than either Group 1 or Group 2 in terms of overall vocabulary scores, with no difference between Groups 1 and 2, suggesting that dictionaries are not helpful on their own, but are useful in the presence of contextual information. The testing instrument used by McCreary and Dolezal consisted of two types of multiple-choice items: synonym identification and antonym identification. It is striking that the mean scores on the two types of items patterned differently (McCreary and Dolezal 1999: Table 1): Group 1 appears to have outperformed Group 2 on the synonym questions (78.3% versus 71.7%, respectively), while on antonym questions the ranking was reversed (43.5% versus 62%). No statistical measures were provided for these intriguing differences or any group by task interactions, but given that Group 1 was the one with access to dictionaries, a possible interpretation might be that AHD2 was especially successful in providing information on synonyms. Since AHD2, like many monolingual dictionaries, tends to rely heavily on synonyms (and synonym-like hyperonyms) in their provision of meaning, it could be that synonym-based test items might have given an advantage to AHD2.

Albus et al. (2001) tried to answer the question of whether access to a monolingual learner's dictionary[8] would improve reading test performance of limited English proficient (LEP) students from the Hmong community attending three urban middle schools in a large metropolitan area of Minnesota, USA. The study found no significant overall effect of dictionary accommodation. When reading proficiency in English was taken into consideration (proficiency level was based on self-reporting), a significant benefit of dictionary accommodation was found for the intermediate-level students, but not for lower or higher level proficiency students. Interestingly, a very high proportion of 95.8% of the Hmong subjects believed that having access to an English dictionary would help them to perform better on a reading test.

[8] *The American Heritage English as a Second Language Dictionary* (AHESLD) was used in the study, referred to by the authors as a "simplified English" dictionary.

Bogaards (2002) tested two monolingual dictionaries for learners of German on a sample of 27 Dutch learners of German attending a pre-university section of a secondary school. The study focused on two aspects: the findability of lexical information and vocabulary comprehension while reading. Bogaards asked his subjects to translate the target words into Dutch. For both dictionaries, statistically significant differences were found indicating that the dictionaries helped students in the understanding of written texts, though there were no significant differences between the two dictionaries. The rate of success for both dictionaries was around 35%. Bogaards (2002) does not directly report the rate of success to compare the performance with the dictionary to performance without dictionary access, but the figures for the individual words (Bogaards 2002: Table 1) suggest that it was probably below 20%.

Szczepaniak (2003; 2004) investigated the comprehension of contextually modified idioms with the help of the third edition of the *Longman Dictionary of Contemporary English* (LDOCE3), using highly advanced Polish learners of English as subjects. Szczepaniak found a mildly positive influence of LDOCE3 on comprehension performance. Because her subjects completed the same task prior to and following dictionary consultation, Szczepaniak obtained data that linked the performance at those two points in the procedure. The results are not exactly encouraging: in 72% of all look-ups, subjects changed their responses following dictionary consultation, but most of these cases (50% of the total) resulted in incorrect responses, and 7% of the total went from a correct response to an incorrect one following dictionary consultation. In 28% of the look-ups, dictionary consultation did not effect a change in the response. Within that number, only 6% of all cases were confirmations of previously correct ideas, and 22% of the total were retentions of previously wrong ideas. On the positive side, the mean success score for all four test items went up from 0.18 prior to dictionary consultation to 0.27 following consultation, an increase by 50%. It is difficult to say, though, how much of the improvement can be ascribed to dictionary consultation, and how much of it occurred as a result of a second examination of the same contexts.

The findings of dictionary-supported comprehension studies discussed above may be taken to indicate that the effectiveness of dictionaries could depend on the users' proficiency level. Advanced learners may not benefit much or at all from dictionary access, while low-proficiency learners may not possess the skills required to locate and extract the relevant information from entries. If this is indeed the case, dictionaries might be of most help to intermediate-proficiency learners. However, such a conclusion remains extremely tentative, because it is based on a small number of studies, and rarely ever is a broad spectrum of learner levels evaluated in one single study. Any differences, then, might be due to numerous factors unrelated to subjects' proficiency level that happened to vary across the different studies. One such aspect is the measure used to assess the degree of success in reading comprehension. Even if structurally similar evaluation

instruments are used, such as the most frequently employed multiple-choice comprehension questions, it would be inappropriate to assume equivalence between the different tests, as items can depend to a greater or lesser degree on overall text comprehension or the comprehension of individual vocabulary items, and such balance is difficult to control.

1.5.1.2 Vocabulary learning

The effectiveness of dictionaries in vocabulary learning is perhaps the most popular strand of research amongst studies dealing with dictionary effectiveness.

Luppescu and Day (1993) found that the use of a bilingual dictionary facilitated the learning of vocabulary by Japanese EFL learners. Similarly, Knight (1994) found that access to bilingual dictionary entries significantly improved delayed (after two weeks) vocabulary recall scores, which may be interpreted as reflecting vocabulary learning.

Hulstijn, Hollander and Greidanus (1996) investigated the effect of dictionary use and marginal glosses on incidental vocabulary learning by Dutch advanced learners of French. A French-Dutch bilingual dictionary was used, and the glosses were also bilingual: they were Dutch equivalents of the French items. Overall, word retention rates in the group of subjects with dictionary access were not significantly higher than in the control group, though marginal glosses did produce significantly higher retention scores. However, subjects in the dictionary group only looked up an average of 12 percent of the target words tested (see section 2.6.1 below for a discussion of the problem of dictionary underuse). Retention rates on those items only for which dictionaries were consulted turned out to be actually higher than for the same words in the marginal glosses group. This could mean that dictionaries are very effective when used, but it could also point to those items that were looked up being special in some way, which might facilitate their retention. Hulstijn, Hollander and Greidanus suggest text relevance might be a factor.

See section 1.5.2 below for a discussion of other studies of the effectiveness of dictionaries in vocabulary learning involving a comparison of two or more dictionary titles or types.

1.5.2 Comparison of the effectiveness of various dictionary types

The current state of research on the effectiveness of various dictionary types is aptly captured by Wingate (2002) in the following statement: “as far as the effectiveness of either the bilingual or monolingual dictionary for learners is concerned, there has been an ongoing discussion with hardly any research basis.”

In a well-controlled experimental study, Oskarsson (1975) compared students’ vocabulary learning success using monolingual glossaries and bilingual

glossaries[9]. Bilingual glossaries were found to aid vocabulary learning to a significantly greater degree than monolingual glossaries. Further, the use of monolingual glossaries "did not result in a better ability to produce the critical words in response to questions derived from the target-language definitions" (Oskarsson 1975: 31).

Summers (1988) tested EFL learners with three types of entries: examples only, abstract definitions, and a combination of definitions and examples, against blank controls, and found no differences between the three types of entries, with all three being significantly more effective for vocabulary comprehension than the control.

In a pretest-posttest study of English vocabulary learning while reading by 52 Swedish undergraduates majoring in English and economics, Krantz (1991) found no difference in the effectiveness of monolingual and bilingual dictionaries. However, the rate of dictionary use was relatively low, with a mean of just 1.7 look-ups per page of reading.

Bogaards (1991) tested 44 Dutch-speaking first-year university students of French under four dictionary conditions: a bilingual dictionary, a learners' monolingual dictionary, a monolingual dictionary for native speakers of French, and no dictionary. On a translation task, the bilingual dictionary group achieved the highest scores, and the control (no dictionary) group the lowest scores. On a delayed vocabulary translation test administered fifteen days later, the learners' dictionary group registered the highest mean score, followed by the bilingual dictionary users, but the differences in scores were minimal. The users of the monolingual dictionary for native speakers ranked third, and subjects with no access to dictionaries scored the lowest of the four groups again.

Laufer (1993), in an extended follow-up study based on Laufer (1992), investigated the effect of dictionary definitions and examples on the comprehension and production of new vocabulary items in a carefully designed experimental set-up. Laufer found that a combined (definition and example) entry was more effective for comprehension than definition alone, which was, in turn, more effective than example alone. Though the present study does not address the contribution of examples to dictionary entries, Laufer's findings indirectly show that dictionary entries can be helpful, or else no differences between the different entry types would have been found.

In a series of studies, Batia Laufer and her collaborators (Laufer 1995; Laufer and Hadar 1997; Laufer and Melamed 1994) compared bilingualized dictionaries against monolingual and bilingual dictionaries. The results of these studies point to an advantage of the bilingualized dictionary over the monolingual dictionary for comprehension. The bilingual dictionary was not significantly

[9] Although glossaries are not the same as dictionaries, the differences are primarily macrostructural. Because this study focuses on the microstructure, and Oskarsson's findings were interesting and relevant, the study was included in this overview.

different from either the bilingualized or the monolingual dictionaries. In production, the bilingual dictionary performed better than the monolingual. Laufer and Melamed (1994) classified their subjects into three groups by dictionary using skills, based on overall scores, and they found two tendencies: bilingual dictionaries resulted in relatively best performance for the least skilled users, and the differences between the dictionaries tended to decrease with growing skill. As in most such studies, Laufer and Melamed (1994) did not control for the actual presentation of information in the three types of entries, as entries from three actual titles were used (Laufer and Melamed 1994: 567). Such a methodology makes it difficult to generalize the results to dictionary types (see 2.6.2 below for discussion).

Raudaskoski (2002) asked twenty Finnish senior secondary school students to complete translation assignments, eight from English into Finnish and eight from Finnish into English, first without the aid of the dictionary, and then with the use of one of two different dictionaries: a large general bilingual dictionary and a much smaller bilingualized dictionary. Raudaskoski reports that the performance of the bilingualized dictionary users improved more between the two rounds of translation than did that of the bilingual dictionary users, but he also warns that the sample was small and "it is impossible to make any universal statements" (Raudaskoski 2002: 3). Unfortunately, the degree or difference in improvement between the two groups is not reported, and no attempt at statistical evaluation is claimed, which is understandable in view of the small sample size. Raudaskoski also reports that both dictionaries helped the English-Finnish translation better than the Finnish-English translation. Somewhat surprisingly, working without the dictionaries, subjects did better on translation into English, their L2, than into Finnish, their L1. The two dictionaries used in the test clearly differed not only in terms of type and size, but also in terms of lexicographic presentation: "[t]he bilingual dictionary, with its dense entries full of symbols and abbreviations, caused difficulties for many students, especially when the necessary headword or equivalent was concealed inside a long entry" (Raudaskoski 2002: 3), so again, the findings cannot be generalized to dictionary types.

Atkins and Varantola (1997; 1998b) monitored dictionary use during translation. Among other measures, subjects were asked to rate the success of the individual look-ups. Of the look-ups in bilingual dictionaries, 64% were judged as successful, 35% as unsuccessful (there was a small number of unspecified responses). Monolingual dictionary look-ups were reported by subjects as successful 48% of the time, while 52% were rated as unsuccessful. However, in many cases monolingual dictionaries were used to follow up on previously unsuccessful searches in bilingual dictionaries. Because of this, monolingual dictionaries may have received a larger proportion of particularly problematic queries, and this effect could have negatively affected the success ratings of monolingual dictionaries. Also, mean success ratings from advanced L2 proficiency subjects were slightly lower than those coming from intermediate subjects and beginners.

Given that the use of monolingual dictionaries is often seen to increase with proficiency level, the ratings of monolingual dictionaries could also have been lowered in this way.

In a similar study using a modified version of the same procedure, Varantola (1998) distinguished between monolingual dictionaries for native speakers and monolingual learners' dictionaries. The two subtypes of monolingual dictionaries yielded very different success rates: 84% of the look-ups with the use of monolingual dictionaries for native speakers were rated as successful, compared to only 44% of the look-ups in monolingual learners' dictionaries, as shown in Table 1 on page 45 below, which is based on Varantola (1998). However, because of the small number of look-ups and subjects and the methodological problems arising from it (see 2.5.1 below for discussion), it would be risky to generalize from the above rates.

Laufer (2000) tested the effect of marginal paper glosses and electronic (on-screen pop-up window) glosses on vocabulary recall. In the electronic glosses condition, subjects could choose the information types displaying on screen. The majority chose L1-translation only glosses, and these achieved a mean immediate recall rate of 83 percent, and a delayed retention rate of 27 percent. Rates of 100 percent on both measures were achieved with three-way entries of L1 translation plus English definition plus example. This might be taken to suggest an advantage of the combined entry. However, a two-element entry structure intermediate between the two, L1 translation plus English definition, scored 90 percent on immediate recall but 0 percent on retention (no words were remembered!). This, plus the fact that the combined entries were selected in a very small minority of cases only, must make us very cautious of the results, and the claim of the superiority of the rich-information entry. Also, since subjects were in control of the types of electronic glosses they wished to see, we must consider seriously the possibility that the few subjects who selected the three-way entries were unusually highly motivated, which might be the true underlying factor responsible for the perfect scores.

Dziapa (2001) compared the effectiveness of vocabulary acquisition through reading by Polish learners of English using bilingual and monolingual dictionaries. Dziapa worked with two groups of learners: beginners and intermediates. Within each group, subjects were randomly assigned to one of two dictionary conditions, a monolingual learner's dictionary and a traditional bilingual dictionary. Overall, results point to an advantage of the bilingual dictionary for beginners, and a relative lack of such advantage for intermediate learners. However, the direction and magnitude of this advantage also depends on the type of lexical tasks: the monolingual dictionary seems to be at its best in tasks requiring users to provide definitions.

Wingate (2002: 50) found incidental vocabulary learning scores by Hong Kong Chinese intermediate learners of German to be higher when using German-English bilingual dictionaries compared with a monolingual German dictionary,

although the difference was only statistically significant for high verbal ability subjects. The sample was very small, so the power of the Kruskal-Wallis test used by Wingate is here small as well, but the effect size is quite considerable, ranging from 30% to 50% depending on verbal ability and type of test. A fairly large difference of about 30% in favour of the bilingual dictionary was also found for reading comprehension scores in low verbal ability subjects, with the p-level just missing the significance threshold at 0.055. However, when Wingate rewrote the definitions to make them maximally helpful, the percentage of successfully looked up words was higher for the new definitions than for the (non-native language) bilingual dictionary, as revealed in a think-aloud study. Still, the total sample was very small, with 6 subjects or less per each dictionary condition, and the difference was not significant. Also, on supply-definition as well as reading comprehension tests it was the bilingual dictionary that registered the highest scores, and the new definitions performed worst. The results of the supply-definition test may be seen as surprising given that the task could be expected to favour definition-supplying dictionary entries[10], which the bilingual dictionary did not offer. Wingate's (2002) findings suggest that learners at the intermediate level may not be in a position to use monolingual dictionaries effectively.

1.6 Preference for early senses

In an experimental set-up using pseudo-words, Tono (1984; see also Tono 2001: Chapter 9) discovered that dictionary users exhibited a tendency to select the early (often initial) senses in a dictionary entry unless clear indications were present within the entry that the early sense was inappropriate. Neubach and Cohen (1988) found in a think-aloud study that subjects (EFL learners) tended to read the first definition only in monolingual dictionaries. A typical comment by one of Neubach and Cohen's subjects was "I didn't go on after the first definition. I thought all the rest were just examples" (Neubach and Cohen 1988: 8). Müllich (1990) also noted a preference for first meanings. Nuccorini (1994) examined written translation by students using bilingual dictionaries and, likewise, found evidence of users stopping after the first sense.

Reading only part or beginning of a dictionary entry was the single most common cause of failure in bilingual dictionary consultation as established by Wingate's (2002: 113) think-aloud study with intermediate learners of German. Furthermore, amongst the successful look-ups, those that did not require the users to look beyond the first sense within an entry were most numerous. The bilingual dictionary used in the study, the *Langenscheidts New College German Dictionary, German-English, English-German* (LNCGD), however, did not include the native language of the dictionary users, which could be an additional factor here. In contrast, reading the early senses only was not found to be a problem in

[10] This relative advantage was also confirmed by Dziapa's (2001) results, as discussed in the previous paragraph.

intermediate learners using monolingual dictionaries in Wingate's sample (Wingate 2002: 115). There was some evidence from Wingate's main think-aloud study, that a cramped entry structure may have contributed to this problem.

Wingate offered the following recommendation for dictionary design based on her findings regarding the preference for early senses:

> The tendency not to read beyond the initial information in dictionary entries could be a common one. If this finding can be confirmed in the main think-aloud study and further research, it has the implication that in dictionary entries, the meanings should always be listed in order of frequency, with the most frequent meanings at the beginning of the entry. This way the learners have the best chance of finding the meaning they are looking for. (Wingate 2002: 118)

The reasoning in the above passage appears to be based on an implicit assumption that dictionary users are most likely to be looking for the most frequent senses. However, that assumption is obviously incorrect in the case of the more advanced learners, who might be expected to have mastered the most frequent senses of the core vocabulary items, and if they need to consult a dictionary for meaning of such items, it will most likely be one of the less frequent senses that will be of interest to them.

Further, in view of the above findings, there may be a price to pay for including too much information under a given sense, because that might discourage the user from going on beyond the first sense. This has been a tentative finding of Tono (1984; 2001: Chapter 9):

> The third implication from the findings on the user's reference skills is that users tend not to appreciate the complexity of dictionary design. This is indicated by the fact that the subjects generally preferred the first definition and that they did not even refer to the second definition when lengthy illustrative examples followed the first definition. Most lexicographers assume that illustrative examples play an important role in the comprehension and production of text, and this is true in many cases. But the findings indicate that most of the subjects who did not major in English did not use examples to find appropriate definitions. On the contrary, examples prevented the subjects from going on to the second definition in many cases (...). These findings add some weight to the argument that the users prefer simplicity and brevity above all. (Tono 2001: 163)

In terms of semantic explanation, simplicity and brevity appears to be more compatible with native language equivalents than with foreign language definitions.

2. Studies of receptive dictionary use: methodological issues

2.1 Literature on the methodology of dictionary use research

Interest in the methodological aspects of dictionary research has so far been relatively small. The two scholars often seen as fathers of the methodology of dictionary use studies are Herbert Ernst Wiegand and Reinhard Rudolf Karl Hartmann.

Wiegand's call in the mid-1970's (Wiegand 1977a; 1977b) to study a "sociology of dictionary use" cleared the way for modern empirical research of dictionary use. Wiegand's extensive theorizing on the subject culminated in his monumental monograph of over 1000 pages (Wiegand 1998).

Reinhard Hartmann is a much more frequently cited author, not the least because he published extensively in English (as well as in German). In an early paper (Hartmann 1983b: 11), he specified a four-fold structure of factors involved in dictionary use. In Hartmann (1987b), he further distinguished four areas of research into dictionary use:

- research into the information categories presented in dictionaries ('dictionary typology')
- research into specific dictionary user groups ('user typology')
- research into the contexts of dictionary use ('needs typology')
- research into dictionary look-up strategies ('skills typology') (Hartmann 1987b: 12)

The stress on *typology* above suggests a taxonomic leaning: perhaps a fitting approach for a newly-developing area of study. Hartmann also bemoaned the scarcity of studies with "statistical correlations" (Hartmann 1987a: 26), and a complete lack of "more complex techniques like controlled experiments" (Hartmann 1987a: 27). Hartmann (1989b) elaborated on the methodological reflection of his previous papers by listing a dozen hypotheses relating to dictionary use and discussing them in the light of contemporary research and delineating seven areas of research into dictionary use.

Hulstijn and Atkins (1998) offered an informative survey of empirical research on dictionary use. Aiming to produce an updated version of Bogaards' (1988) overview, Hulstijn and Atkins conducted a comprehensive search of the literature, bringing together information on "some fifty published papers reporting on empirical investigations in which the dictionary was involved one way or another" (Hulstijn and Atkins 1998: 7). They classified the studies covered in the paper under seven headings. In the second part of their paper, Hulstijn and Atkins put forward a number of suggestions regarding the methodology of research into dictionary use. Among others, they listed fourteen variables that they saw as im-

portant in the use of L2 dictionaries. They suggested that preferably only one variable at a time should be isolated and selected for study. This particular suggestion seems unnecessarily restrictive: such a simple one-way design rules out any investigation of interactions between different variables. Interactions between variables may be both theoretically and practically interesting, so it would be a pity to sacrifice this option a priori, even though doing so does of course generally result in much simpler, more manageable designs. Out of the fourteen variables listed by Hulstijn and Atkins (1998: 12-13), as many as seven are controlled for in the present study. The suggestion to limit the designs to a single variable may be a continuation of the methodological reflection in Atkins et al. (1987). In this earlier publication, the isolation of the variables for testing appears to be a major theme, although the focus of the paper is on individual tests.

The book-length annotated bibliography on language learners as dictionary users by Dolezal and McCreary (1999) was probably to some extent inspired by Zgusta (1988) and preceded by McCreary and Dolezal (1998) and Dolezal and McCreary (1996). Dolezal and McCreary identify three general categories of research on dictionary use by language learners:

- dictionary typology (for instance, the arguments for and against the bilingual dictionary vs. the monolingual dictionary)
- the so-called needs of the user (questions and solutions on informative content, format, accessibility, etc.)
- dictionary choice (which dictionary 'performs' better; which dictionary is more suitable for beginning, intermediate and advanced learners) (Dolezal and McCreary 1999: XII)

In contrast, McCreary and Dolezal (1999: 110) decided to cluster experimental and quasi-experimental studies into two separate groups, one essentially in the methods of reading comprehension research, the other more in the tradition of applied linguistics.

Tono (2001) combines the two classifications of dictionary use research by Hartmann (1989b) and Hulstijn and Atkins (1998) into a single list. Tono categorizes the major empirical studies in dictionary use in terms of Cohen and Manion's (1994) developmental stages of science. He discusses the potential of the following techniques and methods in dictionary use research:

- participant observation
- surveys
- accounts
- correlational research
- experiments, quasi-experiments and single-case studies (Tono 2001: 66-72)

Tono (2001: 72) concludes his discussion of the methodology of dictionary use studies with the suggestion that modern statistical techniques, such as factor analysis, could be fruitfully applied in the area of dictionary use research, for example to reduce the immense volumes of data that can be generated in modern

empirical studies of dictionary use, and thus assist the researcher in the inference process.

Interesting methodological reflection on dictionary use research is offered by Humblé (2001), who is sceptical of experiments and tests and argues for the qualitative, ethnomethodological approach.

As I hope to have shown in this sketchy overview, the literature on the methodology of dictionary use research is hardly comprehensive. Therefore, I will discuss below some methodological issues that I believe to be relevant in the context of the present work. The discussion below is organized by general methodological approach, although some issues are not restricted to a single approach.

2.2 Methods in dictionary use research: introductory

The earliest, and still very popular approach to dictionary use research has been the survey by questionnaire, which was pioneered by Barnhart (1962). Barnhart's questionnaire study was, however, somewhat untypical in that the respondents who were surveyed were not the target dictionary users themselves, but rather a group of teachers of American college students, who were asked to report their beliefs and convictions regarding their students' habits and patterns of dictionary use. With respondents reporting on somebody else's behaviour, Barnhart's procedure added, so to speak, another remove to the questionnaire methodology, compared to the more typical set-up of the respondents speaking for themselves. The first important questionnaire study involving dictionary users directly was the impressive survey by Tomaszczyk (1979), which set the standards for this kind of studies for many years to come.

Questionnaires are at their best in covering the background, or context, of the dictionary consultation situations, but they are also an important source of information on those categories of data whose collection by other methods would not be an acceptable option (Lew 2002a), either because it would be too expensive, too intrusive on ethical grounds, or just plain impossible. Typical examples of such information might include the data on dictionary acquisition and ownership, or frequency of dictionary consultation in the dictionary user's home.

Strictly speaking, the case study is an approach that predates questionnaires. Traditional dictionary reviews can be seen as a subtype of the case study. Though relatively easy to implement, case studies suffer from one basic weakness: they tend to be very subjective and idiosyncratic, which makes their value for systematic study of dictionary use questionable: Hartmann (1989b: 106) calls them "the bottom rung in the hierarchy of scholarly methodology." However, they may be useful as a window into the more private aspects of dictionary use when used in supplement to other methodology and when interpreted with due caution.

In contrast to questionnaires, direct observation studies are best suited for investigating what happens during actual dictionary consultation. As Wingate (2002: 48) aptly puts it, observation protocols are process-oriented.

When it comes to evaluating the results of dictionary consultation, it is tests and experiments that appear to be most useful, with their baggage of accompanying controls and statistical analysis. Wingate (2002: 48) refers to such methodology as result-oriented. In the same vein, questionnaires might be termed context-oriented.

Recently, authors writing on the methodology of dictionary use studies (Hulstijn and Atkins 1998; Nesi 2000b) have begun to advocate the use of more than a single method if only possible. This is exactly what a recent study by Wingate (2002) attempted: she used experimental tests and think-aloud protocols. The novel character of Wingate's study is that she derived hypotheses about the effective and ineffective features of definitions from a think-aloud study, and then attempted to verify them in an experimental design. Unfortunately, in testing the *Langenscheidts Grosswörterbuch Deutsch als Fremdsprache* (LGDaF) definitions against the new definitions written by Wingate herself (NDefs), it is difficult to separate the effects of the individual features of the new definitions from the combined effect of the new entry format treated as a whole. In entries assigned to test a given feature, not all other features were controlled for. There were good technical reasons behind this, but that does not change the fact that the new definitions are a more or less arbitrary cluster of features whose individual effect could not be easily evaluated.

As noted by several authors (Bogaards 1999: 34; Hulstijn and Atkins 1998: 12; Zöfgen 1994: 36), the dictionary user's language proficiency level is an important variable in dictionary use studies. However, very few studies have attempted to cover a broad range of proficiency levels. Battenburg (1989; 1991) placed his learner subjects into three groups according to language level, but, as Nesi (2000b: 10) pointed out, "the three groups did not represent a true language learning continuum, as the intermediate group's responses did not generally bridge the gap between elementary and advanced behaviour." One notable example of a study with careful control of proficiency level was the EURALEX/AILA Research Project on Dictionary Use (Atkins and Varantola 1998a).

In the remainder of this chapter, I will discuss selected methodological issues and problems for the different methodological approaches, concentrating on issues relevant for questionnaires and tests, which reflects the focus of the present study.

2.3 Questionnaires

As Hilary Nesi (2000b: 3) observed, "[q]uestionnaire-based research is perhaps the commonest method of enquiry into the use of English dictionaries," and this observation is probably also true of non-English dictionaries. Questionnaires offer a convenient way of surveying large numbers of subjects. They are indispensable for collecting demographic data.

The first well-known questionnaire study was Barnhart (1962), already mentioned in 2.2 above, which used the indirect methodology of asking teachers about the dictionary needs and habits of their students, as perceived by the teachers themselves. Later studies, starting with Quirk (1974), have tended to approach the target user directly. Nevertheless, the degree of agreement between Barnhart's (1962) results and later studies is quite considerable, which suggests Barnhart's methodology was not entirely without merit.

As already suggested in a general comment under 2.2 above, the inclusion of underlying user variables, especially users' proficiency level, could help improve the accuracy of questionnaire-based investigations. This was pointed out by Zöfgen (1994) in his critique of Béjoint (1981) and Hartmann (1982):

> In Béjoint's case, no distinction was made between the second-, third-, and fourth-year students; in Hartmann's survey averages were taken from groups as different as teachers, university students and pupils from secondary schools. Therefore a most relevant variable, i.e. proficiency in the FL, was neutralized. (Zöfgen 1994: 36)

Another critique of contemporary questionnaire-based studies was that by Bogaards (1988), who did not in principle disapprove of the method as such, and did in fact employ it himself.

In the remainder of this section dealing with the methodology of questionnaire-based studies of dictionary use, I will focus on what I see as two particularly problematic areas central to the methodology of dictionary user surveys. The first area is the fairly well-known issue of the uncertain relationship between reports returned by questionnaires and facts and beliefs that the reports are expected to reflect. The second area of problems concerns the way questions and instructions are communicated to the respondents by the researcher. The language used should be maximally accessible to subjects, so that it can be understood, as closely as possible, in the way the researcher intended, and in a more or less similar way by the individual subjects. At least two linguistic aspects may be within the researcher's control which might seriously affect the comprehensibility of questionnaires: the degree to which technical jargon or obscure metalanguage is used, and the choice between the respondent's native language and a foreign language. The latter aspect in particular has received surprisingly little attention in the literature, although the issue is one that researchers dealing with dictionary use and language learning often find themselves facing.

2.3.1 Reliability of questionnaire reports

The problems of the reliability of questionnaire reports concern the nature of the correspondence between the questionnaire responses and the researchers' expectation of what they indicate, that is, they touch on issues related to construct validity (Trochim 2000). Hatherall's (1984) words are perhaps the most often quoted piece of criticism of questionnaires as a research tool in dictionary use studies: "Are subjects saying here what they do, or what they think they do, or

what they think they ought to do, or indeed a mixture of all three?" (Hatherall 1984: 184). In Lew (2002a), I examined Hatherall's charge in a broader perspective and argued that the methodological problems similar to those pointed out by Hatherall are also inherent in the methodology that he proposed in his paper. Crystal (1986) doubts that questionnaire respondents are in a position to remember the details of dictionary use. Similarly, Nesi (2000b: 8) cautions that some questionnaire items "[depend] rather too much on the students' powers of critical analysis, retrospection and recall." Crystal (1986) also criticizes the constraining nature of many questionnaires, which discourage original responses.

2.3.2 Language in questionnaires

2.3.2.1 Native or foreign language

In the context of foreign language teaching, researchers electing to employ questionnaires may feel an urge to write their questionnaire instructions and questions in the target foreign language. If the researchers are language teachers themselves, the use of a foreign language may be nothing more than a habitual reflex of the test-editing classroom practice, where instructions and questions may be customarily written in the target language. Illustrative material in the foreign language may be present in the questionnaire by design, and the author of the questionnaire may feel that the language of the instructions should match that of the target language. An additional factor that could sway the choice of language might be a preference for one of those language teaching paradigms that see the use of the native language in the classroom as a major offence.

However, in questionnaire instructions and questions it is essential that the intentions of the questionnaire author be communicated as faithfully to the questionnaire respondent as possible. There is little doubt that this role is best served by the native language of the subjects, not by any foreign language that they may be studying. Quite obviously, the risk of communication failure is appreciably greater in the foreign language. For example, Wingate (2002: 48) notes a case where subjects clearly failed to understand questionnaire items in Battenburg's (1991) questionnaire study, in which the questions were written in the foreign language (English, in this case).

A complete misunderstanding of a question or instruction is just one obvious danger of using a foreign language in a questionnaire, but there is also a less obvious one, and it concerns the precision with which the intended meaning is conveyed to the questionnaire respondent. There is good reason to suspect that, even for advanced learners of a foreign language, the conceptual range associated with a foreign language lexical item is less stable across a sample of learners than the conceptual range associated with a similar item in the native language of the subjects. There is no reason why this problem would not affect the conceptualization of even the most common words that may appear unproblematic to questionnaire designers; words such as, for example, *often* or *rarely*. When the sample spans a range of proficiency levels, the resulting lexical vagueness of interpretation could

become very grave. If this is the case, then the use of a foreign language in a questionnaire would introduce additional imprecision, reducing the usefulness of the questionnaire as a research instrument.

Practical considerations may force a researcher to use a foreign language in questionnaires. This could be justified when the sample consists of native speakers of a variety of languages, as can be the case in ESL settings, such as in Battenburg's (1989; 1991) sample, with subjects representing seven native language backgrounds. In such a situation, translations of the research instrument into the native languages of the subjects would have to be produced. But translation is expensive and qualified translators may not be readily available for all the languages involved. In addition, there may be a problem with equivalence between the different translated forms of the questionnaire. The latter problem may in fact be seen as a variation on the problem of subjects interpreting the foreign language questionnaire items, only its milder version, because here the interlingual operation is entrusted with the qualified professional. So, if translators are available and if the number of native languages involved is not overwhelming, it is still in general preferable to use translated forms, as was in fact done in the EURALEX/AILA Research Project on Dictionary Use (Atkins and Varantola 1998a). A decision to use a foreign language would be difficult to defend if the subjects all shared the same native language, as in the present study. It is still less justifiable if the questionnaire author is not the native speaker of the target language but that of his subjects, for example when a Polish teacher of English uses her Polish learners of English as subjects. Should the teacher elect to use English as the language of instruction in her questionnaire in such a case, she has to produce instructions and questions in a foreign language, which generates language problems of its own on top of the ones discussed above.

To conclude the above arguments, due to the increased risk of the subjects failing to understand, completely misinterpreting, partially misunderstanding, or interpreting vaguely the foreign language content of a research instrument, it is the native language of the subjects that should be used in questionnaires if only possible, and as much as possible.

2.3.2.2 Metalanguage

Theoretical lexicography (metalexicography) has now accrued a substantial body of technical concepts and terms that describe the various aspects of dictionary structure and dictionary consultation (e.g. Hartmann and James 1998). When dictionary use researchers entertain and express ideas about dictionaries, they naturally do it in those technical terms. While terminological rigour is certainly helpful in specialist discourse, dictionary users are not dictionary experts, so questionnaires meant for the ordinary dictionary user should steer clear of the technical lingo that is likely to confuse rather than inform potential subjects.

An examination of the existing studies shows that the researchers are not always fully aware of this issue. For example, Nesi (2000b: 10) so criticizes Bat-

tenburg's (1991) use of technical terms in his questionnaire: "It seems unrealistic to expect subjects with elementary English to comment on their use of 'syntactic patterns' and 'derived forms' in dictionaries." And it is somewhat ironic that the author of the famous criticism of the questionnaire methodology had this unnecessarily convoluted piece of English syntax thrown at his subjects in the post-task feedback form (a questionnaire, really): "Do you think the time constraint had any effect on your method of working?" (Hatherall 1984); rather than asking in plain English: "If you had had more time, would you have done anything differently?".

Using technical terms may seem deceptively attractive to the researcher, with its lure of improved precision of questionnaire items, but this is deceptive: precision must be evaluated from the point of view of the subjects interpreting the questionnaire items, rather than the researchers' point of view. Therefore, everyday language should be used, although the dilemma of conveying technical notions in everyday language may be difficult to resolve. Broad consultation and piloting may help in arriving at the optimal wording of questionnaire items. Difficult metalanguage should be avoided as far as possible in questionnaires addressed to dictionary users.

2.3.3 Summary

The chief problems associated with the use of questionnaires in dictionary use research were aptly summarized by Nesi (2000b: 12):

> 1) Results are often a measure of the respondents' perceptions, rather than objective fact. The respondents' desire to conform, their (perhaps unconscious) wish to appear in some way better than they really are, or their inability to recall events in detail may distort the data.
>
> 2) Researcher and respondent do not necessarily share the same terms of reference. Linguistic concepts cannot be accurately expressed without metalanguage, but the specialist terms the linguist uses have no meaning (or a different meaning) for the non-linguist. In large-scale surveys where there is no opportunity for researcher and respondent to negotiate meaning there is likely to be considerable misinterpretation of both the questions and the answers.

While it may be true that "linguistic concepts cannot be accurately expressed without metalanguage," some of those concepts are simply incomprehensible – metalanguage or not – without special training. As for those less esoteric concepts that may be comprehensible to non-linguists, simply replacing the technical jargon with everyday language may do a good enough job. To take an example, asking subjects about "information on what other words to use this word with" will likely invoke as much linguistic awareness as one could reasonably expect of a dictionary user without linguistic training, and will surely fare better than asking about "collocational information."

Finally, useful general advice on aspects of questionnaire design that are not specific to research into dictionary use may be found in existing questionnaire

design manuals written by experts in fields such as sociometry, psychometry, and second language research (Berdie and Anderson 1974; Bradburn, Sudman and Blair 1979; Dörnyei 2003; Oppenheim 1992; Sudman and Bradburn 1982).

2.4 Case studies

Case studies in dictionary use research involve a small number of dictionary users: there could be as few as one subject (e.g. Ronald 2002), or their number could be closer to a dozen (e.g. Nuccorini 1992; Tono 1991). A case study could focus on a detailed analysis of the process of dictionary use during some (naturalistic or artificially imposed) task, such as composition writing (Ard 1982). Case studies may employ observation protocols, self-reports, and interviews for data collection. The advantage of the method is that with a small number of subjects the researcher can afford to allocate significant resources to individuals and may thus be able to capture details that would escape a large-scale study. However, the small number of subjects also means that no statistical generalization to any larger population is generally possible, and this limits the usefulness of case studies.

2.5 Observation protocols

Written and oral protocols have been employed to get insight into what happens during the dictionary consultation process, both in terms of dictionary behaviour and in terms of thought and decision processes of dictionary users. Hatherall (1984) asked his students to fill out protocol forms describing the process of L1→L2 (English-to-German) translation with a dictionary. However, only preliminary information from this study has been reported, and design details are difficult to extract from the article. A newer study with better-documented methodology is Varantola (1998), where four translation majors recorded the details of their dictionary use on special answer sheets.

As many as three recording sheets were being completed by the 211 informants in Harvey and Yuill (1997). The first sheet was a relatively simple eight-item table filled in by dictionary users prior to dictionary consultation. The other two sheets were structured as flowcharts and were filled in following the consultation act. Though carefully designed, the flowcharts were very complex (Harvey and Yuill 1997: 277-278) and there may be realistic concerns about the subjects' ability to complete the charts as intended by the researchers.

Think-aloud protocols have been used to monitor the process of dictionary work (Momoi 1998; Müllich 1990; Neubach and Cohen 1988; Whyatt 2000; Wingate 2002; Yokoyama 1994), and the immediate recall protocol has been useful in the study of dictionary use while reading to assess reading comprehension (Knight 1994).

There is a potential in protocol-based studies for revealing the fine details of the dictionary-consultation process. The difficulty lies in the processing and in-

terpretation of the data collected through protocols, which can be very laborious. This is why only small numbers of subjects usually participate in studies of this type. For example, Ard (1982), who used video-recordings and oral interviews, worked with two subjects only. In some other protocol studies (Momoi 1998; Yokoyama 1994), the sample size may have been somewhat larger, but typically not large enough to warrant generalizations beyond the sample, which is a major weakness of the approach.

A possible solution to the above dilemma may come from data-collection techniques similar to that employed in Nesi and Haill (2002). The data for the study was drawn from the library research skills assignment forming part of the 'Key Academic Skills for International Students' module at Oxford Brookes University. Rather than systematically report the complete details of all their dictionary consultation over a defined stretch of time, students were asked to answer questions about their dictionary consultation of just five selected lexical items. They answered the questions in writing while working on a text of their own choice. The restriction on the volume of data produced with this technique allowed the researchers to collect their data from a larger number of subjects than has usually been the case in protocol designs not so constrained. Overall, 89 contributions, each from a different student, were collected over a three-year period. While such a technique makes it more manageable to collect contributions from a broader sample of subjects, there are trade-offs here as well, though on a different level of sampling. Because each subject selected five specific dictionary searches to report on, questions might be raised as to how representative those five items were of that subject's typical dictionary-using behaviour. Further, since the students were completing the reports as part of an academic assignment, they may have approached the task rather more meticulously than usual. There is a danger that they may have wanted to avoid what they saw as trivial or uninteresting dictionary searches for fear of returning mediocre assignments. Also, they may have felt an urge to submit excessively elaborate descriptions of their work, which might not have truthfully reflected what they actually did with their dictionaries.

Self-reporting by dictionary-using subjects is often used in observation-based studies, just as it was in Nesi and Haill (2002). However, as pointed out by Tono (2001: 54), self-reporting in protocol studies is open to methodological problems similar to those present in questionnaires: subjects may tend to report their beliefs and perceptions rather than facts about the consultation process, or may want to please the researcher. They may be selective in their reporting. The reporting process may interfere with the look-up process itself, especially if the procedure calls for the reports to be completed during actual dictionary work. An interesting example of an alternative to self-reporting is a study by Atkins and Varantola (1998), where written recording sheets (similar to those in Varantola 1998) were filled by lexicography students acting as monitors, each monitor being assigned to a single dictionary user working on a translation task.

2.5.1 Look-ups, subjects, and independent observations

When dictionary users' behaviour is observed and recorded, the dictionary consultation process may be broken down, in a theoretical sense, into searches and look-ups (consultations), where several look-ups may constitute a single search (Atkins and Varantola 1997; 1998b; Scholfield 1982; 1999; Varantola 1998). When results of individual look-ups are being evaluated, an assessment of the data obtained should be made in terms of the independence of the individual look-ups.

To illustrate the point, consider a study of dictionary use in L1→L2 (Finnish to English) translation (Varantola 1998). Three categories of dictionaries were distinguished in the study: bilingual dictionaries, monolingual dictionaries for native speakers and monolingual learners' dictionaries. The mean look-up success rates within the three categories, established on the basis of subjects' evaluation of the individual look-ups, are given in Table 1 (based on Varantola 1998: 185).

Table 1: Dictionary type and success rate in Varantola (1998: 185)

Dictionary type	Successful look-ups	Unsuccessful look-ups	Total look-ups
bilingual	47 (63%)	28 (37%)	75 (100%)
monolingual native	26 (84%)	5 (16%)	31 (100%)
monolingual learners'	7 (44%)	9 (56%)	16 (100%)
Total	80 (66%)	42 (34%)	122 (100%)

Looking at the percentages alone (given in parentheses), the two subcategories of monolingual dictionaries yielded very different overall success rates: 84% of the look-ups with the use of monolingual dictionaries for native speakers were rated as successful, compared to only 44% of the look-ups in monolingual learners' dictionaries. The difference looks impressive, but how much weight should we attach to it? To what extent can we generalize from the above rates?

First, one should note that the number of observations (i.e., individual look-ups) is quite small, particularly in the case of monolingual learners' dictionaries. Varantola (1998) does not test for differences between the different dictionary types. Such a test is attempted in Table 2, using the Chi-square statistic with Yates correction to compensate for low expected frequencies.

Table 2: Yates-corrected Chi-square values and p-levels for dictionary type pairs computed from data in Varantola (1998: 185)

Dictionary type	Yates-corrected Chi-square	p-level
bilingual vs. monolingual native	3.66	0.056
bilingual vs. monolingual learners'	1.25	0.264
monolingual native vs. monolingual learners'	6.32	0.012

The Chi-square test yields only one significant difference: that between the monolingual native dictionaries and monolingual learners' dictionaries, though the difference between bilingual and monolingual native dictionaries approaches the 5% significance level. However, there is a problem with the application of this test: it is a fundamental assumption of the Chi-square test that the individual observations (here, look-ups) are independent. In this case, several look-ups may constitute a single dictionary search; in fact, about 60% of all look-ups were part of multi-look-up searches, as can be estimated from Varantola's (1998: 183) data. There is good reason to claim that multiple look-ups constituting a single search should not be treated as independent.

Even more importantly, the number of subjects in the study was just four, so inevitably many look-ups must have come from a single subject. Obviously, outcomes of multiple look-ups from a single subject cannot be treated as independent observations. In fact, one can easily imagine a situation where a large proportion of the look-ups for a given dictionary type would come from a single dictionary user. It is not at all unlikely that a particular user might have a favourite dictionary, and might tend to rate his or her look-up satisfaction in a markedly idiosyncratic way, either because s/he is a demanding (or otherwise) rater, or simply because s/he is a skilful (or otherwise) dictionary user. Differences such as those reflected in Table 1 could easily have arisen under this – hypothetical but realistic – scenario.

Because dictionary use is such a "private matter" (Nesi and Haill 2002: 277), the above example underscores the importance of, on the one hand, sufficiently large samples, and, on the other hand, careful consideration of issues related to the independence of observations.

2.6 Tests and experiments

It was around the time when the well-known criticism of the early questionnaire studies was voiced (Hatherall 1984) that the experimental approach started to gain some popularity, starting with Bensoussan, Sim and Weiss (1984) and Black (1986). As Nesi (2000b: 12, 31) notes when comparing questionnaire-based and test-based research, an important motivation behind the introduction of tests and experiments into dictionary use research was to obtain more objective and reliable data than was normally achievable in questionnaire-based studies. More rigid design and controlled conditions of data collection were better suited to the experimental paradigm. In terms of the potential of the two approaches, Nesi further observes that "test-based research can enable the researcher to prove or disprove hypotheses in a more conclusive fashion than is possible with questionnaire-based research, which usually generates rather than tests hypotheses" (Nesi 2000b: 31).

One of the earliest and commonest applications of testing techniques in dictionary use studies has been to investigate the effect of dictionary use on learners' performance on reading comprehension tests (Bensoussan, Sim and Weiss

1984; Nesi 2000b, Chapter 2; Nesi and Meara 1991; Tono 1989). Wingate (2002), however, challenged reading comprehension tests as an unreliable instrument to measure dictionary effectiveness; she advocated equivalent-supplying and definition-supplying tasks as a preferable alternative. However, the latter approach carries the risk of the dictionary users copying the equivalents or the definitions, whichever happen to be available in the dictionary entries used. Such a user strategy was observed by Miller and Gildea (1987), where subjects' responses often closely paralleled the illustrative material in the entries, and Miller and Gildea concluded that this did not necessarily indicate successful comprehension.

A popular area of experimental research has been the identification of the headword types that dictionary users would prefer to look up first when trying to locate a multi-word item, such as an idiom or a spaced compound (Atkins and Varantola 1998a; Béjoint 1981; Bogaards 1990; 1991; 1992; Tono 1987). The typical procedure in such studies has involved asking subjects to either underline or note down the word or words that they would choose to look up as the headword. This may or may not have been accompanied by actual dictionary consultation. Designs without the consultation are less natural (and less experimental), since the outcome of a decision as to which item to look for could conceivably be affected by whether the user actually intends to act out the decision in a real look-up act.

Another area of experimental dictionary research centres around the role and importance of various dictionary information types in various tasks. Tono (1984) used custom bilingual dictionary entries in booklet form and introduced nonce words to test the usefulness of eight types of information: grammatical information (Tono included here: part of speech labels, such grammatical labels as "no passive" and information on article use with nouns); verb pattern, noun countability (countable vs. uncountable); gloss (Tono included here synonyms but also collocational indicators following the equivalent in parentheses); collocation; idiom; run-on entries; and illustrative examples. On top of these information types, Tono also manipulated the order of dictionary definitions and found a strong preference by the subjects to select the first definition provided. In later experimental studies, Tono investigated the role of access facilitation devices: entry menus (Tono 1992), and sense access indexes in the form of guidewords and signposts (Tono 1997).

The role of examples in dictionaries has received some degree of experimental study (Black 1986; Laufer 1992; 1993; Laufer and Hadar 1997; Laufer and Melamed 1994; Nesi 1996), although, as in other areas of lexicography, the body of empirical work on examples in dictionaries is far outnumbered by non-empirical studies (e.g. Fox 1987; Humblé 1998; 2001; Jessen 1996; Marello 1987; Van Scherrenburg 1990; Toope 1996, and many others).

Not only has the presence of various lexical information been studied, but also its form, or the way it is presented in dictionaries. In this connection, one

should mention the studies of the effect of different defining styles on performance, both receptive and productive (Cumming, Cropp and Sussex 1994; Kostrzewa 1991; Laufer 1993; Nesi 2000b: Chapter 3; Wingate 2002), as well as studies comparing the effectiveness of authentic and invented examples (Laufer 1992).

The quality of presenting lexical information in a dictionary is not always easy to separate from the users' interpretation of the information offered. A dictionary consultation act involves both the dictionary product and the dictionary user, and so the success or failure of the look-up act may often have competing explanations: first, in terms of the characteristics of the dictionary itself, and second, focusing on the dictionary-using skills of the user. Research into users' reference skills (Atkins 1998; Atkins et al. 1987; Béjoint 1981; Hartmann 1985; Tono 1984; 1988; Turkish 1972) tends to embrace the "user perspective."

In many of the studies cited in this section there was a more or less explicit interest in the effectiveness of dictionaries in general, or in relative differences in effectiveness between various types of dictionaries. Experimental and quasi-experimental set-ups have been applied to test the effectiveness of one or more dictionary products, although statistical evaluation of the differences was not always attempted, and measures of effectiveness varied across studies (Atkins and Varantola 1997; 1998b; Bogaards 1991; Krantz 1991; Laufer 1992; 1993; 1995; 2000; Laufer and Hadar 1997; Laufer and Melamed 1994; McCreary 2002; Oskarsson 1975; Summers 1988; Varantola 1998; Wingate 2002).

2.6.1 The problem of dictionary underuse

A number of carefully designed studies of dictionary use (Atkins and Varantola 1998a; Bensoussan, Sim and Weiss 1984; Hulstijn, Hollander and Greidanus 1996; Krantz 1991) registered disappointingly low rates of dictionary use by subjects in the course of the testing. This phenomenon poses a fundamental methodological problem to dictionary use research, because it presents a threat to a necessary condition underlying such studies: in order to collect information on the details and effects of dictionary use, dictionary use must occur[11]. Furthermore, there may well be a large body of studies where rates of dictionary use are unknown, and which may also be affected with the same methodological problem without the researchers realizing it.

The nature of the problem lies in the fact that when the aim of the study is to investigate the effect of dictionary use, then if the subjects fail to use dictionaries during the study, one cannot expect an effect to follow from a non-occurring condition. If, for example, a study involves a comparison with a control group that has no access to dictionaries, then the real difference between the experimental (dictionary condition) group and the control will be not so much dictionary

[11] The problem does not, of course, affect those studies whose aim is to find out the rate of dictionary use.

consultation, but a mere potential opportunity to consult a dictionary by the experimental group. If dictionary use in the experimental group does occur, but is minimal, then the dictionary-supported behaviour will be statistically overwhelmed by behaviour that does not involve actual dictionary use, so any effect of (actual) dictionary use will be severely watered down and only potentially discoverable in very large samples.

The problem of dictionary underuse is partially related to the problem of the use of vocabulary items that are either known, or just believed to be known, by subjects (see 2.6.4 below for discussion). In such a situation, subjects are less likely to want to invest their time and effort in consulting the dictionary. In the EURALEX/AILA Research Project on Dictionary Use (Atkins and Varantola 1998a), fewer than 25% of the subjects used a dictionary while working on the Dictionary Research Test items. Tono (2001) attributes this low rate of consultation to the difficulty level of the test items, which he says may have been too easy. But similarly low (if not lower) rates have been reported in other studies where the tasks do not seem to be overly easy (e.g. Hulstijn, Hollander and Greidanus 1996), so the such low rates may in fact be truly reflective of the typical dictionary usage in the natural setting. While the dictionary consultation rate is itself of considerable theoretical interest, it is not a good idea, methodologically speaking, to allow a low rate of dictionary consultation to obscure the effect of dictionary use when this effect itself is being evaluated.

2.6.2 The use of actual dictionary entries in testing

In experimental studies in which dictionary effectiveness is evaluated, subjects need to be provided with access to dictionary entries to use as reference support in experimental tasks. These entries may come from actual dictionaries, be it introduced into the experimental setting by the experimenter or brought in by the subjects. Such designs provide information about user behaviour and success with actual existing dictionary products (titles), and provide one element of a natural setting for dictionary use tests, which may at times be important, as in studies of the look-up process (Ard 1982; Mitchell 1983a; Neubach and Cohen 1988; Nuccorini 1992; Tono 1991; Wiegand 1985). The degree of naturalness is greater when subjects are allowed to use their own dictionaries, because they are then working with the dictionaries that they are accustomed to using outside the experimental situation.

Studies which use actual dictionary entries are helpful in the evaluation of existing dictionary products. However, once a difference is found between two dictionary products, the researcher cannot in principle determine what specific features of the two products are accountable for the observed differences: it can only be established that one title as a whole yields better performance than another competing product. Such findings may be useful in marketing research, but they are of limited interest to the metalexicographer, who is more often than not interested in theoretical issues that extend in scope beyond the single dictionary

title, such as whether sentence examples are useful, which types of syntactic information are easier to assimilate, etc. Existing dictionary titles may differ in those respects, and may be used in testing, but they also invariably differ on a number of other dimensions, which remain uncontrolled, and whose influence on subjects' performance cannot be known. These dimensions are too numerous to list here, but we might just mention such apparently trivial matters of typographic presentation as font size, weight and typeface, line spacing, or the ordering of the different information types. When comparing two dictionary products that differ in their lexicographic treatment of a certain focal aspect, it is tempting to attribute to this aspect of interest any differences in performance found in the course of an experiment, but with existing dictionary titles we can never be sure that the difference is not really due to some completely different aspect, such as the ones related to typography, or in fact any complex combination of obvious and entirely unknown elements.

This methodological problem is present in many studies of dictionary effectiveness, but apparently the awareness of its consequences is low. Some authors specifically comment on the incidental features of the titles used in the testing, but nevertheless interpret the results as applying to general types of dictionaries represented by these products. For example, Raudaskoski (2002) used a relatively small bilingualized dictionary and a much larger bilingual dictionary to test the effectiveness of bilingualized and bilingual entries in translation, and complained of the bilingual's "dense entries full of symbols and abbreviations," admitting they "caused difficulties for many students, especially when the necessary headword or equivalent was concealed inside a long entry" (Raudaskoski 2002: 3). Clearly, such unbalanced choice of dictionary titles does not justify generalizations to the dictionary types they are purported to represent.

Tono (2000) is well aware of this problem when he comments on Laufer (1994):

> The different results observed in bilingualised and bilingual dictionaries may not be due to the difference inherent in these two types of dictionaries, but possibly to the matter of quantity of information (equivalents only vs. full information). Not only the dictionary category, bilingual or bilingualised, but also the actual information provided in each type of dictionary must be carefully controlled in order to make the test more valid. (Tono 2000: 23)

The effect of specific types of lexicographic treatment can only be assessed with any certainty if other, incidental aspects that contribute to the dictionary content and presentation are controlled for. This is virtually impossible to achieve with existing dictionary products, because rarely do we find two dictionary products that differ only in a single aspect, even in two titles coming from the same publisher. Sometimes we may be able to find a pair of dictionary products which might differ along a single dimension, such as binding type (paperback versus hardback) or the presence versus absence of a thumb index, but obviously such variation hardly exhausts the list of variables a metalexicographer might be inter-

ested in. To overcome this restriction, a dictionary use researcher can undertake to supply custom-made dictionary entries, where the experimental variables are manipulated by the researcher, and all the other variables are tightly controlled. In such a set-up, special dictionary entries, often in a number of versions, have to be created just for the purpose of the experiment. This was the approach taken by Tono (1984), and a similar approach has been adopted for this study.

2.6.3 Control of dictionary selection and success rates

In general, monolingual dictionaries tend to require greater reference skills as well as higher language proficiency compared with bilingual dictionaries. Inversely, learners with better reference and language skills are more likely to reach for monolingual dictionaries, as is repeatedly shown in almost all relevant studies. This interrelationship poses methodological problems for those designs where dictionary choice is left within the subjects' control.

As an illustration, consider Atkins and Varantola (1998a: Table 17), who report success rates with bilingual and monolingual dictionaries on tasks requiring reference and language skills. The problem is that dictionaries were selected by the subjects themselves. If, as numerous studies show, there is a systematic positive relationship between skill level and the preference for the monolingual dictionary type, then the success rate figures for monolingual dictionaries would automatically become inflated. In other words, if subjects preferring monolingual dictionaries tend to have better reference and language skills, those better skills will contribute to better scores on test measures, quite apart from any benefit of dictionary use.

To remedy this problem, dictionary type should be assigned randomly by the researcher. Unavoidably, some subjects will then end up working with a dictionary that is different from the dictionary they are accustomed to using, which might in turn negatively affect their performance. Random assignment of dictionaries would thus make the dictionary consultation less naturalistic, more removed from real-life experience of the dictionary users, compared to allowing subjects to choose their own dictionaries.

Bensoussan et al. (1984) was one study where both these options regarding the selection of dictionaries were employed. This investigation of the effect of dictionary use on text comprehension scores included three similar but separate studies. Dictionaries were randomly assigned in the Ben Gurion University study, but they were selected by the subjects in the two Haifa University studies. No interesting or significant differences were found in any of the three studies, though it must be noted that the overall rate of dictionary use was low (see 2.6.1 above for a discussion of the problem of dictionary underuse), and the Ben Gurion University sample was much smaller than any of the two Haifa University studies, with 91 subjects against 670 and 740, respectively.

Ultimately, the decision as to how dictionary assignment is to be handled in a particular study of dictionary use should be guided by the goal of the study. If

the investigators want to know how subjects work with the very dictionaries they are normally using, then subjects should be asked to bring their own dictionaries. If, however, comparing the level of helpfulness of the different dictionaries or dictionary types is the primary goal, random assignment seems a better option. There is another related issue here: to fairly compare dictionary types rather than specific dictionary products, specially prepared dictionary entries should be used, as argued in 2.6.2 above. If this solution is adopted, the use of subjects' own dictionaries is out of the question.

2.6.4 Lexical preknowledge

In studying the effect of dictionary consultation on various aspects of lexical behaviour, such as word comprehension or word acquisition, investigators often face the problem of lexical preknowledge. Lexical preknowledge is the knowledge about lexical items which subjects bring with them into the experimental setting and which they draw upon during the conduct of the experiment.

The methodological concern related to lexical preknowledge is that subjects' lexical behaviour may be based not so much on experimental treatment (such as dictionary consultation) as on their pre-existing knowledge of the target lexical items. The extent to which subjects' behaviour is motivated by preknowledge varies individually, and it is difficult to assess, which presents a threat to the internal validity of studies.

There may be several ways of dealing with the problem of lexical preknowledge (other than not addressing the issue at all). Words of appropriately low frequency may be used. One possible problem with this solution is that low-frequency items tend to be less polysemic and more complex in form than high-frequency words. That, however, may not always be a problem, depending on the particular goals of a given study. Such a method of selecting target words may be supplemented – or replaced – by piloting, in which a group of subjects similar to proper study subjects report on their knowledge or ignorance of the target item candidates. Sometimes the problem of lexical preknowledge is simply ignored (e.g. Black 1986; Luppescu and Day 1993), in which case we might also expect the problem of dictionary underuse (see 2.6.1 above) to be aggravated.

An ingenious solution to the problem of lexical preknowledge was adopted by Hulstijn, Hollander and Greidanus (1996) in their excellent study of incidental vocabulary acquisition while reading. Lexical preknowledge of target items was tested explicitly: subjects were confronted with the target words and asked whether they had ever come across them. Subjects' responses were noted, but they were further corrected using the responses to test questions proper, so that if a subject had declared familiarity with a given target word but later gave an incorrect description of its meaning, it was assumed that the subject had been unfamiliar with that target word in the first place, and the lexical preknowledge value for that word was adjusted accordingly. This schema allowed Hulstijn, Hollander and Greidanus to catch and correct – based on subjects' incorrect de-

scriptions of the meaning of target words – some cases of false positives: those where subjects incorrectly declared a prior knowledge of target items. The correction procedure could not, however, have identified those cases of false positives where a correct identification of meaning resulted from vocabulary acquisition during the reading experiment itself. Nor could it have caught any false negatives, that is cases where subjects wrongly declared no prior knowledge of target items, irrespective of the outcome of the meaning test. Another study where subjects were tested for their lexical preknowledge was Knight (1994).

A different approach was adopted by Tono (1984; see also Tono 2001: Chapter 9), who employed pseudo-words as target words. Since pseudo-words are not actual words, there is very little chance of subjects having any prior knowledge of those particular "words," since they could not have possibly come across them in the past[12]. It cannot be ruled out, though, that subjects could formulate some initial hypotheses about pseudo-words based on a perceived analogy with actual words, or from what appears to be a familiar morphological structure. It is difficult to say how such effects could possibly be controlled or eliminated.

Pseudo-words have to be believable to pass off as actual words. If they look suspect because of, say, strange letter sequences, there is a danger that subjects will discover the ruse and modify their behaviour. Some of Tono's pseudo-words (e.g. *lectvus*, *muvitly*) do appear to have letter sequences that are very unusual in English.

2.6.5 Statistical testing

The body of quantitative studies in dictionary research is not particularly impressive, and experimental studies are still rare (Dolezal and McCreary 1999; Hulstijn and Atkins 1998; Tono 2001). It is all the more regrettable that many of those studies that did produce quantifiable results amenable to statistical methods failed to take advantage of the opportunities that inferential statistics offers. In far too many studies, impressionistic claims about theoretically important differences were made on the basis of sample or group means alone. In some cases, other researchers can still use the original data to compute interesting statistics, as did Tono (2001: 48) on Battenburg's (1991) data. Routinely, though, there will not be enough of the original data available to make such a reanalysis possible. It is regrettable when some of the effort that has gone into data collection, often a laborious process, appears to have been partially wasted, and the data have not been utilized to a fuller extent.

For example, it is a little disappointing that the most comprehensive survey of dictionary use to date (Atkins and Varantola 1998a), which took 14 years from conception to the publication of the final report, limits its presentation of results to cross-tabulation of frequencies and means, and conclusions appear to have

[12] On the use of pseudo-words see also Hulstijn (1993; 1997).

been based on an impressionistic assessment of the figures, without any objective statistical testing.

One of the most popular tools of inferential statistics for social sciences, analysis of variance (ANOVA), has been rarely used, though there were notable exceptions, such as the study by Hulstijn, Hollander and Greidanus (1996), a meticulously designed piece of research where multi-way ANOVA was expertly used. In some cases, statistical testing was attempted but apparently without a deeper understanding of the underlying philosophy. Wingate (2002) used two-way ANOVAs in her experiment, but she claims them to be one-way ANOVAs (Wingate 2002: 90). The misunderstanding probably arose from a confusion of the term *level* as the value of the factor (independent variable) with *level* as a synonym for *verbal ability level*, which appears to have been the second independent variable on top of dictionary type, each of the variables having its own two *levels* in the statistical sense.

Also, as already pointed out by Tono (2001: 72), modern statistical techniques, such as multidimensional scaling, could be used to reduce the sometimes mind-boggling complexity of dictionary use research data, but has not yet, to my knowledge, been so used. Further, loglinear analysis, a multi-dimensional generalization of the Chi-square statistic, could conceivably be employed as a data-mining technique to locate patterns in dictionary use questionnaire and test responses, and to verify metalexicographic hypotheses.

To conclude, it appears that statistical techniques have been underused or sometimes misused in dictionary use research, and their potential to reduce the large volumes of data that are often part and parcel of empirical dictionary use studies and assist the researcher in the inference process is underestimated. In the present study, I hope to break with this infamous tradition and attempt to employ statistics wherever it can be of assistance in revealing hidden patterns and replacing intuition in the inference process.

3. The study

The present chapter will outline the research questions (section 3.1), and describe the design (3.2), subjects (3.3) and procedure (3.4) of the study undertaken to answer the research questions. Data on the educational context of the study as revealed by the Teachers' Questionnaire and the Learners' Questionnaire will be reported in sections 3.5 and 3.6, respectively. A detailed description of the Dictionary Effectiveness Test (3.7) will conclude the chapter.

3.1 Research questions

As already indicated in the introduction, the present study was primarily designed as an exploratory enterprise, with a view to revealing patterns between a range of factors related to receptive dictionary use by Polish learners of English (and, likely, language learners more generally) and thus offering partial answers to the many unanswered questions in the area of dictionary use by language learners, as well as suggesting avenues for future research. On top of this general goal, the present study was designed with a number of specific research questions in mind; these questions will be listed below, grouped roughly by area of study within dictionary use research.

The first area concerned the frequency of dictionary use, including different dictionary types and its relationship to learner proficiency level. The research questions in this area were: How frequently do Polish learners consult dictionaries? Do Polish learners, like learners in other published studies, use bilingual dictionaries much more frequently than monolingual dictionaries? Within the bilingual category, are there any significant differences in the frequency of consultation between Polish-English and English-Polish dictionaries? What is the relationship between dictionary consultation frequency and proficiency level? As learners progress, do they start to use monolingual dictionaries more? Do they stop using bilingual dictionaries?

More specific than information on the use of dictionaries in general, is information on dictionary types and specific dictionary products. The research questions in this area focused on the identification of these products and assessing their relative popularity: What specific dictionary titles do Polish learners use? What are their dictionaries of first choice and second choice? What are the dictionary types of first and second choice? How do the preferences vary with users' proficiency level?

The fact that someone uses a dictionary does not necessarily mean that they value it as a tool. In this area, questions were asked about the evaluation of those dictionaries identified in response to previous questions (see above): How do users evaluate specific popular dictionary titles? How do they rate different dictionary types? Are monolingual dictionaries rated more highly than bilingual? Do us-

ers evaluate English-Polish dictionaries differently from Polish-English dictionaries? Are more advanced learners more critical in their ratings, less critical, or about as critical as lower-level learners? Are dictionaries of first choice evaluated more highly than second-choice dictionaries?

When dictionary users refer to dictionary entries, they may be looking for different kinds of information. This aspect of users' reference needs was probed with the help of the following research questions: How often do users refer to the different information categories typically offered in dictionaries: pronunciation, meaning, Polish equivalent, English equivalent, part of speech, syntactic structure, collocation, synonyms, style and register? What is the relative ranking of these different types of information? How do reference needs vary with learner proficiency level?

The focal area of the present study is the effectiveness of dictionaries in lexical tasks, with the following specific research questions: What is the relative effectiveness of monolingual, bilingual and semi-bilingual dictionaries on lexical receptive tasks? How does the relative effectiveness of the different types of dictionaries depend on learner proficiency level? How does the relative effectiveness of the different types of dictionaries depend on the amount of contextual information? Does the provision of definitions on top of equivalents improve the effectiveness of dictionaries? If both definitions and Polish equivalents are given in a dictionary entry, are definitions in English more helpful than definitions in Polish? If both definitions and Polish equivalents are given, how does their ordering affect the effectiveness of a dictionary entry?

As the emphasis of the present study was on lexical semantic aspects, issues of grammar were only treated marginally, and thus limited to one small isolated aspect: the plural inflection marker. The research questions asked under this rubric were: How does the use of different dictionary versions affect the handling of the plural inflection by users? How does learners' proficiency level interact with dictionary version here? How strong is the effect of proficiency level alone on plural inflection accuracy?

Like grammar, sense ordering has not been the primary focus of this study. However, many of the test entries were polysemous, and so the effect of sense ordering was considered with the following research questions in mind: How do users handle polysemous entries? Does the structure of entries predispose users in any way to tend to select either early or late senses? How do learners of different levels fare with the various dictionary versions in this respect?

3.2 Design issues

The present study combined several methods of data collection: teacher survey, learner survey by questionnaire and controlled experiment. Learners of English at a range of proficiency levels served as experimental subjects and also filled in the Learners' Questionnaire (see 3.6 below). Teachers of the learner subjects acted as

respondents for the Teachers' Questionnaire (see 3.5 below). In the experimental part of the study involving the Dictionary Effectiveness Test, six versions of mini-dictionaries were randomly assigned to subjects, as described in detail in 3.7 below. Before I move on to those details, however, I will briefly discuss some design issues.

3.2.1 Control of lexical preknowledge

As discussed in 2.6.4 above, lexical preknowledge may be a problematic issue in dictionary use studies involving the acquisition or comprehension of lexical items. As the present study included lexical tasks meant to be completed with the consultation of dictionary entries, prior knowledge of target items could interfere in two ways. First, if subjects believed they knew the target items, they would be less likely to resort to dictionary consultation. However, dictionary consultation was essential to the design of the study, quite irrespective of whether they really knew those words or just believed they did. The second complication that could result from subjects' prior knowledge of target items would obtain even if subjects did consult the relevant dictionary entries. It would lie in the subjects drawing on their prior knowledge for lexical information, rather than, as intended by design, from the dictionary entries presented in the test dictionaries. For these reasons, pseudo-English words were invented and used as target words in this study.

3.2.2 Choice of session format

As is usual and expected in investigations involving human subjects, the present study had to deal with the hard-to-resolve conflict between how much data the investigator wishes to obtain from subjects and how much data it is feasible to collect.

Subject time is expensive, both in terms of ethics and economics. Since all subjects involved in this study were learners of English at some stage of advancement, and it was during the time normally used for instruction that the testing was conducted, a single class period of 45 minutes (almost universal at Polish schools at most levels) was considered a maximum practically acceptable demand on subjects' time. The ethically relevant question is that any amount of time set aside for testing could not be used for actual instruction. There may, of course, be long-term benefits to learners of English, including those tested, arising from the present research, insofar as it may ultimately contribute to improvements in available dictionaries.

In deciding between single-session collection of data versus repeated-sessions (i.e. a procedure involving more than a single occasion for the same subjects), practical as well as ethical considerations come into play. Moving beyond a single session essentially leaves the experimenter with two choices: either the test instrument is left in the control of the subjects, or else some reliable method is used to keep track of the identities of the subjects from session to session. Both

these solutions raise many serious problems which in this case appear to outweigh the potential gains of extra time for subject responses.

Allowing the subjects to retain the test instrument from one session to another has a number of serious undesirable consequences. Perhaps the least of a problem is the most obvious one: subjects failing to return their questionnaires/tests at all due to forgetfulness or lack of care on their part. Much more insidious are any and all of the ways in which subjects' extended control over their test instruments can affect the responses in those questionnaires and tests that *are* ultimately returned to the experimenter. Having left the test instruments in the hands of the subjects, the experimenter would no longer be able to control the amount of time actually spent by subjects working out the answers to questions. Furthermore, one has to count with subjects tapping various sources for guidance in their responses. Such sources could include all sorts of language resources: not only dictionaries, but also, say, friends and family. A real danger would lie in collaboration between different subjects, especially as in the context of the present study they would normally spend time together anyway in other classes. Since in this study different subjects were assigned to different experimental treatments, collaboration could have been a very serious problem that could have invalidated the results. What is more, given that the present design employed nonce-formations for tighter control of lexical preknowledge, there was a real risk of subjects uncovering the true identity of those items by way of consulting dictionaries or people, with difficult-to-predict consequences for their test behaviour. Sealing the tests at the conclusion of the first session might remedy some, but by no means all of the above problems, but it would involve additional complication, cost and labour.

The other way of extending subject responses beyond one session would be to collect the test instruments at the end of the first session and hand them out again for the subsequent session or sessions; or to split instruments into parts and distribute them to subjects separately. In both these cases, some way of tracking the identities of the subjects would have to be implemented. Whether involving true names, student numbers or nicknames, any such tracking system would inevitably alter the perception of the exercise as a truly anonymous procedure. Having subjects identify themselves in writing would likely have negatively affected the honesty of at least some of the responses, particularly the more sensitive ones.

Another practical aspect was that at least some of the teachers allowing the use of their class period and learners may not have afforded more than a single class period for the purpose of this study. A decision to administer the study over multiple sessions could thus have reduced the sample of subjects, which would have been undesirable for obvious reasons.

All in all, given the above considerations, a decision had been made at a fairly early stage that all data must be collected during a single session. The disadvantage of this solution was the restriction on time available to subjects for

working with their tests and questionnaires, thus limiting the number of items that could be included in the instruments. However, a large number of items is not necessarily better if we consider the potential negative effect of questionnaire fatigue, whereby after some time of continuous filling in, subjects' responses become less and less reliable.

3.2.3 Resolving the problem of dictionary underuse

As discussed in 2.6.1 above, experimental studies of dictionary use often face the problem of dictionary underuse during experimental treatment. Such failure to consult dictionaries masks any potential effects of dictionary consultation. Based on the existing reports, I saw this problem as a serious one, and so I took several steps to minimize the dictionary underuse effect.

First, instructions were placed at the top of the test page in the booklet encouraging subjects to consult the attached dictionary. Second, all words (most of them being pseudo-words) treated in the test dictionary were presented in bold type, and the instructions pointed out that words printed in bold can be found in the dictionary. Third, further encouragement to consult the attached dictionaries was provided by experimenters in oral instructions given during the test sessions. Fourth, the booklet was bound is such a way that the dictionary entries were immediately available on the facing page of the test booklet, obviating the need for any page-turning. Fifth, the test dictionary was only a single page long.

As a consequence of these last two design features, the act of dictionary consultation was simplified compared to a natural act of paper dictionary consultation. The effort involved in reaching for the dictionary and opening it (and also, optionally, selecting the dictionary) was eliminated. Further, access difficulty related to the size of the wordlist was heavily reduced: it was much easier to find the entry sought.

These measures are likely to have made our experimental dictionary consultations easier than is usual in natural consultation acts. As explained above, this was deliberate. The design probably affected such aspects of dictionary consultation as entry access time or rate of success in locating the entry – these aspects cannot be reliably studied with such a design, and they were not studied here. The focus of this study was on the microstructure, and specifically on the semantic information within the entry. Since the stage of locating the entry (headword) in the consultation act is likely to be fairly independent of the stage of locating the sense within the entry and extracting semantic information from it (cf. Scholfield 1982), our research questions should be well-served by the present design.

Subjects' dictionary consultation was not monitored in any way, so the precise extent of dictionary consultation is not known. However, the statistically significant effect of dictionary version (see 4.5 below) testifies to the success of this strategy, because if subjects had not used the dictionaries, it would not have mattered which dictionary version they had at their disposal.

3.3 Subjects

Study subjects were all Polish learners of English and (for the Teachers' Questionnaire) their respective teachers. The 712 learners participating in the study came from 44 organized learner groups (class sections) at 22 different educational institutions around Poland. A complete listing of the participating institutions is given in Appendix 8. Data on subjects comes from the Learners' Questionnaire (see 3.6 below) and – indirectly – from the Teachers' Questionnaire (see 3.5 below).

Profiles of educational institutions and learner groups making up the sample used in the study are discussed in 3.5.1 below. The sample is on the whole opportunistic: systematic sampling was not feasible within the resources available, but the sample does offer a broad selection along the social (age, background), educational (school level, class level, school type), and, to a lesser extent, geographical (region, city/town/country) dimensions.

3.4 Procedure

All data were collected between December 1999 and May 2000. The Learners' Questionnaires and Dictionary Effectiveness Tests were both administered, for a given group of subjects, during a single session (see 3.2.2 above), 45 minutes in duration, under the supervision of one of 20 experimenters. All experimenters except one were English majors at the School of English, Adam Mickiewicz University, enrolled in one of lexicography-related seminars taught by the researcher. The researcher was the one remaining experimenter (for four subject groups). The 19 experimenters received full hands-on training from the researcher with the particular materials used in the study. They also received detailed written instructions to be followed during the sessions (reproduced in Appendix 1, English translation available in Appendix 2).

Teachers completed their Teachers' Questionnaires at their leisure, though they were instructed to read the instructions beforehand and, as far as possible, fill out their questionnaires either during the testing session, or as soon afterwards as practicable, taking notes during the session if necessary.

All test booklets were provided to the experimenters in labelled envelopes, one sealed set for each group of subjects. The labels were generated from a database to maintain consistency and order, and to minimize the risk of any confusion or switching. The forms in the envelopes were also labelled with unique envelope keys. All forms were returned in their original envelopes. Data from all questionnaire and test forms were entered by the researcher into a relational database designed for this particular purpose (Lew 2003 gives the details). The way the questionnaires and tests were constructed, there was little room for subjectivity in interpreting the data from the questionnaire and test forms. Nevertheless, a random selection of about 10% of all papers were read and keyboarded independently by another evaluator, a specially trained lexicography major. This procedure yielded

a 0.993 interrater agreement (measured by Cohen's simple kappa coefficient), and the slight discrepancy was due solely to isolated clerical errors on the part of the second evaluator.

Once the database was complete, SQL queries were written by the researcher to extract the relevant record data into statistical packages, so that the research questions could be addressed with the full support of the descriptive and inferential statistical apparatus.

3.5 Teachers' Questionnaire

The Teachers' Questionnaire forms (Appendix 1, see also Appendix 2 for an English translation) were completed by teachers of each of the subject groups, one form per group. In these questionnaires, teachers were asked to fill in a number of questionnaire items related to the teaching and level of English in the group, and provide details of the experimental session. In the small number of cases where the teacher and the experimenter were different people, the two shared the task of filling in the questionnaire, each taking charge of the relevant section.

3.5.1 Educational institutions

Study subjects came from 44 formal learner groups representing 22 different educational institutions of different levels and profiles. A complete list of educational institutions is given in Appendix 8.

Table 3 provides a list of school types, specifying the number of subjects coming from a particular type, and the numerical level rating assigned to represent the overall English language level for a given type of school in Poland. The ratings are based on the positioning of the particular school types in the hierarchical structure of the Polish educational system, aided to some extent by my personal experience as educator in the Polish EFL setting.

Table 3: School types, number of subjects and assigned numerical EFL level ratings

Type	Polish expansion	approx. English equivalent	Count	Percent	Level rating
G	gimnazjum	middle school	86	12.1	2
LO	liceum ogólnokształcące	high school	284	39.9	3
LZ	liceum zawodowe	trade school	83	11.7	2
NKJO	nauczycielskie kolegium języków obcych	teachers' college	14	2.0	5
SJO	szkoła języków obcych	language school	40	5.6	3
SP	szkoła podstawowa	elementary school	57	8.0	1
SW	szkoła wyższa	university non-English major	52	7.3	4
T	technikum	trade school	62	8.7	2
U	uniwersytet: filologia angielska	university English major	34	4.8	6

A detailed distribution of study subjects by their school's level rating is presented in Table 4 and Figure 1. The modal value of the rating is three, and the distribution is skewed to the right. The skewness should not be seen as a design flaw with respect to distributions related to educational hierarchies: in general, there are more people getting their education at the more basic level than there are those at the more advanced level. This is one respect in which the study sample, though opportunistic, appears to be a realistic reflection of the general population of Polish learners of English.

Table 4: Breakdown of study subjects by level rating of their educational institution

Level rating	Count	Cumulative	Percent	Cumulative
1	57	57	8.0	8.0
2	231	288	32.4	40.4
3	324	612	45.5	86.0
4	52	664	7.3	93.3
5	14	678	2.0	95.2
6	34	712	4.8	100.0

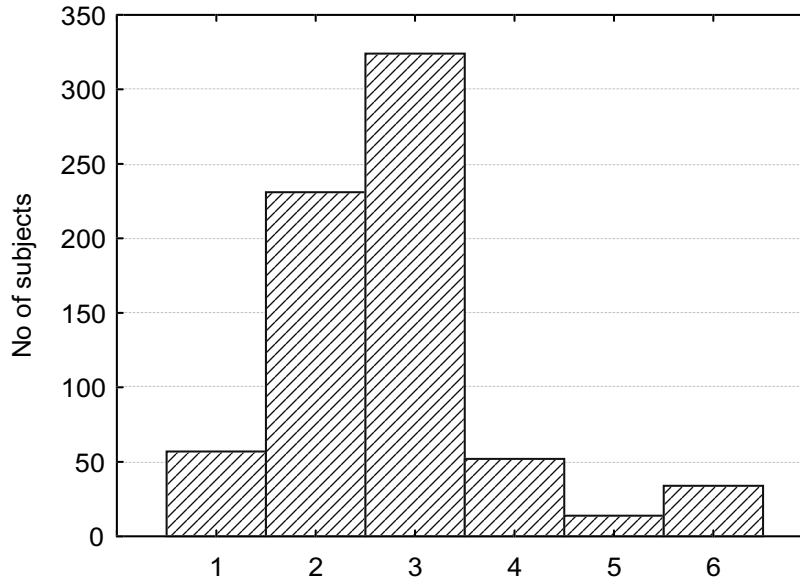

Figure 1: Distribution of study subjects by level rating of their educational institution

3.5.2 Teachers' assessment of proficiency level

In their Teachers' Questionnaires, teachers were asked to assess their learner group's EFL proficiency level, using the following scale: beginner, pre-intermediate, intermediate, upper intermediate, advanced. The breakdown of teachers' responses is given in Table 5, and the distribution is illustrated graphically in Figure 2. The second column of Table 5 gives a numerical code for proficiency level used in the abscissa in Figure 2 as well as throughout this study.

Table 5: Teachers' assessment of proficiency level

Proficiency level	Num	Count	Cumulative	Percent	Cumulative
beginner	1	178	178	25.0	25.0
pre-intermediate	2	255	433	35.8	60.8
Intermediate	3	192	625	27.0	87.8
upper intermediate	4	32	657	4.5	92.3
advanced	5	55	712	7.7	100.0

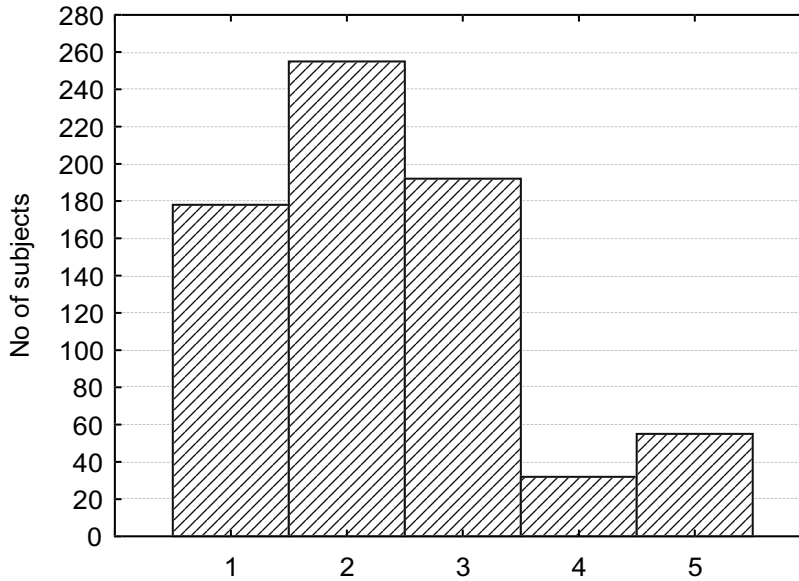

Figure 2: Distribution of teachers' assessment of subject group proficiency level

Comparing Figure 1 and Figure 2, we note that both are skewed to the right. The skewness is more radical in Figure 2, the mode is also lower (with a value of two), and there is a distinctly larger proportion (25% versus 8%) of subjects at the lowest value on the scale. This difference is likely indicative of the fact that there are students whose EFL learning is somewhat delayed when viewed against the educational opportunities currently provided by schools. Possibly, this could be a matter of personal choice: some students might aspire to attend higher-level classes than is warranted by their proficiency level. Alternatively, schools may have expanded by adding higher-level classes to their course offerings which may not yet have been available at the time when the student was opting for a given school and class.

3.5.3 Textbook level

Teachers were also asked about the primary textbook they used with their group for EFL instruction, including information on the text's target proficiency level. The responses are tabulated in Table 6 and their distribution is represented in Figure 3. There is a clear dominance of pre-intermediate textbooks amongst the study groups.

Table 6: Textbook level

Textbook level	Num	Count	Cumulative	Percent	Cumulative
beginner	1	39	39	5.5	5.5
pre-intermediate	2	442	481	62.1	67.6
intermediate	3	132	613	18.5	86.1
upper	4	42	655	5.9	92.0
advanced	5	9	664	1.3	93.3
post-advanced	6	48	712	6.7	100.0

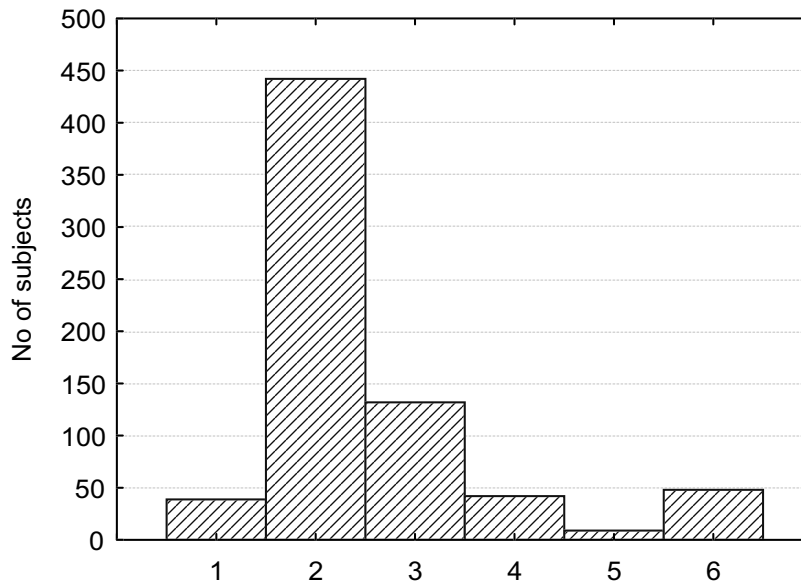

Figure 3: Distribution of textbook level

3.5.4 Learner level

Many of the aspects of dictionary use are potentially sensitive to the users' level of proficiency. In fact, the interaction between the various aspects of dictionary use and learner proficiency is among the more interesting ones, theoretically as well as practically, as dictionary products are more and more often targeted at specific proficiency levels. For example, it is often claimed (Baxter 1980) that monolingual dictionaries are more suited to the needs of the more advanced learners. It was in recognition of the importance of learners' level as a variable, that the sample of the present study was structured so as to represent a broad range of learner groups from various institutions.

For learner level to be introduced as a predictor variable in a study, it needs to be measured. One way to measure subjects' proficiency level would be to test it with a language proficiency test (as done by Atkins and Varantola 1998a). However, an even minimally reliable language proficiency test would require an

amount of time that would make it impossible for such a proficiency test to be accommodated in a single session with the Learners' Questionnaire and the Dictionary Effectiveness Test. On the other hand, as argued in detail in 3.2.2 above, stretching the study into two separate sessions was not an acceptable option.

Furthermore, language proficiency tests themselves suffer from inherent construct validity problems, and measures obtained in this way are polluted to an unknown extent by factors such as test-taking skills, test anxiety, motivation, stress, fatigue, time pressure, and others (Wainer and Braun 1998).

For these reasons, learner level in this study is operationalized as a compound construct derived from a number of items elicited in the Learners' Questionnaire and the Teachers' Questionnaire[13]. The components that go into the computation of learner level are the duration and intensity of EFL instruction (items A and B), peer-relative self-assessment of proficiency (item C), self-assessment of success at language tasks (items D1-D3), teachers' assessment of group level, publisher-declared textbook level, school level (positioning in the hierarchy of the Polish educational system), and class level (grade progression within a school).

Technically, the learner level variable was computed through an SQL query on the database tables holding the test scores and questionnaire results. The query is given in the box below:

```
[A]+[B] AS QEdLevel, IIf(IsNull([C]),2,[C])-
2+IIf(IsNull([D1]),2.5,[D1])+IIf(IsNull([D2]),2,[D2])+IIf(IsNull([D3]),2,[
D3]) AS QEnLevel, [QEdLevel]+[QEnLevel] AS QLevel,
Sessions.TeachLevel, Sessions.TextLevel, [TeachLevel]+[TextLevel] AS
TTLevel, SchoolLevel.SchoolLevel, Sessions.ClassLevel,
3*[SchoolLevel]+[ClassLevel] AS EdLevel,
2*[TTLevel]+[EdLevel]+[QLevel] AS AllLevel,
IIf([AllLevel]<28,1,(IIf([AllLevel]<32,2,(IIf([AllLevel]<38,3,(IIf([AllLeve
l]<46,4,5)))))))) AS [Level]
```

A continuous measure of learner level obtained in this way was then partitioned into five discrete categories for the purpose of ANOVA computations, based on level score intervals. The breakdown of level into the five categories is presented in Table 7 and Figure 4.

[13] Compare MacFarquhar (1983) for the use of questionnaire data to group subjects into proficiency levels, and Albus (2001) for a similar use of self-assessment of proficiency.

Table 7: Breakdown of learner level scores into five discrete categories

Level	Count	Cumulative	Percent	Cumulative
1	145	145	20.4	20.4
2	131	276	18.4	38.8
3	175	451	24.6	63.3
4	154	605	21.6	85.0
5	107	712	15.0	100.0

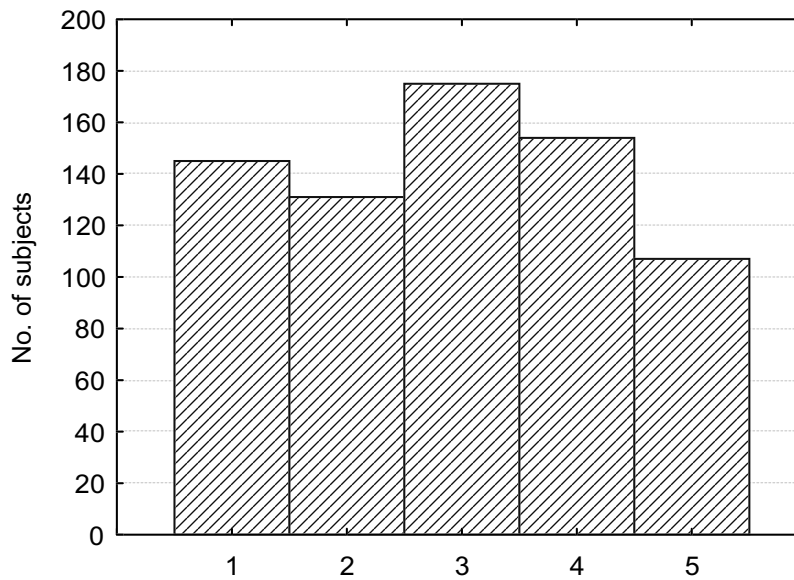

Figure 4: Breakdown of learner level scores into five discrete categories

The five level categories contain roughly similar numbers of scores, with the middle level 3 representing the highest count of 175 subjects (a quarter of the sample), and the top level 5 holding the lowest number of 107 subjects (15 percent of the sample). These differences are a reflection of the clustering of the level values.

In the following pages, whenever learner level is employed as a predictor variable, it is used in the sense of this discrete variable ranging in value from 1 to 5.

3.5.5 Conclusion

The Teachers' Questionnaire provided data on the educational background of the subjects participating in the study, as well as details of the experimental sessions with 44 learner groups coming from 22 educational institutions. The largest section of the sample came from liceum ogólnokształcące, the most popular general-education school type at the secondary level, roughly equivalent to high school in

the United States or comprehensive school in the United Kingdom. In terms of the school level, our sample reveals a right-skewed distribution with a mode of three on a scale of six. Teachers' assessment of subjects' proficiency level also produced a right-skewed distribution with the modal value of two out of five. A similar distribution pattern was found on textbook level data. This suggests that the sample may be a fairly good reflection of the general population of Polish learners of English, with a greater number of learners at the lower proficiency levels than at the top levels. By combining the various data reported by learners and teachers using fitted weighting ratios, a compound measure of learner level was obtained. It was subsequently partitioned into five discrete level groups of roughly similar size. We will now move on to the Learners' Questionnaire.

3.6 Learners' Questionnaire

In this section I report on the results of those items from the Learners' Questionnaire which do not directly relate to dictionary use proper, but rather provide information on individual subjects' educational background and foreign language skills level.

3.6.1 Duration of EFL instruction

Duration of EFL instruction received by subjects, expressed in years, was elicited in item A of the Learners' Questionnaire. A breakdown of responses to question A is given in Table 8 and Figure 5.

Table 8: Years of English instruction

Years	Count	Cumulative	Percent	Cumulative
0-1	62	62	8.7	8.7
2-3	183	245	25.7	34.4
4-5	230	475	32.3	66.7
6+	237	712	33.3	100.0

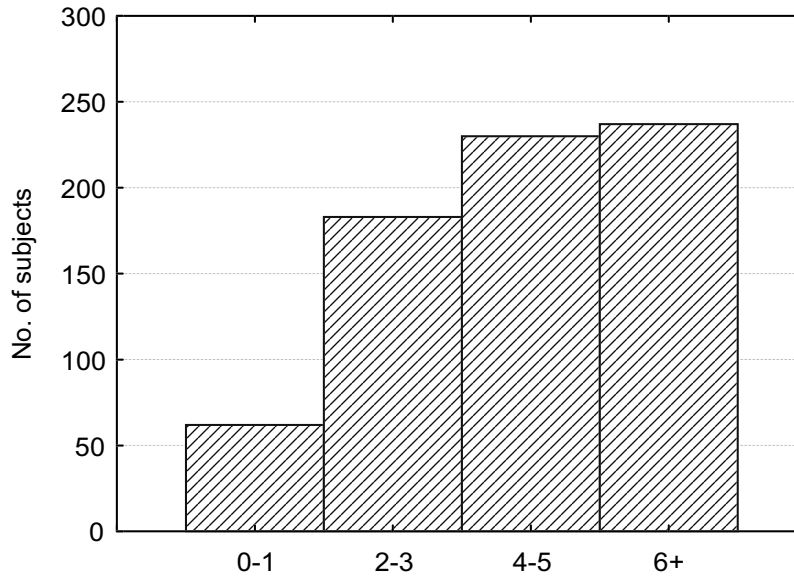

Figure 5: Years of English instruction

The sample is split fairly equally into three groups: those with six or more years of instruction (33.3%), those who have been studying English for a period of four to five years (32.3%), and those with up to three years of English (25.7%+8.7%=34.4%). Within this last category, about a third are complete beginners, reporting only up to one year of EFL instruction.

3.6.2 Number of hours of instruction per week

Question B in the Learners' Questionnaire asked about the number of teaching hours of EFL instruction currently received by student subjects. Teaching hours are understood in the Polish setting as instructional periods of 45 minutes each. A breakdown of responses is given in Table 9 and Figure 6.

Table 9: Hours of EFL instruction per week

Hours	Count	Cumulative	Percent	Cumulative
1	15	15	2.1	2.1
2-3	462	477	64.9	67.0
4-5	163	640	22.9	89.9
6+	72	712	10.1	100.0

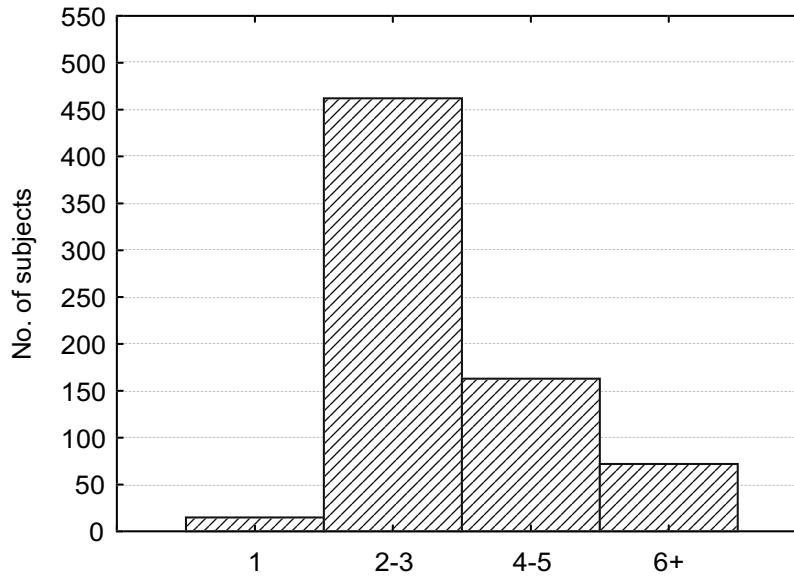

Figure 6: Hours of EFL instruction per week

Nearly two-thirds of the subjects (64.9%) report receiving between two and three hours of EFL instruction per week. A third (33%) take more than four hours a week, with every tenth subject in the sample getting at least six hours of English instruction weekly. Only a very small minority (2.1%) are getting just one hour of English a week.

3.6.3 Peer-relative self-assessment of proficiency

In question C of the Learners' Questionnaire, subjects were asked to assess their proficiency in the English language relative to their classmates. A break-down of responses to question C is presented in Table 10 and Figure 7.

Table 10: Self-assessment of EFL proficiency level

Self-assessment	Count	Cumulative	Percent	Cumulative
below average	82	82	11.5	11.5
average	533	615	74.9	86.4
above average	96	711	13.5	99.9
missing	1	712	0.1	100.0

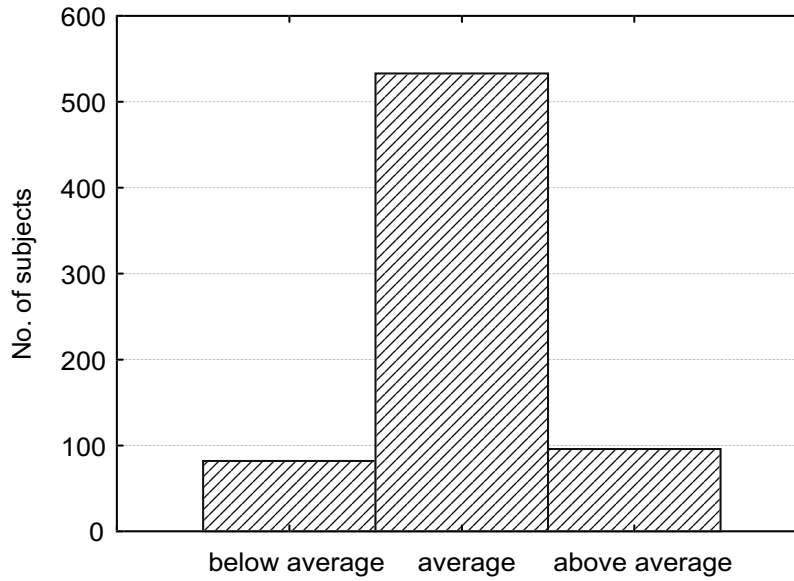

Figure 7: Self-assessment of EFL proficiency level

A strong majority of the subjects rated their proficiency in English as average within their peer groups, and the distribution of those who rated themselves as other-than-average is fairly equally split between *below average* and *above average*.

The symmetry of the distribution (see Figure 7) makes good statistical sense, but we should be wary of the suspiciously low dispersion of responses in what is likely to be perceived as a psychologically sensitive question. Even though responses were anonymous, the social and psychological pressures and feelings of solidarity with the group may well have prevented at least some subjects from rating themselves as different than average. Still, 13.5% of the subjects did report themselves as *above average* while 11.5% gave the *below average* response.

3.6.4 Self-assessment of success at language tasks

Questions D1-D3 of the Learners' Questionnaire asked subjects to assess if they would be capable of performing three language tasks: asking for directions to a station (D1); describing the symptoms of a cold (D2); understanding the lyrics of a song while listening to the song (D3).

The design intention behind items D1-D3 was to provide a range of hypothetical language tasks so as to obtain a compound measure of reported foreign language expertise at a range of ability levels.

A joint breakdown of responses to all three questions (D1-D3) is presented in raw counts and percentages in Table 11.

Table 11: Self-assessment of success on three language tasks: D1=ask for directions to a station; D2=describe the symptoms of a cold; D3=understand a song's lyrics while listening

Can you…?	D1	D1%	D2	D2%	D3	D3%
No	15	2.1	59	8.3	18	2.5
Probably not	87	12.2	201	28.3	118	16.6
Probably yes	288	40.5	271	38.1	405	57.0
Yes	321	45.1	180	25.3	170	23.9
Total	711	100.0	711	100.0	711	100.0

A box plot showing the means and standard deviations for the three questions is shown in Figure 8. To obtain the mean and standard deviation values, responses were recoded to a linear integer scale as follows: *No* as 1; *Probably not* as 2, *Probably yes* as 3, *Yes* as 4.

As expected, asking for directions to a station, a basic communicative task (D1), proved to be the least challenging of the three tasks, resulting in the highest overall ratings of success, with the mean rating somewhat better than *Probably yes* (Figure 8; the mean value is 3.29). Again as expected, describing symptoms of a cold (D2), a productive task, proved to be seen as more difficult on average than asking for directions, with the mean rating of 2.80, that is below *Probably yes*. The receptive task of understanding lyrics of a song while listening was judged, overall, to be of medium difficulty by the average subject, with a mean rating response of *Probably yes* (a numerical value of 3.02). However, the number of unqualified *yeses* for this item was the lowest of all, which may reflect the objective difficulty of getting a task described in D3 completely right. It is also inevitable that respondents did not all conceptualize successful outcomes in exactly the same way.

In terms of the variability of responses, it was the *describe symptoms* task (D2) that caused the responses to be spread rather more evenly over the four options than for the other two tasks. Responses to items D1 and D3 varied less across subjects than did those for D2.

Figure 8: Box plot of means and standard deviations for questions D1-D3, measuring self-assessment of success in three language tasks. Numerical values on the ordinate code responses as follows: 1=No; 2=Probably not, 3=Probably yes, 4=Yes.

A combined histogram of responses to questions D1-D3 is shown in Figure 9. The graph transparently illustrates the relative dominance of high-rating responses (probably yes and yes) for item D1, as well as the comparatively larger spread of responses for item D2.

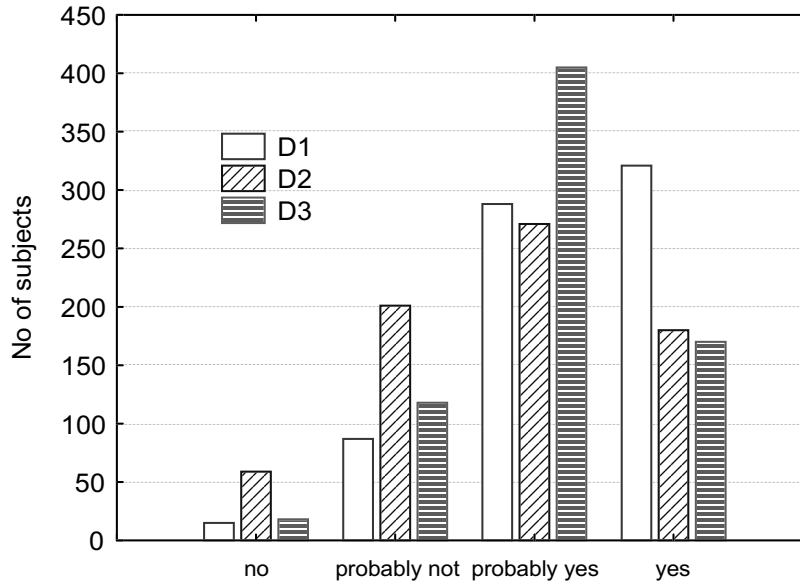

Figure 9: Combined histogram of responses to questions D1-D3

3.6.5 Conclusion

Apart from data on dictionary use, which will be presented and discussed in sections 4.1-4.4, the Learners' Questionnaire provided information on the participating subjects' educational background and proficiency level, including the duration of EFL instruction, the number of hours of instruction received per week, self-assessment of proficiency and success at three language tasks. Two-thirds of the learners in the sample reported having studied English for at least four years. A similar majority of subjects took between two to three hours of instruction per week. Most participating learners rated their proficiency level as average against their class peers. Learners believed they would be fairly successful on the three language tasks listed in the questionnaire, with "asking way to station" achieving the highest success ratings. These data were used in the determination of learner level, as described in 3.5.4 above. We will now move on to the Dictionary Effectiveness Test.

3.7 Dictionary Effectiveness Test

In all subject groups, the Dictionary Effectiveness Test was administered during the same session as the Learners' Questionnaire, directly following the questionnaire at the prompt of the experimenter. The test involved a number of test tasks completed with the help of specially prepared dictionary pages included in the test booklet and presented on the page directly facing the test items. There were six different versions of the dictionary, assigned randomly to subjects.

3.7.1 Test tasks

The Dictionary Effectiveness Test (Appendix 5, see also Appendix 6 for an English translation) consisted of a number of lexical tasks designed to be completed with the assistance of the dictionary entries provided in the same test booklet. Particular emphasis was placed on maximizing dictionary use, as described in 3.2.3 above. The test tasks varied in the amount of textual context accompanying the target lexical items. There was a grading of contextual material from none (isolated lexical items), through sentence-long, up to a short text of several sentences.

3.7.1.1 Out of context

Test section H consisted of two types of items. The first part (H1-H5) involved the semantic matching of a target (pseudo-English) lexical item with a Polish lexical item closest in meaning, that is picking an equivalent, with one best answer and three distractors. It should be pointed out here that in most cases the closest equivalent was deliberately *not* explicitly listed in any dictionary version, to eliminate any straightforward mechanical matching without semantic processing.

The second part of test task H (H6-H7) involved semantic matching of a target (pseudo-English) lexical item with an English item and thus resembled synonym matching. Here again, there were three distractors per item, and all three as well as the target pseudo-word were listed in the attached dictionary.

3.7.1.2 Sentence context

Test sections I and J involved lexical tasks in sentence-long contexts: sentence completion and sentence translation. Section I was a sentence completion exercise, with an empty slot to be filled with one item out of a list of four items provided: one correct item and three distractors, each being pseudo-English words provided in the test dictionary. Section J was based on English-to-Polish translation, with English and pseudo-English target words (again provided in the test dictionary) featuring in each sentence to be translated.

3.7.1.3 Text context

In test section K, subjects were asked to translate a joke from English into Polish, with target words spread out throughout the text of the joke.

3.7.2 Mini-dictionaries

For the test section of the study, subjects were provided with specially prepared mini-dictionaries, bound in the same booklet as the Learners' Questionnaire and the Dictionary Effectiveness Test. Subjects were instructed and encouraged to refer to their dictionaries while engaged in the experimental tasks, but not while completing the Learners' Questionnaire. The dictionaries were laid out so that the entries were immediately accessible on the facing page (without any need for page-turning) during the completion of the experimental tasks. This layout was

designed to overcome the methodological problem of dictionary underuse (see 2.6.1 and 3.2.3 above for a discussion of this issue). In order to address the research questions, six versions of the dictionary were prepared (see Appendix 7), all set in identical font and paragraph styles, and with identical structure, differing only in the type of semantic explanation provided for the senses. The semantic explanation in the six versions was as follows:

1. Polish equivalents
2. definitions in English
3. Polish equivalents followed by definitions in English
4. Polish equivalents followed by definitions in Polish
5. definitions in English followed by Polish equivalents
6. definitions in Polish followed by Polish equivalents

Version 1 thus represents a prototypical bilingual dictionary. Version 2 is a monolingual dictionary, with definitions written in a style most closely approaching that of LDOCE, which was found to be liked by users (MacFarquhar and Richards 1983) as well as relatively effective (Cumming, Cropp and Sussex 1994; Nesi and Meara 1994).[14] Versions 3 and 5 are representative of the semi-bilingual type, and they differ in the relative ordering of the two types of semantic explanation. Versions 4 and 6 are like 3 and 5, respectively, except that the English definitions have been translated into Polish. In addition, in those versions that combined definitions and equivalents, definitions were enclosed in angled brackets (often employed as delimiters for glosses in Polish lexicography), in order to set them apart typographically from the neighbouring Polish equivalents.

Table 12 gives the six versions of the mini-dictionary, contrasted in terms of the types of semantic explanation and their relative ordering.

Table 12: The six dictionary versions used in the Dictionary Effectiveness Test

dictionary version	Polish equivalent	definition	position of definition	language of definition
1	present	absent	N/A	N/A
2	absent	present	sole	English
3	present	present	last	English
4	present	present	last	Polish
5	present	present	first	English
6	present	present	first	Polish

The six versions of the experimental dictionary were assigned randomly to subjects, generating the six experimental groups.

[14] Another reason was that LDOCE has enjoyed wide popularity in Poland. LDOCE2 was reprinted in Poland soon after its original publication and offered at a very competitive price. Therefore, Polish learners may be expected to be more familiar with traditional LDOCE-style definitions than with the COBUILD-style full sentence definitions, which have also been found to be effective in some studies.

4. Results and discussion

In this chapter, I present and discuss subjects' responses to sections E-G of the Learners' Questionnaire and the results of the Dictionary Effectiveness Test. The Learners' Questionnaire items which will be covered here are the ones directly pertaining to dictionary use, as opposed to those dealing with personal and institutional educational context, which are presented and discussed in 3.6 above.

4.1 Frequency of dictionary use

In section E of the Learners' Questionnaire, subjects were asked to specify the frequency with which they consulted three categories of dictionaries: Polish-English (E1), English-Polish (E2) and monolingual English (E3). The split into Polish-English and English-Polish subtypes of the bilingual dictionary could be made in the Polish context because there is a tradition of publishing the two in separate volumes, except of course for the smallest dictionaries. But even for single-volume bilingual dictionaries (combined Polish-English and English-Polish), the question should still present no problems of interpretation, with E1 and E2 referring to the Polish-English and English-Polish sections, respectively, of the combined bilingual dictionary.

The options offered were (English translations given in parentheses): *codziennie* (*daily*), *kilka razy na tydzień* (*a few times a week*), *raz na tydzień* (*weekly*), *rzadziej/wcale* (*less frequently/not at all*). In tables and figures of this section this last option is abbreviated as *less frequently* for reasons of readability, but the *not at all* part was included explicitly after it had been found at the piloting stage that subjects did not necessarily assume *less frequently* to include *not at all*, even though in the researcher's original intention it was meant to cover this possibility.

4.1.1 Polish-English

In item E1 of the Learners' Questionnaire, subjects were asked to assess the frequency with which they consulted Polish-English dictionaries. A detailed breakdown of responses is presented in Table 13, and a histogram of responses is shown in Figure 10.

Table 13: Frequency of Polish-English dictionary consultation

Consultation frequency	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
less frequently	147	147	20.8	20.8	20.6	20.6
weekly	252	399	35.6	56.4	35.4	56.0
a few times a week	287	686	40.6	97.0	40.3	96.3
daily	21	707	3.0	100.0	2.9	99.3
missing	5	712	0.7	N/A	0.7	100.0

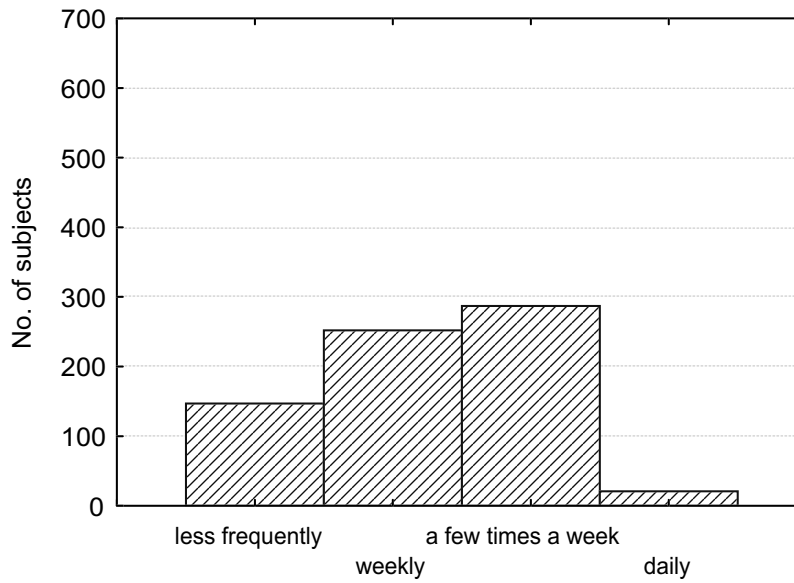

Figure 10: Frequency of Polish-English dictionary consultation

About 40% of the subjects consult a Polish-English dictionary a few times a week. In a close second place, weekly use is reported by a third of the subjects. Every fifth subject consults a Polish-English dictionary less frequently than once a week. Only 3% – a surprisingly low figure – use a Polish-English dictionary with daily regularity.

4.1.2 English-Polish

In item E2, subjects reported the frequency with which they consulted an English-Polish dictionary. Table 14 gives a detailed breakdown of responses to this item, and the histogram in Figure 11 presents the distribution of responses in graphical form.

Table 14: Frequency of English-Polish dictionary consultation

Consultation frequency	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
less frequently	133	133	18.8	18.8	18.7	18.7
weekly	228	361	32.2	51.1	32.0	50.7
a few times a week	312	673	44.1	95.2	43.8	94.5
daily	34	707	4.8	100.0	4.8	99.3
missing	5	712	0.7	N/A	0.7	100.0

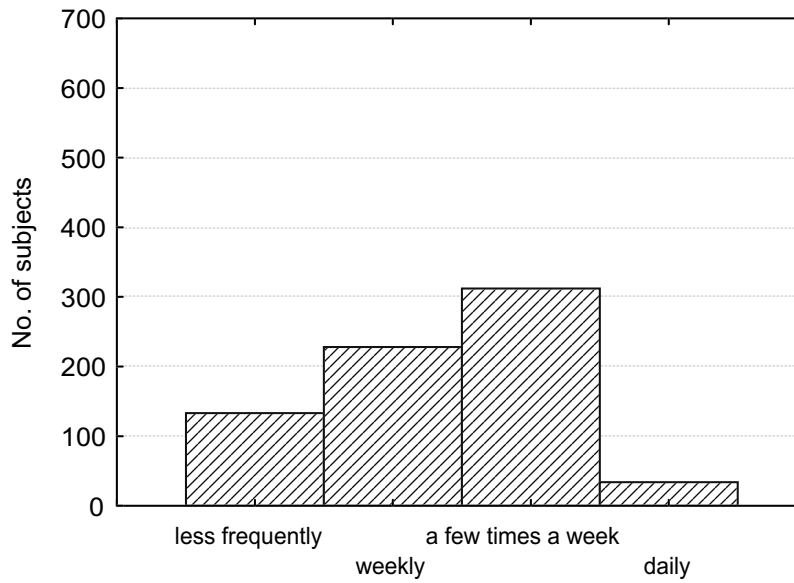

Figure 11: Frequency of English-Polish dictionary consultation

Basically, the distribution of responses for the English-Polish dictionaries is quite similar to that for Polish-English dictionaries. The response *a few times a week* dominates here even more, accounting for no less than 44% of all responses. 32% of the subjects report weekly use of English-Polish dictionaries, that is slightly less than for Polish-English. Close to 19% claim to consult English-Polish dictionaries less frequently than once a week. However, the proportion of daily users is noticeably higher for English-Polish dictionaries than for Polish-English dictionaries, with 5% and 3% of responses, respectively.

4.1.3 Monolingual English

In item E3 of the Learners' Questionnaire, subjects were asked to assess the frequency with which they consulted monolingual English dictionaries. A breakdown of responses is presented in Table 15, and a histogram of responses is shown in Figure 12.

Table 15: Frequency of monolingual English dictionary consultation

Consultation frequency	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
less frequently	583	583	84.0	84.0	81.9	81.9
weekly	35	618	5.0	89.0	4.9	86.8
a few times a week	45	663	6.5	95.5	6.3	93.1
daily	31	694	4.5	100.0	4.4	97.5
missing	18	712	2.6	N/A	2.5	100.0

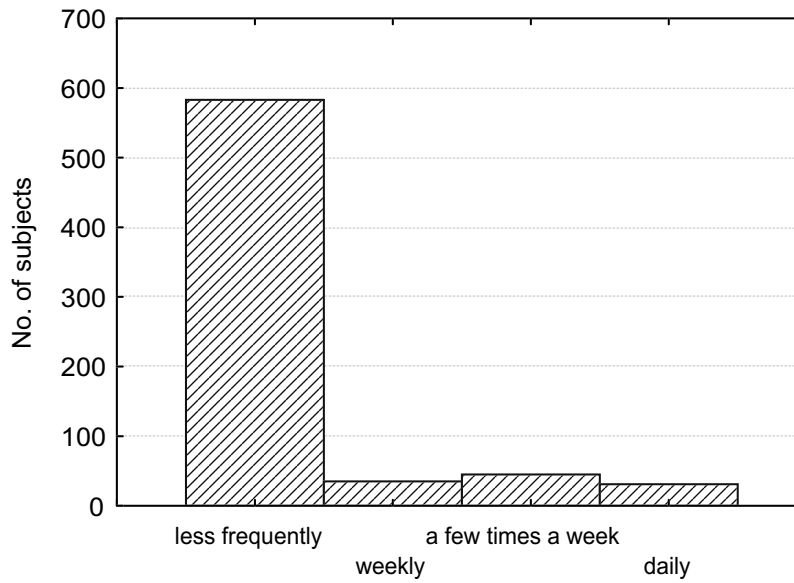

Figure 12: Frequency of monolingual English dictionary consultation

The distribution of reported consultation frequency for monolingual English dictionaries looks radically different from either of the bilingual dictionary types. As much as 84% of the subjects claimed they used monolingual English dictionaries less often than once a week. The three remaining frequency bands are split into roughly equal parts of around 5% each.

4.1.4 Bilingual versus monolingual

The contrast between the distributions for the three dictionary types shows up clearly in a combined histogram given in Figure 13.

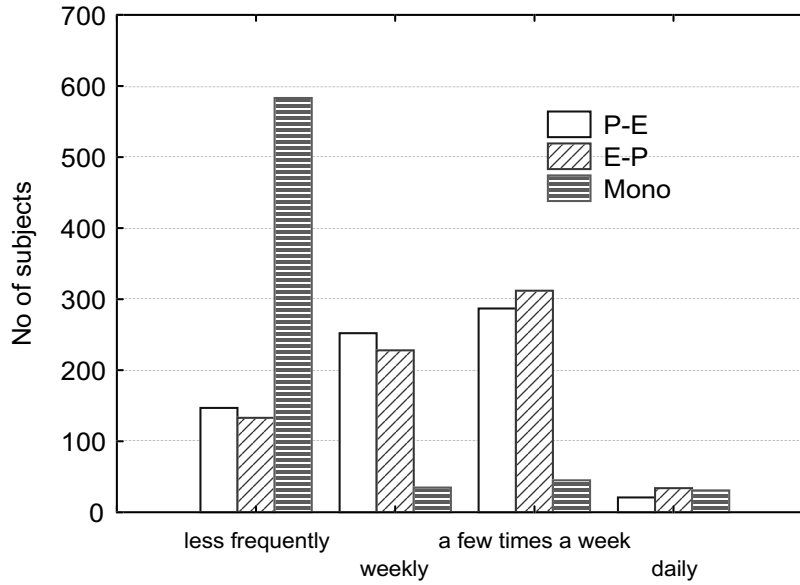

Figure 13: Reported frequency of dictionary consultation for Polish-English, English-Polish, and monolingual English dictionaries

The pattern in the histogram suggests a strong difference between the look-up frequencies for monolingual dictionaries versus (any section of) bilingual dictionaries. To verify this impression, textual responses were assigned numerical values as follows: *daily* = 4, *a few times a week* = 3, *weekly* = 2, *less frequently/not at all* = 1. These values were used to compute ranks and means of responses.

To test whether there is any difference between how frequently subjects, based on their own reports, consulted the three types of dictionaries (i.e. monolingual dictionaries and the two sections of bilingual dictionaries), a Friedman ANOVA was computed for repeated measures on the Polish-English, English-Polish, and monolingual English ranks. The Friedman statistic here compares the ranking of responses for the three types of dictionaries, and it does so separately for each individual subject, which makes it a suitably powerful statistic for the purpose at hand, even though it only requires measurement on a rank-order (ordinal) scale. The Friedman ANOVA Chi-square value is a very high 664.13 (at N = 691, df = 2) and it is highly significant at $p<0.0001$. A detailed Friedman ANOVA table is presented in Table 16, which, for completeness, also includes mean and standard deviation values.

Table 16: Friedman ANOVA table for a comparison of reported consultation frequency between Polish-English, English-Polish, and monolingual dictionary

Dictionary	Valid N	Average Rank	Sum of Ranks	Mean	Std.Dev.
Polish-English	707	2.27	1569.5	2.26	0.81
English-Polish	707	2.37	1640.0	2.34	0.83
Monolingual	694	1.36	936.5	1.31	0.77

Based on the Friedman ANOVA test results we can conclude that there are significant differences between the consultation frequencies for Polish-English, English-Polish, and monolingual English dictionaries. It is abundantly clear that a huge portion of this variation is due to the much less frequent consultation of monolingual dictionaries versus bilingual dictionaries.

An impressionistic comparison of the distributions of consultation frequencies reported for Polish-English and English-Polish dictionaries suggests a large degree of similarity between the two. However, the Friedman ANOVA for repeated measures on the Polish-English and English-Polish ranks only, yields a Chi-square value of 16.33 (N = 703, df = 1), and it is highly significant at $p<0.0001$. A relevant Friedman ANOVA table is presented in Table 17.

Table 17: Friedman ANOVA table for a comparison of reported consultation frequency between Polish-English and English-Polish parts of a bilingual dictionary

Dictionary	Average Rank	Sum of Ranks
Polish-English	1.46	1025.5
English-Polish	1.54	1083.5

Based on the Friedman ANOVA results for Polish-English vs. English-Polish dictionaries, we can conclude that even though the patterns of consultation frequency for the two bilingual dictionary parts are fairly similar, the difference between them is nevertheless highly significant. It is the English-Polish section that has the higher mean value of the two. The tendency for this English-Polish part to be more frequently consulted can also be discerned in the combined histogram in Figure 13, where the E-P bars are taller than P-E bars for the top two frequency values (*daily* and *a few times a week*), but shorter than P-E for the bottom two frequency values (*weekly*, *less frequently/not at all*).

The higher frequency of English-Polish versus Polish-English dictionaries may be due to the primacy of decoding needs of the subjects over their encoding needs. The effect is consistent with Tomaszczyk's (1979) findings, which included a more frequent use of L2→L1 dictionaries compared with L1→L2 dictionaries. It will be interesting to examine how the preference for English-Polish versus Polish-English relates to learner level.

4.1.5 Dictionary type versus level

In this section the relationship between dictionary type and learner level will be examined. There is widespread belief among language teachers and some researchers (e.g. Baxter 1980) that monolingual dictionaries are better suited for the more advanced learners, and appeals to learners to move from bilingual to monolingual dictionaries are often heard in the classroom as well as from the monolingual learner dictionary publishing houses.

Table 18 lists the mean and standard error values for reported frequency of dictionary consultation for Polish-English, English-Polish and monolingual English dictionaries broken down by learner level.

Table 18: Means and standard errors by level for reported consultation frequency for Polish-English, English-Polish and monolingual English dictionaries

Level	P-E Mean	P-E Std.Err.	E-P Mean	E-P Std.Err.	Mono Mean	Mono Std.Err.	N
1	2.22	0.07	2.15	0.07	1.10	0.05	143
2	2.17	0.07	2.22	0.07	1.08	0.05	123
3	2.13	0.06	2.27	0.06	1.08	0.05	169
4	2.32	0.07	2.39	0.07	1.13	0.05	152
5	2.55	0.08	2.80	0.08	2.49	0.06	104

Mean values of reported consultation frequency for Polish-English, English-Polish and monolingual English dictionaries are plotted against level in Figure 14.

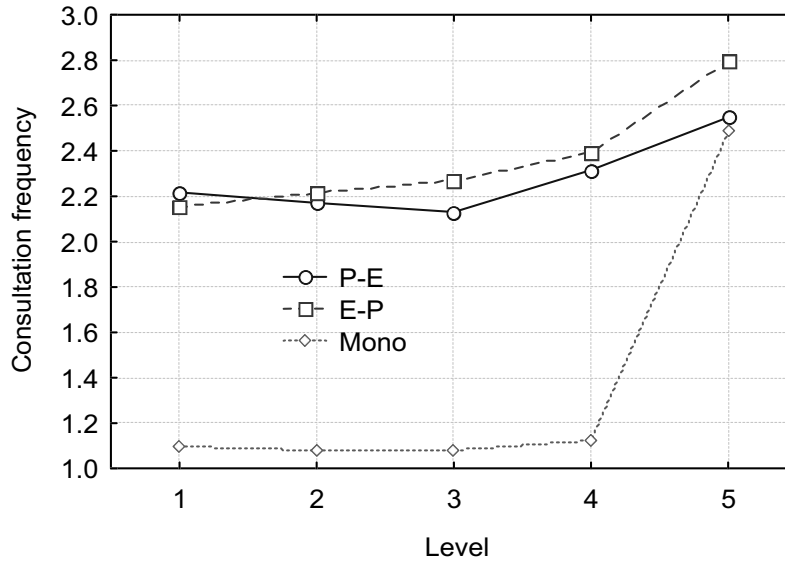

Figure 14: Reported frequency of dictionary consultation for Polish-English, English-Polish and monolingual English dictionaries as a function of level

As seen in Table 18 and Figure 14, there appears to be a tendency for frequency of dictionary consultation to increase with level for all three dictionary categories for which this measure was elicited. However, there are differences across the three types as to how sharp the increase is and the exact level range over which it is observed. For English-Polish dictionaries, the upward tendency is distributed across the whole level range, with the slope of the curve also increasing with level. For Polish-English dictionaries, the increase in frequency appears to be restricted to levels 4 and 5. This last tendency may be interpreted as reflecting an emergence of encoding use for learners between levels 3 and 4, and a subsequent growth of that use in intensity.

A very different picture emerges for the monolingual dictionary: here the reported frequency stays uniformly flat and very low all the way up to level 4, with a sharp peak at level 5, at which the frequency mean catches up with the Polish-English dictionary. The patterning of the frequency profiles for monolingual and bilingual dictionaries suggests that monolingual dictionary use is minimal except for the highest level, where the frequency of consultation approaches that for bilingual dictionaries. Interestingly, the monolingual dictionary does not appear to be used as a replacement for the bilingual dictionary by the advanced learners, but rather in addition to the bilingual dictionary, a finding that differs from that by Atkins and Varantola (1998a), where the use of bilingual dictionaries decreased more or less in step with the use of monolingual dictionaries.

A General Linear Model (GLM) analysis reveals a highly significant overall (multivariate) effect of level (Wilks lambda = 0.56760, $F_{(12,\ 1810)}$=36.003, p<0.0001). Details of this analysis are presented in Table 19.

Table 19: GLM multivariate tests of significance for reported consultation frequency by level for Polish-English, English-Polish and monolingual English dictionaries

Effect	Wilks lambda	F	Effect df	Error df	p
Intercept	0.073585	2870.452	3	684.000	<0.0001
Level	0.567596	36.003	12	1809.985	<0.0001

A univariate GLM analysis presented in Table 20 reveals highly significant univariate effects of level for all three types of dictionaries (Polish-English, English-Polish and monolingual English).

Table 20: GLM univariate tests of significance for reported consultation frequency by level for Polish-English, English-Polish and monolingual English dictionaries

Effect	df	P-E SS	P-E MS	P-E F	P-E p	E-P SS	E-P MS	E-P F	E-P p	M SS	M MS	M F	M p
Intercept	1	3479	3479	5382	<0.0001	3764	3764	5735	<0.0001	1268	1268	3618	<0.0001
Level	4	13	3.3	5.1	0.0005	29.8	7.4	11.3	<0.0001	172	43	123	<0.0001
Error	686	443	0.6			450	0.7			240	0.4		
Total	690	456				480				412			

The above results would suggest that there is indeed a significant effect of learner level on the overall frequency of dictionary consultation (as reported by subjects), as well as an effect of level for each of the three dictionary categories (Polish-English, English-Polish and monolingual English) separately.

To examine more closely the difference between the E-P and P-E sections of bilingual dictionaries, the difference between the numerical values of E2 and E1 (E-P and P-E frequencies, respectively) was computed for each subject who reported both pieces of data. The difference is highly significant (F(4,698)=4.79; p=0.0008) and the means of this difference variable are given in Table 21, broken down by learner level.

Table 21: Mean differences between reported English-Polish and Polish-English consultation frequencies by level

Level	EP-PE
1	-0.06
2	0.03
3	0.15
4	0.08
5	0.26

The distribution of the differences between reported consultation frequencies for English-Polish and Polish-English dictionaries (EP-PE) within the five learner levels is given in Table 22.

Table 22: Distribution of consultation frequency differences between English-Polish and Polish-English dictionaries (EP-PE) by level

EP-PE	Level 1	Level 2	Level 3	Level 4	Level 5	Row Total
-3	0	0	0	0	1 (1%)	1
-2	3 (2%)	4 (3%)	1 (1%)	1 (1%)	0	9
-1	26 (18%)	16 (12%)	9 (5%)	8 (5%)	5 (5%)	64
0	93 (65%)	83 (64%)	133 (76%)	121 (80%)	67 (64%)	497
1	19 (13%)	24 (19%)	25 (14%)	22 (14%)	29 (28%)	119
2	2 (1%)	2 (2%)	6 (3%)	0	3 (3%)	13
Total	143 (100%)	129 (100%)	174 (100%)	152 (100%)	105 (100%)	703 (100%)

Most subjects reported using the E-P and P-E sections of bilingual dictionaries with equal frequency (0 values in the EP-PE row). For level 5, however, as many as 29 out of 105 subjects (28%) reported using English-Polish dictionaries more often than Polish-English dictionaries. This is interesting in view of the sharp increase in the frequency of use of monolingual dictionaries for that level. One might expect that the use of the monolingual dictionary would tend to be an alternative for the English-Polish dictionary more readily than for the Polish-English dictionary, since the monolingual dictionary is in principle relatively more suitable for decoding tasks than it is for encoding. The observed effect could be due to an overall increase in decoding tasks at level 5, with corresponding greater use of both these types of dictionaries. Another explanation could be that the monolingual dictionary is heavily used for encoding, e.g. composition, not so much to locate a target lexical item, but rather to learn how the word could be used in context. Some support for this last explanation may come from the reported increased interest of level 5 subjects in collocational information in a dictionary (see 4.4.11 below). This type of reference need was probably, at the time of the study, better met by the monolingual learners' dictionaries than by the bilingual dictionaries available in Poland, though of course this does not have to be true of these types of dictionaries in principle. The particular interaction between consultation frequencies for the two sides of the bilingual dictionary does not tally well with Tomaszczyk's (1979: 106) claim that L2→L1 dictionaries are used more often than L1→L2 dictionaries for all groups of dictionary users. However, it appears that all of Tomaszczyk's language learner subjects were at the college level, so his sample does not seem to have covered the lower levels of language proficiency that are also represented in this study.

It should be remembered that whenever a difference in frequency of consultation between two dictionary types is revealed by the questionnaire, it might be

expected to be of a non-trivial magnitude, since only four discrete categories for consultation frequency were available to subjects. This means that any non-zero EP-PE value as given in Table 22 indicates a difference of at least one category (e.g. *once a week* versus *a few times a week*). The Pearson Chi-square statistic for the frequencies in Table 22 is 52.84 at 20 df, and is highly significant at $p<0.0001$.

4.1.6 Maximum frequency

Different subjects have different preferences and habits when it comes to using the three dictionary types (that is, Polish-English, English-Polish and monolingual English dictionaries). In this section, the maximum frequency reported by each subject was noted, out of the three types of dictionaries for which consultation frequency reports were elicited. The measure presented in this section is a better indication of the frequency of dictionary consultation in general than the categorized frequency figures given so far. It will also be useful in the determination of absolute frequency with which various information types are consulted, based on relative frequency data directly reported by subjects (see 4.4.11.2). A breakdown of maximum frequency responses obtained in this manner is presented in Table 23, and a histogram of responses is shown in Figure 15.

Table 23: Maximum frequency of dictionary consultation

Consultation frequency	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
less frequently	92	92	12.9	12.9	12.9	12.9
weekly	226	318	31.8	44.7	31.7	44.7
a few times a week	331	649	46.6	91.3	46.5	91.2
daily	62	711	8.7	100.0	8.7	99.9
missing	1	712	0.1	N/A	0.1	100.0

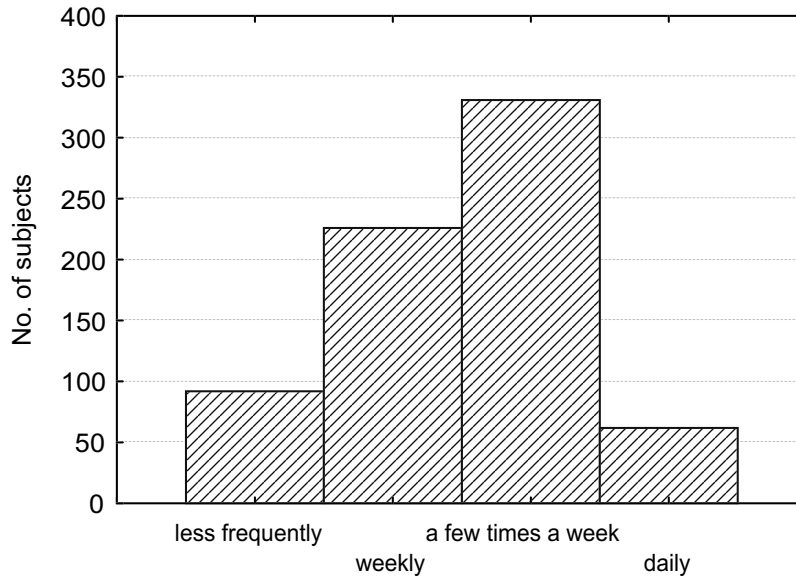

Figure 15: Maximum frequency of dictionary consultation

The modal value apparent in Table 23 and Figure 15 indicates that the largest section of the subjects, nearly half of the sample, claim to consult their most-used dictionary with a frequency of *a few times a week*. The second most popular response, given by close to a third of the subjects, is *weekly*. It is followed by *less frequently*, which was the frequency reported by about 13% of the sample. Only about 9% of the subjects reported consulting their most-frequently used dictionaries on a *daily* basis. Speaking in absolute terms, one might characterize the consultation frequency data as moderately high, with 87% of the subjects reporting that they consult their preferred dictionary at least weekly.

Let us now look at the relationship between maximum consultation frequency and learner level. For this analysis, dictionary type (or, in tied scores, dictionary types) for which the maximum consultation frequency was reported was noted and entered as a variable with maximum consultation frequency in a GLM analysis.

In Figure 16, maximum reported consultation frequency is plotted against learner level. For the purpose of this portion of the analysis, the maximum frequency variable is treated as a numerical variable, and the four levels of the frequency variable are assigned numerical values of 1, 2, 3, and 4. This procedure might be seen as somewhat problematic, given how the original variable was measured: we are here dealing with an ordered scale, but not exactly an interval scale. The presentational value of Figure 16 and Table 24 that employ the numerical values should remain unaffected, but the accompanying significance test results should be treated with some caution.

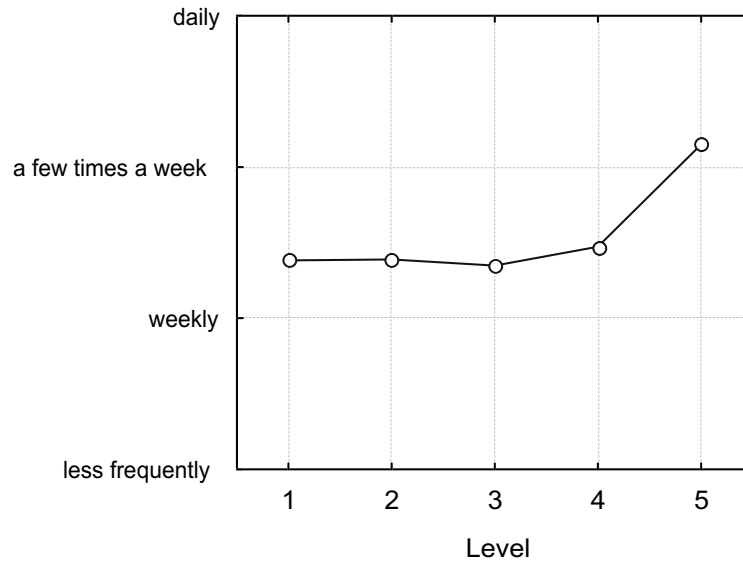

Figure 16: Maximum frequency of dictionary consultation by level

Figure 16 plots the distribution of mean maximum frequency[15] of dictionary consultation, computed from numerical frequency values as explained above.

Table 24 provides numerical values of mean maximum frequency of dictionary consultation with their standard errors and 95% confidence intervals.

Table 24: Maximum frequency of dictionary consultation by level with standard error and 95% confidence intervals

Level	MaxFreq Mean	MaxFreq Std.Err.	MaxFreq -95%	MaxFreq +95%	N
1	2.38	0.07	2.25	2.51	144
2	2.39	0.07	2.25	2.52	131
3	2.35	0.06	2.23	2.46	175
4	2.47	0.06	2.35	2.60	154
5	3.15	0.08	3.00	3.30	107

GLM analysis yields a highly significant effect of level on maximum consultation frequency ($F_{(4, 706)}=21.5$, $p<0.0001$). Trend analysis with linear polynomial contrast confirms that there is a highly significant increase in consultation frequency with level ($F_{(1, 706)}=54$; $p<0.0001$). The graph in Figure 16 suggests that this effect is primarily due to an increase in consultation frequency at the highest

[15] Where 'mean' refers to averaging over all subjects of a given level, whereas 'maximum' refers to the selection of the top consultation frequency reported for any of the three dictionary categories.

learner level. Indeed, post-hoc pair-wise contrast analysis reveals that it is level 5 that is significantly different from any other level in terms of consultation frequency, but there are no statistically significant differences between the frequencies for the remaining levels. However, results of the above analysis must be viewed with caution, given the character of the maximum frequency variable.

Because of the problematic status of the maximum frequency variable in the above model, an alternative GLM analysis was conducted with this variable as a discrete predictor of the level variable. Here, mean levels were computed within groups of subjects who reported each of the four respective maximum consultation frequencies. The results are listed in Table 25 and plotted in Figure 17. The effect is highly significant ($F_{(3, 707)}$=25.2, p<0.0001).

Table 25: Mean level as a function of maximum frequency of consultation

Consultation frequency	Mean Level	N
less frequently	2.73	92
weekly	2.62	226
a few times a week	2.95	331
daily	4.19	62

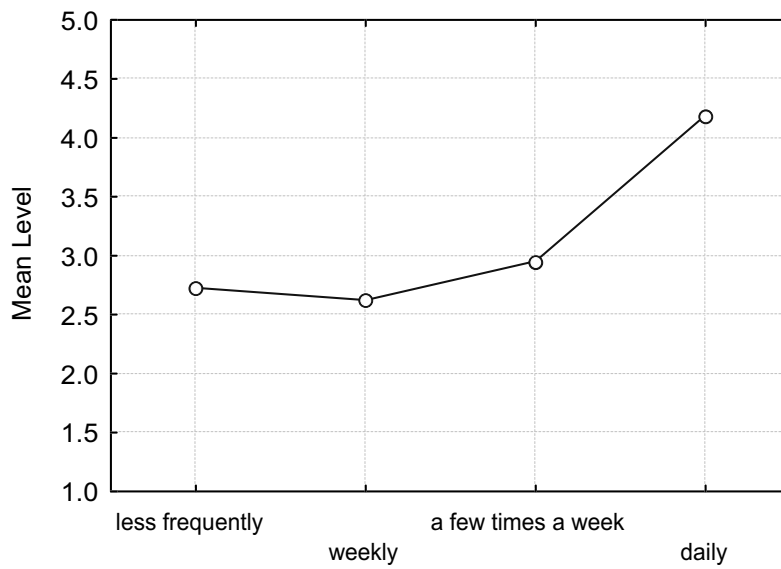

Figure 17: Mean level as a function of maximum frequency of consultation

In the present analysis, the three pairwise contrasts with *daily* are all highly significant (p<0.0001). The contrast between *a few times a week* and *weekly* is significant (p<0.003), and no other contrasts are significant. The upward trend (linear polynomial contrast) is highly significant (p<0.0001).

The above results suggest that the more frequent consultations are more characteristic of the higher-level learners, and this is especially true of *daily* consultations. However, the difference between *weekly* and *less frequent* consultations of the learners' preferred dictionaries is independent of their level.

4.1.7 Conclusion

In terms of the reported frequency of consultation, *a few times a week* was the most popular response for bilingual dictionaries, Polish-English as well as English-Polish. The situation looks very different for monolingual dictionaries, where the lowest-frequency option *less frequently/not at all* accounts for 84% of all responses. A Friedman ANOVA confirms that learners refer (according to their reports) significantly more frequently to bilingual dictionaries than they do to monolingual dictionaries. The same statistic also reveals a significant difference in reported consultation frequency between Polish-English and English-Polish dictionaries, even though impressionistically the distribution for the two appears to be quite similar. English-Polish dictionaries (or dictionary sections, as the case may be) are consulted more often than the reverse direction Polish-English ones.

When proficiency level is factored in, the consultation frequency for monolingual dictionaries is seen to rise sharply at the highest level 5, but remains virtually flat, and very low, throughout the remaining levels. The use of bilingual Polish-English dictionaries tends to increase from level 3 upwards, while consultation frequency for bilingual English-Polish dictionaries rises steadily across the whole level range, though the increase tends to become steeper at the higher levels. These findings are not compatible with the scenario where advanced learners are replacing bilingual dictionaries with monolinguals. Rather, the data suggest that highly advanced learners keep on using their bilingual dictionaries with an intensity at least as high as that at lower proficiency levels, but they supplement their active set of lexical reference works with monolingual dictionaries. One likely interpretation of this could be that the monolingual dictionary may be used for encoding in conjunction with the Polish-English bilingual dictionary: the bilingual dictionary could be used to locate a potential English equivalent, while the monolingual dictionary could be consulted for examples and details of usage, collocation, and grammar, and to confirm that an appropriate equivalent was found in the bilingual dictionary. Finally, highest-level learners tend to consult their dictionaries daily, more frequently than learners at any of the lower levels.

4.2 Dictionary preference

In section F of the Learners' Questionnaire, subjects were asked to name two dictionaries they used most frequently. The two spaces in the questionnaire were numbered to stress the point that respondents were meant to identify their dictionary of first choice at number one (F1), followed by their dictionary of second

choice at number two (F2). Subjects were also asked to rate the dictionaries they listed on a five-point scale (see 4.3 below).

4.2.1 Dictionary identification

A reasonable-guess policy was adopted by the researcher in decisions involving dictionary identification, largely because subjects' recall of just about any features of their dictionaries was very poor. This problem had been anticipated, and thus users were encouraged in the instructions to try to give as many details as they could recall (title, publisher, author, edition) to maximally aid the positive identification of the dictionary product subjects meant in each case. Unfortunately, such positive identification was rarely fully possible.

Thus, for example, if a subject reported "żółto-niebieska okładka z dużą literką L" ('yellow-and-blue cover with a large L'), such a response was interpreted as *Langenscheidt's Pocket English Dictionary*, since, to the researcher's knowledge, at the time when the data was collected, this was the only dictionary on the Polish market with these particular physical characteristics.

In this section, dictionaries identified by subjects are often represented with their abbreviated database dictionary codes. These codes are expanded in Appendix 9. Since section F of the Learners' Questionnaire was open-ended, the degree of detail given depended largely on the subjects, although they were encouraged to provide as much detail as they could. It is quite possible that in many cases learners did not recall any further useful details beyond those provided. As a result, the degree of specificity to which positive identification was possible varies, and the responses recorded here are a reflection of this. Thus, we find here codes that represent actual identifiable dictionary titles (products), such as LDOCE, author/editor names, names of publishers, partial titles, as well as codes for broad categories for any bilingual, any monolingual, or any other dictionary. It is quite likely that some dictionaries hidden inside the broad "wastebasket" categories also feature elsewhere in some narrower category, but it is hard to do anything about it, given the notoriously imperfect recollection of dictionary details by users and the open-ended question format. On the other hand, a multiple choice format was not a viable option here because of the staggering number of potential choices and the space they would have taken. And even if an extensive list of dictionary titles had been provided to choose from, subjects would often probably have misattributed their dictionaries anyway, misled by deceptively similar titles or for other reasons.

Information given in part E was sometimes used to aid the decision as to which dictionary was meant in part F. It will be recalled that in part E subjects were asked about the relative frequency of use of specific types of dictionaries. Since part F attempted to elicit the one or two most frequently used dictionaries, it is reasonable that the types identified in F and E should correspond in most cases. Thus, for example, if a subject reported "Longman" as the dictionary most frequently used in F (and rated it) and previously in E the same subject reported

the frequency of use for the two bilingual dictionary types as *several times a week* but *less frequently/not at all* for monolingual dictionaries, then *Longman podręczny słownik angielsko-polski, polsko-angielski* was entered as the dictionary identified in F, since, to my knowledge, it was the only (Polish) bilingual Longman dictionary available at the time the study was conducted.

It is possible that in some isolated cases the above triangulation procedure may have led to incorrect results, such as when subjects did not remember any details of the dictionary they used most frequently and instead decided to report on some other dictionary which they happened to recall better, even though they did not use the latter dictionary as often as the former. The same objection, however, could also be levelled at part F of the Learners' Questionnaire alone. Combining the data reported in sections E and F made it possible to reconstruct much useful information about how subjects rated dictionaries, and specifically to at least identify the types (bilingual vs. monolingual) rather than be forced into using the catch-all category of "other" more often than was absolutely necessary. However, if it was not quite clear from the two responses which dictionary type was meant, the more general option was conservatively entered.

In one case, "TP Dictionary, Robert Lew" was obviously copied from the footer of the attached test dictionary, reported as the second most frequently used dictionary, and rated as "good". This response was rejected (ignored) as it was likely meant in jest, and was also in conflict with the information reported in part E, where the least common use of the three types of dictionaries was *weekly*, and thus a one-time use of the alleged "TP Dictionary" could not possibly have been the second most frequent.

If subjects reported equal frequency at E1 and E2, but only gave a single rating identifiable as bilingual at F, then it was assumed the rating referred to the bilingual dictionary as a whole, and not just to one part of it, Polish-English or English-Polish, even if the dictionary's description at F only made explicit mention of one part. It appears to be a common informal abbreviatory convention in Polish to refer to a bilingual dictionary by explicitly mentioning its one part only, and so it was assumed here that the subjects were following this convention.

4.2.2 Dictionary of first/second choice

There were 645 valid responses to item F1, which means that 91% of the subjects specified their dictionary of first choice. A much lower number of 363 subjects (51%) provided information on their second-choice dictionary. This might be taken to suggest that approximately half of the subjects only use one English dictionary. All subjects who reported items for F1 or F2 also provided their ratings of the dictionaries, yielding a total of 1008 ratings: 645 first-choice ratings and 363 second-choice ratings.

In Table 26, categorized responses to items F1 and F2 are combined. For each dictionary code (see Appendix 9 for code expansions), the table reports the number of times a given dictionary or dictionary category was listed across the

whole sample as one of the two most frequently used dictionaries: first-choice only, second-choice only, a total of the two, and, in the final column, a total expressed as percentage of all responses. The rows are sorted by total frequency first, then by code.

Table 26: Dictionaries most often used: total, first-choice and second-choice

Dictionary code	1st choice	2nd choice	Total	Percent
Bi	157	46	203	20.1
BiEP	53	81	134	13.3
BiPE	84	45	129	12.8
Langenscheidt	71	16	87	8.6
BGW	52	15	67	6.6
Other	28	36	64	6.3
STAG	11	14	25	2.5
Jaworska	18	4	22	2.2
OxWord	16	4	20	2.0
Grzebieniowski	18	1	19	1.9
BGWEP	9	8	17	1.7
BGWPE	7	9	16	1.6
BGWCD	8	7	15	1.5
Kiesz	8	6	14	1.4
Podr	10	4	14	1.4
LongPodr	11	2	13	1.3
LDOCE	7	4	11	1.1
ALD	7	3	10	1.0
Mono	2	8	10	1.0
Kałuża	7	2	9	0.9
LDLC	8	0	8	0.8
LongMono	4	4	8	0.8
CIDE	6	1	7	0.7
Cobuild	2	5	7	0.7
NewHot	2	5	7	0.7
Uniwersalny	2	4	6	0.6
Saloni	4	1	5	0.5
Electronic	2	2	4	0.4
Oxford	3	1	4	0.4
OxMono	2	2	4	0.4
StanP	4	0	4	0.4
BBI	0	3	3	0.3
Collins	3	0	3	0.3
ElBi	2	1	3	0.3
Etranslator	1	1	2	0.2
LTT	1	1	2	0.2

Dictionary code	1st choice	2nd choice	Total	Percent
Mizgalski	2	0	2	0.2
OxPocket	1	1	2	0.2
OxQuick	1	1	2	0.2
Translator	1	1	2	0.2
Wtranslator	1	1	2	0.2
AHD	0	1	1	0.1
BBC	1	0	1	0.1
CobuildCD	0	1	1	0.1
Coll	0	1	1	0.1
ElEP	1	0	1	0.1
ElPE	0	1	1	0.1
Idioms	0	1	1	0.1
Idiomy	0	1	1	0.1
LASD	1	0	1	0.1
LLA	0	1	1	0.1
LongPocket	1	0	1	0.1
LPD	1	0	1	0.1
Mini	1	0	1	0.1
OxCon	0	1	1	0.1
OxStud	0	1	1	0.1
Penguin	1	0	1	0.1
Piotrowski	1	0	1	0.1
RHC	0	1	1	0.1
SOED	0	1	1	0.1
Tematyczny	1	0	1	0.1
Turystyczny	0	1	1	0.1
Webster	0	1	1	0.1
Total	645	363	1008	100

To take an example, in the top row of the table, the code Bi stands for any bilingual dictionary that has not been positively identified as belonging to a more specific dictionary category. This happens to have been the most frequent response, with a total count of 203 (or 20.1% of all responses), of which 157 were given as first choice (F1) and 46 as second choice (F2). The two runners-up are also unspecified bilingual dictionaries, but here the source language/target language combination was indicated. These three catch-all categories account for nearly half of all responses.

The first identifiable title on the list is the *Langenscheidt's Pocket English Dictionary*, with 87 occurrences throughout. The Collins-BGW family of dictionaries lists next with 67 occurrences. However, one of the two popular Collins-BGW titles was published in two separate volumes, and these separate English-Polish and Polish-English parts register 16 and 17 occurrences, respec-

tively. When these are added to the 67, then Collins-BGW gets ahead of Langenscheidt, and even more so if the CD edition of the Collins-BGW dictionary is also included in the count, with a further 15 occurrences.

Bilingual dictionaries clearly occupy the top-ranking positions to the exclusion of monolingual dictionaries. This is consistent with the results of Section E of the Learners' Questionnaire, as discussed in 4.1.4 above. It is the differences in the distribution of monolingual and bilingual dictionaries that we turn to next.

4.2.3 Monolingual vs. bilingual

Most of the dictionary categories listed in section 4.2.2 were classified as either monolingual or bilingual, and an appropriate two-way type identifier of dictionary type was included in the dictionary table of the database. The individual values of the identifier are listed in Appendix 9. Using this dictionary type identifier, comparisons involving the monolingual-bilingual contrast can be made.

Table 27 lists the frequencies with which monolingual and bilingual dictionaries were reported by subjects as the dictionaries they used most frequently and second-to-most frequently. The table also provides column percents, which indicate the proportion of responses that each dictionary type accounts for within either choice.

Table 27: Monolingual versus bilingual dictionaries given as 1st and 2nd choice; column percents are given in parentheses

Dictionary type	1st choice	2nd choice	Total
bilingual	565 (93%)	283 (87%)	848 (91%)
monolingual	43 (7%)	41 (13%)	84 (9%)
Total	608 (100%)	324 (100%)	932 (100%)

The total number of responses of 932 is smaller here than that in Table 26, because for the present analysis it was necessary to ignore those responses which could not be unequivocally assigned to either the monolingual or bilingual dictionary type. Looking at the results, we note that monolingual dictionaries only account for about 9% of all dictionaries reported, which confirms the findings from section 4.1, indicating an overall low incidence of monolingual dictionary use. We also note a tendency for monolingual dictionaries to be relatively dispreferred as 1st choice and preferred as 2nd choice. A Chi-square test shows this effect to be significant (χ^2=8; p=0.005). This may be taken to indicate that those learners that do use monolingual dictionaries tend to treat them as their second-choice dictionary, rather than the primary language reference.

4.2.4 Dictionary of first/second choice by level and type

The frequency data for first- and second-choice dictionaries can be further broken down by learner level, as shown in Table 28.

Table 28: Breakdown of dictionaries by choice, type, and level

Dictionary type	Choice	Level					Row Totals
		1	2	3	4	5	
bilingual	1	121	113	152	120	59	565
	2	69	67	52	51	44	283
	Total	190	180	204	171	103	848
monolingual	1	0	0	0	4	39	43
	2	3	0	2	4	32	41
	Total	3	0	2	8	71	84

The table clearly shows that the overwhelming majority of cases where monolingual dictionaries were named come from level-5 subjects (71 out of a total of 84, or 85%). This confirms the findings from section E of the Learners' Questionnaire, indicating a very low frequency of use of monolingual dictionaries for all proficiency levels except level 5 (cf. Figure 14). Not a single monolingual dictionary was named by subjects at level 2.

4.2.5 Conclusion

Subjects were asked to identify their dictionaries of first choice and second choice and to rate them on a five-point scale. As expected and usual in collecting this type of information from informants, the details provided by subjects were largely incomplete and vague, and so in many cases they were insufficient to make a positive identification of a specific title, let alone edition. To help with the identification, data provided by subjects in section E of the questionnaire was used for triangulation. Over 90% of the subjects gave details on their dictionary of first choice, but only about 50% on their second-choice dictionary, suggesting that roughly half of the subjects only used a single dictionary title with any regularity. A ranking table produced from the data allowed the identification of the most popular dictionary titles. Where the positive identification of dictionary title proved impossible, an attempt was made to at least identify the dictionary type: bilingual or monolingual. Bilingual dictionaries accounted for 91% or all responses, while monolinguals for 9% only. Monolingual dictionaries were dispreferred as first-choice, suggesting that those learners that do use monolingual dictionaries tend to use them in as their secondary lexical reference works, with the bilingual dictionary serving as the primary lexical aid. An analysis by level revealed that monolingual dictionaries were hardly listed at all by subjects at levels lower than level 5.

4.3 Dictionary evaluation

In section F of the Learners' Questionnaire, subjects were asked to rate on a five-point scale the dictionary (or dictionaries) they named as the one (or two) they used most frequently. Subjects were asked to tick one of the five boxes with

graded verbal descriptions in Polish as follows (English glosses and corresponding numerical codes are given in parentheses): *świetny* (‘*excellent*’, 5), *dobry* (‘*good*’, 4), *ujdzie* (‘*OK*’, 3), *kiepski* (‘*poor*’, 2), *dno* (‘*awful*’, 1). Numerical codes were used for coding the five responses in a manner similar to that in Tomaszczyk (1979), and are used throughout this section.

A total of 1008 ratings were obtained, with a breakdown as shown in Table 29 and in the histogram in Figure 18.

Table 29: Breakdown of subjects’ ratings of the dictionaries named as frequently used

Rating	Count	Cumulative	Percent	Cumulative %
1	8	8	0.8	0.8
2	23	31	2.3	3.1
3	187	218	18.6	21.6
4	590	808	58.5	80.2
5	200	1008	19.8	100.0

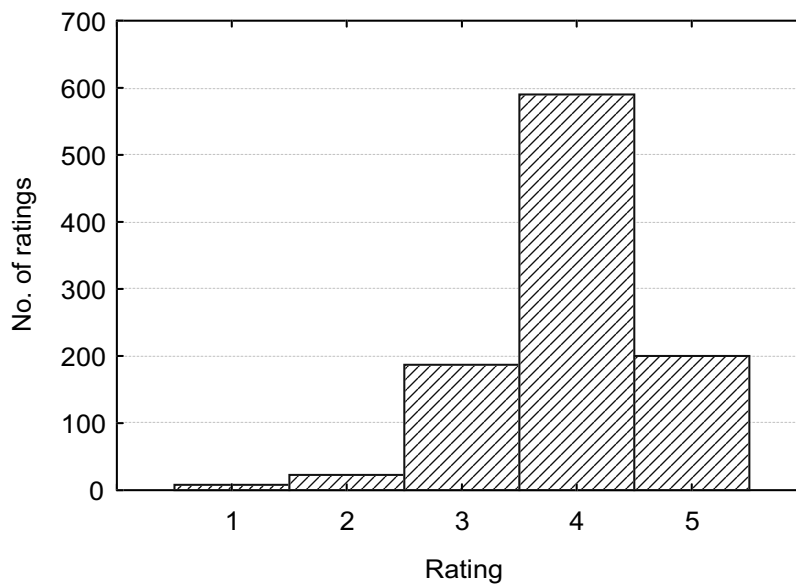

Figure 18: Histogram of subjects’ ratings of the dictionaries named as frequently used

Table 29 and Figure 18 indicate that the ratings exhibit a well-defined modal value of 4 (‘good’) and a left-skewed distribution. Since subjects were probably mostly rating the dictionaries they themselves preferred to consult, it should not be surprising that the ratings tend to be mostly positive, with 78% of the responses representing better-than-neutral evaluations, and only 4% worse than neutral. Negative ratings could reflect a variety of underlying situations that might lead to dissatisfaction with the dictionaries most frequently used: the user

might prefer another (better) dictionary, but the dictionary choice is beyond the user's control (school policy, school copies, a gift); the user might not be able to afford a better dictionary; a better dictionary might be impractical to use, such as when most dictionary use takes place at school and a better dictionary is too heavy to carry, or is installed on the user's home desktop computer; a better dictionary might not be available on the market, or the user has not been able to obtain one; the user may be unfavourably disposed to dictionaries in general, but is nevertheless forced to use them. Overall, however, such hypothetical cases are probably a minority, since the general picture that emerges from subjects' ratings is one of a fairly high level of satisfaction with dictionaries.

4.3.1 Ratings by level

Let us now examine how learners of different levels rate their dictionaries. Table 30 gives mean ratings for subjects within the five level groups. Figure 19 plots mean ratings as a function of level.

Table 30: Mean dictionary rating by learner level

Level	Mean rating	N
1	3.94	210
2	3.89	201
3	3.86	221
4	4.00	196
5	4.04	180

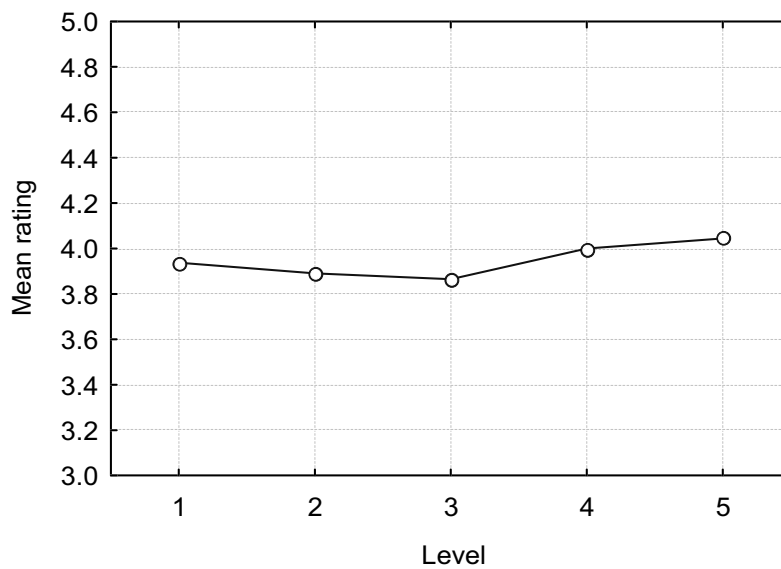

Figure 19: Mean dictionary rating by learner level

Table 30 indicates that the contributions of the five learner levels to ratings data are fairly equal, with about 200 ratings from subjects of each level, ranging from 180 to 221.

The tendency that emerges from Figure 19 appears to be for the dictionary evaluations to decline from the lowest learner level to level 3, and then climb for levels 4 and 5. This effect, however, is not significant ($F_{(4, 1003)}=2$; $p=0.09$). Perhaps the upward tendency for levels 4 and 5 is due to the increased involvement of monolingual dictionaries, which tend to be rated more highly than bilingual dictionaries.

4.3.2 Evaluation of individual dictionary categories

In Table 31, mean subject ratings and their standard errors are given for the dictionary categories introduced in section 4.2.2. A detailed distribution of ratings for each dictionary category is also provided.

Table 31: Mean rating and breakdown of rating responses by dictionary category

Dictionary code	Mean rating	Std. Err. of mean	Rating					N
			1	2	3	4	5	
Bi	3.96	0.05	2	4	31	129	37	203
BiEP	3.78	0.06	0	4	36	80	14	134
BiPE	3.78	0.06	1	6	29	78	15	129
Langenscheidt	3.82	0.08	3	3	14	54	13	87
BGW	4.16	0.09	1	0	9	34	23	67
Other	3.78	0.09	1	2	15	38	8	64
STAG	3.60	0.14	0	0	12	11	2	25
Jaworska	3.95	0.15	0	0	5	13	4	22
OxWord	4.40	0.16	0	0	0	12	8	20
Grzebieniowski	3.74	0.16	0	0	6	12	1	19
BGWEP	4.41	0.17	0	0	1	8	8	17
BGWPE	4.50	0.18	0	0	1	6	9	16
BGWCD	4.13	0.18	0	0	2	9	4	15
Kiesz	3.71	0.19	0	1	3	9	1	14
Podr	3.86	0.19	0	0	5	6	3	14
LongPodr	3.92	0.20	0	0	3	8	2	13
LDOCE	4.55	0.21	0	0	0	5	6	11
ALD	4.50	0.22	0	0	0	5	5	10
Mono	4.40	0.22	0	0	1	4	5	10
Kałuża	3.89	0.24	0	0	2	6	1	9
LDLC	4.63	0.25	0	0	0	3	5	8
LongMono	4.38	0.25	0	0	1	3	4	8
CIDE	4.29	0.27	0	0	0	5	2	7
Cobuild	4.29	0.27	0	0	0	5	2	7
NewHot	3.00	0.27	0	2	3	2	0	7

Dictionary code	Mean rating	Std. Err. of mean	Rating					N
			1	2	3	4	5	
Uniwersalny	3.83	0.29	0	0	1	5	0	6
Saloni	4.20	0.32	0	0	0	4	1	5
Electronic	4.00	0.35	0	0	0	4	0	4
Oxford	5.00	0.35	0	0	0	0	4	4
OxMono	4.50	0.35	0	0	0	2	2	4
StanP	4.00	0.35	0	0	0	4	0	4
BBI	4.67	0.41	0	0	0	1	2	3
Collins	3.67	0.41	0	0	1	2	0	3
ElBi	4.00	0.41	0	0	0	3	0	3
Etranslator	4.00	0.50	0	0	0	2	0	2
LTT	4.50	0.50	0	0	0	1	1	2
Mizgalski	3.00	0.50	0	1	0	1	0	2
OxPocket	4.00	0.50	0	0	1	0	1	2
OxQuick	4.00	0.50	0	0	0	2	0	2
Translator	3.50	0.50	0	0	1	1	0	2
Wtranslator	4.50	0.50	0	0	0	1	1	2
AHD	4.00	0.71	0	0	0	1	0	1
BBC	4.00	0.71	0	0	0	1	0	1
CobuildCD	4.00	0.71	0	0	0	1	0	1
Coll	4.00	0.71	0	0	0	1	0	1
ElEP	3.00	0.71	0	0	1	0	0	1
ElPE	4.00	0.71	0	0	0	1	0	1
Idioms	5.00	0.71	0	0	0	0	1	1
Idiomy	4.00	0.71	0	0	0	1	0	1
LASD	5.00	0.71	0	0	0	0	1	1
LLA	5.00	0.71	0	0	0	0	1	1
LongPocket	4.00	0.71	0	0	0	1	0	1
LPD	5.00	0.71	0	0	0	0	1	1
Mini	3.00	0.71	0	0	1	0	0	1
OxCon	4.00	0.71	0	0	0	1	0	1
OxStud	3.00	0.71	0	0	1	0	0	1
Penguin	4.00	0.71	0	0	0	1	0	1
Piotrowski	3.00	0.71	0	0	1	0	0	1
RHC	4.00	0.71	0	0	0	1	0	1
SOED	5.00	0.71	0	0	0	0	1	1
Tematyczny	4.00	0.71	0	0	0	1	0	1
Turystyczny	4.00	0.71	0	0	0	1	0	1
Webster	5.00	0.71	0	0	0	0	1	1
Total	3.94	0.02	8	23	187	590	200	1008

The mean ratings given in the table are an indication of how learners in the sample rated a specific dictionary or dictionary category. The overall mean rating is 3.94. If the number of subjects who rated a particular dictionary category is small (column N), then the indication is not very reliable. Since Table 31 is sorted by descending N, this is certainly the case for the items towards the bottom of the table: they are given here for completeness, but no great weight should be attached to them. The ratings are most reliable (note the standard error values) for the most frequently named dictionary categories, which come towards the top of Table 31. These dictionary categories also tend to be of greater interest, because they typically represent the more popular dictionaries. It is for the above reasons that I have decided not to sort the table by rating: a rating that only comes from one or two raters is not worth much. Instead, I will proceed from high to low ratings in my discussion of the results in the remainder of this section: first for monolingual dictionaries, then for bilingual dictionaries, as there is a clear tendency for monolingual dictionaries to receive higher ratings than bilingual dictionaries (and I will examine this tendency more closely in section 4.3.3).

At the top of the list for categories with at least four ratings there is the category 'Oxford', which has earned top ratings from all raters, and an average rating of 5.00. However, this category is so vague, not much can be said of it, except perhaps that the positive image and authority behind the name 'Oxford' might be partially responsible for the high ratings.

LDLC and LDOCE, the two Longman dictionaries (*Longman Dictionary of English Language and Culture* and *Longman Dictionary of Contemporary English*) are rated very highly, with mean ratings of 4.63 and 4.55, respectively. ALD (*Oxford Advanced Learner's Dictionary of Current English*), a line of titles with the longest tradition in English pedagogical lexicography, is not far behind Longman with a mean rating of 4.50. The same high rating is held by OxMono, a code which stands for any unspecified monolingual dictionary identified as 'Oxford'. A wastebasket category of 'Mono' follows with a rating of 4.40. The category 'LongMono', representing any unspecified monolingual dictionary identified as 'Longman' has a rating of 4.38. Cobuild and CIDE have identical mean ratings of 4.29, with the same number and distribution of ratings.

'OxWord' (*Oxford Wordpower. Słownik angielsko-polski z indeksem polsko-angielskim*) is a special case here. Because it uses Polish for semantic explanation by providing Polish equivalents, it is classified as bilingual in the coarse two-way taxonomy. More specifically, though, it is a *bilingualized* dictionary (and thus a *semi-bilingual* one). Its average rating is very high at 4.40, a value more typical of monolingual dictionaries in the sample.

Amongst the most frequently named bilingual dictionaries that can be positively identified from subjects' responses, the BGW family of dictionaries (coded as BGW, BGWEP, BGWPE, BGWCD) earned the highest grades from learners in the sample, with a mean rating of 4.24 across the four codes.

'Saloni', which may represent a number of dictionary titles co-authored by Zygmunt Saloni, earns a high score of 4.20, although that is a mean of only five ratings. 'StanP', which represents any of the numerous editions of *Podręczny słownik angielsko-polski* and/or *Podręczny słownik polsko-angielski* from Wiedza Powszechna, gets a mean rating of 4.00.

'Bi', a catch-all category for an unspecified bilingual dictionary and, in fact, the single most numerous category, earned a mean rating of 3.96.

'Jaworska', which, as far as I have been able to establish, refers to any of a number of compact dictionaries published as *Słownik angielsko-polski, polsko-angielski WNT* with Teresa Jaworska as author, received an average rating of 3.95.

The new *Podręczny słownik angielsko-polski, polsko-angielski* ('LongPodr') carrying the Longman brand name received a mean rating of 3.92. Close behind are 'Kałuża' (*Słownik angielsko-polski, polsko-angielski* by Jan Kałuża, published by Exlibris), with 3.89, and the generic pseudo-category 'Podr' (a mean rating of 3.86), which could refer to any of a number of dictionaries with the word *podręczny* in the title.

The code 'Uniwersalny' stands for *Słownik uniwersalny angielsko-polski, polsko-angielski* published by Harald G., Andrzej Kaznowski's adaptation of Tadeusz Grzebieniowski's older work. This dictionary received an average rating of 3.83. Very close behind, and quite appropriately, with a mean rating of 3.82, is Grzebieniowski's prototype of 'Uniwersalny', revived after several decades as 'Langenscheidt'. This dictionary is the second most popular category among the more specific dictionary categories, second to the BGW, although the latter category actually represents not one, but two popular paper titles (excluding the BGWCD). There is yet another category, 'Grzebieniowski', rated at 3.74, which could refer to any of the titles otherwise included under 'Uniwersalny', and 'Langenscheidt', but there was not enough information to classify them as such, since subjects would only give the single name in those cases. 'Uniwersalny', 'Langenscheidt' and 'Grzebieniowski' jointly account for 112 occurrences, which is more than the 100 for paper BGW dictionaries, but somewhat less than the 115 for BGW including the CD version.

The three large catch-all categories: 'Other', 'BiEP' and 'BiPE' all show mean user ratings of 3.78. What is interesting here is the remarkable convergence of these values, and the difference between the 'BiEP' and 'BiPE' ratings against the more general 'Bi' category at 3.96. This difference is statistically significant (t-value=2.80 at df=464, p=0.005). It is difficult at present to interpret this difference in any meaningful way.

If we ignore the dubious 'NewHot' category (mean rating 3.00), which is not a typical dictionary, but rather a glossary attached to a coursebook used by a specific group of learners, then the two bilingual dictionary categories with lowest ratings are 'STAG' and 'Kiesz'. 'STAG' refers to Jan Stanisławski's *Wielki*

słownik angielsko-polski and *Wielki słownik polsko-angielski*, an immensely popular reference work which has served generations of users since the mid-sixties. This dictionary earned a disappointingly low mean rating of 3.60, perhaps because it is now very much out of date, or perhaps the users' expectations of it are higher compared to the smaller dictionaries. 'Kiesz', which stands for *Kieszonkowy słownik angielsko-polski i polsko-angielski* by Janina Jaślan and Jan Stanisławski, received a mean rating of 3.71.

It appears that, overall, learners evaluate monolingual dictionaries more highly than bilingual dictionaries. Amongst the dictionaries most frequently used, the BGW and Langenscheidt dictionaries get the best grades. The difference in ratings between monolingual and bilingual dictionaries will be the focus of the next section.

4.3.3 Monolingual vs. bilingual

A detailed analysis of ratings for individual dictionary categories in section 4.3.2 above suggests that monolingual dictionaries tend to receive higher ratings from subjects than do bilingual dictionaries. This effect will be examined next[16]. Table 32 gives the mean ratings, standard errors, and 95% confidence intervals for bilingual and monolingual dictionaries.

Table 32: Subjects' ratings of bilingual and monolingual dictionaries

Type	Mean Rating	Std. Error	-95%	+95%	N
bilingual	3.90	0.02	3.86	3.95	848
monolingual	4.40	0.08	4.25	4.56	84

The number of ratings, 932, is here somewhat smaller than the total 1008, since those ratings for which the monolingual-bilingual status could not be determined had to be rejected. The results show that the typical rating for a monolingual dictionary (4.40) is higher than that for a bilingual dictionary (3.90) by exactly half a grade. This effect is highly significant by one-way ANOVA ($F_{(1,\ 930)}=36.8$, $p<0.0001$). The finding is consistent with that of Tomaszczyk (1979), who also found that Polish dictionary users value monolingual dictionaries more highly than they do bilingual dictionaries. Similar observations have been reported by other researchers (Baxter 1980; Béjoint 1981; Kharma 1985).

The effect is open to several interpretations. It could be that learners express greater satisfaction with monolingual dictionaries simply because monolingual dictionaries are objectively more effective as tools than bilingual dictionaries. This interpretation, however, appears to be contradicted by the results of section 4.5 below, although the results may not hold for all types of tasks. It could be that the monolingual dictionaries available on the Polish market represent, on average, a better professional level of lexicography than do the available bilingual

[16] Some of the data and analysis appearing in this section were presented at the Euralex 2004 Congress and are to be included in the proceedings volume published by Euralex.

dictionaries. There is much anecdotal evidence to support this as at least a partial explanation. One should consider the extensive resources available to the major publishers of monolingual dictionaries, and the fact that until fairly recently most bilingual dictionaries were several decades out of date. In terms of the coverage of current terms and senses, such older dictionaries must be at a disadvantage. Even with these considerations, there is one important consideration that puts this interpretation into question: if subjects really believed that monolingual dictionaries are more useful, why is it that so many of them were unwilling to use them, and even if they did, they tended to use them as second choice?

Another interpretation is that Polish learners see monolingual dictionaries as better because the level of their language and reference skills may make it easier for them to register the failures of bilingual dictionaries than of monolingual dictionaries. Because semantic explanation in monolingual dictionaries takes the form of a (sometimes complex) syntactic construction in the foreign language (or rather a special metalanguage based on the foreign language), users may tend to lay the blame on their lack of foreign language skills whenever they experience problems with interpreting a dictionary entry and fitting it to the textual context, rather than blame the dictionary itself. In contrast, they my feel more in a position to find fault with bilingual dictionaries, where much of the content is provided in their native language.

Yet another aspect of dictionary evaluation by users is the impact of the opinions of various authorities on learners' evaluations: to what extent are learners' opinions a reflection of their own experience with dictionaries, and to what extent are they a passive restatement of the opinions of others? Anecdotally, Polish language teachers often speak of monolingual dictionaries making it possible for learners to "think in the foreign language." The transition from bilingual to monolingual dictionaries is seen as a sign of progress (and, in a way, it is progress, of course, as also confirmed by some results of this study), and thus renders an aura of superiority to monolingual dictionaries, which may well filter into subjects' ratings.

4.3.4 Ratings by choice

It was expected that subjects would tend to give higher ratings to the dictionary they named as the most frequently used (first choice) than to the one they named as the second most frequently used (second choice). Mean ratings broken down by choice, together with standard error values and 95% confidence intervals are given in Table 33.

Table 33: Subjects' ratings of first-choice and second-choice dictionaries

Choice	Mean Rating	Std. Error	-95%	+95%	N
1st	3.98	0.03	3.93	4.04	645
2nd	3.87	0.04	3.79	3.95	363

The mean rating for the dictionary of first choice is 3.98, for second choice it is 3.87. Although the effect size is small, only a tenth of a grade, it is statistically significant (one-way ANOVA, $F_{(1, 1006)}=5.6$, $p=0.02$).

When only those responses are included that can be unambiguously assigned to either the bilingual or monolingual type, the corresponding data would look as in Table 34.

Table 34: Subjects' ratings of first-choice and second-choice dictionaries, restricted to bilingual or monolingual only

Choice	Mean Rating	Std. Error	-95%	+95%	N
1st	4.25	0.06	4.14	4.36	608
2nd	4.03	0.06	3.91	4.15	324

Mean ratings are here somewhat higher than in Table 33, the effect is more significant ($F_{(1, 928)}=7.3$, $p=0.007$), and the effect size is twice as large. A possible explanation for the increase in mean ratings is that the dictionaries that the subjects cannot even identify as bilingual or monolingual tend not to be the dictionaries that learners care much about, hence perhaps the lower evaluations in Table 33.

4.3.5 Ratings by type and choice

Mean subjects' ratings broken down by type and choice are given in Table 35.

Table 35: Subjects' ratings by dictionary type and choice

Dictionary Type	Choice	Mean Rating	N
bilingual	1st	3.95	565
bilingual	2nd	3.82	283
monolingual	1st	4.56	43
monolingual	2nd	4.24	41

The type by choice interaction is not significant ($F_{(1, 928)}=1.2$, $p=0.27$), so the main effects hold irrespective of the particular combination of factor levels.

4.3.6 Ratings by choice and level

Table 36 gives mean ratings for all combinations of choice and level.

Table 36: Subjects' ratings by choice and learner level

Choice	Level	Mean Rating	N
1st	1	4.01	130
1st	2	3.89	121
1st	3	3.93	161
1st	4	3.97	132
1st	5	4.18	101
2nd	1	3.82	80
2nd	2	3.89	80
2nd	3	3.70	60
2nd	4	4.06	64
2nd	5	3.87	79

A General Regression Model (GRM) ANOVA with rating as dependent variable and level and choice as categorial predictors produces a significant whole model ($p=0.005$), significant main effects of choice ($F_{(1, 998)}=6.65$, $p=0.01$) and level ($F_{(4, 998)}=2.64$, $p=0.03$) but a non-significant choice by level interaction ($F_{(4, 998)}=2.23$, $p=0.06$). The main effects have already been discussed above. The choice by level interaction, although not significant at the 5% level, approaches this level closely and is quite interesting (see Figure 20).

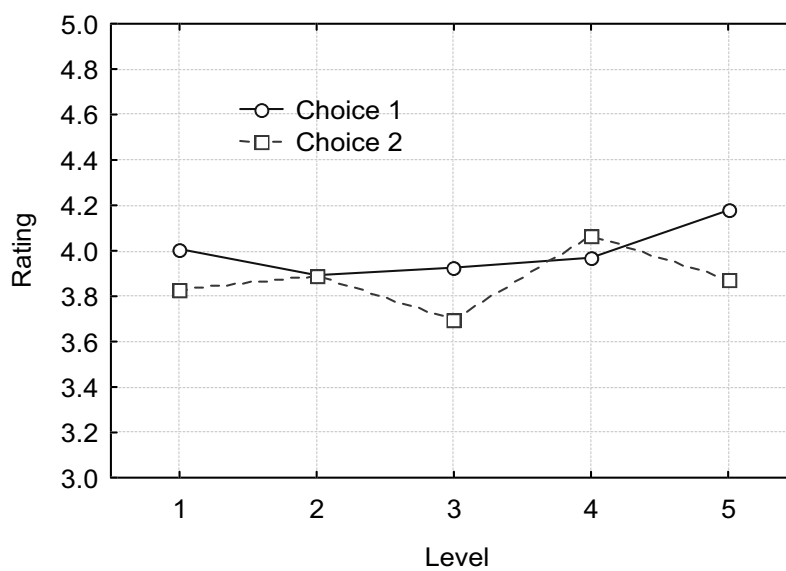

Figure 20: Subjects' ratings by choice and learner level

There appears to be a tendency for first-choice dictionary ratings to decline for lower-range proficiency level learners and then to go up for higher-level learners. For second-choice dictionaries, it is difficult to observe any regular tendencies.

The mean ratings for first-choice dictionaries are higher than those for second-choice dictionaries at levels 1, 3 and 5. The ratings for the two choices are identical at level 2, and at level 4 the mean rating for second-choice dictionaries is actually higher than for first-choice dictionaries, which is somewhat surprising. At first sight, one could try to explain this effect by the influence of monolingual dictionaries, which receive consistently higher ratings and are at the same time more likely to be named as second choice. However, a careful examination of Table 28 reveals that this explanation is unlikely, as very few level 4 subjects named monolingual dictionaries at all. Simply, there are too few ratings given for monolingual dictionaries at level 4 to make a difference here. In order to examine the relationships more closely, a three-way interaction must be analyzed, taking into account choice, level, and dictionary type at the same time.

4.3.7 Ratings by dictionary type, choice and level

Ratings broken down by choice, level and dictionary type are listed in Table 37.

Table 37: Breakdown of subjects' ratings by dictionary type, choice and learner level

Dictionary Type	Choice	Level	Mean Rating	N
bilingual	1st	1	4.02	121
bilingual	1st	2	3.90	113
bilingual	1st	3	3.94	152
bilingual	1st	4	3.94	120
bilingual	1st	5	3.92	59
bilingual	2nd	1	3.84	69
bilingual	2nd	2	3.87	67
bilingual	2nd	3	3.69	52
bilingual	2nd	4	4.04	51
bilingual	2nd	5	3.59	44
monolingual	1st	4	4.50	4
monolingual	1st	5	4.56	39
monolingual	2nd	1	3.67	3
monolingual	2nd	3	4.50	2
monolingual	2nd	4	4.50	4
monolingual	2nd	5	4.25	32

Some combinations with monolingual dictionaries are missing from the table, because they did not occur in the sample: for example, no subject at level 1 gave a monolingual dictionary as their first choice, and no subject at level 2 named a monolingual dictionary as any choice (see also Table 28). The data are somewhat difficult to assimilate in tabular form, and so for easier viewing the mean ratings are plotted in Figure 21.

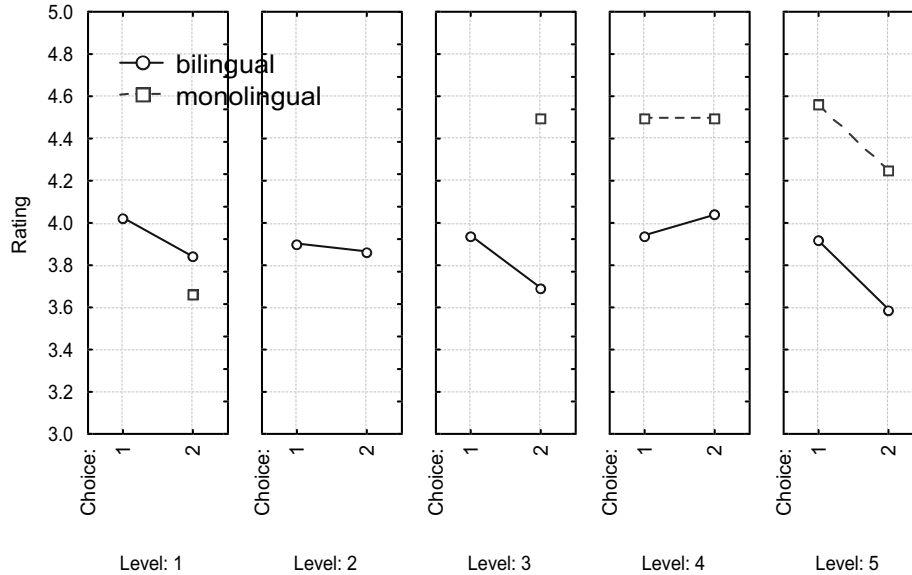

Figure 21: Subjects' dictionary ratings plotted by dictionary type, choice, and learner level

As in Table 28, so in Figure 21 some of the mean ratings for monolingual dictionaries are not estimable because these particular combinations of levels did not occur in the sample, and so they are missing from the plots. For this reason, a complete model with three predictors (type, choice, and level) cannot be evaluated statistically. Looking at the plots, however, two interesting tendencies emerge. Firstly, a tendency for monolingual dictionaries to be rated more highly is reconfirmed here, with the exception of level 1, although this last mean only represents three data points, so it should not be taken too seriously. Secondly, first-choice dictionaries appear to be evaluated more highly than second-choice dictionaries, as one might reasonably expect, with the exception of level 4 subjects, where the reverse tendency is observed, as already noted in 4.3.6 above. There, it was suggested that this counterintuitive tendency cannot be explained by the involvement of monolingual dictionary ratings, and the graph for level 4 clearly shows that the tendency actually comes from the ratings for *bilingual* dictionaries, which tend to be granted lower ratings by level 4 subjects when named as first choice compared to second choice. A close inspection of the individual responses of level 4 subjects does not shed any further light on why these subjects, when choosing between two bilingual dictionaries, would rate their second choice more highly. Perhaps there are some external factors in the context of the dictionary consultation act that make those learners refer more frequently to a dictionary which is their second favourite, such as when substantial dictionary work is done in the classroom while the best liked dictionary is reserved for the less frequent home use, but this is mere speculation.

4.3.8 Conclusion

Subjects rated the dictionaries they named as the most frequently used. The most frequent response was *good*, corresponding to a 4 on a scale of 1 to 5, which indicates a fairly high overall level of satisfaction with dictionaries. The ratings do not vary significantly by raters' proficiency level. Between individual titles, where these could be positively identified, ratings vary quite substantially, as given in detail in Table 31. This data may be of practical relevance to publishers, teachers, learners, and other would-be dictionary users. Researchers may also find it worthwhile to take a closer look at the different titles featured in the table and perhaps try to consider which possible features of the products were highly valued, and which were disliked. Out of the dictionaries most frequently named, the best ratings were given to the Collins-BGW and Langenscheidt bilingual dictionaries.

Monolingual dictionaries received significantly higher ratings than bilingual dictionaries. Various interpretations of this effect have been considered. The interpretation that monolingual dictionaries are simply more helpful to subjects must be viewed with scepticism in the light of the experimental results presented in 4.5 below. Instead, the difference in ratings has been ascribed to the relative difficulty with which failings of monolingual dictionaries can be registered, and the positive image of the monolingual dictionary, perhaps fostered by many teachers.

As could be predicted, dictionaries of first choice received higher ratings than dictionaries of second choice. However, when proficiency level is factored in, the reverse is true for level 4. A three-way analysis reveals this unexpected effect to be due to lower ratings being awarded to bilingual dictionaries when used as first choice. It is difficult to give a definitive interpretation of this effect.

4.4 Reference needs: Information sought

In section G, subjects were asked about the frequency with which they were looking for specific types of information in their dictionaries. Options in the Learners' Questionnaire were phrased and presented in such a way as to elicit *relative* frequency, which was meant to give an indication of the frequency of consultation for the particular types of information relative to any other types of information available in a dictionary, with the formulation *most often* (see Appendix 3 for the original formulations of questionnaire items) intended to indicate the type of information looked up more often than, or at least as often as, any other type of information. Whether the formulations were in all cases interpreted as intended by the researcher is, of course, not known, but it is hoped that the majority of the subjects did interpret them in this way, so there is no systematic error in the results.

One type of information featuring in some other similar surveys for which frequency information was not elicited, given the space and time constraints of

the study, was spelling. The reason for omitting spelling is that, unlike any of the other types of information, spelling information is in any case inherently and implicitly present in dictionary entries. Since spelling representation is the primary mode of representing headwords, it is the cornerstone of just about any dictionary's access structure. Because of this special status of spelling, it is its role in access structure that is of primary importance in dictionary consultation, and that indexical role should guide decisions regarding the salience of the spelling representation. Access structure is not my main concern here, yet because of the special status of spelling representation in access structure, a question like whether spelling information can be excluded from dictionaries is clearly not of paramount practical interest.[17]

Below, results for the nine individual types of information in turn are presented in tables and histograms. Then, a comparison of the average reported consultation frequencies is attempted. Next, we look at the influence of level on the relative and absolute consultations frequencies for the nine information types. Finally, the degree of interdependence between the nine information types is assessed by computing mutual correlations.

4.4.1 Pronunciation

The first item out of the nine types of information elicited in this study was pronunciation (item G1 of the Learners' Questionnaire). Learners were asked how often they were looking in their dictionaries for information on how a word was pronounced (see Appendix 3 for the original formulations of this and the remaining questionnaire items). The breakdown of responses is presented in Table 38, with raw frequencies given for each of the four questionnaire options, followed by percentages of subjects in parentheses. The *missing* column indicates the number (and percentage) of the subjects who did not tick any box for that particular questionnaire item.

Table 38: Breakdown of responses to item G1: Pronunciation

never	rarely	often	most often	missing
179 (25.1%)	336 (47.2%)	160 (22.5%)	35 (4.9%)	2 (0.3%)

The corresponding histogram is given in Figure 22. Missing data are omitted from the histogram, because the number is here very low and it would not be visible, nor does it significantly contribute to the overall picture.

[17] Interestingly, contrary to many teachers' expectations, the use of dictionaries appears to lower students' ability to detect spelling errors (McNaughton, Hughes and Clark 1997).

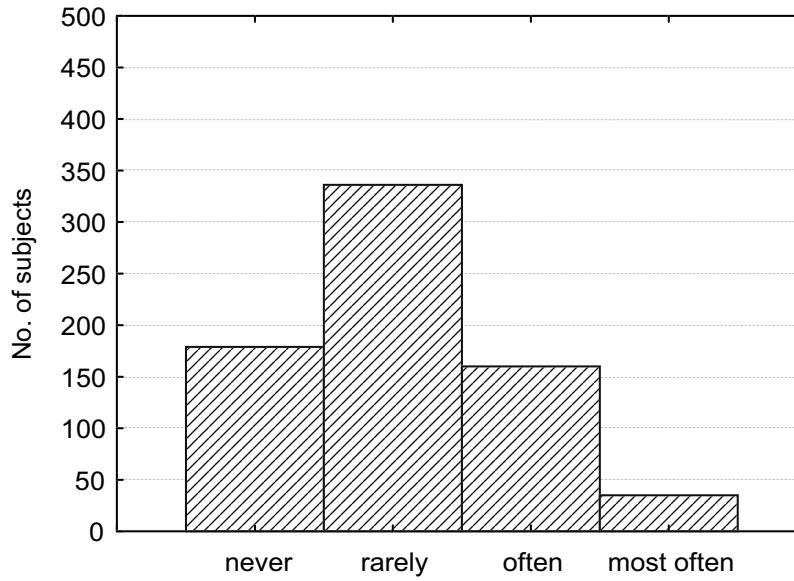

Figure 22: Histogram of responses to item G1: Pronunciation

Nearly half of the subjects said that they looked up pronunciation *rarely*. A quarter of the sample *never* consult their dictionaries for pronunciation. Another quarter look up pronunciation at least *often*.

4.4.2 Meaning

In item G2 of the Learners' Questionnaire, subjects were asked how often they looked for word meaning in their dictionaries. It must be pointed out that an attempt was made in this study to distinguish between consultations for word meaning (semasiological consultation acts) and consultations for linguistic forms (onomasiological consultation acts). Consultations for Polish linguistic forms are covered in item G3, and consultations for English linguistic forms are covered in item G4 of the questionnaire. The three items have been presented close together in the questionnaire so as to draw the attention of the subjects to the distinction. However, it must be realized that the distinction between wanting to know what the English word means and wanting to know how that English word translates into Polish, while pefectly valid to a linguist or metalexicographer, may not be valid at all to an ordinary dictionary user. Due to the primacy of the native language linguistic system (and perhaps also, as Baxter (1980) would see it, due to the bilingual dictionary habit), the questions "What does this word mean?" and "What's this word in Polish" may actually be two sides of the same coin for the typical Polish learner of English: the meaning of the foreign word may well be perceived as its Polish equivalent. Further, even if the concepts of meaning and L1 translation are separate for a lay dictionary user, the separateness may not be apparent at the level of consciousness sufficient to be helpful in completing a

questionnaire. Further still, a search for meaning of a foreign lexical item may well be combined with a search for native language equivalent in what is effectively a single consultation act, thus integrating the two reference needs practically, even if not theoretically.

All the above reservations notwithstanding, the nontrivial differences between the distributions of responses to items G2 and G3 presented in this section and the following one, suggest that subjects, on the whole, had been successful in telling apart a search for meaning from a search for Polish equivalents, perhaps thanks to the two items (G2 and G3) being presented together. Nevertheless, some caution is advised in evaluating these results.

The breakdown of responses to item G2 is presented in Table 39, and the corresponding histogram is given in Figure 23 in the same format as discussed above for item G1.

Table 39: Breakdown of responses to item G2: Meaning

never	rarely	often	most often	missing
14 (2.0%)	52 (7.3%)	168 (23.6%)	477 (67.0%)	1 (0.1%)

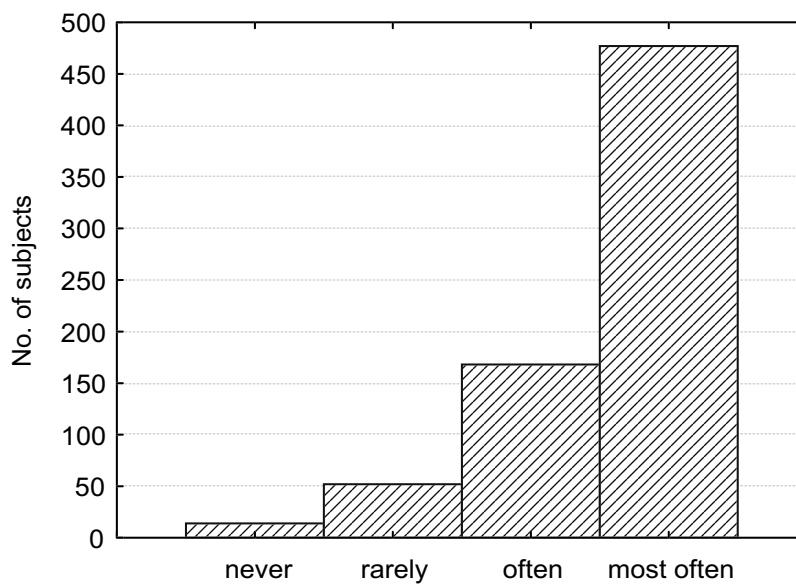

Figure 23: Histogram of responses to item G2: Meaning

The results indicate that as many as two thirds of the subjects look for meaning *most often*. The second most popular option was *often*, which was chosen by nearly 24%. As a result, these two options jointly account for over 90% of the responses, with few subjects admitting to consult dictionaries for meaning *rarely* (7%), and even fewer (a mere 2%) claiming they *never* look for meaning in a dictionary.

The present results would tally well with those of Quirk (1974), where a similar proportion 67% of the native speaker subjects gave meaning as the primary reason for look-up in their monolingual dictionaries. Other studies of reference needs of native speakers (Greenbaum 1977; Kipfer 1987) also point to the primacy of meaning amongst the look-up reasons. In studies of reference needs of EFL learners, the dominance of meaning is at least as pronounced, with the following percentages of use of dictionaries for meaning reported by the respective researchers: Tomaszczyk (1979) 85%, Béjoint (1981) 87%, Nuccorini (1992) 75%. These figures, however, cannot be compared directly, as they result from various survey procedures, which differ in terms of how many options a subject is allowed to check at a time. Also, what is reported as "meaning" in the various studies may be elicited in different ways, sometimes including definitions and equivalents.

4.4.3 Polish equivalent

In item G3 of the Learners' Questionnaire, subjects were asked how often they looked for Polish equivalents in their dictionaries. As discussed in 4.4.2 above, this item was meant to elicit responses relating to dictionary consultations for Polish equivalents, independently from searches for meaning. Problems with the independence of these two types of reference needs have also been discussed above.

The breakdown of responses is presented in Table 40 and the corresponding histogram is given in Figure 24

Table 40: Breakdown of responses to item G3: Polish equivalent

never	rarely	often	most often	missing
18 (2.5%)	111 (15.6%)	291 (40.9%)	291 (40.9%)	1 (0.1%)

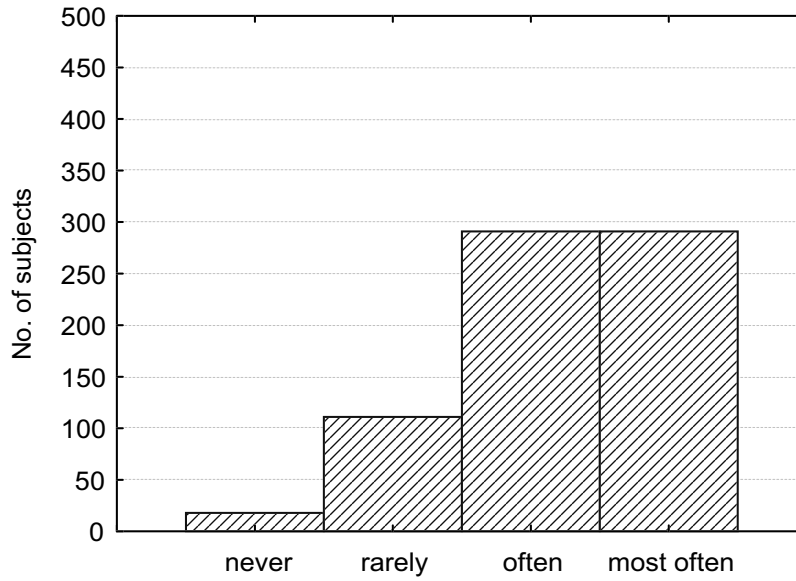

Figure 24: Histogram of responses to item G3: Polish equivalent

An identical number of subjects (41% each) ticked the options *often* and *most often*. 16% of the subjects *rarely* look up Polish equivalents, with one in twenty claiming they *never* refer to a dictionary in search of a Polish equivalent. It should be recalled that in their responses to the meaning item (4.4.2 above) almost three times as many subjects chose *most often* than did *often*, whereas here these options have received equal votes. It may be assumed that looking up Polish equivalents is a reference need most characteristic of decoding activities. A Polish equivalent would serve as a key to meaning (for a linguistically untrained person, meaning and native language expression are virtually one and the same thing). A Polish equivalent could also be sought in a situation when the user basically understands the English item but cannot think of a Polish equivalent, or wants to double-check on it, such as during an L2→L1 translation task.

4.4.4 English equivalent

In item G4, subjects were asked to indicate how often they consulted their dictionaries for English equivalents. The breakdown of responses is presented in Table 41 and the corresponding histogram is given in Figure 25.

Table 41: Breakdown of responses to item G4: English equivalent

never	rarely	often	most often	missing
15 (2.1%)	79 (11.1%)	283 (39.7%)	334 (46.9%)	1 (0.1%)

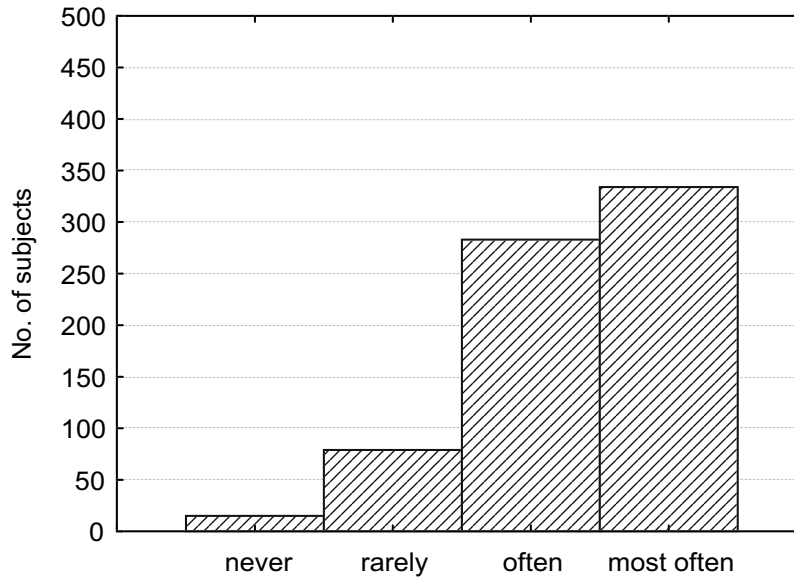

Figure 25: Histogram of responses to item G4: English equivalent

The modal value of the distribution of responses for English equivalents is *most often*, chosen by close to half of the subjects (47%), the second most popular option is *often*, with 40% of the subjects' votes. The other two options, *rarely* and *never* are not nearly as popular, accounting for only 11% and 2% of the responses, respectively. English equivalents are a type of information that Polish learners of English would typically access when engaged in encoding activities, as part of an onomasiological search.

4.4.5 Part of speech

In item G5, subjects were asked how often they had a need to look for part-of-speech information in their dictionaries. The breakdown of responses for this questionnaire item is presented in Table 42 and the corresponding histogram is given in Figure 26.

Table 42: Breakdown of responses to item G5: Part of speech

never	rarely	often	most often	missing
254 (35.7%)	383 (53.8%)	55 (7.7%)	18 (2.5%)	2 (0.3%)

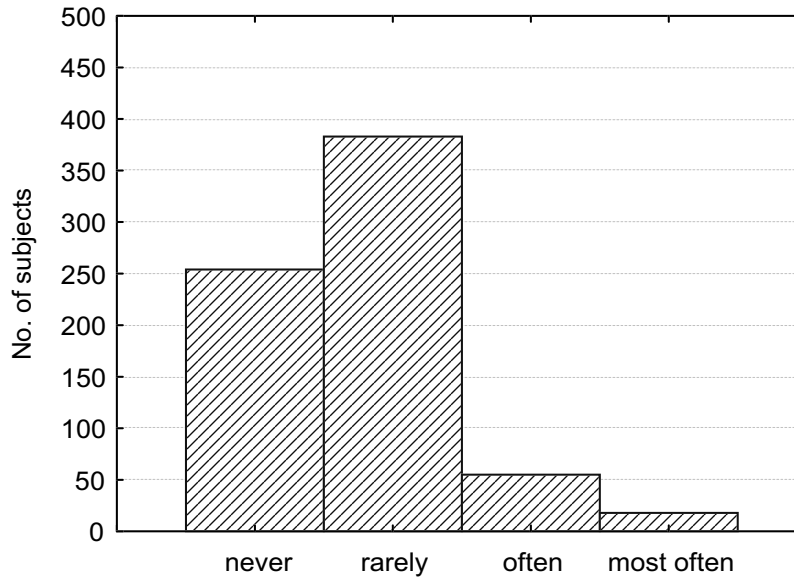

Figure 26: Histogram of responses to item G5: Part of speech

Over half (54%) of the subjects said they looked for part-of-speech information only *rarely*. Another 36% *never* see the need to consult their dictionaries for part of speech. The two options jointly account for close to 90% of all responses. Only 8% of the subjects look for part-of-speech information *often*, and less than 3% *most often*.

The above results suggest that syntactic categorization (part of speech) is a type of information that is only rarely, if at all, sought by Polish learners of English. Another type of syntactic information was covered by item G6.

4.4.6 Syntactic structure

While item G5 discussed above referred to part-of-speech information, this next item asked subjects how often they would look in their dictionaries for guidance on how to construct a sentence. The breakdown of responses is presented in Table 43 and the corresponding histogram is given in Figure 27.

Table 43: Breakdown of responses to item G6: Syntactic structure

never	rarely	often	most often	missing
207 (29.1%)	339 (47.6%)	147 (20.6%)	18 (2.5%)	1 (0.1%)

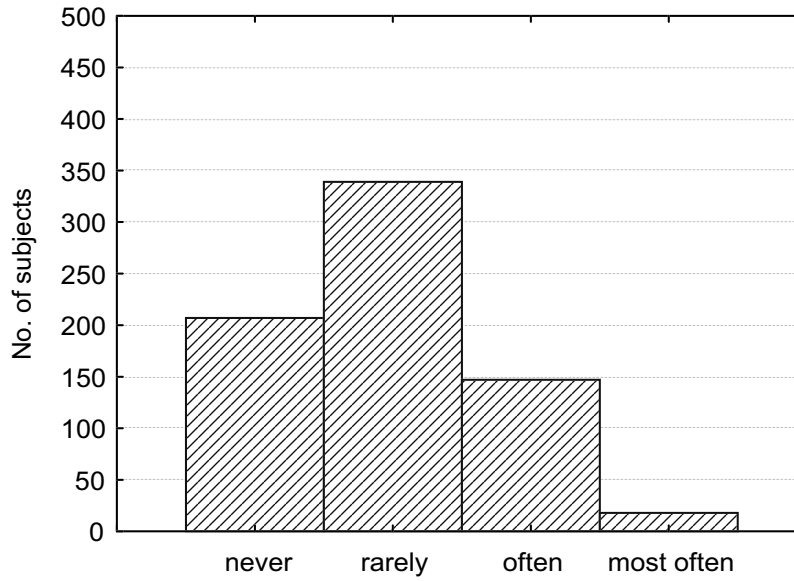

Figure 27: Histogram of responses to item G6: Syntactic structure

Close to half (48%) of the subjects look for syntactic structure information *rarely*, and 29% *never* need such information. A substantial fifth of the sample (21%) said they *often* used dictionaries for sentence structure information. Only a small margin (less than 3%) opted for *most often*.

Previous studies with which I am familiar did not distinguish between basic syntactic categorization (part of speech) and more detailed syntactic information (syntactic subcategorization, complementation). Perhaps partly for this reason, previous studies produced varying results as regards the need for syntactic information, with some studies reporting syntactic information to be quite popular (e.g. Béjoint 1981), and others finding that it is hardly used at all (e.g. Battenburg 1991).

When responses are compared for item G5 above and item G6, the most striking difference is in the popularity of the *often* option: nearly three times as many subjects chose this option for syntactic structure than did for part of speech, suggesting that the interest in detailed syntactic information helpful in constructing sentence structure may be greater than in the basic part-of-speech information.

4.4.7 Collocation

In item G7, subjects were asked about their use of dictionaries for finding typical word combinations. The breakdown of responses to this item is presented in Table 44 and the corresponding histogram is given in Figure 28.

Table 44: Breakdown of responses to item G7: Collocation

never	rarely	often	most often	missing
174 (24.4%)	312 (43.8%)	199 (27.9%)	25 (3.5%)	2 (0.3%)

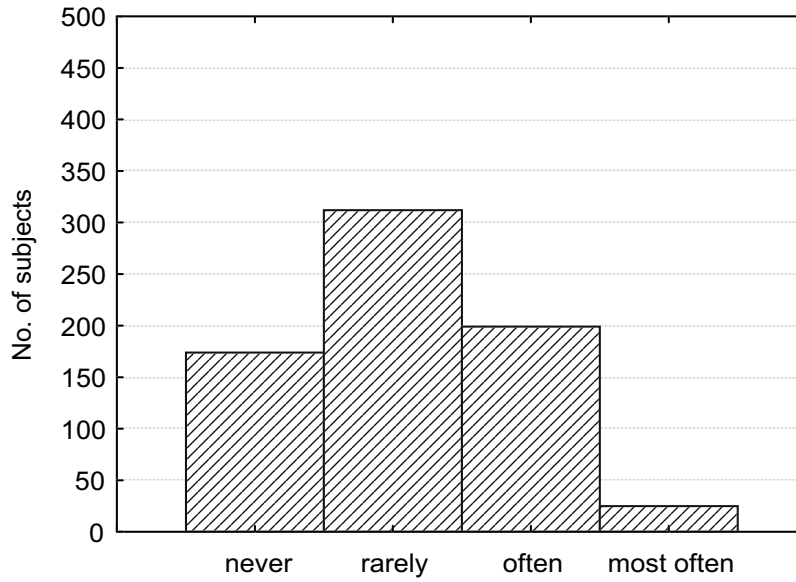

Figure 28: Histogram of responses to item G7: Collocation

The most popular response for this item was *rarely*, picked by 44% of the subjects, followed by *often* (28% of the responses). A quarter of the subjects said they *never* looked for collocational information in their dictionaries. Less than 4% reported collocation to be the type of information they looked for *most often*.

4.4.8 Situation

In item G8, subjects were asked how often they consulted dictionaries to check in what types of situations a word was typically used. This item was designed to elicit information on the need of the subjects for information on style and register, though the technical phrasing was avoided as a rule. The breakdown of responses is presented in Table 45 and the corresponding histogram is given in Figure 29.

Table 45: Breakdown of responses to item G8: Situation

never	rarely	often	most often	missing
128 (18.0%)	310 (43.5%)	246 (34.6%)	27 (3.8%)	1 (0.1%)

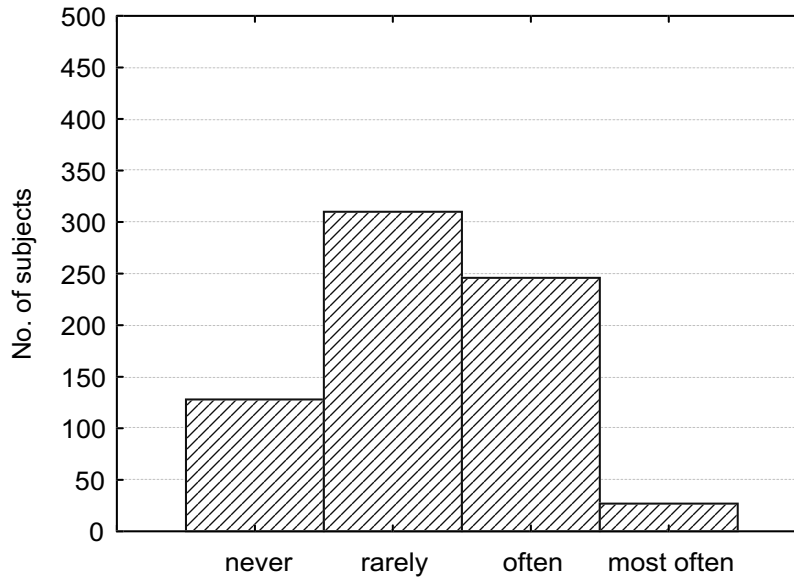

Figure 29: Histogram of responses to item G8: Situation

The top option for item G8 was *rarely*, accounting for 44% of the responses. The second most common response was *often* (35%), followed by *never* (18%). Only 4% of the subjects ticked *most often*. The distribution for this item is quite similar to the one for collocational information in 4.4.7 above, except that the *often* option was here somewhat more popular at the cost of *never*.

4.4.9 Synonyms

In the final item in section G of the Learners' Questionnaire, subjects were asked how often they looked for synonyms ('words with similar meaning') in their dictionaries. The breakdown of responses is presented in Table 46 and the corresponding histogram is given in Figure 30.

Table 46: Breakdown of responses to item G9: Synonyms

never	rarely	often	most often	missing
119 (16.7%)	306 (43.0%)	257 (36.1%)	29 (4.1%)	1 (0.1%)

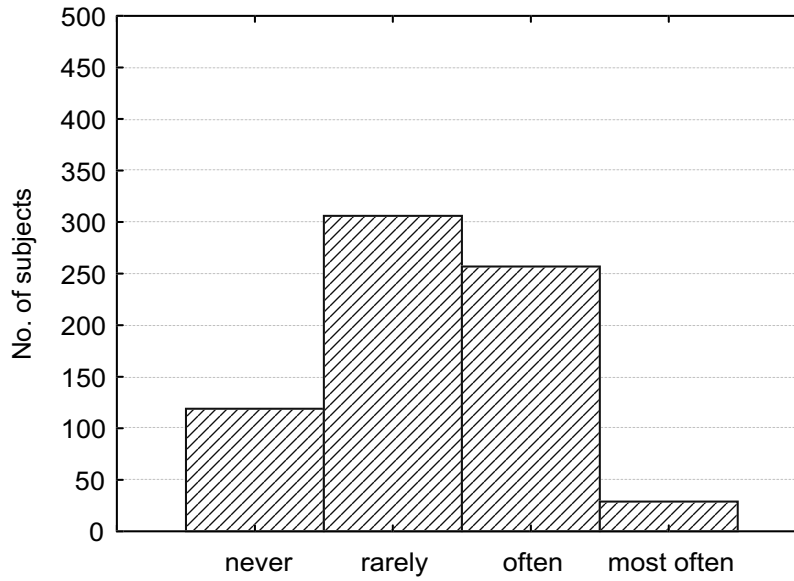

Figure 30: Histogram of responses to item G9: Synonyms

The modal value in the distribution of responses falls on the *rarely* option, which accounts for 43% of the subjects' responses. *Often* is a close second with 36% of the subjects in the sample selecting this response. 17% of the subjects said they *never* looked for synonyms in their dictionaries. The least popular response was *most often*, with only 4% of the responses. Overall, the distribution of responses to this item is remarkably similar to the one for style and register (4.4.8 above). There is a possibility that this might be due to subject fatigue.

4.4.10 Ranking of consultation frequencies

To more directly compare subjects' reported frequency with which they looked for the nine information types in dictionaries, the four responses were assigned numerical values from 1 to 4 as follows: *never* = 1, *rarely* = 2, *often* = 3, *most often* = 4. These numerical values were used to compute mean consultation frequencies for each information type. The information types were then ranked by mean consultation frequency, and the ranking is presented in Table 47.

Table 47: Ranking of the nine information types by mean consultation frequency

Rank	Information Type	Mean	Std.Dev.	Valid N
1	G2: Mng	3.56	0.72	711
2	G4: Eng	3.32	0.75	711
3	G3: Pol	3.20	0.79	711
4	G9: Syn	2.28	0.79	711
5	G8: Sitn	2.24	0.79	711
6	G7: Coll	2.11	0.81	710
7	G1: Pron	2.07	0.82	710
8	G6: Synt	1.97	0.77	711
9	G5: POS	1.77	0.70	710

According to the ranking, the type of information most frequently sought by the subjects is meaning, with a mean frequency score of 3.56. This is followed closely by English equivalent (3.32) and Polish equivalent (3.20). Such a configuration of top entries suggests that users in the study are most interested in semantic information for known word forms (semasiological consultation acts) and in locating word forms, either to express specific meanings (onomasiological consultation acts, using Polish words as indices for access), or to serve as translations for forms in the other language. These three types of information appear to form the core of users' reference needs, with all the other information categories being more or less subsidiary or marginal.

Beyond the three central types of information that are accessed most frequently, synonyms are the most popular (2.28), followed close behind by style and register information (2.24). Collocational information is next with a frequency score of 2.11, and pronunciation is the next favourite (2.07). Syntactic information is the least popular of all those addressed in the questionnaire, with sentence structure information scoring 1.97, the second lowest score. Part-of-speech information is at the very bottom of the ranking list with a low score of just 1.77.

4.4.11 Effect of level on consultation frequency

It is interesting to learn how the consultation frequencies for the different information types of information vary by learner level. The design of the present study allows such an analysis, and this is the focus of this section[18]. First, I look at the relative frequency scores computed as in 4.4.10 above. However, since the overall consultation frequency increases with level (see 4.1.6 above), absolute frequency scores will also be computed, taking into account the overall frequency of dictionary consultation.

[18] Preliminary results of this part of the study appeared in Lew (2002b).

4.4.11.1 Relative frequency

Relative consultation frequency scores for all nine information types covered in the questionnaire, computed as in 4.4.10 above, are plotted against learner level in Figure 31. A GLM analysis on relative frequency scores shows the effect of learner level on relative consultation frequencies to be highly significant (Wilks lambda=0.73, $F_{(36, 2606.2)}$=6.44, p<0.0001). Univariate tests are also highly significant (p<0.0001) for all nine information types.

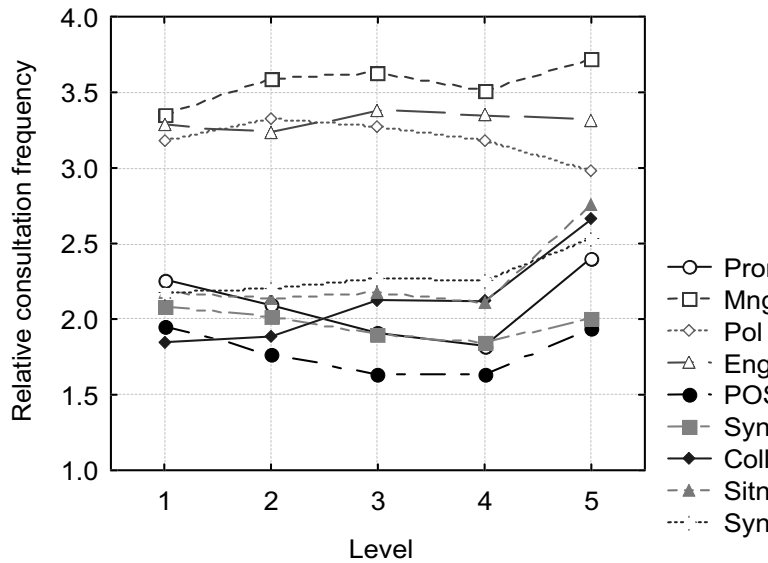

Figure 31: Relative consultation frequency by learner level

The graph indicates that for all levels except the highest level five, information types appear to cluster into two groups. Meaning, English equivalents and Polish equivalents form the high-frequency cluster, whereas the remaining information types constitute the low-frequency group of information types. This clustering into core and peripheral types of information has already been revealed in the information types ranking discussed in section 4.4.10 above.

Something that has not been revealed in the ranking by mean only, without reference to learner level, is the breakdown of the two-way clustering at level 5: consultation frequencies for peripheral non-semantic types of information shoot up to almost catch up with the three core information types. It is just the two syntactic types of information (sentence structure and part of speech) that appear to lag behind the others at level 5.

The blurring of the basic division into core and peripheral information types at level 5 indicates advanced learners' increased interest in the less central types of information provided in dictionaries. This could be indicative of a transition in the way dictionaries are used: from a fairly superficial, semantic-centred use

typical of non-advanced users, to a more sophisticated, more comprehensive mode of dictionary use.

It is also interesting to examine the dynamics of the changes in consultation frequencies for the different information types across the range of levels, but this is better expressed with absolute frequency scores, to which I now turn.

4.4.11.2 Absolute frequency

The way section G of the Learners' Questionnaire was constructed, information on consultation frequencies was elicited as relative frequency. For example, if a learner used a dictionary once a week on average, and used an English-Polish dictionary to look up Polish equivalents, that learner would have ticked the *most often* option for the Polish equivalents item in section G, and would at the same time indicate *once a week* in item E2 (frequency of consultation for English-Polish dictionaries). Another learner using the same type of dictionary in exactly the same way, only more frequently, say daily, would have ticked the same option in section G, but would have indicated *daily* under item E2. It is in this sense that I claim responses to item G to reflect *relative* frequencies of consultation.

To obtain a measure of absolute frequency with which specific information type was sought, its corresponding relative value reported by each subject (items G1-G9) was multiplied by the maximum value reported for frequency of dictionary consultation (i.e. the largest of E1-E3; see 4.1.6) for this subject. A measure computed in this way is likely to correspond more directly to absolute consultation frequencies for the different information types.

A plot of absolute consultation frequency scores versus learner level for the nine information types is given in Figure 32. In a GLM analysis on absolute scores, the effect of learner level on absolute consultation frequencies is again highly significant (Wilks lambda=0.66, $F_{(36,\ 2602.5)}=8.62$, $p<0.0001$), and so are univariate tests ($p<0.0001$) for all nine information types.

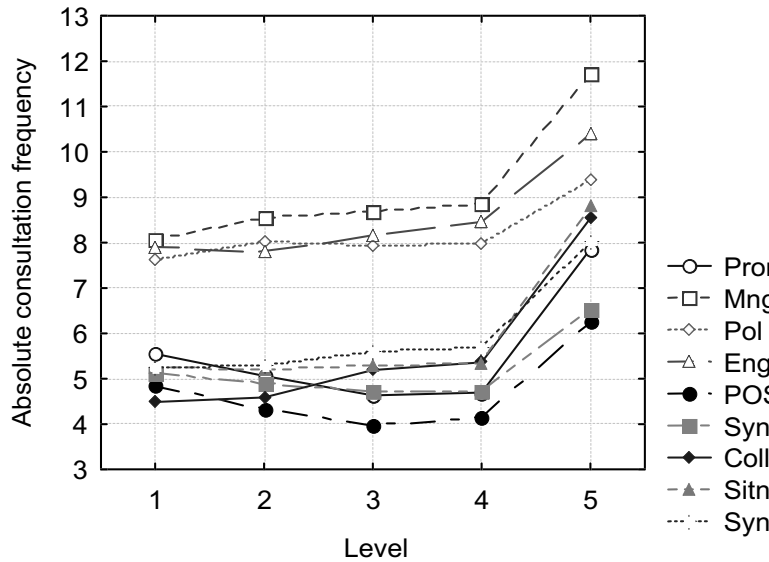

Figure 32: Absolute consultation frequency by learner level

Compared with the graph in Figure 31, we note the higher slope values for absolute scores, especially for level 5, which is probably primarily related to an increase in general consultation frequency at this level. It is worth noting that the increase is only evident at the highest level. Previous studies have returned highly inconsistent results on the relationship between proficiency level and dictionary use. Perhaps part of the reason for this lack of consistency lies in the incompatibility between the level ranges of the subjects covered in the different samples.

While consultation frequencies for all peripheral information types surge at level 5, they do not all behave alike at lower levels. Some types of information, such as synonyms, collocation, style and register enjoy a steady rise across the whole level range. In contrast, pronunciation, sentence structure, and part-of-speech information decline from level 1 through 4, only to jump up at level 5. This effect is especially salient for pronunciation. This could indicate that learners lose their initial interest in looking up pronunciation at a fairly early stage of their learning, maybe because they feel their control of English phonetics to be sufficient for that stage.

4.4.12 Correlation between types of information

To investigate the relative independence between the nine types of information for which consultation frequency was elicited, Pearson correlation coefficients were computed for all pairs of information types. A matrix of Pearson coefficients is given in Table 48. For easier reference, coefficient values are given as a

complete square matrix, even though the values are symmetrical about the main diagonal.

Table 48: Pearson correlation coefficients for relative consultation frequency scores for the nine information types

	Pron	Mng	Pol	Eng	POS	Synt	Coll	Sitn	Syn
G1: Pron	1.00	0.14	-0.00	0.09	0.34	0.26	0.26	0.29	0.19
G2: Mng	0.14	1.00	0.44	0.41	-0.00	0.14	0.19	0.22	0.12
G3: Pol	-0.00	0.44	1.00	0.57	0.05	0.19	0.05	0.10	0.01
G4: Eng	0.09	0.41	0.57	1.00	0.07	0.20	0.19	0.19	0.08
G5: POS	0.34	-0.00	0.05	0.07	1.00	0.42	0.30	0.31	0.18
G6: Synt	0.26	0.14	0.19	0.20	0.42	1.00	0.42	0.38	0.23
G7: Coll	0.26	0.19	0.05	0.19	0.30	0.42	1.00	0.60	0.40
G8: Sitn	0.29	0.22	0.10	0.19	0.31	0.38	0.60	1.00	0.40
G9: Syn	0.19	0.12	0.01	0.08	0.18	0.23	0.40	0.40	1.00

The strongest correlation in the matrix obtains between collocation and style/register (0.60). This means that there is a marked degree of kinship between the two types of information, in the sense that learners in the sample tend to look them up with similar frequencies. It is worth pointing out that both categories are amongst the less basic types of information offered in dictionaries, and both are primarily useful in encoding tasks. Also, synonym is fairly strongly correlated (0.40) with both collocation and style/register.

The second highest correlation coefficient (0.57) was obtained for the Polish equivalent – English equivalent pair. These two, in contrast to the previous pair, are among the core information types. They would both be typically used in translation tasks, and would likely represent the primary usage mode of a bi-directional bilingual dictionary, especially by the unsophisticated user.

Relatively high correlation coefficients were also obtained between meaning and Polish equivalent (0.44), as well as meaning and English equivalent (0.41). That the former is higher is understandable in view of the close practical and theoretical affinity, for the Polish learner, between meaning and Polish equivalent. The latter correlation is probably mainly due to the joint membership in the core information cluster.

A high correlation coefficient of 0.42 was found between syntax and part of speech, on the one hand, but also between syntax and collocation, on the other. The first correlation can be accounted for by the fact that both information types provide syntactic information of some kind. The second pair may be associated through productive dictionary use, where the two types of information address two aspects of text construction: the syntactic and semantic conventions of word combination.

In terms of the items with the lowest correlation coefficients, pronunciation and Polish equivalent exhibit a near-zero value, as do meaning and part of

speech, as well as Polish equivalents with synonym. Very low values were found for Polish equivalents and collocation (0.05), English equivalents and part-of-speech information (0.07), English equivalents and synonyms (0.08), and English equivalents and pronunciation (0.09). In terms of statistical significance, correlations lower or equal to 0.07 are not significant at the 5% level, those of 0.08 or higher are significant at this level.

In the next section, I supplement the present analysis of the relationships between the different information types. Rather than simply list the Pearson correlation values, factor analysis and cluster analysis will be used to explore the multidimensional aspects of the information types.

4.4.13 Information types: extraction of factors and clusters

In order to investigate more closely how the different information types position themselves in relation to one another, a factor analysis was performed using the PCA (Principal Components Analysis) extraction method. Projection of the nine information types on the factor plane of two factors with the greatest loadings is presented in Figure 33.

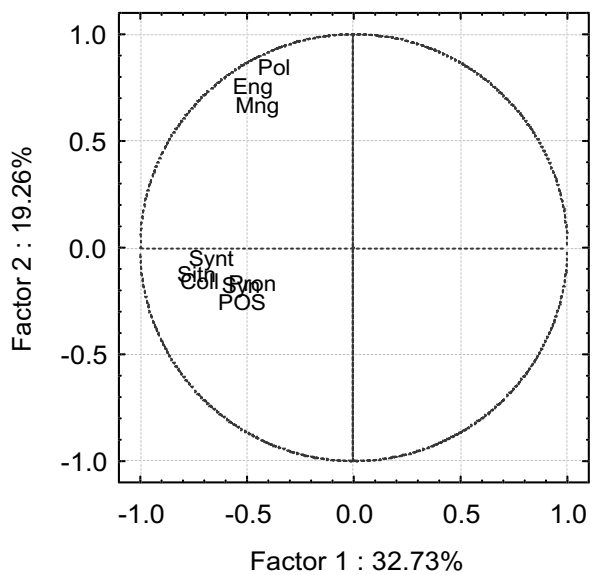

Figure 33: Factor analysis (PCA) on consultation frequencies for information types: projection of the nine types on the factor plane

The two principal factors account for 52% of the total variance. The graph in Figure 33 indicates that Factor 1 is mainly composed of the six information types previously identified as the peripheral types (Synt, Sitn, Coll, Syn, POS, Pron: some of the labels are a little hard to read in the graph, since their coordinates overlap so closely); for Factor 2, the main contribution comes from the three in-

formation types identified as core types (meaning, Polish equivalent, English equivalent).

In order to investigate further the nature of the clustering of the information types, a cluster analysis was performed on the consultation frequencies for the nine information types, with Euclidean distances used as metric and single linkage. The resulting vertical icicle plot is given in Figure 34.

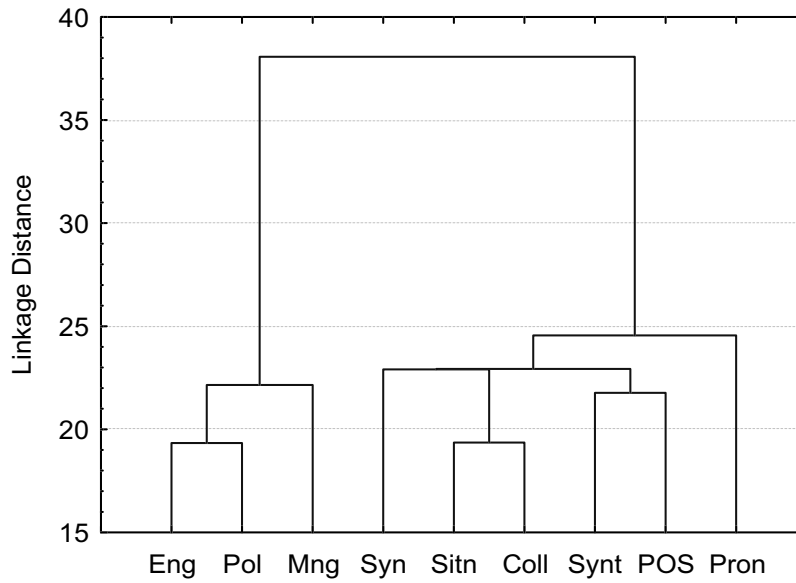

Figure 34: Cluster analysis on consultation frequencies for information types

The icicle plot in Figure 34 reveals two clear clusters: the core Eng-Pol-Mng triad against the peripheral group of the remaining six information types again, just as was identified by means of the factor analysis above. However, the plot in Figure 34 clearly shows the clustering configuration of the information types. We see the first-level clustering of English with Polish equivalents within the core main cluster on the one hand, and style/register with collocation in the peripheral group of information types. We also note the affinity between the two types of syntactic information (Synt and POS), and the relative autonomy of pronunciation information within its peripheral cluster. This last effect has not been apparent in the analysis so far.

4.4.14 Conclusion

In this study, the lexicographic information types cluster into two groups in terms of how often subjects consult them. The three core types of information, which are looked up most frequently, include: meaning, English equivalents, and Polish equivalents. The remaining six types of information (of those covered in this study), looked up less often, form the peripheral group: synonyms, style and register, collocation, sentence structure, part of speech, pronunciation. Syntactic in-

formation is looked up the least. At the most advanced level, learners become more interested in the peripheral types of information, and the gap between the two clusters narrows. However, they still retain their interest in the basic types of information.

It would be interesting to see how the need for the different types of information depends on the type of task a dictionary user is engaged in. The design of the present study did not, however, allow for such information to be elicited.

4.5 Lexical dictionary effectiveness

In this part of the study, the effectiveness of six dictionary versions (see 3.7.2 above for a detailed description of the six versions, which are reproduced in Appendix 7) is assessed in a number of lexical tasks with varying amount of context. The effects of dictionary version and of learner level will be evaluated.

4.5.1 Overall lexical effectiveness

Overall lexical effectiveness of the test dictionaries was measured by the mean subject scores across all tasks. Effects of level (with version controlled) and version (with level controlled) on lexical effectiveness scores will be discussed below, followed by the interaction effect of level by version. All three effects have been found to be highly significant by GLM analysis. A detailed ANOVA table is given in Table 49.

Table 49: ANOVA table for overall lexical effectiveness scores

Effect	SS	df	MS	F	p
Intercept	123387.1	1	123387.1	20148.33	<0.0001
Level	2558.1	4	639.5	104.43	<0.0001
Version	843.3	5	168.7	27.54	<0.0001
Level*Version	246.6	20	12.3	2.01	0.0055
Error	4176.5	682	6.1		

4.5.1.1 Effect of level

It was expected that subjects' lexical performance on the Dictionary Effectiveness Test would improve steadily with learner level, based on the rationale that subjects' reference skills would tend to be generally positively correlated with their level. This expectation was fully confirmed by the experimental results. Mean values for each of the five levels with their respective 95% confidence intervals are presented in Table 50.

Table 50: Overall lexical effectiveness by level

Level	Score	Std. Error	-95%	+95%	N
1	10.29	0.32	9.66	10.92	145
2	12.68	0.25	12.17	13.18	131
3	13.89	0.21	13.47	14.31	175
4	14.93	0.16	14.62	15.24	154
5	15.87	0.12	15.63	16.10	107
Total	13.46	0.12	13.21	13.70	712

The effect size of level on overall lexical performance scores measured by the Dictionary Effectiveness Test is considerable, with an over fifty percent increase in mean scores across the whole range of learner levels. The rate of increase is fairly uniform, with a slightly steeper climb from level 1 to level 2. This is clearly seen in a plot of the level effect presented in Figure 35.

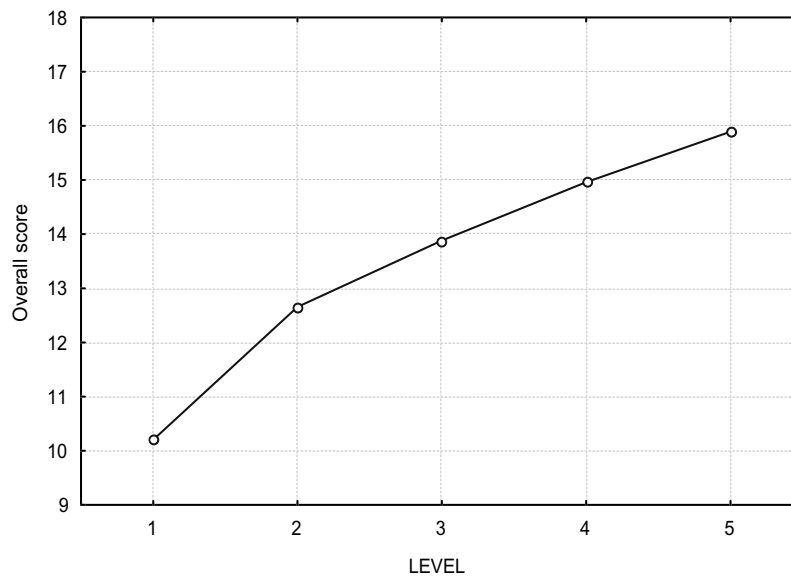

Figure 35: Plot of overall lexical effectiveness by level

The effect of learner level on lexical scores is highly significant ($F_{(4, 682)}$=104.43, p<0.0001). An examination of the 95% confidence intervals given in Table 50 reveals that the confidence intervals for the consecutive levels never overlap, with each interval range being placed higher than the one for the previous level. This is an indication that the population means increase monotonically with increasing levels, in a manner similar to that shown in Figure 35.

4.5.1.2 Effect of dictionary version

One of the primary aims of the present study has been to test the effectiveness of various dictionary types in lexical tasks completed with the aid of dictionary entries. Here, the effectiveness of the six versions of dictionary entries used in this study will be compared. Table 51 lists mean lexical scores for each of the six versions, and their 95% confidence intervals.

Table 51: Overall lexical effectiveness by version
P=Polish; E=English; eq=equivalent; df=definition; +=followed by

Version	Score	Std. Error	-95%	+95%	N
1.Peq	14.11	0.24	13.62	14.59	119
2.Edf	11.01	0.41	10.21	11.81	121
3.Peq+Edf	14.02	0.22	13.59	14.45	116
4.Peq+Pdf	13.95	0.26	13.44	14.46	118
5.Edf+Peq	13.90	0.27	13.36	14.44	118
6.Pdf+Peq	13.80	0.29	13.23	14.38	120
Total	13.46	0.12	13.21	13.70	712

In the table, dictionary versions are identified by number (from 1 to 6). For easier identification, abbreviated types of semantic information have also been provided as follows: P=Polish; E=English; eq=equivalent; df=definition; +=followed by; e.g. Peq+Edf=Polish equivalent followed by English definition, etc.

The size of the effect of version is substantial, with the top-performing versions achieving scores 28% higher than the bottom-scoring version 2. A plot of means for the dictionary version effect is shown in Figure 36.

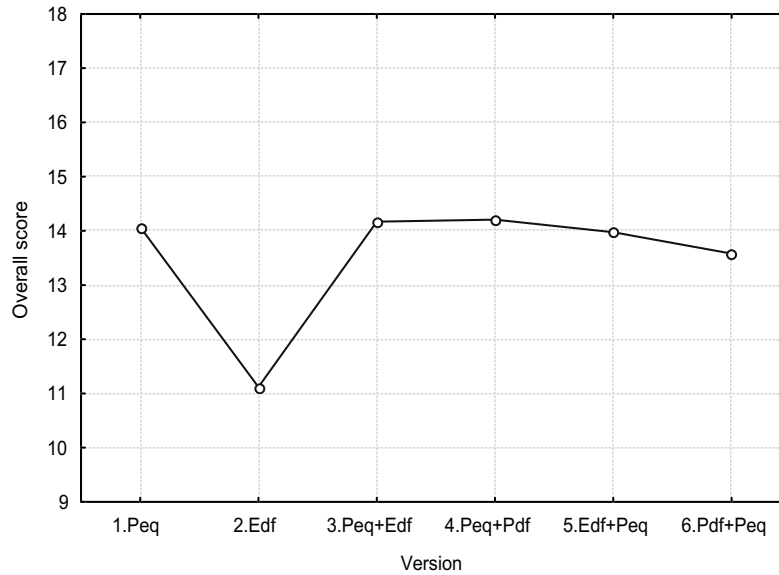

Figure 36: Plot of overall lexical effectiveness by version

The effect of version on scores is highly significant ($F_{(5, 682)}$=27.54, p<0.0001). The one dictionary type that clearly stands out from the rest is version 2 (the monolingual dictionary with English definitions). The mean lexical effectiveness score for this dictionary version is outperformed by any other version by about one third. A Tukey HSD (Honest Significant Difference) post hoc test for differences between the individual versions finds a highly significant difference between version 2 and every other version, but no significant differences between any other versions (see Table 52; please note that the matrix is symmetrical about the main diagonal, with the values given twice each for easier reference).

Table 52: Post hoc test on overall lexical effectiveness scores for dictionary versions: Tukey HSD probabilities; significant differences are starred

Version	1	2	3	4	5	6
1		*0.00002	0.99980	0.99711	0.98883	0.93620
2	*0.00002		*0.00002	*0.00002	*0.00002	*0.00002
3	0.99980	*0.00002		0.99996	0.99927	0.98606
4	0.99711	*0.00002	0.99996		0.99999	0.99730
5	0.98883	*0.00002	0.99927	0.99999		0.99964
6	0.93620	*0.00002	0.98606	0.99730	0.99964	

The Tukey test splits the set of six versions into two homogenous groups; the first group only consists of version 2, while the second group holds the other five versions of the test dictionary. The above clearly points to subjects' performance

with version 2 dictionaries being much worse than with any other version. Since version 2 is the only one that lacks the Polish equivalent in its semantic information, this is what has to be singled out as the underlying factor for the significant difference in performance obtaining between version 2 and the remaining versions.

4.5.1.3 Interaction of level by version

In this section, the interaction effect of level by version on lexical effectiveness scores will be examined. Mean values for each of the 30 combinations of the five levels and six versions are presented in Table 53. 95% confidence intervals are also included in the table.

Table 53: Overall lexical effectiveness by level and version

Version	Level	Score	Std. Error	-95%	+95%	N
1.Peq	1	11.00	0.74	9.45	12.55	22
1.Peq	2	13.17	0.41	12.32	14.02	24
1.Peq	3	15.07	0.27	14.52	15.62	28
1.Peq	4	15.24	0.32	14.58	15.90	25
1.Peq	5	15.88	0.25	15.36	16.39	20
2.Edf	1	6.36	0.83	4.64	8.08	25
2.Edf	2	9.83	0.81	8.15	11.52	21
2.Edf	3	10.93	0.73	9.43	12.42	28
2.Edf	4	13.09	0.53	12.00	14.19	27
2.Edf	5	15.38	0.26	14.83	15.92	20
3.Peq+Edf	1	11.72	0.47	10.77	12.68	29
3.Peq+Edf	2	13.41	0.41	12.54	14.28	17
3.Peq+Edf	3	14.44	0.32	13.77	15.10	24
3.Peq+Edf	4	15.08	0.25	14.56	15.60	30
3.Peq+Edf	5	16.19	0.25	15.66	16.71	16
4.Peq+Pdf	1	11.73	0.54	10.64	12.83	30
4.Peq+Pdf	2	13.00	0.55	11.86	14.14	25
4.Peq+Pdf	3	14.50	0.46	13.55	15.45	23
4.Peq+Pdf	4	15.60	0.24	15.09	16.11	20
4.Peq+Pdf	5	16.20	0.21	15.76	16.64	20
5.Edf+Peq	1	10.69	0.76	9.10	12.28	21
5.Edf+Peq	2	13.50	0.54	12.38	14.62	25
5.Edf+Peq	3	14.02	0.42	13.15	14.88	31
5.Edf+Peq	4	15.50	0.28	14.93	16.07	28
5.Edf+Peq	5	16.15	0.32	15.46	16.84	13
6.Pdf+Peq	1	9.72	0.92	7.77	11.67	18

Version	Level	Score	Std. Error	-95%	+95%	N
6.Pdf+Peq	2	13.03	0.58	11.80	14.25	19
6.Pdf+Peq	3	14.34	0.42	13.50	15.18	41
6.Pdf+Peq	4	15.25	0.34	14.54	15.96	24
6.Pdf+Peq	5	15.56	0.41	14.68	16.43	18
Total	–	13.46	0.12	13.21	13.70	712

Performance scores for the different combinations of dictionary type and learner level are spread over a wide span, ranging from a low of about 6 to a high of around 16. The nature of the interaction effect is easier to assimilate from a graphical plot presented in Figure 37. The effect is statistically significant ($F_{(20, 682)}=2.01$, $p<0.006$).

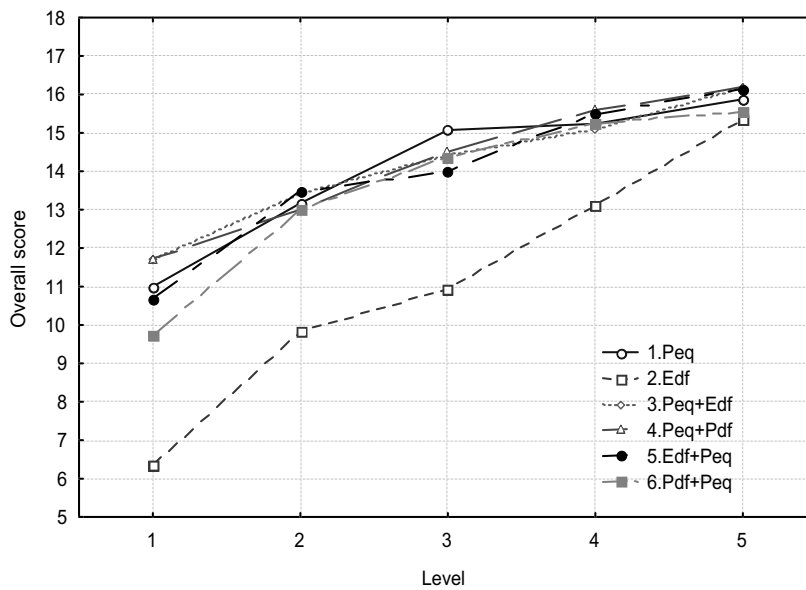

Figure 37: Interaction plot of overall lexical effectiveness by level and version

The plot reveals a steady increase in scores as a function of level, and this effect is observed for all dictionary versions. The second feature showing up clearly in the graph is the striking gap between scores obtained with the aid of the monolingual dictionary and those for all the other dictionary versions: it appears that the monolingual dictionary is much less effective on the lexical tasks tested. This substantial gap is largest for levels 1 to 3. It becomes somewhat smaller at level 4, to finally disappear at the highest level 5. One should, however, note, that the performance achieved with the monolingual dictionary by level 5 is only about as good as that of level 3 subjects using the bilingual dictionary: the respective mean scores are 15.38 and 15.07, with confidence intervals largely overlapping.

Foreign language teachers often hold and express very assured opinions about the relative advantage of the monolingual dictionary over the bilingual dictionary for advanced learners (Thompson 1987). In terms of the results of the present study, we do indeed find the performance difference between the monolingual and bilingual dictionary to be related to the proficiency level of the user, and the direction of the interaction effect is properly predicted. However, the nature of this relationship as revealed by this study is such that the *disadvantage* of the monolingual dictionary decreases with growing learner level, but it never really turns into an advantage in absolute terms. This is surely not the kind of advantage that proponents of the early weaning of learners onto monolingual-only dictionaries have in mind. The present results also add some empirical substance to learners' apparent desire to keep using their bilingual dictionaries even at an advanced level.

In the following sections, I will examine results for individual tasks, which vary in the amount of lexical context provided, to investigate whether the measured effectiveness of the different dictionary versions is sensitive to how much contextual information is provided, and if so, what the nature of the relationship is.

4.5.2 Out of context: word match

Out-of-context lexical effectiveness was assessed with the aid of the word match task (part H of the Dictionary Effectiveness Test), wherein subjects were instructed to match the word with the closest meaning to the target word. Just as for overall lexical effectiveness, effects of level (with version controlled) and version (with level controlled) on lexical effectiveness scores are of interest here, as well as the interaction effect of level by version. GLM analysis reveals all three effects to be significant, with the main effects being highly significant. A detailed ANOVA table is given in Table 54.

Table 54: ANOVA table for out-of-context lexical effectiveness scores

Effect	SS	df	MS	F	p
Intercept	18658.09	1	18658.09	10865.90	<0.0001
Level	374.14	4	93.53	54.47	<0.0001
Version	124.89	5	24.98	14.55	<0.0001
Level*Version	62.44	20	3.12	1.82	0.0158
Error	1171.08	682	1.72		

4.5.2.1 Effect of level

Table 55 gives mean scores for each of the five levels and their respective 95% confidence intervals. We find performance improving steadily with level, as for overall results, and the confidence intervals are non-overlapping, indicating a clear separation of levels in terms of estimated population word-match task

scores. The table indicates that the mean score for level 5 is over 50% higher than that for the lowest level.

Table 55: Word match task scores by level

Level	Score	Std. Error	-95%	+95%	N
1	4.05	0.11	3.83	4.26	145
2	4.91	0.12	4.68	5.13	131
3	5.26	0.10	5.06	5.45	175
4	5.79	0.11	5.58	6.00	154
5	6.28	0.13	6.03	6.54	107
Total	5.21	0.06	5.10	5.33	712

The shape of the level by word match task score relationship can be appreciated in Figure 38. The slope remains fairly constant throughout the whole range of levels present in the sample, with score values increasing steadily. The plot does not appear to depart markedly from that obtained for overall results. The effect is highly significant ($p<0.0001$).

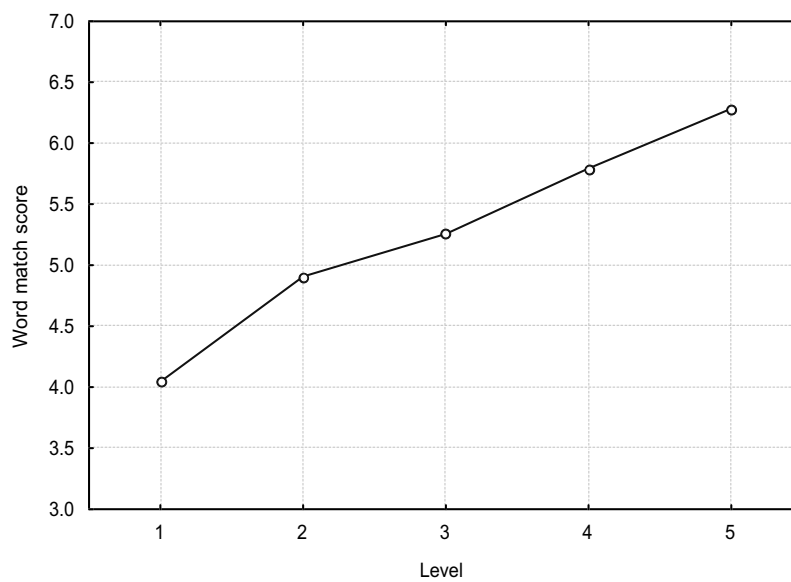

Figure 38: Plot of word match task scores by level

4.5.2.2 Effect of dictionary version

Mean scores with 95% confidence intervals for the six dictionary versions are presented in Table 56.The confidence intervals for versions 1, 3, 4, 5 and 6 overlap with each other, but the score for version 2 is markedly lower, similarly as for overall results. The top-scoring version outperforms version 2 by 29%, so the size of the effect of version is practically the same as for overall scores.

Table 56: Word match task scores by version

Version	Score	Std. Error	-95%	+95%	N
1.Peq	5.21	0.12	4.98	5.45	119
2.Edf	4.37	0.12	4.13	4.61	121
3.Peq+Edf	5.34	0.13	5.09	5.59	116
4.Peq+Pdf	5.64	0.12	5.40	5.88	118
5.Edf+Peq	5.39	0.13	5.14	5.64	118
6.Pdf+Peq	5.58	0.13	5.34	5.83	120
Total	5.21	0.06	5.10	5.33	712

Figure 39 illustrates the means for the individual versions in the form of a line plot. Just as was seen in overall results, version 2 clearly lags behind the other versions, but there is somewhat less of a uniformity here in the scores for the remaining versions than was evident for overall results (cf. Figure 36). The effect is highly significant ($p<0.0001$).

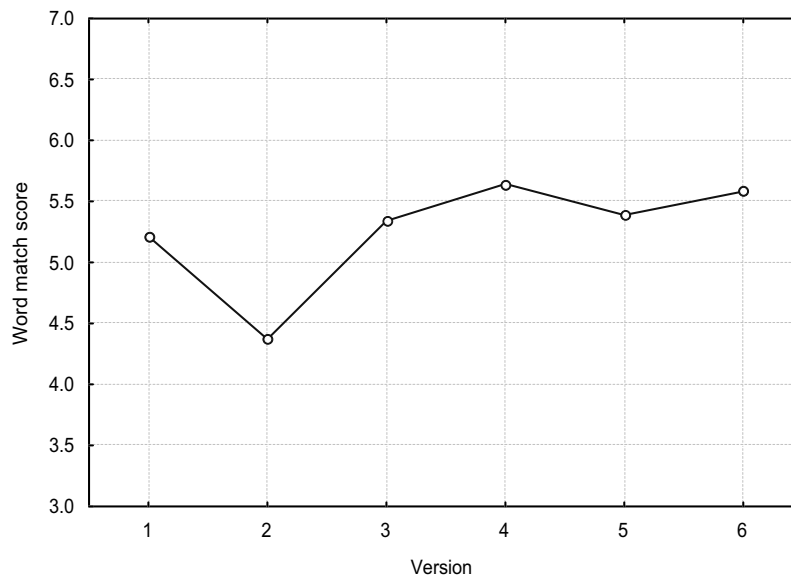

Figure 39: Plot of word match task scores by version

The zigzag pattern in the graph may suggest that versions 4 and 6 hold some advantage on this task. These two versions are different from the rest in that they contain definitions in Polish alongside Polish equivalents as part of the semantic information. Given the above emerging pattern, two contrast analyses were performed, excluding version 2, because it lacks Polish equivalents, whose presence has already been established as the strongest predictor of success on the word match task as well as overall.

In the first contrast analysis, versions 4 and 6 were compared against versions 1, 3 and 5, effectively testing subjects' performance on the word match task under two conditions: with Polish definitions and without Polish definitions. All subjects had access to Polish equivalents, and some subjects in the second group had access to English definitions. The set of contrast coefficients for this analysis is given in Table 57.

Table 57: Contrast coefficients for contrast analysis of versions 4 and 6 versus 1, 3 and 5, word match task scores

Version	N	Contrast coefficient
1	119	-2
2	121	0
3	116	-2
4	118	3
5	118	-2
6	120	3

Table 58 gives significance test results for the coefficients set as given in Table 57, showing this contrast to be significant at p=0.009, and suggesting that the Polish definitions were found to be helpful in raising subjects' scores on the word match task in the presence of Polish equivalents.

Table 58: Test of significance for contrast analysis of versions 4 and 6 versus 1, 3 and 5, word match task scores

	SS	df	MS	F	p
Effect	11.89	1	11.89	6.92	0.009
Error	1171.08	682	1.72		

In the above contrast analysis, the group that did not have access to Polish definitions was not uniform in what it had available in place of the Polish definitions: some had access to English definitions, others did not have such access. This could pose interpretation problems, and so a second contrast analysis was performed, in which versions 4 and 6 were compared with versions 3 and 5 only, excluding version 1 (as well as version 2). Such a test provides a more balanced design, with one group having access to Polish definitions and the other group to English definitions, on top of Polish equivalents. The set of contrast coefficients for this analysis is given in Table 59.

Table 59: Contrast coefficients for contrast analysis of versions 4 and 6 versus 3 and 5, word match task scores

Version	Cell N	Contrast coefficient
1	119	0
2	121	0
3	116	-1
4	118	1
5	118	-1
6	120	1

Significance test results using the coefficients from Table 59 are reported in Table 60, revealing this contrast to be significant at p=0.05.

Table 60: Test of significance for contrast analysis of versions 4 and 6 versus 3 and 5, word match task scores

	SS	df	MS	F	p
Effect	6.75	1	6.75	3.93	0.05
Error	1171.08	682	1.72		

The contrast analysis would thus suggest that subjects who had access to Polish definitions as well as Polish equivalents performed better on the word match tasks than did subjects who had access to English definitions as well as Polish equivalents.

This last effect will now be examined in greater detail in an alternative factorial analysis, using the underlying factors of language of definition (DefLang) in conjunction with position of definition (DefPos), both of which have been introduced in Table 12. The analysis excludes version 2, for which the position of definition is undetermined (there is no second semantic information element with which the definition could switch places), and version 1, for which both DefPos and DefLang are undetermined (as in version 2 above, plus there is no definition at all). An ANOVA table for this analysis is given in Table 61.

Table 61: ANOVA table for definition position and definition language, word match task scores

Effect	SS	df	MS	F	p
Intercept	13961.68	1	13961.68	7132.02	<0.0001
DefPos	0.87	1	0.87	0.44	0.51
DefLang	9.30	1	9.30	4.75	0.03
DefPos*DefLang	0.02	1	0.02	0.01	0.93
Error	916.16	468	1.96		

The only effect found to be significant is the main effect of DefLang (language of definition, p=0.03). Group means with 95% confidence intervals are presented in Table 62.

Table 62: Word match task scores by language of definition

DefLang	Score	Std. Error	-95%	+95%	N
Pol	5.58	0.09	5.40	5.76	238
Eng	5.30	0.09	5.12	5.48	234

The results indicate that the mean word-match task score under the Polish definitions condition is higher than the mean score under the English definitions condition, all in the presence of Polish equivalents. The difference, though statistically significant, is not a great one, with Polish definitions outperforming English definitions by some 5%.

Finally, we will briefly examine the advantage of Polish definitions over English definitions as a function of learner level. The scores on the word match task under the two conditions are plotted by level in Figure 40.

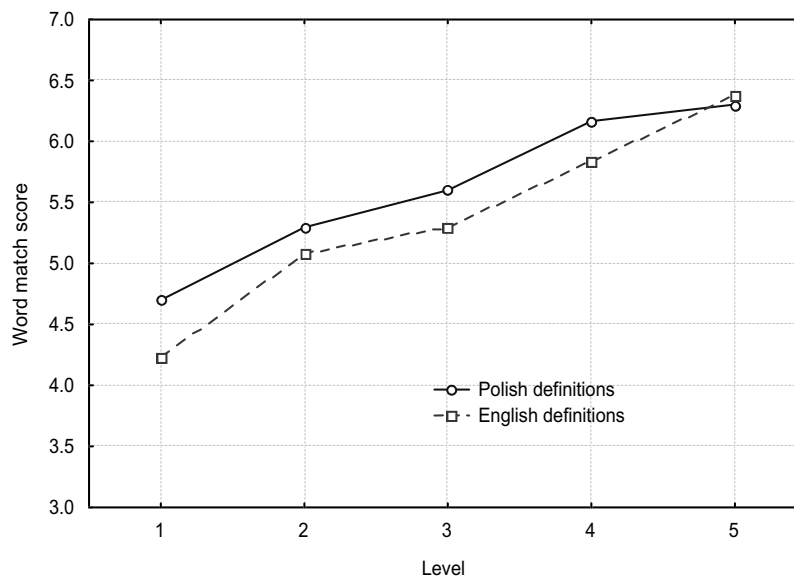

Figure 40: Word match score by level and language of definition

The two lines in the graph run in parallel for levels 1-4, and converge at level 5. This would indicate that in the presence of Polish equivalents, Polish definitions are providing dictionary users more additional help than English definitions for all levels except the most advanced one, at which the users' competence in English is presumably high enough to compensate for the gap.

The intriguing question in view of these findings is why Polish definitions should be giving users the extra boost on this out-of-context task. This is a question that is not an easy one to answer under the present design, but I will nevertheless address it briefly, although what follows is highly speculative and should be treated with caution. It will be helpful to examine the plots of individual scores for all H items, which are given in Figure 41.

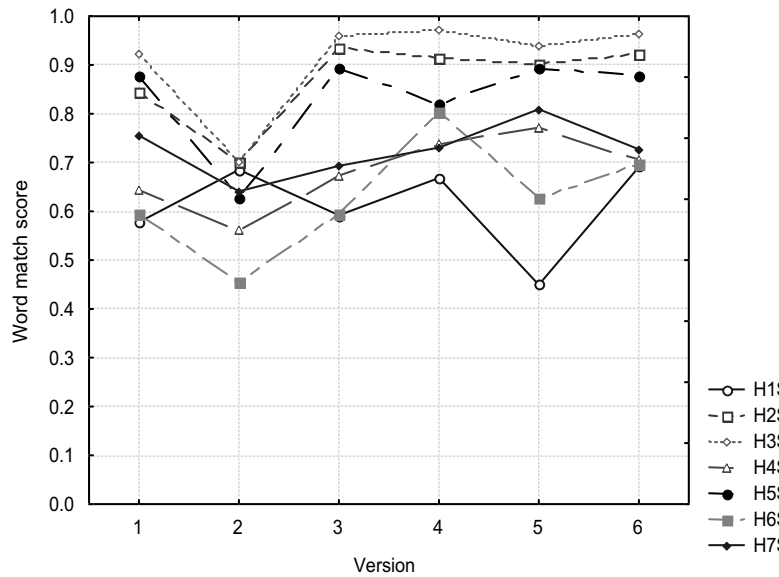

Figure 41: Word match scores by version for individual task H items

A close analysis of the graph suggests that the boost effect for versions 4 and 6 of the dictionary should be attributed primarily to items H1 and H6 of the out-of-context task. In item H6, the element that might be responsible for the improved performance of entries with Polish definition is the presence of the Polish lexical item *zabronić* in the Polish definition of the synonym option word *ban*: the same word *zabronić* is the Polish equivalent of the target pseudo-word *strod*. Some users might therefore match the two entries by identifying the same Polish word in the broader semantic information part of the two entries, which is perhaps not entirely unlike the kidrule strategy for production (Miller and Gildea 1987; Mitchell 1983b)[19].

Item H1 appears to be helped by Polish definition in a rather more subtle way. The target pseudo-word *helk* has two senses. The sense that is matched with the Polish synonym is the second sense. When the meaning is only given as the Polish equivalent, the polysemy of the Polish equivalent *delikatny* comes into play. It is possible that the Polish definition is helpful here as a sense indicator of the Polish equivalent. It would not be surprising if the polysemy issues came to

[19] Wingate (2002: 304) reports evidence of the kidrule strategy in comprehension.

the surface in a type of task where no context was provided for the target words, because textual clues present in contextually richer tasks would be missing. However, the above is highly speculative; perhaps the issue can be clarified in future research involving a more naturalistic research set-up.

4.5.2.3 Interaction of level by version

In this section, the interaction effect of level by version on out-of-context lexical task scores will be examined. Two-way tables of means (like Table 53 for overall scores) will not be reported for individual tasks, as they take up a lot of space[20]. Such interaction effects will only be presented in graphical form. Figure 42 shows the plot of word match task scores for different dictionary versions and learner levels.

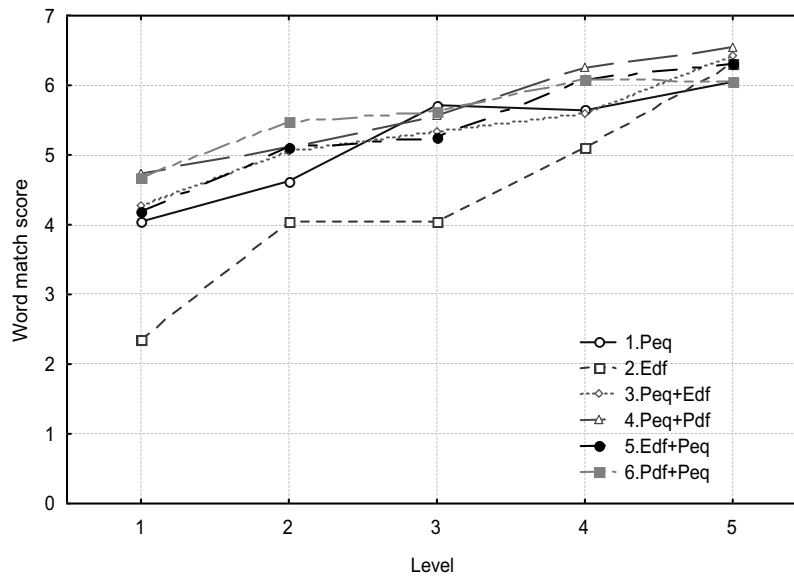

Figure 42: Interaction plot of word match task scores by level and version

The interaction effect is significant (p=0.016). The graph bears a general resemblance to that for overall scores (cf. Figure 37). Two points are worthy of note, though. The first is the slightly poorer performance of the bilingual-only dictionary vis-à-vis the other versions (except the monolingual-only version 2, of course), though level 3 appears to be unaffected by this. The second point is the break in an otherwise monotonic increase of the version-2 line between levels 2 and 3. It appears that for this level range there is no improvement in performance with the monolingual-only dictionary on the word match out-of-context task. However, if we connect the little squares for level 1 and 3, the resulting trend line

[20] Approximate values can be read off the plot; contact the author if you need exact numerical values.

will be approximately straight throughout the whole range of levels, so this slope irregularity appears to be related to a large improvement in performance between levels 1 and 2, followed by no change between levels 2 and 3.

4.5.3 Sentence-length context: lexical gap completion

Sentence-length context lexical effectiveness was assessed with two different tasks: lexical gap completion (multiple choice), and sentence translation. The lexical gap task is covered in this section, the sentence translation task is treated in section 4.5.4 below, and overall results for sentence-length context tasks are summarized under 4.5.5 below.

In the lexical gap completion task (part I of the Dictionary Effectiveness Test) subjects were instructed to pick a suitable (pseudo-)word for a sentence frame. In a GLM analysis for this task, the main effect of version is found to be significant, and level is highly significant. The version by level interaction effect, however, is not significant. A detailed ANOVA table is presented in Table 63.

Table 63: ANOVA table for sentence-length context, lexical gap completion effectiveness scores

Effect	SS	df	MS	F	p
Intercept	1365.179	1	1365.179	4051.395	<0.0001
Level	69.092	4	17.273	51.260	<0.0001
Version	4.256	5	0.851	2.526	0.0281
Level*Version	5.939	20	0.297	0.881	0.6119
Error	229.810	682	0.337		

4.5.3.1 Effect of level

Table 64 gives mean lexical gap completion task scores with 95% confidence intervals for the five learner levels.

Table 64: Lexical gap completion scores by level

Level	Score	Std. Error	-95%	+95%	N
1	0.93	0.05	0.84	1.03	145
2	1.18	0.05	1.08	1.29	131
3	1.48	0.04	1.39	1.56	175
4	1.67	0.05	1.58	1.77	154
5	1.84	0.06	1.73	1.95	107
Total	1.41	0.02	1.36	1.46	712

The scores are increasing with level, just as was the case for overall results. The confidence intervals are generally non-overlapping, so the scores for each consecutive level are clearly separated. The size of the effect of level is large for the lexical gap completion task, with level 5 scores nearly 100% higher than level 1 scores, and the effect is highly significant ($p<0.0001$).

Figure 43 presents the plot of lexical gap completion task scores as a function of learner level.

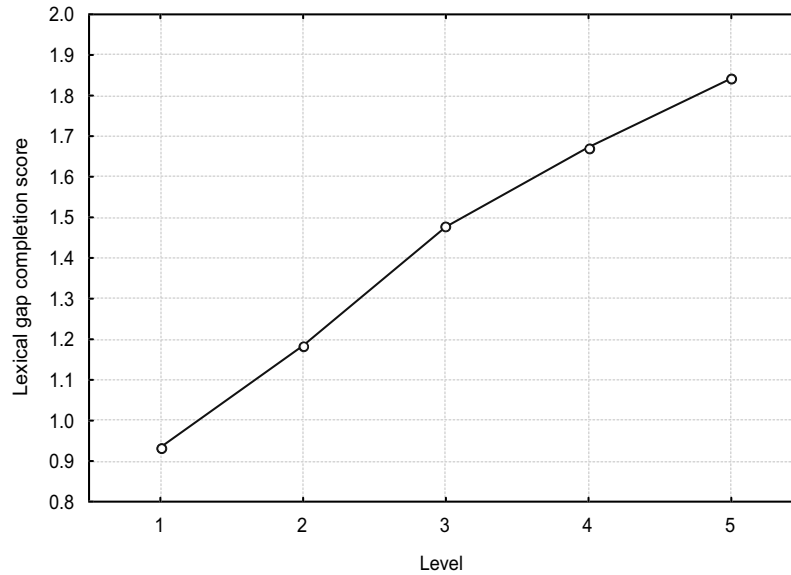

Figure 43: Plot of lexical gap completion scores by level

Note the high degree of constancy of the slope, with the plot line almost perfectly straight. This indicates that performance on this task improves steadily with learner level.

4.5.3.2 Effect of dictionary version

Mean lexical gap completion task scores with 95% confidence intervals for the six dictionary versions are listed in Table 65. Here the best-performing versions appear to be versions 1 and 2; versions 4, 5, 6 resulted in lowest scores, and version 3 earned moderate scores. The confidence intervals tend to overlap, indicating smaller differences between the versions than for overall scores. The top-scoring version outperformed the worst-performing version by only 14%, so the size of the effect of version is smaller than for overall scores, and so is the statistical significance of the effect, with p=0.03.

Table 65: Lexical gap completion scores by version

Version	Score	Std. Error	-95%	+95%	N
1.Peq	1.53	0.05	1.43	1.64	119
2.Edf	1.52	0.05	1.41	1.62	121
3.Peq+Edf	1.43	0.06	1.32	1.54	116
4.Peq+Pdf	1.34	0.05	1.24	1.45	118
5.Edf+Peq	1.37	0.06	1.26	1.48	118
6.Pdf+Peq	1.34	0.06	1.23	1.45	120
Total	1.41	0.02	1.36	1.46	712

The plot of mean lexical gap completion task scores for the six dictionary versions is given in Figure 44.

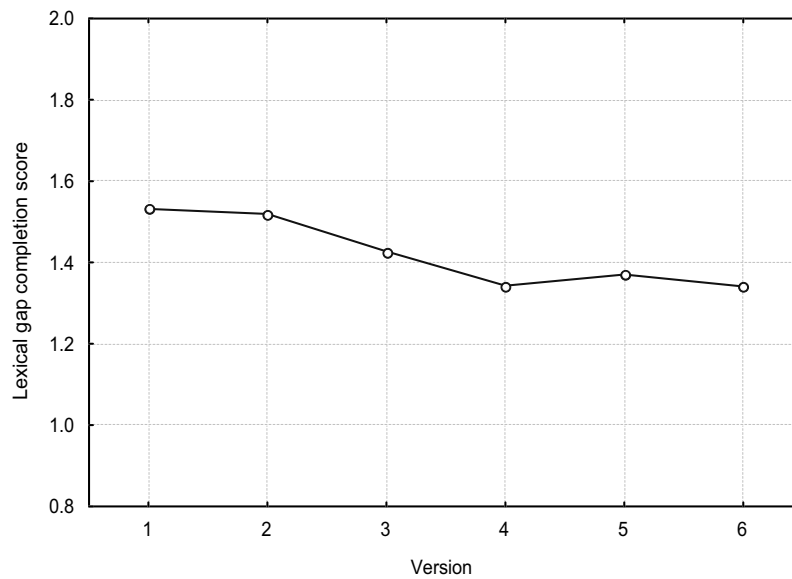

Figure 44: Plot of lexical gap completion scores by version

The shape of the plot is radically different from that for overall results (cf. Figure 36). The two striking differences are the surprisingly good performance of version 2 subjects and the low scores for versions 4, 5 and 6. This is an interesting phenomenon and it calls for a closer inspection. An analysis of individual items uncovers an unusual pattern for item I2 of the Dictionary Effectiveness Test. The breakdown of responses for this item is given in Table 66, where *scolb* was the response awarded one point on the test, the other responses did not earn any points.

Table 66: Breakdown of item I2 responses by version

Version	scolb	marbish	cullen	gerd	blank	N
1.Peq	77	37	0	2	3	119
2.Edf	85	24	6	5	1	121
3.Peq+Edf	65	48	1	1	1	116
4.Peq+Pdf	49	64	3	1	1	118
5.Edf+Peq	57	53	5	2	1	118
6.Pdf+Peq	58	58	2	0	2	120
Total	391	284	17	11	9	712

Comparing the results for the different dictionary versions, we note that version 2 (English definitions only) resulted in the highest number of correct responses for *scolb*. My preferred interpretation for this untypical helpfulness of the English definition is the presence of the lexical item *car* in the definition of the nonce word *scolb*: the same word *car* was used in the test sentence of item I2.

Further light on the issue may be thrown by examining mean I2 scores as a function of version and level (Figure 45).

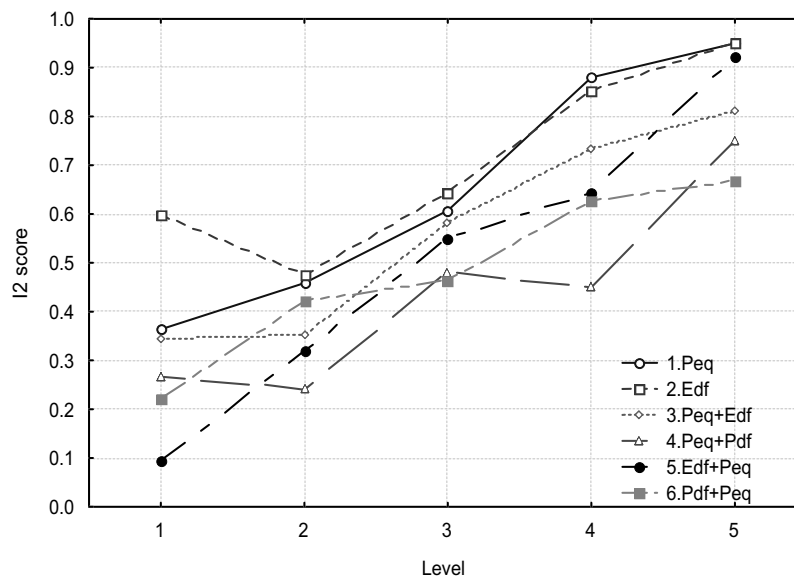

Figure 45: Item I2 scores plotted by level and version

The graph indicates that the relative advantage of version 2 is most pronounced for level 1 subjects (beginning learners of English). In fact, level-1 subjects using version 2 entries outperform level-2 subjects and are almost as good as level-3 subjects. In view of this finding, it becomes all the more likely that what we are dealing with here is the mechanical matching of individual familiar words, another variation on the kidrule strategy (Miller and Gildea 1987; Mitchell 1983b),

rather than sophisticated use of reference skills. In this particular case, the mechanical matching produces fairly satisfactory results: if we trace the line for version 2 on the plot, we find that it performs just as well as the bilingual dictionary for levels 2 through 5, and for level 1 it produces superior results because the kidrule-type familiar word matching wins out in the face of the poor language and reference skills of these beginner subjects.

Regarding the equivalent + definition versions, these complex semantic information entries appear to be related to lower performance for this item than the simplex versions 1 and 2. Let us examine the results of the Tukey HSD post hoc test for differences between the individual versions, given in Table 67.

Table 67: Post hoc test on I2 item scores for dictionary versions: Tukey HSD probabilities; significant differences are starred

Version	1	2	3	4	5	6
1		0.93651	0.69411	*0.00135	0.06396	0.06279
2	0.93651		0.15894	*0.00004	*0.00286	*0.00275
3	0.69411	0.15894		0.14701	0.78921	0.78887
4	*0.00135	*0.00004	0.14701		0.86525	0.86105
5	0.06396	*0.00286	0.78921	0.86525		1.00000
6	0.06279	*0.00275	0.78887	0.86105	1.00000	

The Tukey test finds significant differences between version 2 and each of 4, 5 and 6; and between version 1 and version 4.

For those versions that combine the Polish equivalent with the English definition, the presence of the Polish equivalent may render the English definition less salient, probably making it less likely for subjects to resort to the mechanical pattern matching strategy. The addition of the Polish definition to the equivalent also appears to be making matters worse for the users, perhaps by overloading them with too much semantic information, or distracting them in some way.

It is interesting to note that the combination of two of the most successful elements in isolation, English definition (version 2) and Polish equivalent (version 1), in this order in a single entry (version 5) results here in significantly poorer performance, suggesting that, in at least some combinations of semantic information in dictionary entries, less may be more.

4.5.3.3 Interaction of level by version

The interaction effect of level by version on lexical gap completion task is plotted in Figure 46.

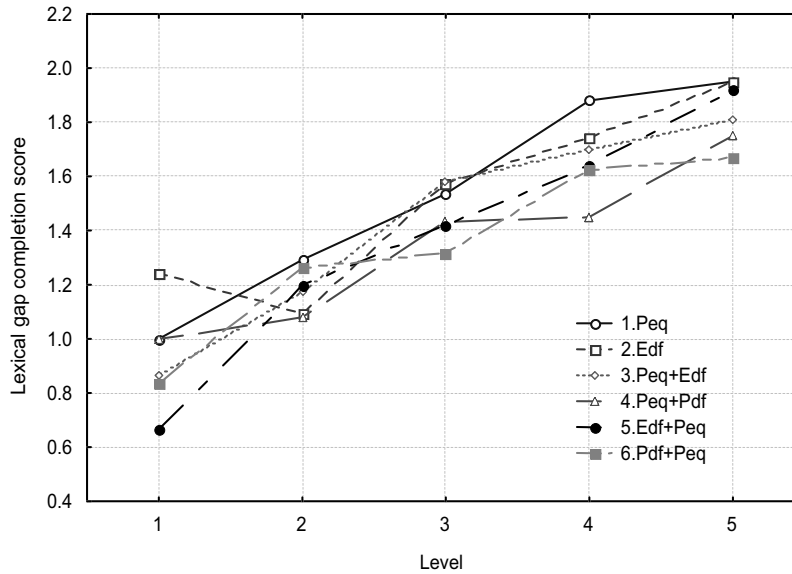

Figure 46: Interaction plot of lexical gap completion scores by level and version

Unlike for overall scores, the interaction effect here is not significant (p=0.61), which shows up in the graph as the six lines for the six versions tending to cluster together and run in parallel. The irregularity exhibited at level 1, version 2, is due to item I2 and has already been accounted for in 4.5.3.2 above.

4.5.4 Sentence-length context: translation

Part J of the Dictionary Effectiveness Test consisted of a sentence translation (English to Polish) task, which was the second lexical task with sentence-length context. Only lexical choice for target words was evaluated. GLM analysis for this task reveals the two main effects of level and version, as well as the interaction between the two, to be highly significant. A detailed ANOVA table is given in Table 68.

Table 68: ANOVA table for sentence-length context, translation lexical effectiveness scores

Effect	SS	df	MS	F	p
Intercept	13631.05	1	13631.05	15797.92	<0.0001
Level	104.44	4	26.11	30.26	<0.0001
Version	178.69	5	35.74	41.42	<0.0001
Level*Version	46.56	20	2.33	2.70	<0.0001
Error	588.46	682	0.86		

4.5.4.1 Effect of level

Mean sentence translation task scores for the five levels and their 95% confidence intervals are provided in Table 69.

Table 69: Sentence translation task scores by level

Level	Score	Std. Error	-95%	+95%	N
1	3.76	0.08	3.61	3.91	145
2	4.46	0.08	4.30	4.63	131
3	4.59	0.07	4.45	4.73	175
4	4.77	0.08	4.62	4.92	154
5	4.89	0.09	4.71	5.07	107
Total	4.49	0.04	4.41	4.57	712

As elsewhere, the scores go up with level, and the effect is highly significant ($p<0.0001$), though confidence intervals overlap for the higher levels. The influence of level on scores is clearly represented in Figure 47.

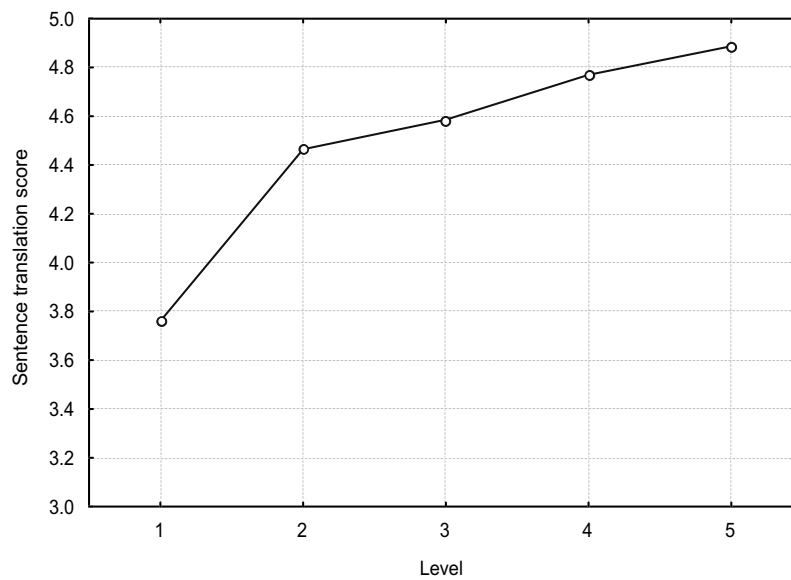

Figure 47: Plot of sentence translation task scores by level

The scores go up sharply between levels 1 and 2, then climb at a much slower rate for the remainder of the level range. The total gain in performance from level 1 to level 5 is 30%, which is less than the 50% for overall scores.

4.5.4.2 Effect of dictionary version

Mean sentence translation task scores with 95% confidence intervals for the six dictionary versions are listed in Table 70.

Table 70: Sentence translation task scores by version

Version	Score	Std. Error	-95%	+95%	N
1.Peq	4.80	0.09	4.63	4.97	119
2.Edf	3.41	0.09	3.24	3.57	121
3.Peq+Edf	4.83	0.09	4.66	5.01	116
4.Peq+Pdf	4.71	0.09	4.54	4.88	118
5.Edf+Peq	4.79	0.09	4.61	4.96	118
6.Pdf+Peq	4.43	0.09	4.25	4.60	120
Total	4.49	0.04	4.41	4.57	712

The size of the effect of version on this task is relatively high, with the top-scoring version being 42% better than the worst-performing version. This is higher than the 28% for overall scores. The effect is highly significant ($p<0.0001$). Pairwise post hoc tests (Tukey HSD) reveal version 2 to be significantly different from every other version, but no other pairs of means are significantly different.

The means for the six versions are plotted in Figure 48.

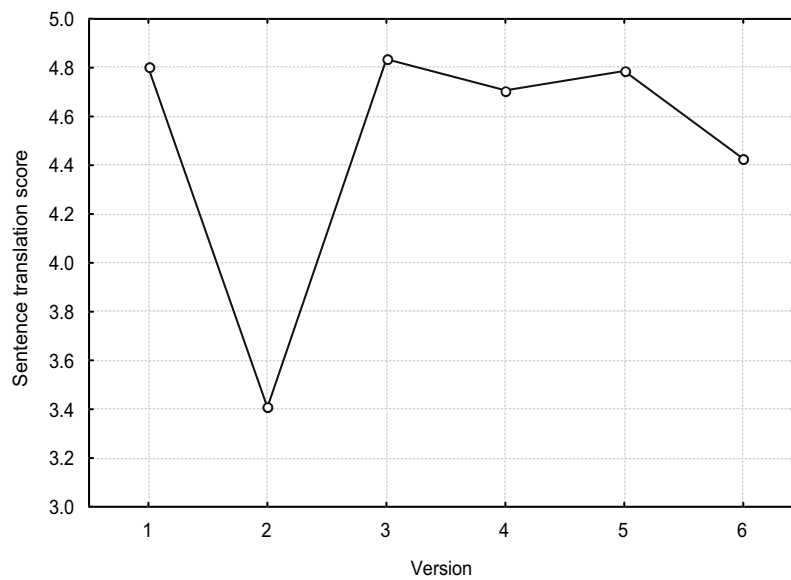

Figure 48: Plot of sentence translation task scores by version

The graph displays a marked dip at version 2, showing the English-only version to be inferior for this task. Of the other versions, version 6 with Polish definitions preceding Polish equivalents earned a lower mean score than the other versions with Polish equivalents, but these differences did not reach significance in the corrected post hoc test. The shape of the graph is in general similar to that for overall scores, but version 2 lags behind the others more on this task than overall.

4.5.4.3 Interaction of level by version

Figure 49 illustrates the interaction effect of level and version on the sentence translation task scores. The effect is highly significant ($p<0.0001$).

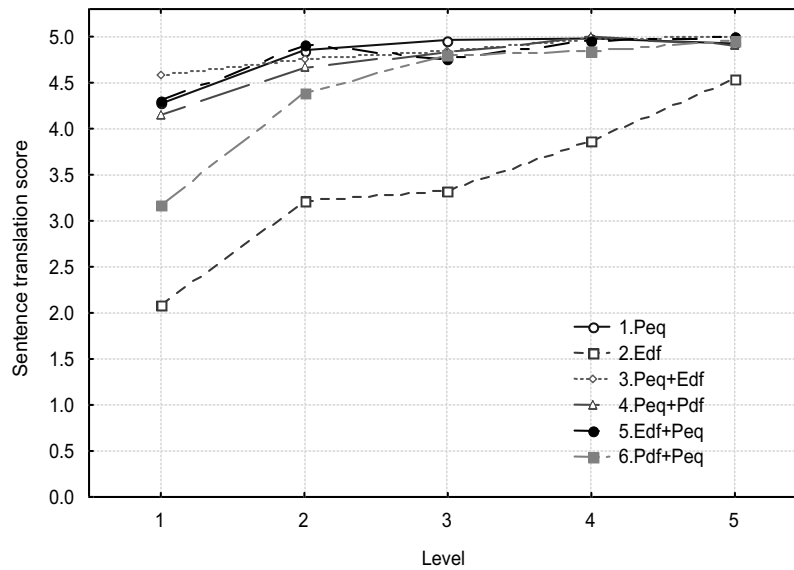

Figure 49: Interaction plot of sentence translation task scores by level and version

The graph clearly shows the disadvantage of version 2, which for this task persists even at the advanced level 5. The other versions cluster close together, with the exception of version 6 for beginners, where the delaying of the Polish equivalent past the Polish definition apparently causes problems for low-proficiency subjects. I suspect that those beginning users are able to skip the English definition and get right to the equivalent in version 5, but get stalled if the definition is in Polish.

4.5.5 Sentence-length context: overall

The results from sentence-length context lexical tasks I and J are conflated here. Effects of level (with version controlled) and version (with level controlled) on lexical effectiveness scores are revealed to be highly significant by GLM analysis. The interaction effect of version by level, however, is not significant, though it does approach the conventional 5% threshold with $p=0.07$. Detailed Univariate ANOVA results are presented in Table 71.

Table 71: ANOVA table for sentence-length context overall lexical effectiveness scores

Effect	SS	df	MS	F	p
Intercept	23623.82	1	23623.82	17096.18	<0.0001
Level	332.68	4	83.17	60.19	<0.0001
Version	160.07	5	32.01	23.17	<0.0001
Level*Version	41.73	20	2.09	1.51	0.0709
Error	942.40	682	1.38		

4.5.5.1 Effect of level

Table 72 gives the mean sentence-length context effectiveness scores for the five levels, and their 95% confidence intervals.

Table 72: Sentence-length context effectiveness scores by level

Level	Score	Std. Error	-95%	+95%	N
1	4.69	0.10	4.50	4.89	145
2	5.65	0.10	5.45	5.85	131
3	6.06	0.09	5.89	6.24	175
4	6.44	0.10	6.25	6.63	154
5	6.73	0.12	6.50	6.95	107
Total	5.90	0.05	5.79	6.01	712

Scores get better with increasing level, and the confidence intervals for the individual levels tend to be non-overlapping. The score for the highest level is 43% higher than the score for level 1, and the effect is highly significant ($p<0.0001$).

Figure 50 gives a plot of mean score versus level on the sentence-length context tasks.

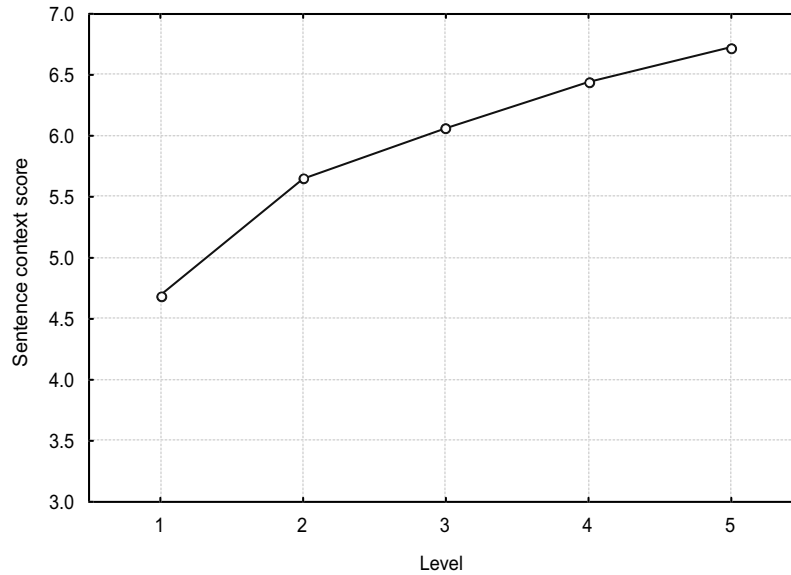

Figure 50: Plot of sentence-length context effectiveness scores by level

The scores increase monotonically with level, with the fastest rate of increase occurring between levels 1 and 2.

4.5.5.2 Effect of dictionary version

Mean scores with their 95% intervals for sentence-length context tasks broken down by dictionary version are given in Table 73.

Table 73: Sentence-length context effectiveness scores by version

Version	Score	Std. Error	-95%	+95%	N
1.Peq	6.33	0.11	6.12	6.54	119
2.Edf	4.93	0.11	4.71	5.14	121
3.Peq+Edf	6.26	0.11	6.04	6.48	116
4.Peq+Pdf	6.05	0.11	5.84	6.27	118
5.Edf+Peq	6.16	0.11	5.93	6.38	118
6.Pdf+Peq	5.77	0.11	5.55	5.99	120
Total	5.90	0.05	5.79	6.01	712

The effect of version on sentence-length context effectiveness scores is highly significant ($p<0.0001$). The highest mean score (version 1) outperforms the lowest score (version 2) by 28%, so the effect size is the same here as for overall scores.

A graphical plot of score means by version is presented in Figure 51.

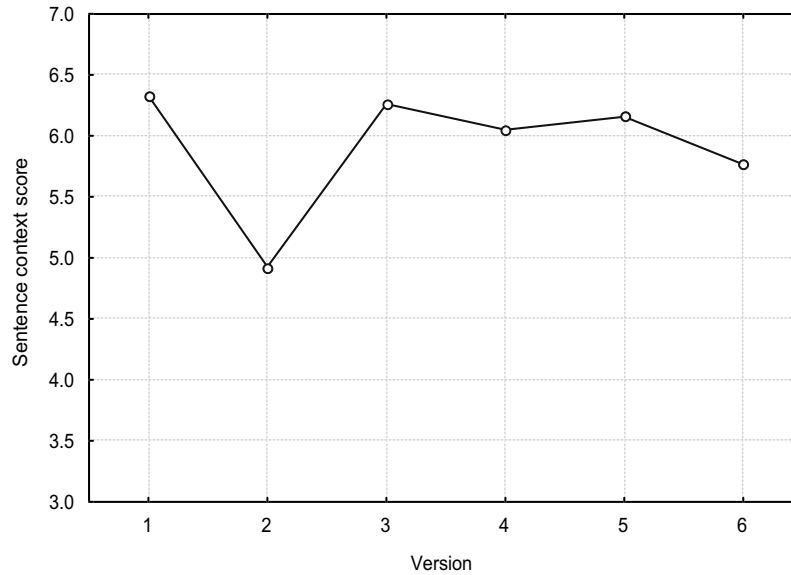

Figure 51: Plot of sentence-length context effectiveness scores by version

There is a distinct dip in performance at version 2 (English definition). Also, there appears to be some disadvantage to preceding the Polish equivalent with a definition, as indicated by the Tukey HSD post hoc test (Table 74), which shows version 6 to be significantly less effective than version 1 (Polish equivalent only).

Table 74: Post hoc test on sentence-length context scores for dictionary versions: Tukey HSD probabilities; significant differences are starred

Version	1	2	3	4	5	6
1		*0.00002	0.97836	0.13919	0.74605	*0.02269
2	*0.00002		*0.00002	*0.00002	*0.00002	*0.00002
3	0.97836	*0.00002		0.52712	0.98868	0.16917
4	0.13919	*0.00002	0.52712		0.88959	0.98689
5	0.74605	*0.00002	0.98868	0.88959		0.50723
6	*0.02269	*0.00002	0.16917	0.98689	0.50723	

4.5.5.3 Interaction of level by version

The interaction effect of level by version on sentence-length context effectiveness scores is plotted in Figure 52.

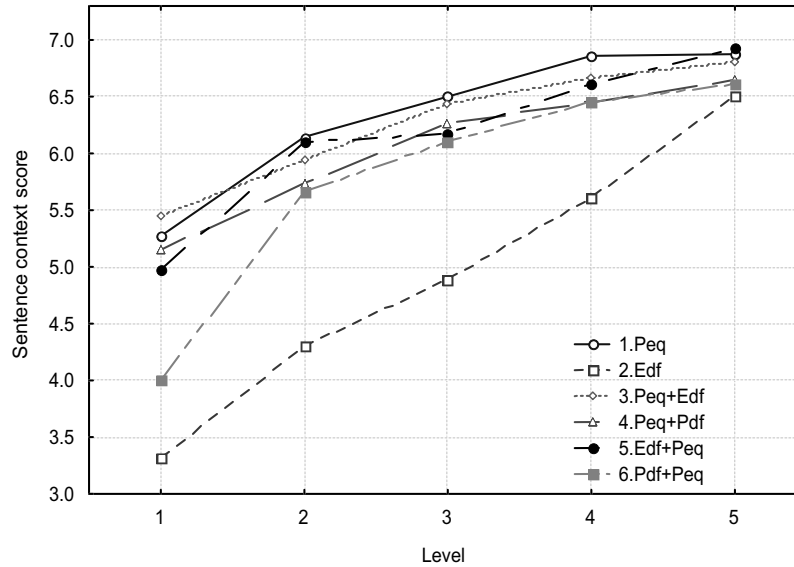

Figure 52: Interaction plot of sentence-length context effectiveness scores by level and version

The interaction effect of level by version is not significant for sentence-length task scores (p=0.07). There is little line crossing in the plot, and the lines tend to run in parallel.

4.5.6 Text context

Text context lexical effectiveness of the dictionary entries was assessed through the text translation task (part K of the Dictionary Effectiveness Test), in which again subjects' lexical choices were evaluated. As shown in Table 75, the two main effects (level and version) are highly significant, but the interaction effect (level by version) is not significant.

Table 75: ANOVA table for text context lexical effectiveness scores

Effect	SS	df	MS	F	p
Intercept	3717.338	1	3717.338	5430.854	<0.0001
Level	176.102	4	44.025	64.319	<0.0001
Version	47.241	5	9.448	13.804	<0.0001
Level*Version	14.832	20	0.742	1.083	0.3620
Error	466.819	682	0.684		

4.5.6.1 Effect of level

Table 76 gives mean text context effectiveness scores with their 95% confidence intervals for the five learner levels.

Table 76: Text context effectiveness scores by level

Level	Score	Std. Error	-95%	+95%	N
1	1.47	0.07	1.33	1.60	145
2	2.10	0.07	1.96	2.24	131
3	2.56	0.06	2.44	2.69	175
4	2.73	0.07	2.59	2.86	154
5	2.88	0.08	2.72	3.04	107
Total	2.34	0.04	2.27	2.42	712

The means increase with level, and the score for level 5 is almost twice the score for level 1. The effect is very large, and highly significant ($p<0.0001$).

Figure 53 plots the mean scores on the text context task as a function of learner level.

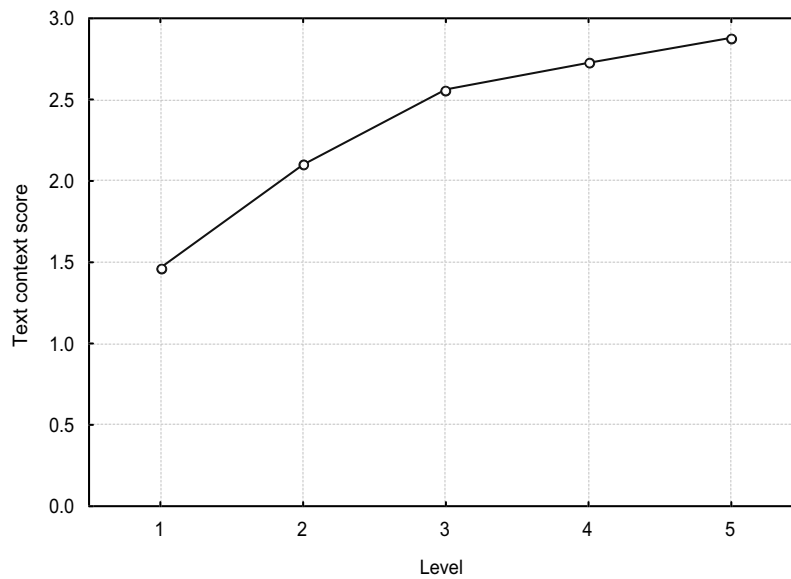

Figure 53: Plot of text context effectiveness scores by level

The slope of the plot is steeper for the lower levels (from 1 to 3), and somewhat flatter for the higher levels (from 3 to 5), indicating that an increase in level is related to a larger gain in performance for the less advanced learners, and a smaller gain for the more advanced learners.

4.5.6.2 Effect of dictionary version

Mean text context effectiveness scores for the six dictionary versions and their 95% confidence intervals are given in Table 77.

Table 77: Text context effectiveness scores by version

Version	Score	Std. Error	-95%	+95%	N
1.Peq	2.52	0.08	2.38	2.67	119
2.Edf	1.82	0.08	1.67	1.97	121
3.Peq+Edf	2.57	0.08	2.41	2.72	116
4.Peq+Pdf	2.51	0.08	2.36	2.66	118
5.Edf+Peq	2.43	0.08	2.27	2.58	118
6.Pdf+Peq	2.23	0.08	2.07	2.38	120
Total	2.34	0.04	2.27	2.42	712

Again, version 2 produced the lowest mean score, and was outperformed by the top-scoring version 3 by 42%, so the size of the effect of version is substantial. The effect is highly significant ($p<0.0001$).

A plot of mean scores by version is shown in Figure 54.

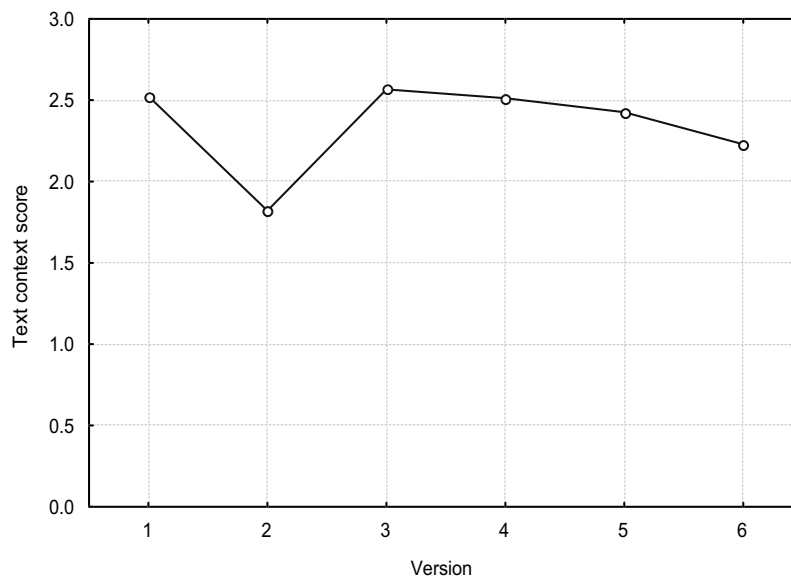

Figure 54: Plot of text context effectiveness scores by version

Apart from the clearly inferior performance of version-2 subjects, there appears to be a disadvantage to the Polish equivalent being delayed in the presentation of semantic information within a dictionary entry. This tendency is more marked when the intervening element is a definition in Polish (version 6), compared to a definition in English (version 5), and version 6 mean score is indicated to be significantly lower than version 1 score by the Tukey post hoc test (Table 78).

Table 78: Post hoc test on text context scores for dictionary versions: Tukey HSD probabilities; significant differences are starred

Version	1	2	3	4	5	6
1		*0.00002	0.99618	0.27884	0.68773	*0.02084
2	*0.00002		*0.00002	*0.00002	*0.00002	*0.00002
3	0.99618	*0.00002		0.59634	0.93194	0.09238
4	0.27884	*0.00002	0.59634		0.98700	0.91328
5	0.68773	*0.00002	0.93194	0.98700		0.55276
6	*0.02084	*0.00002	0.09238	0.91328	0.55276	

4.5.6.3 Interaction of level by version

The interaction effect of level by version on text context scores is illustrated in Figure 55.

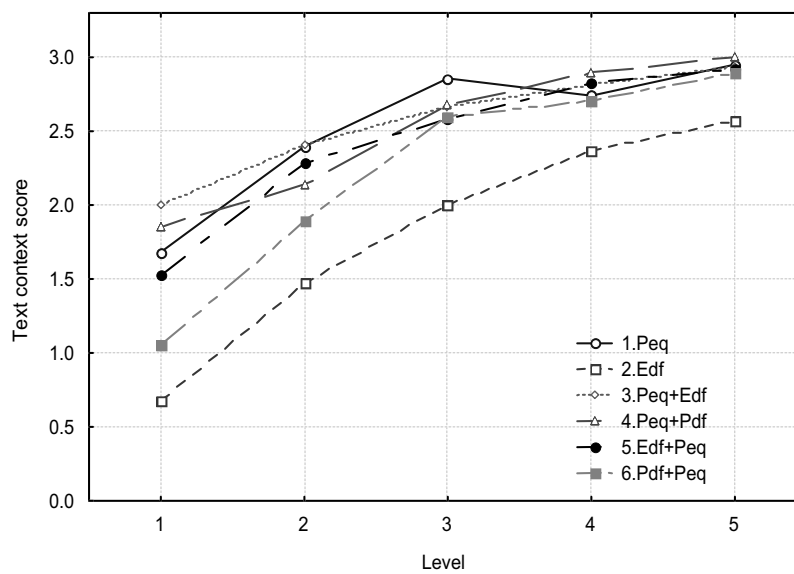

Figure 55: Interaction plot of text context effectiveness scores by level and version

This effect is not significant for text context scores (p=0.36). The lines are near-hyperbolic in shape, and tend to run in parallel to each other.

4.5.7 Position and language of definition

Four of the six versions of the mini-dictionary used in the study (versions 3, 4, 5 and 6) contained two components with semantic information: definition and Polish equivalent. The Polish equivalents were all the same for the four versions (and identical to version 1 as well), but the definition components varied along two dimensions: in terms of its language and in terms of its position relative to

the Polish equivalent. As shown in Table 12 on page 76, the language of the definition takes up two values: Polish or English. With regard to its relative position, the definition may precede the Polish equivalent, or follow the Polish equivalent. These two positions will be referred to briefly as *first* and *last*, respectively.

Here, I will briefly present an analysis of overall dictionary effectiveness scores in terms of these two underlying variables of definition position and language, as well as the learner level variable. Only the four dictionary versions for which the variables can be defined are included in the present analysis (see 4.5.2.2 for further discussion). A complete factorial ANOVA table for this analysis is presented in Table 79.

Table 79: ANOVA table for level, definition position and definition language, overall scores

Effect	SS	df	MS	F	p
Intercept	85934.93	1	85934.93	16374.79	<0.0001
Level	1368.36	4	342.09	65.18	<0.0001
DefPos	18.67	1	18.67	3.56	0.0599
DefLang	3.47	1	3.47	0.66	0.4169
Level*DefPos	38.88	4	9.72	1.85	0.1178
Level*DefLang	10.00	4	2.50	0.48	0.7533
DefPos*DefLang	5.10	1	5.10	0.97	0.3248
Level*DefPos*DefLang	6.66	4	1.67	0.32	0.8663
Error	2372.10	452	5.25		

As it turns out, none of the effects except the level effect (already extensively covered above) is statistically significant. The only other effect that approaches significance is DefPos, the position of definition. The score means for this effect with their 95% confidence intervals are listed in Table 80.

Table 80: Breakdown of overall dictionary effectiveness scores by position of definition

DefPos	Score	Std. Error	-95%	+95%	N
Last	14.19	0.15	13.89	14.49	234
First	13.78	0.16	13.47	14.08	238
Total	13.92	0.13	13.66	14.17	472

There is a tendency for a definition following the equivalent to produce somewhat higher scores, but, as noted above, this tendency does not reach significance. In any case, the difference between the two means is only about 3%.

A plot of means in Figure 56 illustrates how the position of definition effect interacts with learner level.

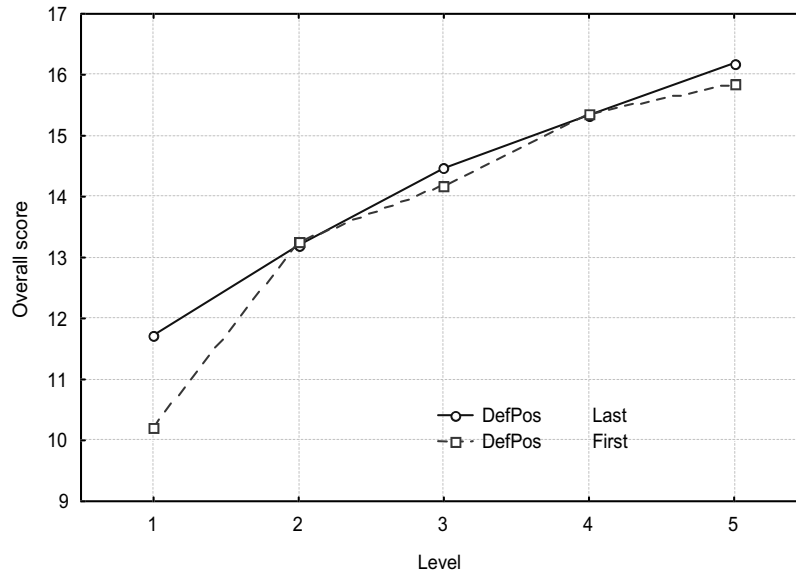

Figure 56: Plot of overall score by definition position and level

The plot in Figure 56 reveals that a major share of the position effect comes from level 1 subject scores, where the difference related to the two positions is quite substantial, about 15%. In fact, contrast analysis test of significance for level 1 turns out to be highly significant ($F_{(1, 462)}$=9.99; p=0.0017). It would appear that the definition, when placed in front of the Polish equivalent, may make it more difficult for beginners to locate the equivalent, the single most helpful element in the entry.

Finally, for a more complete picture, let us look at the effect of the language of definition, even though this effect does not even approach significance. The means and confidence intervals are given in Table 81.

Table 81: Breakdown of overall dictionary effectiveness scores by language of definition

DefLang	Score	Std. Error	-95%	+95%	N
Pol	13.88	0.19	13.50	14.26	238
Eng	13.96	0.17	13.62	14.30	234
Total	13.92	0.13	13.66	14.17	472

As one might expect, the means are hardly affected by the language of the definitions, the difference between the two mean scores being less than 0.6%. The mean scores are plotted against level in Figure 57.

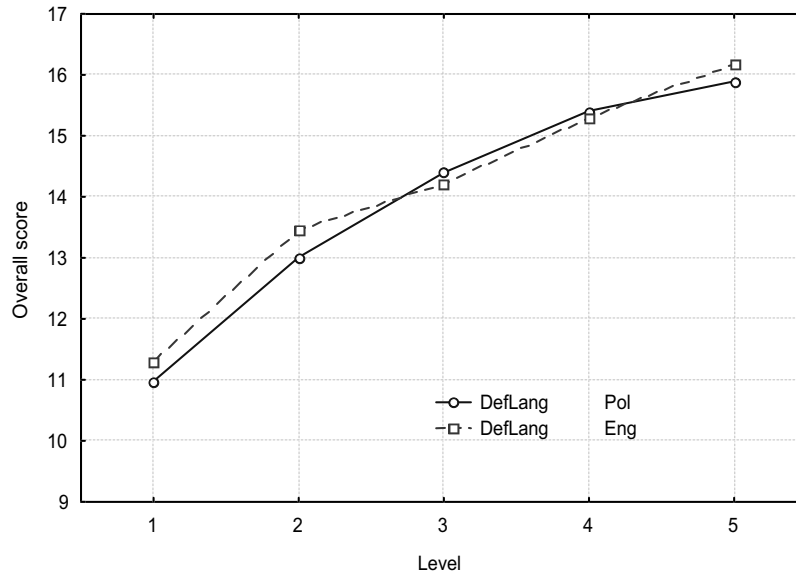

Figure 57: Plot of overall score by definition language and level

The two lines run very close together, which is consistent with the high p-level for this interaction effect (p=0.75). A likely interpretation of this almost perfect indifference of scores to the language of the definition is that nobody actually reads the definition when the Polish equivalent is available. The only way that a definition seems to matter here is that it gets in the way of the equivalent for lower-level subjects. The slightly higher scores for the English definition at levels 1 and 2 in Figure 57 may be a reflection of the English definition being easier to separate from the Polish equivalent, and thus easier to skip when present in front of the equivalent: we have already seen data patterns indicating that skipping the definition to get to the equivalent may present something of a problem to the beginning learners in our sample.

4.5.8 Conclusion

Throughout this section (4.5), I have presented and interpreted the results of the Dictionary Effectiveness Test, which assessed the relative success of subjects at different levels, using six versions of the dictionary to complete a range of lexical tasks with varying amount of context. I attempt to visualize the complex interaction of the effects of level and dictionary version under varying amount of context in Figure 58.

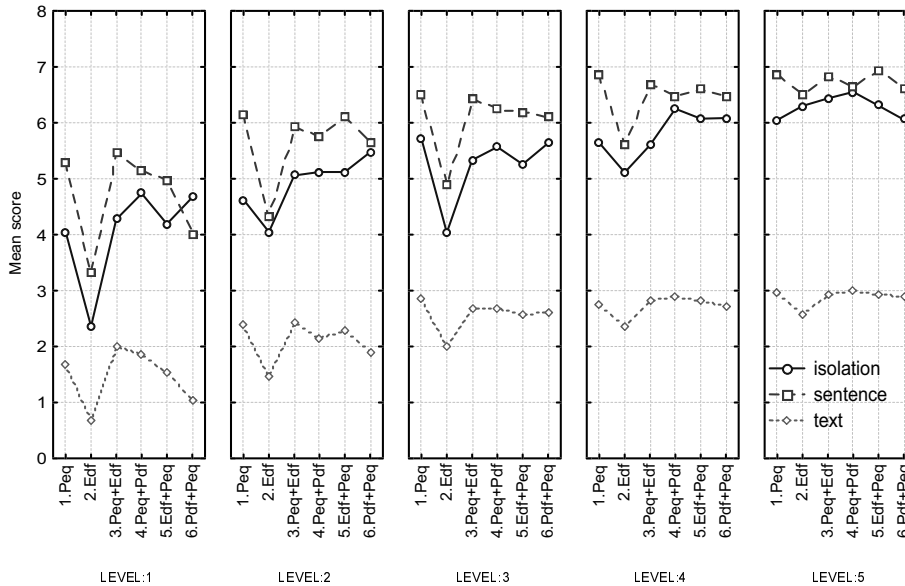

Figure 58: Dictionary effectiveness scores by context, version and level

The five segments of the graph represent subjects at the five levels. Within each segment, the six dictionary versions are represented on the abscissa, and the ordinate indicates test scores. The three types of context are represented with a continuous line (words in isolation), dashed line (sentence-length context), and dotted line (text context).

The main findings of this section can be summarized as follows:

- Scores go up with level. This effect is consistent and expected: learner level probably correlates fairly strongly with both reference skills and language skills, both of which would normally make dictionary consultations more effective.
- Version 2 is consistently the least effective of all. Clearly, the absence of a Polish equivalent is a major factor negatively affecting dictionary effectiveness.
- The disadvantage of version 2 is more serious for the lower-level users. As learners progress, so do their foreign language skills, and their strict reliance on Polish equivalents becomes less of a necessity.
- The disadvantage of version 2 affects all types of contexts: words in isolation, sentence-length context, and text context.
- Inserting a definition in front of a Polish equivalent affects performance negatively for bottom-level subjects in sentence and text contexts. For beginners, dictionary entries have to be kept simple: a surplus of information distracts and confuses the users. However, an English definition

may be slightly easier to skip in this case, because it is more distinct from the Polish equivalent, being in a different language.

- When a definition is present as well as a Polish equivalent, there is some advantage in the definition being in Polish rather than in English for words in isolation. This advantage disappears at level 5.
- Dictionary users, and especially beginners, appear to resort sometimes to a strategy of mechanical matching of lexical items between source context and dictionary definition.

4.6 Morphological accuracy: plural inflection

The Dictionary Effectiveness Test was primarily meant to test the lexical effectiveness of the different versions of dictionary entries, which was the focus of section 4.5 above. The test has not been designed to systematically measure grammatical accuracy. However, there was one lexical item in task K, *remestians*, which was given in the plural form, but all dictionary versions (as would the great majority of actual dictionary products in paper format) employed a singular citation form for entry headword, here *remestian*. In this section, I would like to examine to what extent subjects were able to transfer the plural feature to the Polish translation of the lexical item, which reflects their performance on the item *remestians* with the aid of the mini-dictionary entries provided. As above, the effects of dictionary version and level on subjects' performance will be examined.

At this point a decision needs to be taken on how to treat those cases where subjects have not succeeded in translating the target word. The first option is to exclude such cases from the analysis. The second option is to treat such cases as failures to effectively extract the morphological information from the dictionary. Here, the second option was selected, because the first option would lead to misleading results, as will be shown briefly towards the end of section 4.6.2. A 2-way ANOVA table for the second option is presented in Table 82.

Table 82: ANOVA table for level and version, plural inflection score

Effect	SS	df	MS	F	p
Intercept	149.0293	1	149.0293	755.8425	<0.0001
Level	38.7239	4	9.6810	49.0997	<0.0001
Version	0.2905	5	0.0581	0.2947	0.9159
Level*Version	3.1164	20	0.1558	0.7903	0.7271
Error	134.4698	682	0.1972		

Only the level effect is significant; the main effect of version and the interaction of level by version do not even approach significance, with F ratios being less than 1. One should note at this juncture that the alternative analysis (first option

with the missing cases excluded) produces roughly similar results, that is with just the level effect being significant.

I will now take up the three effects in turn, starting with the significant effect of level.

4.6.1 Effect of level

Mean plural inflection scores with their 95% confidence intervals for the five levels are given in Table 83.

Table 83: Plural inflection score by level

Level	Score	Std. Error	-95%	+95%	N
1	0.14	0.04	0.06	0.21	145
2	0.29	0.04	0.21	0.37	131
3	0.47	0.03	0.40	0.54	175
4	0.57	0.04	0.49	0.64	154
5	0.89	0.04	0.80	0.97	107
Total	0.46	0.02	0.42	0.49	712

The scores increase monotonically with level. They can be interpreted here as mean success rate, i.e. a score of 0.5 represents a situation where about half of the subjects in the group have supplied the correct plural inflectional marker. The effect is highly significant ($p<0.0001$), and its size is very substantial, with the mean for level 5 being about six times the mean for level 1. The size of this effect is much higher than for any *lexical* effectiveness measures noted thus far. This should not surprise us: after all, an ability to correctly interpret the English plural ending might be expected to correlate highly with foreign language proficiency; more so than lexical-semantic interpretation of nonce words. In the alternative analysis excluding the missing cases, the effect size is, understandably, smaller, but still impressive, with an over threefold increase in scores.

Plural inflection score by level is plotted in Figure 59.

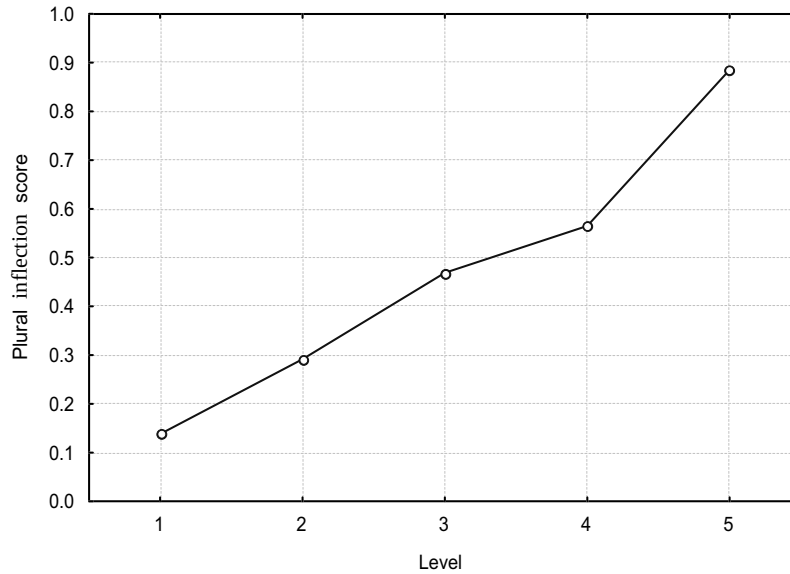

Figure 59: Plot of plural inflection score by level

The plot reveals that the slope of the line is steep and fairly constant, with a slight slowdown for level 4, which is compensated for at level 5.

4.6.2 Effect of dictionary version

Mean plural inflection scores for the six dictionary versions with their 95% confidence intervals are given in Table 84.

Table 84: Plural inflection score by version

Version	Score	Std. Error	-95%	+95%	N
1	0.44	0.04	0.36	0.52	119
2	0.45	0.04	0.37	0.53	121
3	0.48	0.04	0.39	0.56	116
4	0.46	0.04	0.38	0.54	118
5	0.50	0.04	0.42	0.59	118
6	0.49	0.04	0.40	0.57	120
Total	0.46	0.02	0.42	0.49	712

The values for the different versions are close, and all confidence intervals overlap, which is consistent with the high p-level (p=0.92). The corresponding plot of means is presented in Figure 60.

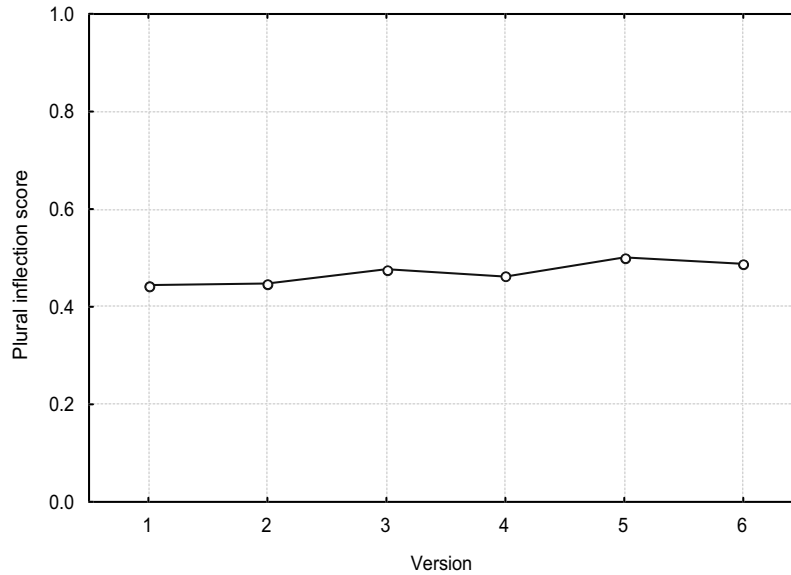

Figure 60: Plot of plural inflection score by version

It is interesting to note that the two lowest mean scores were achieved with the simplex entry versions, with just a single semantic information element. However, we should bear in mind that the main effect of version does not approach significance, and the difference is very slight. Also, a contrast analysis for versions 1 and 2 versus combined versions 3-6 produces a nonsignificant result ($F_{(1, 682)}$=1.02, p=0.31). We conclude, then, that dictionary version does not appear to influence plural inflection scores in any way.

The above analysis has included those cases where translations of the target word are missing, interpreting this as evidence of failure. An alternative analysis would be to exclude such cases from analysis, as explained in 4.6 above. For this type of analysis, Table 84 would be replaced by Table 85.

Table 85: Plural inflection score by version with missing cases excluded

Version	Score	Std. Error	-95%	+95%	N
1	0.46	0.05	0.37	0.56	104
2	0.62	0.06	0.51	0.73	78
3	0.54	0.05	0.44	0.63	97
4	0.52	0.05	0.43	0.61	97
5	0.54	0.05	0.45	0.64	103
6	0.58	0.06	0.48	0.69	98
Total	0.56	0.02	0.52	0.60	577

If we compare the values in the two tables, we note the higher score for version 2 in Table 85, which now becomes the top-scoring version. This difference is baf-

fling at first. The explanation for this effect is found in the N rows of the two tables: note that the number of contributing subjects falls dramatically, much more so than for any other version. This is because so many version 2 subjects were unsuccessful with this item, which is the true reason for the inflated success rate for version 2 on the alternative analysis. It is not so much, as Table 85 might appear to imply, that version 2 subjects were more readily able to supply the plural inflection, but rather the less successful users with access to version 2 of the dictionary were unable to supply the translation at all, whereas with the other dictionary versions subjects were at least able to get the lexical item right, but missed out on the plural feature. Based on this rationale, the alternative analysis is rejected as misleading, and the analysis including the missing cases is preferred. In any event, the version effect is not significant on the alternative analysis either.

4.6.3 Interaction of level by version

The interaction effect of level by version on plural inflection scores is illustrated in Figure 61.

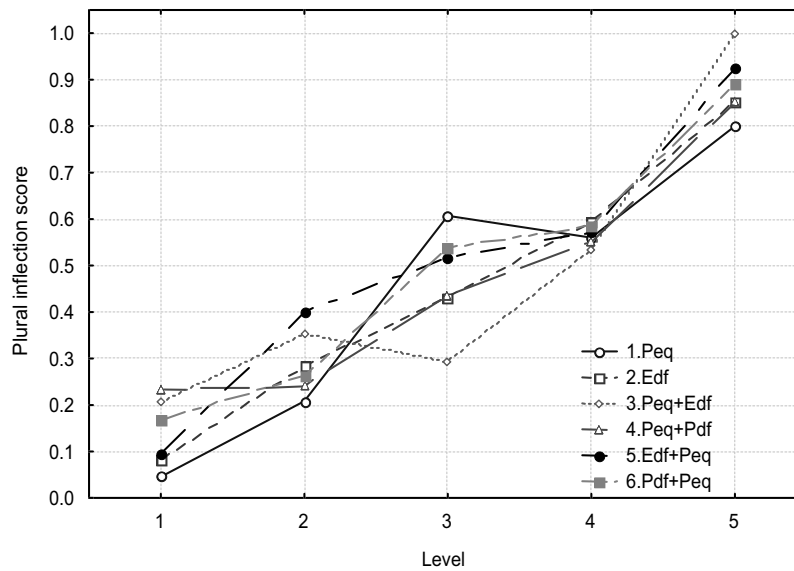

Figure 61: Interaction plot of plural inflection score by level and version

This effect is not significant (p=0.73). It may be visually intriguing that the six versions appear to converge so closely at level 4, and diverge at level 3, but contrast analysis does not find any statistical significance here either. It may be concluded that dictionary version does not affect plural inflection scores, either on its own, or in combination with learner level.

4.7 Sense ordering

In his pioneering study, Tono (1984) discovered that his dictionary users tended to select the first sense listed in a dictionary entry unless clear indications were present that this sense was not appropriate. Although investigating the effect of sense ordering was not the primary aim of this study, some polysemous items were present in the test dictionaries, and target senses were balanced for order, meaning that in polysemous entries, half of the target (correct) responses were included as initial (early) senses, the other half as late senses. The items were not systematically balanced for difficulty; to do that would have involved a further proliferation of dictionary versions, with orders of senses switched for particular items. Such a design solution had been considered but rejected, because the design was already fairly complex. Nevertheless, these details should be borne in mind when evaluating the results presented in this section.

The effect of sense ordering is here operationalized as the so-called early-late differential score. This measure was computed as the difference between the total score on the early-sense target items and the total score on the late-sense target items, and should reflect the relative advantage of early placement (the first sense) of target information within the entry versus late placement (later senses, here usually the second, and last, sense). A differential score of zero would thus suggest a zero effect of sense ordering, a positive differential score would indicate an advantage of early sense placement within an entry (first sense in a polysemous entry), whereas a negative score would suggest an advantage of late placement. However, the two sets of items are different items, rather than the same items with senses switched, so the early and late items are not perfectly balanced for difficulty. Despite efforts to avoid any such bias, it is possible that the early items are on average more – or less – difficult than the late items.

The overall early-late differential score turns out to be –0.05, which is very close to zero. In fact, the 95% confidence interval includes the value of zero (see the bottom row of Table 87 below), so we cannot reject the null hypothesis of the score being generally unaffected by sense placement, although the reservations discussed in the previous paragraphs still hold.

While the overall early-late differential score may be unreliable because of the lack of systematic control over item difficulty, the influence of level and version on the differential should be less sensitive to this design problem. I will now proceed with a two-way factorial GLM ANOVA on differential scores for level and version – just as I have done for absolute dictionary effectiveness measures. An ANOVA table for this analysis is presented in Table 86.

Table 86: ANOVA table for level and version, early-late differential scores

Effect	SS	df	MS	F	p
Intercept	0.9373	1	0.9373	1.2252	0.2687
Level	8.3846	4	2.0962	2.7401	0.0278
Version	11.1264	5	2.2253	2.9089	0.0131
Level*Version	12.9984	20	0.6499	0.8496	0.6527
Error	521.7225	682	0.7650		

The analysis reveals significant main effects of level and version. The interaction of level by version is not significant (and so is the intercept, which confirms the lack of general sense ordering effect). The two main effects and the interaction will be taken up in turn below.

4.7.1 Effect of level

The values of early-late differential score for each level with their 95% confidence intervals are listed in Table 87.

Table 87: Early-late differential scores by level

Level	Early-Late Differential	Std. Error	-95%	+95%	N
1	0.10	0.07	-0.04	0.25	145
2	-0.10	0.08	-0.25	0.05	131
3	-0.20	0.07	-0.33	-0.07	175
4	-0.04	0.07	-0.18	0.10	154
5	0.05	0.09	-0.12	0.22	107
Total	-0.05	0.03	-0.11	0.02	712

All values are very close to zero, and their 95% confidence intervals include zero, with the exception of the level 3 differential score. This is clearly visible in the plot in Figure 62.

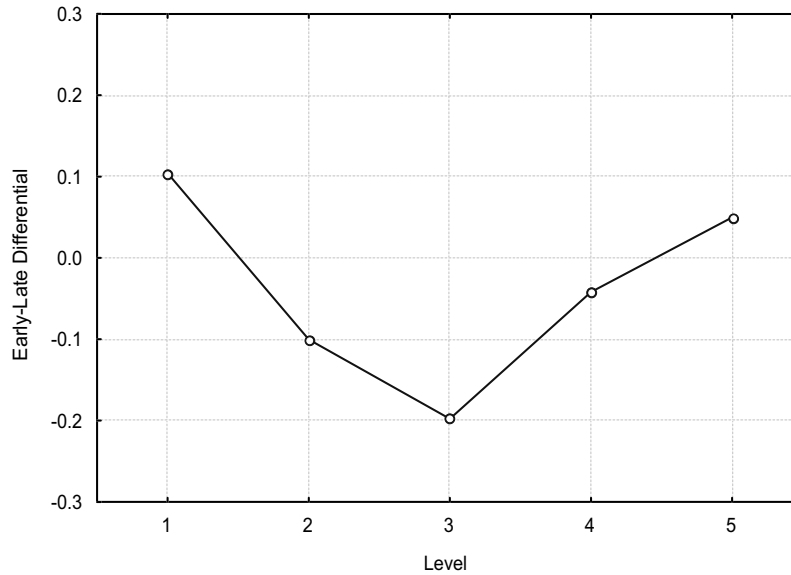

Figure 62: Plot of early-late differential scores by level

Judging from the plot, there is a tendency for the differential values to be lowest for the intermediate level subjects, suggesting a slight but significant (p=0.03) relative advantage of late-placed senses for this range of levels. It is difficult to offer a definitive interpretation for such a relationship, but one possibility might be that learners at this level range are especially meticulous in reading dictionary entries, and are careful to read polysemous entries through to the end. Such an interpretation is highly speculative, though, and would have to be verified by further evidence, preferably from direct-observation studies.

4.7.2 Effect of version

Mean early-late differential scores for the six dictionary versions with their 95% confidence intervals are given in Table 88.

Table 88: Early-late differential scores by version

Version	Early-Late Differential	Std. Error	-95%	+95%	N
1	-0.04	0.08	-0.20	0.12	119
2	-0.22	0.08	-0.38	-0.07	121
3	0.01	0.08	-0.16	0.17	116
4	-0.05	0.08	-0.21	0.11	118
5	0.20	0.08	0.04	0.37	118
6	-0.12	0.08	-0.28	0.05	120
Total	-0.05	0.03	-0.11	0.02	712

The values cluster around zero, and there is relatively little variation, although not as little as for level-related variation. Versions 1, 2, 4, and 6 show negative values, of which 1 and 4 are small. Version 3 shows a small positive value, version 5, a larger positive value. Versions 2 and 5 are the two whose confidence intervals do not include a value of zero. The values are plotted in Figure 63.

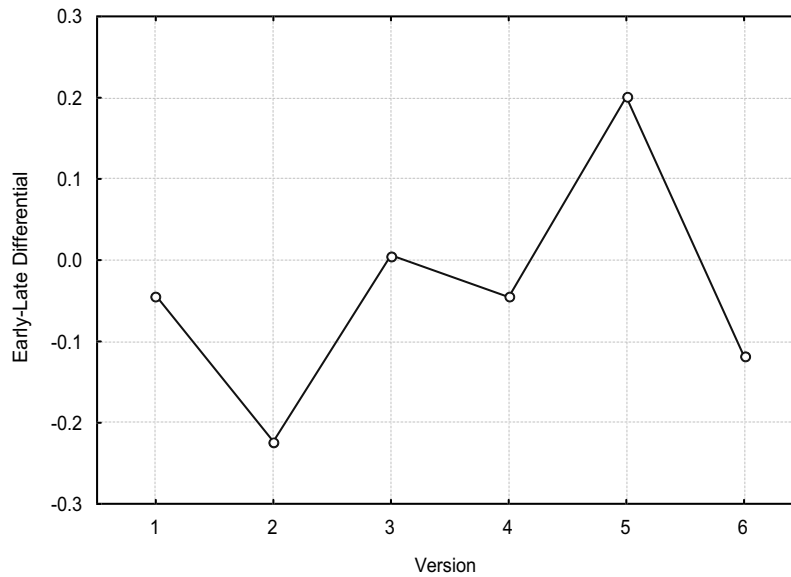

Figure 63: Plot of early-late differential scores by version

The two dictionary versions that differ most from each other are 2 and 5. It is interesting to note that both these versions include English definitions within their semantic information. In version 5, however, the English definition is followed by the Polish equivalent for each sense. It would appear that such an entry structure, common in bilingualized dictionaries, is related to the greatest degree of preference for the early sense in polysemic entries. In view of the difference between versions 2 and 5, an interpretation suggests itself that the typical users in the sample exhibit different behaviour patterns when presented with the two versions, having to read the English definition (as there is nothing else) in version 2, but skipping over it in version 5. Here, however, users are apparently unwilling to skip over the second English definition to locate another Polish equivalent, which would account for their relative preference for the early senses. Quite the reverse seems to be the case in version 2, which exhibits the highest degree of preference for the late senses. Here, users struggling with the English definition in the absence of the Polish equivalent, appear to be making a special effort to continue reading the entry in order to make the most of the rather unhelpful English-language entry. The above hypothetical scenario for version 5 seems to get further support from results for version 6. Version 6 differs from version 5 only in that the definition that precedes the Polish equivalent is in Polish, rather than

in English. This difference alone results in a marked relative preference for the early senses in version 5 vis-à-vis version 6. It would thus appear that either users find it harder to skip over the English definition, which is more difficult to distinguish from the equivalent (also in Polish), or else they actually read the Polish definition rather than skip over it.

4.7.3 Interaction of level by version

The interaction effect of level by version on early-late differential scores did not turn out to be significant (p=0.65). However, let us review the pattern of the early-late differential scores plotted against the two factors, as shown in Figure 64, in the hope that it will allow us a more accurate interpretation of the main effects of level and version.

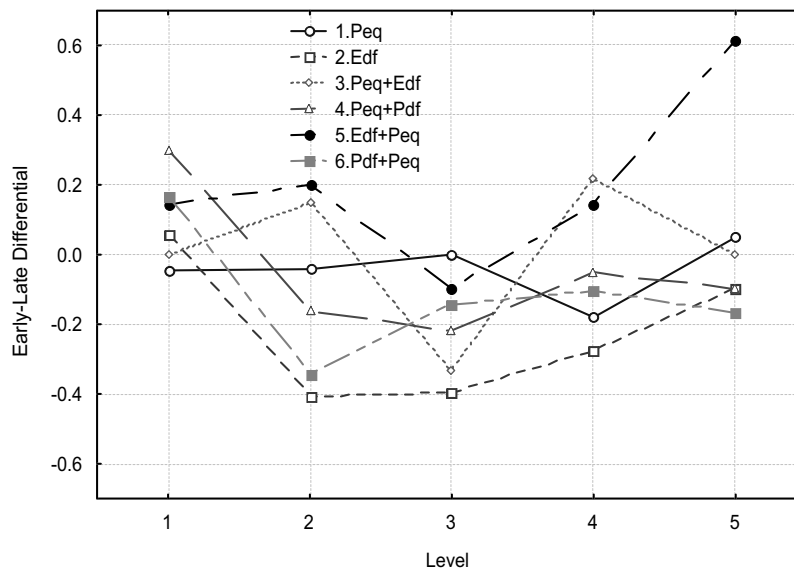

Figure 64: Interaction plot of early-late differential scores by level and version

In terms of the effect of level, we observe that the dip at level 3 noted in 4.7.1 above is mostly the result of versions 3 and 5, that is the two versions that combine Polish equivalents with English definitions. This combination, in whatever order, appears to make intermediate learners, in contrast to learners of other levels, more likely to look at the later senses.

As regards version 5 and its preference for early senses, we find level 5 subjects to be the chief contributors to this effect. This suggests that advanced subjects are rather unwilling to read beyond the first sense of an entry when working with a semi-bilingual dictionary (English definition followed by Polish equivalent), more than for any other dictionary type, and more than any lower-level subjects. The latter difference appears to be especially surprising and difficult to account for. One possible explanation is that advanced learners have been "con-

ditioned" to look at the early entries through their exposure to modern learners' dictionaries, which almost universally order senses based on frequency. This explanation is questionable on two grounds. Firstly, if that were the case, one would expect the effect to be mimicked for monolingual dictionaries, version 2, but that does not appear to be the case. Secondly, advanced learners using learners' dictionaries would normally find it necessary to delve deeper into polysemous entries, because they would often be familiar enough with the most frequent senses that tend to come early on in an entry. Clearly, this issue calls for further study.

4.7.4 Conclusion

This study found no indication of a preference for early senses over late senses, as the differential score between early senses and late senses averaged to zero across the whole sample. However, the effect might be masked as the difficulty of items was not balanced in the present design. Also, the polysemous entries that were used in the study were rather short, up to three senses in length. The results might have been different with longer entries.

The influence of level and version on the early-late differential score was examined. The two main effects were found to be significant, but not the interaction effect. A small relative preference for late senses was noted with intermediate-level subjects, as opposed to beginners and advanced learners, and test dictionary versions 3 and 5 were the strongest contributors to this effect. A preference for early senses was revealed for users of version 5, especially advanced ones, suggesting that English definitions may be blocking access to the (more effective) Polish equivalents in semi-bilingual dictionaries. In monolingual dictionaries (version 2), a tendency was found for users to persevere to the later sense. This last finding is consistent with the results of a think-aloud study by Wingate (2002: 113-115), who found that reading only the beginning of an entry was a problem for bilingual dictionaries, but not for monolingual dictionaries. On the other hand, Neubach and Cohen (1988) found monolingual dictionaries to be affected by the prefernce for early sense. Perhaps the difference between versions 5 and 2 is here due to the advanced users expecting monolingual entries to take more time reading and consequently allowing themselves more time when faced with such entries.

5. Conclusions and suggestions

5.1 Review of the findings

In the present study, I have tried to probe a range of questions related to receptive dictionary use by Polish learners of English by combining questionnaire surveys and controlled experimental tests with the same subjects.

The average Polish learner appears to be using a bilingual dictionary *a few times a week*. This holds for Polish-English as well as English-Polish dictionaries, though, as in Tomaszczyk (1979), it is the latter that are used somewhat more frequently. In contrast, monolingual dictionaries are used very rarely by learners at all levels except the highest. The more intensive use of monolingual dictionaries by most advanced learners is not accompanied by a decreased use of bilingual dictionaries; on the contrary, an increase in the frequency of use is observed in bilingual dictionaries at the highest proficiency level, though relatively speaking not as dramatic as for monolingual dictionaries. The patterns of frequency of use at different levels suggest that monolingual dictionaries at the highest level supplement bilingual dictionaries rather than supplanting it.

The present findings are in general agreement with the results of those studies which noted an increase in the frequency of dictionary use with level (Hatherall 1984; Knight 1994; Wingate 2002), rather than those that observed a reverse tendency (Atkins and Varantola 1998a; Tomaszczyk 1979). This discrepancy in the findings of previous studies could be due to the different proficiency ranges covered in those investigations, since this study, covering a very broad span of proficiency levels, found very little overall increase in the consultation frequency with level except at the highest proficiency ranges. Therefore, if a given study did not have sufficiently advanced subjects in the sample, an increase may not have been noted. Furthermore, the effect of level did not affect all types of dictionaries in the same way. The rise in the use of monolingual dictionaries is only noticeable, though very pronounced, at the top level 5. As far as bilingual dictionaries go, the use of Polish-English dictionaries begins to increase at the intermediate level, while the frequency of consultation of English-Polish dictionaries exhibits a more or less steady rise throughout the whole proficiency range. The concomitant rise in the frequency of use of monolingual dictionaries and Polish-English dictionaries may suggest that monolingual dictionaries may be used in conjunction with Polish-English dictionaries in productive tasks, the bilingual dictionary being used to locate the English equivalent, and monolingual to obtain examples and guidance on usage, collocation and grammar.

The investigation of dictionary preferences revealed that while most subjects in the sample named their dictionary of first choice, only half of them gave a dictionary of second choice. As expected, the identification of specific titles was often impossible, as users' recollection of the details of the dictionaries they used

tends to be fragmentary and unreliable. The analysis was conducted in such a way as to make maximum use of even this incomplete information. Only 9% of all responses gave monolingual dictionaries as their preferred dictionaries, with the remaining 91% of responses naming bilingual dictionaries. It was primarily highest-proficiency learners who named monolingual dictionaries. Further, monolingual dictionaries, if named at all, tend to have been given as second-choice rather than first-choice dictionaries, which confirms their secondary role for advanced students.

The average rating that learners gave their dictionaries was *good*, a 4 on a scale of 1 to 5, which suggests that the overall degree of satisfaction with dictionaries was fairly high. Of the dictionaries most frequently named, users valued the Collins-BGW and Langenscheidt dictionaries the highest. There were substantial differences in the ratings across the different titles and brands. These details may be of interest to lexicographers, publishers, teachers and dictionary users. Learners' ratings do not appear to be influenced much by their proficiency level.

As in Tomaszczyk's study (1979), our learners gave higher ratings to monolingual than to bilingual dictionaries. In view of the poor performance of monolingual dictionaries on the Dictionary Effectiveness Test, this effect is best seen as reflecting the positive image of the monolingual dictionary. Also, it may be easier for learners' to register problems with bilingual dictionaries than with monolingual dictionaries, since monolingual entries tend to be more difficult to understand and so would be more likely to be given the benefit of the doubt.

As expected, dictionaries named as first choice received higher ratings from their users than dictionaries of second choice. This is not true across the whole range of proficiency levels, though, with level-4 learners evaluating second-choice dictionaries higher, on average, than first-choice dictionaries. A partial explanation of this rather puzzling finding is afforded by a three-way analysis, which revealed that lower ratings for first-choice bilingual dictionaries are responsible for this effect. It is difficult to provide a principled interpretation of why this should be so, but perhaps the answer lies in the specific context of dictionary use at this level, such as when access to dictionary is motivated by circumstance more than personal preference.

The types of information that learners seek from dictionaries tended to group into two clusters, as established by factor analysis. A high-frequency cluster was formed by three types of lexicographic information: meaning, English equivalents and Polish equivalents. The dominance of these three information types suggests that the basic reference needs are those of finding word meanings (meaning and Polish equivalents) and finding word forms (English equivalents and, to a lesser degree, Polish equivalents). The other types of information investigated in this study were synonyms, style and register, collocation, sentence structure, part of speech, and pronunciation; they were sought less often and formed the peripheral information type cluster. At the most advanced level, how-

ever, the gap between the central and peripheral clusters narrows, as advanced learners' interest is aroused in the more "exotic" categories of information. Nevertheless, advanced learners do not seem to lose their interest in the central information types.

The Dictionary Effectiveness Test, the core part of this study, has provided a number of insights into the effectiveness of dictionary versions offering different types of semantic explanation. Looking at learners of different levels, we find a consistent pattern of lexical task scores increasing with level. This is understandable and expected, as learner level certainly goes hand in hand with reference skills and language skills, both of which are likely to improve the effectiveness of dictionary consultation. More interestingly, the monolingual version 2, which offered just the English definition by way of meaning provision, was consistently the least effective of all – for words in isolation, in sentence-length context, and in text context alike. The absence of a Polish equivalent had a strongly negative effect on scores. It was the lower-level learners that the monolingual dictionary placed at the greatest disadvantage: as learners' skills progress, their reliance on Polish equivalents should become less essential. The negative effect of the missing Polish equivalent did however persist all the way to the highly advanced learners at level 5. Adding a definition in front of a Polish equivalent had a depressing effect on scores of lower-level subjects in sentence and text contexts. This negative effect was observed for definitions in English and in Polish alike. English definitions fared somewhat better than Polish definitions here, although they certainly did not help, as advocates of the semi-bilingual dictionary type might hope. My best interpretation of the data is that those entries which combined two ways of meaning provision turned out to be too crowded and thus confusing to learners at the lower levels, when compared with the simpler bilingual entry. The fact that adding an English definition in front of the Polish equivalent lowered the effectiveness scores to a lesser extent than adding a Polish definition in the same position can be explained in terms of the English definition obscuring the Polish equivalent less than a Polish definition does, and especially so for Polish speakers. Further support for this interpretation comes from the results of the two configurations with definitions following the Polish equivalent, which produced better scores, as the equivalent was immediately available, and the following material could be easily ignored. For words in isolation, when a definition was presented along with a Polish equivalent, Polish-language definition worked better than English definition, although this difference vanished at level 5. An analysis of individual test items reveals evidence of a mechanical matching strategy of similar lexical items between source context and dictionary definition. This behaviour is similar to the kidrule strategy noted in the literature, and was here found to be particularly typical of the beginner-level users.

The single point of grammar investigated here, namely the accuracy in the provision of the plural inflection marker, revealed the rather predictable effect of level: the higher the learner's proficiency level, the more accurate the provision

of the plural marker. The size of this effect was very strong. It turned out that the accuracy of the plural marker remained unaffected by dictionary version, with all six versions performing similarly. There was also no interaction effect of level and version.

Unlike in Tono (1984), no evidence was found in this study of a generalized preference for early senses over late senses. The early-late differential score that was adopted as a measure of this preference turned out to be very close to zero for the sample as a whole. However, all polysemous entries used in this study were relatively short (three senses at the most), as would be real entries for infrequent words, for which our pseudo-words posed. Although no general preference was noted, effects of level and version on the early-late differential score turned out to be significant. Intermediate-level learners, as opposed to both beginners and advanced learners, exhibited a slight preference for late senses, especially when using versions 3 and 5. On the other hand, version 5 correlated with a preference for early senses, especially for advanced learners. This effect suggests that English definitions may to some extent block access to the more effective Polish equivalent in semi-bilingual entries. Interestingly, users of the monolingual version 2 did go on to the later senses. Perhaps this is a question of the users being prepared to allow themselves more time when reading the (longer) monolingual entries.

5.2 Lexicographic recommendations

This study shows that learners' clear and consistent preference for bilingual dictionaries over monolingual dictionaries has a good practical justification: the bilingual dictionary type simply offers the greatest help in lexical tasks, independent of the amount of context. Perhaps learners are instinctively aware of the high effectiveness of bilingual entries, even though, when asked, they award higher ratings to monolingual dictionaries.

The advantage of bilingual dictionaries slowly diminishes with growing proficiency level. Wingate commented in the following way on the level that learners have to be at in order to benefit from monolingual dictionaries:

> Even though this finding has to be validated by further research, it illustrates a major deficit in the discussion about the effectiveness of bilingual versus monolingual dictionaries (...). The prevailing argument that monolingual dictionaries are more effective for learners can be misleading as long as it is not clear what proficiency level learners must have reached. However, no previous study has pinpointed the stage in the process of language acquisition when learners start to be able to understand monolingual definitions. This research provides evidence that this ability is not reached by all intermediate learners. (Wingate 2002: 229)

According to the findings of the present study, this ability is only reached by learners at a very high proficiency level, but even at this high level, the monolin-

gual entry still fails to hold an absolute advantage over the bilingual entry. When Wingate's findings are interpreted, two important facts must be borne in mind. Firstly, actual dictionaries were used in the comparison, and the monolingual dictionaries used in the study were learners' dictionaries that held a number of advantages over the traditional bilingual dictionary. One of the two monolingual dictionaries used by Wingate was a commercial learner's dictionary (*Langenscheidts Grosswörterbuch Deutsch als Fremdsprache,* LGDaF), and the other was carefully put together by Wingate, who used all the available insights of modern lexicography to make it maximally comprehensible to intermediate learners, but had little regard for space-saving requirements (Wingate's new entries were longer than the LGDaF ones by about 56%). Secondly, and perhaps more importantly, the bilingual dictionary used in Wingate's study was a German-English dictionary with the native language of the subjects being Chinese, so it did not explain the foreign language (German) words in the users' native language (Chinese), but rather in what can at best be called their second language (English). If these two facts are taken into account, and especially in view of the findings of this study, we must question the validity of the recommendation so popular amongst educators of the presumed superiority of the monolingual dictionary. There is hardly any empirical evidence available to support that supposed superiority, and what little relevant evidence is available, points to the bilingual dictionary as the more effective dictionary for reception. The present study provides further evidence of this type.

Wingate suggests that the bilingualized dictionary might be the ideal solution for her subjects, since it "could provide learners with the security of equivalents, while at the same time introducing them to monolingual word explanations" (2002: 230). Not much can be said about the use of semi-bilingual dictionaries by Polish learners, because their popularity in Poland is still minimal, as shown by the questionnaire responses. However, it should be stressed that the suggested advantages of the semi-bilingual entry were not confirmed by the results of the present study. The Dictionary Effectiveness Test did not find semi-bilingual dictionary entries to be significantly more helpful than bilingual ones, though they did perform much better than monolingual dictionaries thanks to the presence of the Polish equivalent.

The findings of this study suggest that there is much to be said for the simplicity of the entry. Providing too much information appears to have a confusing effect on users, particularly the lower-level learners, probably through making it harder for them to quickly locate the Polish equivalent.

On the other hand, advanced learners display an interest in the peripheral types of lexicographic information such as synonyms, style and register, collocation, sentence structure, part of speech, and pronunciation. Of these secondary lexical reference needs, traditional bilingual dictionaries have typically provided systematic information on only two of the above: part of speech and pronunciation. And yet there is no principled reason why bilingual dictionaries should not

offer to those advanced, more demanding users a richer range of information about lexical items that are often found in the modern monolingual learners' dictionaries. Such calls have in fact been made in the past (e.g. Atkins 1996; 2002; Tomaszczyk 1983; Worsch 1999), and they are beginning to be heard at last, as indicated by the recent publishing of the new *Longman słownik współczesny angielsko-polski, polsko-angielski* (LSW), which seems to do just this, by offering a range of lexical information that a modern monolingual dictionary would offer, but without the space-consuming and often distracting definitions.

5.3 Paper versus electronic dictionaries

The present study used paper copies of dictionaries for the testing procedure. Some comments are due, though, on the possible relevance of this study's finding to electronic dictionaries, which are no doubt gaining in popularity and are expected to grow so as to dominate the dictionary scene (Koren 1997; Meijs 1990; Nesi 2000a; de Schryver 2003; Sharpe 1995).

It has been claimed (Cumming, Cropp and Sussex 1994) that with the move from paper to online dictionaries, restrictions of space would disappear. That, however, is a simplification at best. While storage space may indeed become irrelevant, there are still severe restrictions as to how much information can be displayed at a time. In fact, even the best currently available display devices are still easily beaten by the old-fashioned printed paper in terms of visual resolution. So space-saving issues will still be with us[21] for at least as long as the visual modality is primarily used for information transfer from dictionary to user (as opposed to some kind of direct uplink to the brain which, popular as it may be in science fiction literature and films, is not yet anywhere on the drawing board).

The results of this study may be largely relevant for electronic dictionaries for as long as on-screen display of dictionary entries basically simulates paper presentation (as they still do, de Schryver 2003). Conversely, on-screen presentation of entries has been used to investigate some aspects of dictionary use which were then generalized to paper dictionaries (Hulstijn 1993; Knight 1994; Nesi and Meara 1991). The use of computer forms in this capacity has much to offer to the researcher by way of convenience, including a potential to log responses automatically, thus obviating the need for the laborious paperwork and keyboarding at the data entry stage, as well as allowing "unobtrusive observation" (Hulstijn 1993: 139). However, the equivalence of the two forms (on-screen and paper) should not be taken for granted, as Laufer (2000) found significant and substantial differences in word recall scores between marginal paper glosses and on-screen pop-up window glosses. By analogy, one should be cautious about extending the findings of the experimental section of this study to electronic dictionaries.

[21] See also Corréard (2002).

5.4 Limitations of this study

The claims that can be made by this study are limited by the methodology used. Potential problems with questionnaires were pointed out by Hatherall (1984). The look-up situations in this study were not entirely naturalistic in the sense that users were encouraged to consult the enclosed dictionary entries by emphasizing the target words and by explicit instructions to use the dictionary, so the study did not exactly replicate the conditions under which subjects might naturally wish to consult such items. The purpose behind adopting this methodology was to counteract the insidious problem of dictionary underuse (see 2.6.1 above), which is potentially destructive to any otherwise carefully designed study of dictionary use. The target items were not actual English words but rather nonce words pretending to be English words, in an attempt to control for lexical preknowledge. Real book-size dictionaries were not used, but rather artificially constructed mini-dictionaries. Surely, this must have influenced the look-up behaviour of subjects with regard to the dictionary macrostructure, making it easier for them to get to the target entry. However, as the focus of the study was entry content, or microstructure, rather than macrostructure, the facilitation of entry access was in fact a desirable design feature as part of an attempt to offset the dictionary underuse problem discussed in 2.6.1.

Even though the sample used in the study covered a broad range of educational institutions, no systematic sampling frames were used to obtain a sample representative of "Polish learners of English," so there is bound to be some areal and demographic bias in the structure of the present sample. The methodological decision here was motivated by pragmatic considerations, since systematic sampling would have been unacceptably expensive; as well as by fundamental reasons, since it would be difficult to systematically discriminate between "Polish learners of English" and "Polish non-learners of English" in a principled, non-arbitrary way, given that a lot of language learning occurs incidentally in today's world of satellite television and the Internet.

The present study investigated receptive dictionary use, so no direct claims can be made regarding production in the foreign language with the use of the dictionary. Also, the study did not address any long-term learning or retention effects that may follow receptive dictionary use, and which are theoretically and practically distinct (Scholfield 1999: 30) from the more immediate goals of getting lexical support from dictionary consultation while engaged in lexical problem solving tasks.

No single study can, of course, be reasonably expected to answer all the interesting questions. Dictionary user research is a relatively new area, and it still faces more questions than it can offer answers. The net was cast fairly broadly in this study, but it is hoped that it will provide some inspiration for more closely focused research. I believe it will take many years of study by numerous scholars using a variety of methods before we can honestly say that the highly complex

and private activity of dictionary use has been described satisfactorily. I do hope, though, that this study has made a valuable contribution to this description.

References

Dictionaries

The American Heritage Dictionary of the English Language (AHD2). Second College Edition. Berube, Margery S. (ed.). Boston: Houghton Mifflin. 1982.

The American Heritage English as a Second Language Dictionary (AHESLD). Boston: Houghton Mifflin. 1998.

Cambridge International Dictionary of English (CIDE). Procter, Paul (ed.). Cambridge: Cambridge University Press. 1995.

Collins COBUILD English Language Dictionary (COBUILD). Sinclair, John (ed.). London - Glasgow: Collins. 1987.

Collins Cobuild Students' Dictionary. Bridge Bilingual English-Portuguese (CobuildBBEP). Sinclair, John (ed.). London: HarperCollins. 1995.

Collins English-Polish Dictionary (BGWEP). Fisiak, Jacek (ed.). Warsaw: HarperCollins-BGW. 1996.

Collins English-Polish, Polish-English Dictionary (BGWCD). Fisiak, Jacek (ed.). Warsaw: HarperCollins-BGW-YDP. 1997.

Collins Gem English Dictionary (CGED). McLeod, William T. (ed.). London - Glasgow: Collins. 1987.

Collins Polish-English Dictionary (BGWPE). Fisiak, Jacek (ed.). Warsaw: HarperCollins-BGW. 1996.

Dictionnaire de la comptabilité (DC). Sylvain, Fernand (ed.). Toronto: ICC. 1982.

Langenscheidts Grosswörterbuch Deutsch als Fremdsprache (LGDaF). Götz, Dieter, Günther Haensch, and Hans Wellmann (eds.). Berlin - München: Langenscheidt. 1993.

Langenscheidts New College German Dictionary, German-English, English-German (LNCGD). Brough, Sonia (ed.). Berlin - München: Langenscheidt. 1995.

Longman Active Study Dictionary (LASD). Harlow: Longman. 1983.

Longman Dictionary of Contemporary English (LDOCE3). Third edition. Summers, Della (ed.). Harlow: Longman. 1995.

Longman Language Activator (LLA). Rundell, Michael (ed.). Harlow: Longman. 1993.

Longman Lexicon of Contemporary English (LLCE). McArthur, Tom (ed.). Harlow: Longman. 1981.

Longman słownik współczesny angielsko-polski, polsko-angielski (LSW). Fisiak, Jacek, Arleta Adamska-Sałaciak, Mariusz Idzikowski, Ewelina Jagła, Michał Jankowski, and Robert Lew (eds.). Harlow: Pearson Education. 2004.

Podręczny słownik angielsko-polski, polsko-angielski (LongPodr). Fisiak, Jacek, Arleta Adamska-Sałaciak, Mariusz Idzikowski, and Michał Jankowski (eds.). Harlow: Longman. 1999.

Random House Word Menu (RHWM). Glazier, Stephen (ed.). New York: Random House. 1992.

Wielki słownik polsko-angielski (STAG). Stanisławski, Jan (ed.). 1969.

Other works

Al-Ajmi, Hashan

1992 The use of monolingual English and bilingual Arabic-English dictionaries in Kuwait: An experimental investigation into the dictionaries used and reference skills deployed by university students of arts and science. Ph.D. diss., University of Leeds.

Albus, Deb, John Bielinski, Martha Thurlow, and Kristin Liu

2001 The effect of a simplified English language dictionary on a reading test (LEP Projects Report 1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from the WWW on 21 March 2004: http://education.umn.edu/nceo/OnlinePubs/LEP1.html

Al-Kasimi, Ali M.

1984 The interlingual/translation dictionary. In *Lexicography: Principles and practice*, Hartmann, Reinhard R.K. (ed.), 153-162. London: Academic Press.

Altarriba, Jeanette, and Katherine M. Mathis

1997 Conceptual and lexical development in second language acquisition. *Journal of Memory and Language* 36: 550-568.

Anderson, Peter, and Richard C. Freebody

1981 Vocabulary knowledge. In *Comprehension and teaching: Research reviews*, Guthrie, John T. (ed.), 77-117. Newark DE: International Reading Association.

Ard, Josh

1982 The use of bilingual dictionaries by ESL students while writing. *ITL Review of Applied Linguistics* 58: 1-27.

Atkins, Beryl T. Sue

1985 Monolingual and bilingual learners' dictionaries: A comparison. In *Dictionaries, lexicography and language learning*, Ilson, Robert F. (ed.). Oxford: Pergamon Press.

1996 Bilingual dictionaries - past, present and future. In *EURALEX '96 Proceedings*, Gellerstam, Martin, Jerker Jarborg, Sven-Göran Malmgren, Kerstin Noren, Lena Rogström, and Catarina Röjder Papmehl (eds.), 515-546. Göteborg: Department of Swedish, Göteborg University.

2002 Bilingual dictionaries: Past, present and future. In *Lexicography and natural language processing. A festschrift in honour of B.T.S. Atkins*, Corréard, Marie-Hélene (ed.), 1-29. EURALEX.

———— (ed.)

1998 *Using dictionaries. Studies of dictionary use by language learners and translators*. (Lexicographica Series Maior 88.) Tübingen: Niemeyer.

Atkins, Beryl T. Sue, Hélene Lewis, Della Summers, and Janet Whitcut

1987 A research project into the use of learners' dictionaries. In *The dictionary and the language learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 Apr. 1985*, Cowie, Anthony Paul (ed.), 29-43. (Lexicographica Series Maior 17.) Tübingen: Niemeyer.

Atkins, Beryl T. Sue, and Krista Varantola
1997 Monitoring dictionary use. *International Journal of Lexicography* 10 (1): 1-45.
1998a Language learners using dictionaries: The final report on the EURALEX/AILA Research Project on Dictionary Use. In *Using dictionaries. Studies of dictionary use by language learners and translators*, Atkins, Beryl T. Sue (ed.), 21-81. (Lexicographica Series Maior 88.) Tübingen: Niemeyer.
1998b Monitoring dictionary use. In *Using dictionaries. Studies of dictionary use by language learners and translators*, Atkins, Beryl T. Sue (ed.), 83-122. (Lexicographica Series Maior 88.) Tübingen: Niemeyer.

Bareggi, Carla
1989 Gli studenti e il dizionario: Un'inchiesta presso gli studenti di inglese del Corso di Laure in Lingue e Letterature Straniere della Facolta di Lettere di Torino. In *Dal dizionario ai dizioneri: orientamento e guida all'uso per studenti di lingua inglese*, Zabregelsky, Prat, and Maria Teresa (eds.), 155-190. Turin: Tirrenia Stampatori.

Barnhart, Clarence L.
1962 Problems in editing commercial monolingual dictionaries. In *Problems in lexicography*, Householder, Fred W., and Sol Saporta (eds.), 161-181. Bloomington: Indiana University.

Battenburg, John D.
1989 A study of English monolingual learner's dictionaries and their users. Ph.D. diss., Purdue University.
1991 *English monolingual learners' dictionaries: A user-oriented study*. (Lexicographica Series Maior 39.) Tübingen: Niemeyer.

Baxter, James
1980 The dictionary and vocabulary behavior: A single word or a handful? *TESOL Quarterly* 14 (3): 325-336.

Béjoint, Henri
1981 The foreign student's use of monolingual English dictionaries: A study of language needs and reference skills. *Applied Linguistics* 2 (3): 207-222.
1988 Psycholinguistic evidence and the use of dictionaries by L2 learners. In *ZüriLEX '86 proceedings*, Snell-Hornby, Mary (ed.), 139-148. Tübingen: Francke Verlag.

Béjoint, Henri, and André Moulin
1987 The place of the dictionary in an EFL programme. In *The dictionary and the language learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 Apr. 1985*, Cowie, Anthony Paul (ed.), 97-114. (Lexicographica Series Maior 17.) Tübingen: Niemeyer.

Benson, Morton, Evelyn Benson, and Robert F. Ilson
1986 *Lexicographic description of English*. Amsterdam: John Benjamins.

Bensoussan, Marsha, and Batia Laufer
1984 Lexical guessing in context. *Journal of Research in Reading* 7 (1): 15-32.

Bensoussan, Marsha, Donald Sim, and Razelle Weiss
1984 The effect of dictionary usage on EFL test performance compared with student and teacher attitudes and expectations. *Reading in a Foreign Language* 2 (2): 262-275.

Berdie, Douglas R., and John F. Anderson
1974 *Questionnaires: Design and use*. Metuchen, N.J.: Scarecrow Press.

Bergenholtz, Henning, and Joachim Mugdan (eds.)
1985 *Lexikographie und Grammatik*. (Lexicographica Series Maior 3.) Tübingen: Niemeyer.

Bishop, Graham
1998 Research into the use being made of bilingual dictionaries by language learners. *Language Learning Journal* 18: 3-8.

Black, Alison
1986 *The effects on comprehension and memory of providing different types of defining information for new vocabulary: A report on two experiments conducted for Longman ELT Dictionaries and Reference Division*. MRC Applied Psychology Unit.

Bogaards, Paul
1988 A propos de l'usage du dictionnaire de langue étrangere. *Cahiers de Lexicologie* 52 (1): 131-152.
1990 Oú cherche-t-on dans le dictionnaire? *International Journal of Lexicography* 3 (2): 79-102.
1991 Dictionnaires pédagogiques et apprentissage du vocabulaire. *Cahiers de Lexicologie* 59 (2): 93-107.
1992 French dictionary users and word frequency. In *EURALEX '92 Proceedings*, Tommola, K. Hannu, Krista Varantola, T. Salami-Tononen, and J. Schopp (eds.), 51-59. Tampere: Department of Translation Studies, University of Tampere.
1993 Models of dictionary use. *Toegepaste Taalwetenschap in Artikelen* 47-48: 17-28.
1999 Research on dictionary use: An overview. In *Dictionaries in language learning. Recommendations, national reports and thematic reports from the Thematic Network Project in the Area of Languages, sub-project 9: dictionaries*, Hartmann, Reinhard R.K. (ed.), 32-35. Berlin: Freie Universität Berlin.
2002 The use of the *de Gruyter Wörterbuch Deutsch als Fremdsprache* for receptive purposes. In *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des de Gruyter Wörterbuchs Deutsch als Fremdsprache*, Wiegand, Herbert Ernst (ed.), 647-660. Tübingen: Niemeyer.

Bradburn, Norman M., Seymour Sudman, and Edward Blair
1979 *Improving interview method and questionnaire design*. (The Jossey-Bass social and behavioral science series.) San Francisco: Jossey-Bass.

Calzolari, Nicoletta
1977 An empirical approach to circularity in dictionary definitions. *Cahiers de Lexicologie*: 118-128.

Carter, Ronald
1987 *Vocabulary. Applied linguistic perspectives*. London: Allen & Unwin.

Carter, Ronald, and Michael McCarthy (eds.)
1988 *Vocabulary and language teaching*. London: Longman.

Chen, Hsuan-Chih
1990 Lexical processing in a non-native language: Effects of language proficiency and learning strategy. *Memory and Cognition* 18: 279-288.

Chen, Hsuan-Chih, and Man-Lai Ng
1989 Semantic faciliation and translation priming effects in Chinese-English bilinguals. *Memory and Cognition* 17 (4): 454-462.

Cohen, Louis, and Lawrence Manion
1994 *Research methods in education*. London - New York: Routledge.

Corréard, Marie-Hélene
2002 Are space-saving strategies relevant in electronic dictionaries? In *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 12-17, 2002, Vol.2*, Braasch, Anna, and Claus Povlsen (eds.), 463-470. Copenhagen: Center for Sprogteknologi, Copenhagen University.

Cowie, Anthony Paul
1999 *English dictionaries for foreign learners: A history*. Oxford: Clarendon Press.

Crystal, David
1986 The ideal dictionary, lexicographer and user. In *Lexicography: An emerging international profession*, Ilson, Robert F. (ed.), 72-81. Manchester: Manchester University Press.

Cumming, Geoff, Samantha Cropp, and Roland Sussex
1994 On-line lexical resources for language learners: Assessment of some approaches to word formation. *System* 22 (3): 369-377.

Dolezal, Frederic Thomas, and Don R. McCreary
1996 Language learners and dictionary users: Commentary and an annotated bibliography. *Lexicographica* 12: 125-165.
1999 *Pedagogical lexicography today. A critical bibliography on learners' dictionaries with special emphasis on language learners and dictionary users*. (Lexicographica Series Maior 96.) Tübingen: Niemeyer.

Dörnyei, Zoltán
2003 *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Drosdowski, Günter, Helmut Henne, and Herbert Ernst Wiegand (eds.)
1977 *Nachdenken über Wörterbücher*. Mannheim: Bibliographisches Institut.

Dziapa, Edyta
2001 The relative effectiveness of vocabulary acquisition through reading by Polish learners of English using two types of dictionaries: monolingual and bilingual. M.A. diss., Adam Mickiewicz University.

Ellis, Nick C., and Alan Beaton
1995 Psycholinguistic determinants of foreign language vocabulary learning. In *Lexical issues in language learning*, Harley, Birgit (ed.), 107-165. (The Best of Language Learning Series.) Ann Arbor - Amsterdam - Philadelphia: John Benjamins.

Fontenelle, Thierry (ed.)
1998 *EURALEX '98 Actes/Proceedings*. Liege: Université Départements d'Anglais et de Néerlandais.

Fox, Gwyneth
1987 The case for examples. In *Looking up: An account of the COBUILD project in lexical computing*, Sinclair, John (ed.), 137-149. London - Glasgow: Collins.

Galisson, Robert
1983 Image et usage du dictionnaire ches des étudiants (en langue) de niveau avancé. *Études de Linguistique Appliquée* 49: 5-88.

Gerard, Linda D., and Don L. Scarborough
1989 Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition* 15: 305-315.

Greenbaum, Sidney
1977 The linguist as experimenter. In *Current trends in linguistics*, Eckman, Fred R. (ed.). Washington, D.C.: Hemisphere.

Greenbaum, Sidney, Charles F. Meyer, and John Taylor
1984 The image of the dictionary for American college students. *Dictionaries: The Journal of the Dictionary Society of North America* 6: 31-52.

de Groot, Annette M.B., and Rineke Keijzer
2000 What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign language vocabulary learning and forgetting. *Language Learning* 50: 1-56.

de Groot, Annette M.B., and Judith F. Kroll (eds.)
1997 *Tutorials in bilingualism: Psycholinguisic perspectives*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

de Groot, Annette M.B., and Gerard L.J. Nas
1991 Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language* 30: 90-123.

Hanks, Patrick
1987 Definitions and explanations. In *Looking up: An account of the COBUILD project in lexical computing*, Sinclair, John (ed.), 116-136. London - Glasgow: Collins.

Harley, Birgit (ed.)
1995 *Lexical issues in language learning*. (The Best of Language Learning Series.) Ann Arbor - Amsterdam - Philadelphia: John Benjamins.

Harris, Richard J. (ed.)
1992 *Cognitive processing in bilinguals*. (Advances in Psychology.) Amsterdam: North-Holland.

Hartmann, Reinhard R.K.
1982 Das zweisprachige Wörterbuch im Fremdsprachenerwerb. *Germanistische Linguistik* 3-6 (80): 73-86.
1983a *Lexicography: Principles and practice*. London: Academic Press.
1983b On theory and practice. In *Lexicography. Principles and practice*, Hartmann, Reinhard R.K. (ed.), 3-11. London: Academic Press.
1985 Surveying the needs and skills of the dictionary users. In *The cultivated Australian*, Clark, P.E. (ed.), 289-297. Hamburg: Buske.

1987a Dictionaries of English: the user's perspective. In *Dictionaries of English: Prospects for the record of our language*, Bailey, Richard W. (ed.), 121-135. Ann Arbor: University of Michigan Press.

1987b Four perspectives on dictionary use: A critical review of research methods. In *The dictionary and the language learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 Apr. 1985*, Cowie, Anthony Paul (ed.), 11-28. (Lexicographica Series Maior 17.) Tübingen: Niemeyer.

1989a The dictionary as an aid to foreign-language teaching. In *Wörterbücher/Dictionaries/Dictionnaires. An international encyclopedia of lexicography, Vol.1*, Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand, and Ladislav Zgusta (eds.), 181-188. Berlin: Walter de Gruyter.

1989b Sociology of the dictionary user: Hypotheses and empirical studies. In *Wörterbücher/Dictionaries/Dictionnaires. An international encyclopedia of lexicography, Vol.1*, Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand, and Ladislav Zgusta (eds.), 102-111. Berlin: Walter de Gruyter.

1994 Bilingualised versions of learners' dictionaries. *Fremdsprachen Lehren und Lernen* 23: 206-220.

1999a Case study: The Exeter University survey of dictionary use. In *Dictionaries in language learning. Recommendations, national reports and thematic reports from the Thematic Network Project in the Area of Languages, sub-project 9: dictionaries*, Hartmann, Reinhard R.K. (ed.), 36-52. Berlin: Freie Universität Berlin.

2001 *Teaching and researching lexicography*. Harlow: Pearson Education.

———— (ed.)

1999b *Dictionaries in language learning. Recommendations, national reports and thematic reports from the Thematic Network Project in the Area of Languages, sub-project 9: dictionaries*. Berlin: Freie Universität Berlin.

Hartmann, Reinhard R.K., and Gregory James

1998 *Dictionary of lexicography*. London: Routledge.

Harvey, Keith, and Deborah Yuill

1997 A study of the use of a monolingual pedagogical dictionary by learners of English engaged in writing. *Applied Linguistics* 18 (3): 253-278.

Hatherall, Glyn

1984 Studying dictionary use: Some findings and proposals. In *LEXeter '83 Proceedings: Papers from International Conference on Lexicography at Exeter, 9-12 Sept. 1983*, Hartmann, Reinhard R.K. (ed.), 183-189. (Lexicographica Series Maior 1.) Tübingen: Niemeyer.

Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand, and Ladislav Zgusta (eds.)

1989-1991 *Wörterbücher/Dictionaries/Dictionnaires. An international encyclopedia of lexicography*. Berlin: Walter de Gruyter.

Hausmann, Franz Josef, and Herbert Ernst Wiegand

1989 Component parts and structures of general monolingual dictionaries: A survey. In *Wörterbücher/Dictionaries/Dictionnaires. An international encyclopedia of lexicography, Vol.1*, Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand, and Ladislav Zgusta (eds.), 328-360. Berlin: Walter de Gruyter.

Herbst, Thomas
1996 On the way to the perfect learners' dictionary: A first comparison of OALD5, LDOCE3, COBUILD2 and CIDE. *International Journal of Lexicography* 9 (4): 321-357.

Hosenfeld, Carol
1977 A preliminary investigation of the reading strategies of successful and nonsuccessful second language learners. *System* 5: 110-123.

Hulstijn, Jan H.
1993 When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *Modern Language Journal* 77 (2): 139-147.

Hulstijn, Jan H., and Beryl T. Sue Atkins
1998 Empirical research on dictionary use in foreign-language learning: Survey and discussion. In *Using dictionaries. Studies of dictionary use by language learners and translators*, Atkins, Beryl T. Sue (ed.), 7-19. (Lexicographica Series Maior 88.) Tübingen: Niemeyer.

Hulstijn, Jan H., and Robert DeKeyser
1997 Second-language acquisition research in the laboratory: Possibilities and limitations. *Studies in Second Language Acquisition* 19 (2): 131-143.

Hulstijn, Jan H., Merel Hollander, and Tine Greidanus
1996 Incidental vocabulary learning by advanced foreign-language students: The influence of marginal glosses, dictionary use, and reoccurrence of unfamiliar words. *Modern Language Journal* 80 (3): 327-339.

Humblé, Philippe
1998 The use of authentic, made-up and 'controlled' examples in foreign language dictionaries. In *EURALEX '98 Actes/Proceedings*, Fontenelle, Thierry (ed.), 593-599. Liege: Université Départements d'Anglais et de Néerlandais.
2001 *Dictionaries and language learners*. Franfurt am Main: Haag und Herchen.

Iannucci, James E.
1957 Meaning discrimination in bilingual dictionaries: A new lexicographical technique. *Modern Language Journal* 41: 272-281.

Ilson, Robert F.
1984 Towards a taxonomy of dictionary definitions. In *A spectrum of lexicography*, Ilson, Robert F. (ed.), 61-73. Amsterdam - Philadelphia: John Benjamins.

Jakubowski, Marcin
2001 The use of dictionaries by high school learners: The place of the monolingual and bilingual dictionary in the learning process. M.A. diss., Adam Mickiewicz University.

Jansen, Jacques, J.P. Mergeai, and J. Vanadroye
1987 Controlling LDOCE's controlled vocabulary. In *The dictionary and the language learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 Apr. 1985*, Cowie, Anthony Paul (ed.), 78-94. (Lexicographica Series Maior 17.) Tübingen: Niemeyer.

Jessen, Annette
1996 The presence and treatment of terms in general dictionaries. M.A. diss., University of Ottawa.

Jiang, Nan
2000 Lexical representation and development in a second language. *Applied Linguistics* 21: 47-77.

Jin, Yi
1990 Effects of concreteness on cross-language priming in lexical decisions. *Perceptual and Motor Skills* 70: 1139-1154.

Kharma, Nayef N.
1985 Wanted: A brand-new type of learners' dictionary. *Multinlingua* 4: 85-90.

Kipfer, Barbara Ann
1987 Dictionaries and the intermediate student: Communicative needs and the development of user reference skills. In *The dictionary and the language learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 Apr. 1985*, Cowie, Anthony Paul (ed.), 44-54. (Lexicographica Series Maior 17.) Tübingen: Niemeyer.

Kirkpatrick, Betty
1985 A lexicographical dilemma: Monolingual dictionaries for the native speaker and for the learner. In *ELT documents 120: lexicography and language learning*, Ilson, Robert F. (ed.). Oxford: Pergamon Press/The British Council.

Knight, Susan
1994 Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal* 78 (3): 285-299.

Koren, Shira
1997 Quality versus convenience: Comparison of modern dictionaries from the researcher's, teacher's and learner's points of view. *TESL Electronic Journal* 2 (3): 1-16.

Kostrzewa, Frank
1991 Merkmale verstehens - und behaltensförderender kontextueller Bedeutungserklärungen. Ph.D. diss., Universität Bielefeld.

Krantz, Gösta
1991 *Learning vocabulary in a foreign language: A study of reading strategies*. (Gothenburg studies in English, vol. 63.) Gothenburg: Acta Universitatis Gothoburgensis.

Kroll, Judith F.
1993 Accessing conceptual representations for words in a second language. In *The bilingual lexicon*, Schreuder, Robert, and Bert Weltens (eds.), 53-82. (Studies in Bilingualism 6.) Amsterdam: John Benjamins.

Kroll, Judith F., and Alexandra Sholl
1992 Lexical and conceptual memory in fluent and nonfluent bilinguals. In *Cognitive processing in bilinguals*, Harris, Richard J. (ed.), 191-203. (Advances in Psychology 83.) Amsterdam: North-Holland.

Kroll, Judith F., and Erika Stewart
1994 Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* 33: 149-174.

Kroll, Judith F., and Natasha Tokowicz
2001 The development of conceptual representation for words in a second language. In *One mind, two languages*, Nicol, Janet L. (ed.), 49-71. (Explaining linguistics.) Oxford - Malden: Blackwell.

Kroma, Natalia
2001 Using marginal glosses to develop students' circumlocution skills. M.A. diss., Adam Mickiewicz University.

Kromann, Hans-Peder, Theis Riiber, and Paul Rosbach
1991 'Active' and 'pasive' bilingual dictionaries: The Scerba concept reconsidered. In *Wörterbücher/Dictionaries/Dictionnaires. An international encyclopedia of lexicography, Vol.3*, Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand, and Ladislav Zgusta (eds.), 2711-2728. Berlin: Walter de Gruyter.

Kühn, Peter
1996 Langenscheidts Großwörterbuch Deutsch als Fremdsprache und die deutschen Wörterbücher. In *Perspektiven der pädagogischen Lexikographie des Deutschen*, Wiegand, Herbert Ernst (ed.). (Lexicographica Series Maior 86.) Tübingen: Niemeyer.

Landau, Sidney I.
2001 *Dictionaries: The art and craft of lexicography*, 2nd edition. Cambridge: Cambridge University Press.

Laufer, Batia
1992 Corpus-based as against lexicographer examples in comprehension and production of new words. In *EURALEX '92 Proceedings*, Tommola, K. Hannu, Krista Varantola, T. Salami-Tononen, and J. Schopp (eds.), 71-76. Tampere: Department of Translation Studies, University of Tampere.

1993 The effect of dictionary definitions and examples on the use and comprehension of new L2 words. *Cahiers de Lexicologie* 63 (2): 131-142.

1995 A case for a semi-bilingual dictionary for productive purposes. *Kernerman Dictionary News* 3: 2-4.

2000 Electronic dictionaries and incidental vocabulary acquisition: Does technology make a difference? In *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany*, Heid, Ulrich, Stefan Evert, Egbert Lehmann, and Christian Rohrer (eds.), 849-853. Stuttgart: Institut for maschinelle Sprachverarbeitung, Universität Stuttgart.

Laufer, Batia, and Linor Hadar
1997 Assessing the effectiveness of monolingual, bilingual, and "bilingualised" dictionaries in the comprehension and production of new words. *Modern Language Journal* 81: 189-196.

Laufer, Batia, and Michal Kimmel
1997 Bilingualised dictionaries: How learners really use them. *System* 25 (3): 361-369.

Laufer, Batia, and Linor Melamed
1994 Monolingual, bilingual and 'bilingualized' dictionaries: Which are more effective, for what and for whom? In *EURALEX '94 Proceedings*, Martin, Willy, Willem Meijs, Margreet Moerland, Elsemiek Ten Pas, Piet Van Sterkenburg, and Piek Vossen (eds.), 565-576. Amsterdam: Vrije Universiteit.

Lew, Robert
2002a Questionnaires in dictionary use research: A reexamination. In *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 12-17, 2002, Vol.1*, Braasch, Anna, and Claus Povlsen (eds.), 267-271. Copenhagen: Center for Sprogteknologi, Copenhagen University.

2002b A study in the use of bilingual and monolingual dictionaries by Polish learners of English: A preliminary report. In *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 12-17, 2002, Vol.2*, Braasch, Anna, and Claus Povlsen (eds.), 759-763. Copenhagen: Center for Sprogteknologi, Copenhagen University.

2003 Designing relational database structures for storing and processing language questionnaire data: Example from a study in dictionary use. In *Research and scholarship in integration processes: Poland - USA - EU*, Oleksy, Elżbieta H., and Barbara Lewandowska-Tomaszczyk (eds.), 355-363. Łódź: Łódź University Press.

Li, Lan
1998 Dictionaries and their users in Chinese universities: With special reference to ESP learners. In *Lexicography in Asia. Selected papers from the Dictionaries in Asia Conference, Hong Kong, and other papers*, McArthur, Tom, and Ilan Kernerman (eds.), 61-79. Tel Aviv: Password Publishers.

Luppescu, Stuart, and R. Richard Day
1993 Reading, dictionaries, and vocabulary learning. *Language Learning* 43 (2): 263-287.

Lyons, John
1977 *Semantics*. Cambridge: Cambridge University Press.

MacFarquhar, Peter D., and Jack C. Richards
1983 On dictionaries and definitions. *RELC Journal* 14 (1): 111-124.

Mackintosh, Kristen
1995 An empirical study of dictionary use in version. M.A. diss., University of Ottawa.

MacWhinney, Brian
1997 Second language acquisition and the competition model. In *Tutorials in bilingualism: Psycholinguistic perspectives*, de Groot, Annette M.B., and Judith F. Kroll (eds.), 113-142. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Magay, Tamás, and Judit Zigány (eds.)
1990 *Budalex '88 Proceedings*. Budapest: Akadémiai Kiadó.

Marello, Carla
1987 Examples in contemporary Italian bilingual dictionaries. In *The dictionary and the language learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 Apr. 1985*, Cowie, Anthony Paul (ed.), 224-237. (Lexicographica Series Maior 17.) Tübingen: Niemeyer.

1989 *Dizionari bilingui*. Bologna: Zanichelli.

McArthur, Tom, and Ilan Kernerman (eds.)
1998 *Lexicography in Asia. Selected papers from the Dictionaries in Asia Conference, Hong Kong, and other papers*. Tel Aviv: Password Publishers.

McCawley, James D.
1993 How to achieve lexicographic virtue through selective and judicious sinning. *Dictionaries: The Journal of the Dictionary Society of North America* 14: 120-129.

McCreary, Don R.
2002 American freshmen and English dictionaries: 'I had *Aspersions* of becoming an English teacher'. *International Journal of Lexicography* 15 (3): 181-205.

McCreary, Don R., and Frederic Thomas Dolezal
1998 Language learners and dictionary users: Bibliographic findings and commentary. In *EURALEX '98 Actes/Proceedings*, Fontenelle, Thierry (ed.), 611-618. Liege: Université Départements d'Anglais et de Néerlandais.

1999 A study of dictionary use by ESL students in an American university. *International Journal of Lexicography* 12 (2): 105-144.

McNaughton, David, Charles Hughes, and Karen Clark
1997 The effect of five proofreading conditions on the spelling performance of college students with learning disabilities. *Journal of Learning Disabilities* 30 (6): 643-651.

Meara, Paul, and Fiona English
1988 *Lexical errors and learners' dictionaries*. University College Swansea: Centre for Applied Language Studies Group ERIC ED 654 321.

Meijs, Wim J.
1990 Morphology and word-formation in a machine-readable dictionary: Problems and possibilities. *Folia Linguistica* 24 (1-2): 45–71.

Miller, George A., and Patricia M. Gildea
1985 How to misread a dictionary. *AILA Bulletin* 1985: 13-26.

1987 How children learn words. *Scientific American* September 1987: 86-91.

Mitchell, Evelyn
1983a *Search-do reading: Difficulties in using a dictionary*. Aberdeen College of Education Formative Assessment of Reading Working Paper 21.

1983b *Search-do reading: Using a dictionary: a preliminary analysis*. Aberdeen College of Education Formative Assessment of Reading Working Paper 20.

Momoi, Hidetomo
1998 ESL learners' behavior in using bilingual dictionaries while writing an English composition. M.Ed. diss., Tokyo Gakugei University.

Müllich, Harald
1990 *Die Definition ist blöd! Herübersetzen mit dem einsprachigen Wörterbuch. Das französische und englishe Lernerwörterbuch in der Hand der deutschen Schüler*. Tübingen: Niemeyer.

Nakamoto, Kyohei
1995 Monolingual or bilingual, that is *not* the question: The 'bilingualised' dictionary. *Kernerman Dictionary News* 2: 2-4.

Nesi, Hilary
1994 The effect of language background and culture on productive dictionary use. In *EURALEX '94 Proceedings*, Martin, Willy, Willem Meijs, Margreet Moerland, Elsemiek Ten Pas, Piet Van Sterkenburg, and Piek Vossen (eds.), 577-585. Amsterdam: Vrije Universiteit.
1996 The role of illustrative examples in productive dictionary use. *Dictionaries: The Journal of the Dictionary Society of North America* 17: 198-206.
2000a Electronic dictionaries in second language vocabulary comprehension and acquisition: The state of the art. In *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany*, Heid, Ulrich, Stefan Evert, Egbert Lehmann, and Christian Rohrer (eds.), 839-841. Stuttgart: Institut for maschinelle Sprachverarbeitung, Universität Stuttgart.
2000b *The use and abuse of EFL dictionaries*. (Lexicographica Series Maior 98.) Tübingen: Niemeyer.

Nesi, Hilary, and Richard Haill
2002 A study of dictionary use by international students at a British university. *International Jounal of Lexicography* 15 (4): 277-305.

Nesi, Hilary, and Paul Meara
1991 How using dictionaries affects performance in multiple-choice EFL tests. *Reading in a Foreign Language* 8 (1): 631-643.
1994 Patterns of misinterpretation in the productive use of EFL dictionary definitions. *System* 22 (1): 1-15.

Neubach, Abigail, and Andrew D. Cohen
1988 Processing strategies and problems encountered in the use of dictionaries. *Dictionaries: The Journal of the Dictionary Society of North America* 10: 1-19.

Nicol, Janet L. (ed.)
2001 *One mind, two languages*. (Explaining linguistics.) Oxford-Malden: Blackwell.

Nuccorini, Stefania
1992 Monitoring dictionary use. In *EURALEX '92 Proceedings*, Tommola, K. Hannu, Krista Varantola, T. Salami-Tononen, and J. Schopp (eds.), 89-102. Tampere: Department of Translation Studies, University of Tampere.
1994 On dictionary misuse. In *EURALEX '94 Proceedings*, Martin, Willy, Willem Meijs, Margreet Moerland, Elsemiek Ten Pas, Piet Van Sterkenburg, and Piek Vossen (eds.), 586-597. Amsterdam: Vrije Universiteit.

Oppenheim, Abraham Naftali
1992 *Questionnaire design, interviewing, and attitude measurement*. London - New York: Pinter Publishers.

Oskarsson, Mats
1975 On the role of the mother tongue in learning foreign language vocabulary: An empirical investigation. *ITL Review of Applied Linguistics* 27: 19-32.

Padrón, Yolanda N., and Hersh C. Waxton
1988 The effect of ESL students' perceptions of their cognitive strategies on reading achievement. *TESOL Quarterly* 22 (1): 146-150.

Piotrowski, Tadeusz
1989 Monolingual and bilingual dictionaries: Fundamental differences. In *Learners' dictionaries: State of the art*, Tickoo, Makhan L. (ed.), 72-83. Singapore: SEAMEO Regional Language Centre.
1994 *Problems in bilingual lexicography*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Quirk, Randolph
1974 The image of the dictionary. In *The linguist and the English language*, Quirk, Randolph (ed.), 148-163. London: Edward Arnold.

Raudaskoski, Seppo
2002 Translation, the key or the equivalent? *Kernerman Dictionary News* 10: 2-3.

Ronald, James
2002 L2 lexical growth through extensive reading and dictionary use: A case study. In *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 12-17, 2002, Vol.2*, Braasch, Anna, and Claus Povlsen (eds.), 765-771. Copenhagen: Center for Sprogteknologi, Copenhagen University.

Rundell, Michael
1988 Changing the rules: Why the monolingual learner's dictionary should move away from the native speaker tradition. In *ZüriLEX '86 proceedings*, Snell-Hornby, Mary (ed.), 127-137. Tübingen: Francke Verlag.

Van Scherrenburg, Daniel
1990 The arrangement of information in the general bilingual dictionary entry. M.A. diss., University of Ottawa.

Scholfield, Phil
1982 Using the English dictionary for comprehension. *TESOL Quarterly* 16 (2): 185-194.
1999 Dictionary use in reception. *International Journal of Lexicography* 12 (1): 13-34.

Schreuder, Robert, and Bert Weltens (eds.)
1993 *The bilingual lexicon*. (Studies in Bilingualism 6.) Amsterdam: John Benjamins Publishing Company.

de Schryver, Gilles-Maurice
2003 Lexicographers' dreams in the electronic-dictionary age. *International Jounal of Lexicography* 16 (2): 143-199.

Sharpe, Peter
1995 Electronic dictionaries with particular reference to the design of an electronic bilingual dictionary for English-speaking learners of Japanese. *International Journal of Lexicography* 8 (1): 39-54.

Shcherba, Lev Vladimirovich
1995 Towards a general theory of lexicography. *International Journal of Lexicography* 8 (4): 314-350.

Sinclair, John (ed.)
1987 *Looking up: An account of the COBUILD project in lexical computing*. London - Glasgow: Collins.

Snell-Hornby, Mary
1987 Towards a learner's bilingual dictionary. In *The dictionary and the language learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 Apr. 1985*, Cowie, Anthony Paul (ed.), 159-170. (Lexicographica Series Maior 17.) Tübingen: Niemeyer.

Sobkowiak, Włodzimierz
1999 *Pronunciation in EFL machine-readable dictionaries*. Poznań: Motivex.

Stein, Gabrielle
1990 From the bilingual to the monolingual dictionary. In *Budalex '88 Proceedings*, Magay, Tamás, and Judit Zigány (eds.), 401-407. Budapest: Akadémiai Kiadó.

Steiner, Roger J.
1989 The absence of text: The bilingual dictionary as an index. *International Jounal of Lexicography* 2 (3): 249-257.

Sudman, Seymour, and Norman M. Bradburn
1982 *Asking questions: A practical guide to questionnaire design*. (Jossey-Bass series in social and behavioral sciences.) San Francisco: Jossey-Bass.

Summers, Della
1988 The role of dictionaries in language learning. In *Vocabulary and language learning*, Carter, Ronald, and Michael McCarthy (eds.), 111-125. London: Longman.

Szczepaniak, Renata
2003 What users do with dictionaries in situations of comprehension deficit: An empirical study. *Studia Anglica Posnaniensia* 39: 191-232.
2004 An empirical study into the use of a monolingual learner's dictionary: The effects on the comprehension of idiom variation for advanced students of English. Ph.D. diss., Adam Mickiewicz University.

Thompson, Geoff
1987 Using bilingual dictionaries. *English Language Teaching Journal* 41 (4): 282-286.

Tomaszczyk, Jerzy
1979 Dictionaries: Users and uses. *Glottodidactica* 12: 103-119.
1983 On bilingual dictionaries: The case for bilingual dictionaries for foreign language learners. In *Lexicography: Principles and practice*, Hartmann, Reinhard R.K. (ed.), 41-51. London: Academic Press.
1987 FL learners' communication failure: Implications for pedagogical lexicography. In *The dictionary and the language learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 Apr. 1985*, Cowie, Anthony Paul (ed.), 136-145. Tübingen: Niemeyer.

Tommola, K. Hannu, Krista Varantola, T. Salami-Tononen, and J. Schopp (eds.)
1992 *EURALEX '92 Proceedings*. Tampere: Department of Translation Studies, University of Tampere.

Tono, Yukio
1984 On the dictionary user's reference skills. B.Ed. diss., Tokyo Gakugei University.

1987 Which word do you look up first? A study of dictionary reference skills. M.Ed. diss., Tokyo Gakugei University.

1988 Assessment of the EFL learner's dictionary-using skills. *JACET Bulletin* 19: 103-126.

1989 Can a dictionary help you read better? On the relationship between EFL learners' dictionary, reference skills and reading comprehension. In *Lexicographers and their works*, James, Gregory (ed.), 192-200. (Exeter Linguistic Studies 14.) Exeter: Exeter University Press.

1991 A good dictionary user: What makes the difference? In *Recent Studies on English Language Teaching*, Ito, Kaichi, Ken Kanatani, and Tetsuyu Noda (eds.). Yumi Shobou.

1992 The effect of menus on EFL learners' look-up processes. *Lexicos* 2: 230-253.

1997 Guide Word or Signpost? An experimental study on the effect of meaning access indexes in EFL learners' dictionaries. *English Studies* 28: 55-77.

2000 On the effects of different types of electronic dictionary interfaces on L2 learners' reference behaviour in productive/receptive tasks. In *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany*, Heid, Ulrich, Stefan Evert, Egbert Lehmann, and Christian Rohrer (eds.), 855-861. Stuttgart: Institut for maschinelle Sprachverarbeitung, Universität Stuttgart.

2001 *Research on dictionary use in the context of foreign language learning: Focus on reading comprehension*. (Lexicographica Series Maior 106.) Tübingen: Niemeyer.

Toope, Michael

1996 Examples in the bilingual dictionary. M.A. diss., University of Ottawa.

Trochim, William M.

2000 *The research methods knowledge base*, 2nd edition. Cincinnati, OH: Atomic Dog Publishing.

Turkish, Marion P.

1972 A study of dictionary skills used by pupils in grades four, five, and six. Ph.D. diss., Fordham University.

Tzelgov, Joseph, and Sigal Eben-Ezra

1992 Components of the between-language semantic priming effect. *European Journal of Cognitive Psychology* 4: 253-272.

Varantola, Krista

1998 Translators and their use of dictionaries. In *Using dictionaries. Studies of dictionary use by language learners and translators*, Atkins, Beryl T. Sue (ed.), 179-192. (Lexicographica Series Maior 88.) Tübingen: Niemeyer.

Wainer, Howard, and Henry I. Braun (eds.)

1998 *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Whyatt, Bogusława

2000 A psycholinguistic investigation into the processes of comprehension and production: A decision making approach towards the preservation of meaning in translation. Ph.D. diss., Adam Mickiewicz University.

Wiegand, Herbert Ernst
1977a Einige grundlegende semantisch-pragmatische Aspekte von Wörterbucheinträgen. Ein Beitrag zur praktischer Lexikologie. *Kopenhagener Beiträge zur Germanistischen Linguistik* 12: 59-149.
1977b Nachdenken über Wörterbücher: Aktuelle Probleme. In *Nachdenken über Wörterbücher*, Drosdowski, Günter, Helmut Henne, and Herbert Ernst Wiegand (eds.), 71-102. Mannheim: Bibliographisches Institut.
1985 Fragen zur Grammatik in Wörterbuchbenutzungs-protokollen. Ein Beitrag zur empirischen Erforschung der Benutzer einsprachiger Wörterbücher. In *Lexikographie und Grammatik*, Bergenholtz, Henning, and Joachim Mugdan (eds.), 20-98. (Lexicographica Series Maior 3.) Tübingen: Niemeyer.
1998 *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatiesierung der Lexicographie*. Berlin: Walter de Gruyter.

Wierzbicka, Anna
1985 *Lexicography and conceptual analysis*. Ann Arbor: Karoma.
1993 What are the uses of theoretical lexicography? *Dictionaries: The Journal of the Dictionary Society of North America* 14: 44-78.

Wingate, Ursula
2002 *The effectiveness of different learner dictionaries. An investigation into the use of dictionaries for reading comprehension by intermediate learners of German*. (Lexicographica Series Maior 112.) Tübingen: Niemeyer.

Worsch, Wolfgang
1999 Recent trends in publishing bilingual learners' dictionaries. In *Dictionaries in language learning. Recommendations, national reports and thematic reports from the Thematic Network Project in the Area of Languages, sub-project 9: dictionaries*, Hartmann, Reinhard R.K. (ed.), 99-107. Berlin: Freie Universität Berlin.

Yokoyama, T.
1994 A study of the EFL learners' dictionary look-up strategies in the case of Japanese-English dictionary. B.Ed. diss., Tokyo Gakugei University.

Zabregelsky, Prat, and Maria Teresa (eds.)
1989 *Dal dizionario ai dizioneri: Orientamento e guida all'uso per studenti di lingua inglese*. Turin: Tirrenia Stampatori.

Zgusta, Ladislav
1988 *Lexicography today. An annotated bibliography of the theory of lexicograpy*. (Lexicographica Series Maior 18.) Tübingen: Niemeyer.

Zgusta, Ladislav, V. Cerný, Zdenka Hermanová-Novotná, and Danuška Heroldová
1971 *Manual of lexicography*. (Janua linguarum. Series maior, 39.) Den Haag: Mouton.

Zöfgen, Ekkehard
1994 *Lernerwörterbücher in Theorie und Praxis: Ein Beitrag zur Metalexikographie mit besonderer Berücksichtigung des Französischen*. (Lexicographica Series Maior 59.) Tübingen: Niemeyer.

Appendices

Appendix 1. Teachers' Questionnaire

Formularz nauczyciela z instrukcją

Zapoznaj się z tym formularzem na kilka dni przed badaniem i ponownie na kilka godzin przed badaniem. Jeden taki formularz towarzyszy każdej testowanej grupie. Zatem np. jeśli badane są trzy klasy, wypełnij trzy formularze.

Przebieg badania:

Jeśli jest taka możliwość, rozsadź badanych tak, aby utrudnić odpisywanie (jak w czasie klasówki, kartkówki). Poproś o spakowanie/odłożenie słowników, książek, jeśli jest taki zwyczaj.
Wprowadzenie (tekst w ramkach należy odczytać lub powiedzieć):

Wasza klasa/grupa została wytypowana do badań nad nauczaniem języka angielskiego prowadzonych na Uniwersytecie Adama Mickiewicza. Badanie jest w pełni anonimowe, nie podpisujcie się nigdzie, nie wpłynie w żaden sposób na niczyją ocenę indywidualną. Cel – to poprawa jakości nauczania języka angielskiego. Żeby cel osiągnąć w pełni, konieczna jest wasza współpraca - starajcie się szczerze i skrupulatnie podejść do tego. Nie pomijajcie żadnej odpowiedzi, niezależnie od tego, czy jesteście jej pewni, czy nie. Nie przepisujcie od sąsiada, ważne są odpowiedzi, jakie da każdy osobiście. Dlatego nie wolno też się porozumiewać. Teraz rozdam test, ale proszę jeszcze nie zaczynać, wyjaśnię kilka rzeczy.

Rozdaj test, po jednym każdej osobie, bezpośrednio po rozdaniu mów dalej:

Badanie składa się z trzech stron: kwestionariusza, testu językowego i słowniczka. Na pierwszej stronie mamy kwestionariusz, gdzie należy zakreślić krzyżykiem po jednym i tylko jednym kwadraciku w każdym poziomym rzędzie. Na ten kwestionariusz wystarczy 10 minut. Reszta czasu przeznaczona jest na test językowy, który jest na następnej stronie. Są to cztery krótkie ćwiczenia. Wystąpi w nich trochę trudniejszych – zapewne nieznanych wam – słów angielskich, ale nie zniechęcajcie się - każde słowo angielskie wydrukowane pogrubionymi literkami znajdziecie w specjalnym słowniczku, który jest na sąsiedniej stronie. Korzystajcie z tego słownika. Natomiast nie wolno korzystać z żadnych innych słowników ani materiałów. Test językowy zabierze 30 minut, czyli łącznie ok. 40 minut, skończymy wszyscy razem, więc proszę się nie spieszyć. Można już zacząć – zacznijcie od pierwszej strony.

Niewykorzystane arkusze włóż z powrotem do koperty. Po upływie 10 minut przypomnij:

Należy już dokończyć kwestionariusz na pierwszej stronie i przejść do testu na drugiej stronie.

Pilnuj, aby osoby nie porozumiewały się, nie ściągały i nie korzystały z żadnych materiałów. W przypadku pytań o technikę wypełniania możesz krótko wyjaśnić lub pomóc; natomiast na pytania typu "Co oznacza to pytanie/słowo" odpowiedz: Nic więcej nie wolno mi wyjaśniać ani podpowiadać; tak jak rozumiesz, tak wypełnij.
Po upłynięciu 40 minut zbieramy wszystkie testy i wkładamy do właściwej koperty. Wszystkie testy – wykorzystane lub nie – muszą wrócić w oryginalnej kopercie. Niniejszą kartkę również włóż do tej koperty.

Informacje o przebiegu badania i grupie badanej:

W chwili spokoju wypełnij następujące informacje (jeśli nie ma takiej chwili, to jak najszybciej po teście):
Data przeprowadzenia badania:……
Nazwa i dokładny typ szkoły:.....
Nazwa i dokładny typ klasy/grupy:.....
Jakiego głównego podręcznika do angielskiego używa ta klasa/grupa (tytuł/wydawca/autor/poziom)?
.....
Liczba osób w tej grupie/klasie biorących udział w badaniu:……
Średni poziom tej grupy/klasy w ocenie własnej nauczyciela (zaznacz ☒):

beginner	pre-intermediate	intermediate	upper intermediate	advanced
☐	☐	☐	☐	☐

Appendix 2. Teachers' Questionnaire: English translation

Teacher's form and instructions

Familiarize yourself with this form a few days prior to questionnaire and test administration; reread it a few hours prior to administration. One form goes with each group tested; so if, for example, you are testing three groups, complete three forms.

Procedure:

If possible, seat subjects so as to make cheating difficult (as you would for a quiz). Ask them to put away dictionaries and books, if that is what you would normally do.

Introduction (read out or say aloud the text in the box):

> Your class/group has been selected to participate in a test study on English language teaching done at the University of Adam Mickiewicz. The study is anonymous, so do not sign your name. The results will not affect your grade. The general goal is to improve the quality of English language teaching. To meet the goal, your cooperation is essential, so please approach the task with honesty and care. Do not skip items, even if you are not sure of your answer. Do not copy responses from others; it is vital that you give your own answer. You may not communicate with others. I will now distribute the test, but do not begin before I've explained a few more things.

Distribute one test booklet to each subject and continue speaking:

> The test has three pages: a questionnaire, a language test, and a dictionary. The first page is the questionnaire: here, cross only one box in each line. You have ten minutes to complete the questionnaire, use the remaining time for the test, which is found on the following page. It includes four short tasks.You will come across a few fairly difficult English words, which will probably be unknown to you. Do not worry, though: you will find each English word printed in bold in a special dictionary on the facing page. Use this dictionary, but do not use any other dictionaries or other aids. The test will take 30 minutes, that's about 40 minutes in all. We will all finish at the same time, so please do not rush. You can get started on the first page now.

Put unused test booklets back in the envelope. After 10 minutes, prompt the subjects:

> You should be done by now with the questionnaire on page one and move on to the test on page two.

Make sure that subjects do not communicate, cheat, or use any external materials. Should there be questions regarding the technique of filling in the questionnaire or test, offer a brief clarification or help; however, to questions of the type "What does this item/word mean?", say: I am not allowed to say anything more about this; please do it the way you understand it.

After 40 minutes, collect all booklets and put them back in the envelope. All sheets – used and unused – must be returned to me in their original envelopes. Put this sheet in the same envelope as well.

Data on the procedure and subjects:

When you have a quite moment during the testing, fill in the following (if there is no good opportunity, do it as soon as you can after the test administration):

Date when test administered:……

Name and detailed type of school:.....

Name and detailed type of class/section:.....

What is the principal EFL textbook used by this class/section (title/publisher/author/level)?

.....

Number of people in this class/section participating in the study:……

Teacher's assessment of the average level of this class/section (tick a box ☒):

beginner	pre-intermediate	intermediate	upper intermediate	advanced
☐	☐	☐	☐	☐

Appendix 3. Learners' Questionnaire

Instrukcja: odpowiedz szczerze (test jest w pełni anonimowy) na **wszystkie** pytania, zakreślając krzyżykiem (w ten sposób: ☒) dokładnie **jeden** kwadracik w każdym poziomym rzędzie. Na tę część masz ok. 10 minut.

A. *Ile lat (w sumie) uczysz się języka angielskiego?*

0-1	2-3	4-5	6 lub więcej
☐	☐	☐	☐

B. *Ile godzin zajęć z angielskiego masz w tygodniu (w sumie – szkoła, kursy, lekcje prywatne)?*

1	2-3	4-5	6 lub więcej
☐	☐	☐	☐

C. *Swoją znajomość angielskiego oceniasz na tle grupy jako:*

☐ ponad przeciętną ☐ przeciętną ☐ poniżej przeciętnej

D. *Czy potrafisz po angielsku:*

	tak	raczej tak	raczej nie	nie
1. dowiedzieć się o drogę na stację?	☐	☐	☐	☐
2. opisać objawy przeziębienia?	☐	☐	☐	☐
3. zrozumieć tekst piosenki ze słuchu?	☐	☐	☐	☐

E. *Jak często korzystasz ze słowników następujących typów:*

	codziennie	kilka razy na tydzień	raz na tydzień	rzadziej/wcale
1. polsko-angielski?	☐	☐	☐	☐
2. angielsko-polski?	☐	☐	☐	☐
3. angielsko-angielski?	☐	☐	☐	☐

F. *Z jakich dwóch słowników korzystasz najczęściej (podaj typ, tytuł, wydawcę, autora, wydanie – na ile pamiętasz) i jak je oceniasz?*

tytuł/wydawca/autor/wydanie	świetny	dobry	ujdzie	kiepski	dno
1.	☐	☐	☐	☐	☐
2.	☐	☐	☐	☐	☐

G. *Zaznacz, jak często poszukujesz następujących informacji w słowniku:*

	najczęściej	często	rzadko	nigdy
1. jak się to wymawia? ----------	☐	☐	☐	☐
2. co ten wyraz oznacza? ----------	☐	☐	☐	☐
3. jak to będzie po polsku? ----------	☐	☐	☐	☐
4. jak to będzie po angielsku? ----------	☐	☐	☐	☐
5. jaka to część mowy? ----------	☐	☐	☐	☐
6. jak zbudować zdanie? ----------	☐	☐	☐	☐
7. z jakimi innymi słowami się używa? --	☐	☐	☐	☐
8. w jakich sytuacjach się używa? --------	☐	☐	☐	☐
9. inny wyraz o podobnym znaczeniu? --	☐	☐	☐	☐

Instrukcja: po wypełnieniu tej strony (sprawdź, czy masz wszystkie odpowiedzi), przewróć kartkę.

Appendix 4. Learners' Questionnaire: English translation

Instructions: answer truthfully (the test is fully anonymous) **all** the questions by crossing (like this: ☒) exactly **one** box in each row. You have 10 minutes for this part.

A. *How many years (all told) have you been learning English?*

0-1	2-3	4-5	6 or more
☐	☐	☐	☐

B. *How many hours of English a week are you taking (at school, courses, private tutoring)?*

1	2-3	4-5	6 or more
☐	☐	☐	☐

C. *How would you rate your level of English against your classmates:*

☐ above average ☐ average ☐ below average

D. *Would you be able to do the following in English:*

	yes	probably yes	probably not	no
1. ask for directions to the station?	☐	☐	☐	☐
2. describe the symptoms of a cold?	☐	☐	☐	☐
3. understand song lyrics when listening?	☐	☐	☐	☐

E. *How often do you use the following types of dictionaries:*

	daily	a few times a week	weekly	less freq./not at all
1. Polish-English?	☐	☐	☐	☐
2. English-Polish?	☐	☐	☐	☐
3. English-English?	☐	☐	☐	☐

F. *What two dictionaries do you use most often (give type, title, publisher, author, edition – as far as you remember), and how do you rate them?*

title/publisher/author/edition	excellent	good	OK	poor	awful
1.	☐	☐	☐	☐	☐
2.	☐	☐	☐	☐	☐

G. *How often do you look for the following information in a dictionary:*

	most often	often	rarely	never
1. how do you pronounce that?----------	☐	☐	☐	☐
2. what does this word mean? ----------	☐	☐	☐	☐
3. how do you say it in Polish?----------	☐	☐	☐	☐
4. how do you say it in English? ----------	☐	☐	☐	☐
5. what part of speech is this? ----------	☐	☐	☐	☐
6. how to form a sentence?----------	☐	☐	☐	☐
7. what words to use this word with?-----	☐	☐	☐	☐
8. in what situations is this word used? --	☐	☐	☐	☐
9. another word with similar meaning? --	☐	☐	☐	☐

Instruction: when you are done with this page (double-check that you've answered all items), turn the page.

Appendix 5. Dictionary Effectiveness Test

Instrukcja: zrób poniższe ćwiczenia. Wystąpi w nich trochę trudniejszych słów angielskich, ale nie zniechęcaj się - każde słowo wydrukowane **takimi literkami** znajdziesz w słowniku na sąsiedniej stronie. Korzystaj z tego słownika.

H. *W każdej z linijek zakreśl* (☒) ***jeden*** *kwadracik przy najbardziej podobnym pod względem znaczenia wyrazie polskim (linijki 1-5) lub angielskim (6 i 7) spośród czterech podanych:*

1. **helk**	oznacza:	☐ trafny	☐ właściwy	☐ miły	☐ taktowny
2. **sulk**	oznacza:	☐ obrażać	☐ marzyć	☐ dąsać się	☐ kląć
3. **vab**	oznacza:	☐ recenzować	☐ krytykować	☐ poprawiać	☐ oskarżać
4. **woozy**	oznacza:	☐ zmęczony	☐ zawiany	☐ niecierpliwy	☐ malutki
5. **spating**	oznacza:	☐ belka	☐ ławka	☐ rynna	☐ tynk
6. **strod**	oznacza:	☐ **legalize**	☐ **ban**	☐ **punish**	☐ **introduce**
7. **luton**	oznacza:	☐ **hat**	☐ **dress**	☐ **vase**	☐ **cover**

I. *Zakreśl* ***jeden*** *kwadracik przy tym z czterech podanych słów angielskich, które najlepiej pasuje znaczeniem do reszty zdania:*

1. He would always …… at night, and the noise sometimes frightened his wife.
☐ **lenate** ☐ **sturton** ☐ **mirk** ☐ **reaff**

2. Her car must have very good …… if she can get from Berlin to Poznań in just three hours.
☐ **gerd** ☐ **marbish** ☐ **scolb** ☐ **cullen**

J. *Przetłumacz zdania na język polski:*

1. The **boss** plans to **batenate** Mary tomorrow.
⇘ .....
2. John has **tentrophilia** and will not come to work for a few days.
⇘ .....
3. I think this **blouse** is too **crelly**.
⇘ .....

K. *Przetłumacz krótki dowcip na język polski:*

Doctor to Edward: This is really **amazing**! It looks like you may be the only **feltous** man in history!
Edward: Oh, this is terrible! What will my **remestians** say to this?

⇘ .....
.....
.....
.....

Appendix 6. Dictionary Effectiveness Test: English translation

Instruction: complete the following tasks. There are some fairly difficult words in it, but do not get discouraged: all words printed **in bold** are explained in the dictionary on the facing page. Use this dictionary.

H. *In each line below, cross (☒) **one** box out of four, choosing a Polish (lines 1-5) or English (lines 6 and 7) word most similar in meaning to the first word:*

1. **helk**	means:	☐ trafny	☐ właściwy	☐ miły	☐ taktowny
2. **sulk**	means:	☐ obrażać	☐ marzyć	☐ dąsać się	☐ kląć
3. **vab**	means:	☐ recenzować	☐ krytykować	☐ poprawiać	☐ oskarżać
4. **woozy**	means:	☐ zmęczony	☐ zawiany	☐ niecierpliwy	☐ malutki
5. **spating**	means:	☐ belka	☐ ławka	☐ rynna	☐ tynk
6. **strod**	means:	☐ **legalize**	☐ **ban**	☐ **punish**	☐ **introduce**
7. **luton**	means:	☐ **hat**	☐ **dress**	☐ **vase**	☐ **cover**

I. *Cross **one** box next to that English word, of the four words given, which best fits the meaning of the rest of the sentence:*

1. He would always …… at night, and the noise sometimes frightened his wife.
☐ **lenate** ☐ **sturton** ☐ **mirk** ☐ **reaff**

2. Her car must have very good …… if she can get from Berlin to Poznań in just three hours.
☐ **gerd** ☐ **marbish** ☐ **scolb** ☐ **cullen**

J. *Translate the sentences into Polish:*

1. The **boss** plans to **batenate** Mary tomorrow.
⇘ .....
2. John has **tentrophilia** and will not come to work for a few days.
⇘ .....
3. I think this **blouse** is too **crelly**.
⇘ .....

K. *Translate the short joke into Polish:*

Doctor to Edward: This is really **amazing**! It looks like you may be the only **feltous** man in history!
Edward: Oh, this is terrible! What will my **remestians** say to this?

⇘ .....
.....
.....

Appendix 7. Mini-dictionaries

Version 1

Słownik angielsko-polski

amazing [əˈmeɪzɪŋ] *adj.* **1.** zdumiewający **2.** niewiarygodny **3.** doskonały
ban [bæn] *v.* **1.** zdelegalizować
batenate [ˈbætəneɪt] *v.* **1.** podsmażyć **2.** awansować
blouse [blaʊz] *n.* bluzka
boss [bɒs] *n.* szef, szefowa
cover [ˈkʌvə] *n.* **1.** osłona **2.** pokrywka, pokrywa
crelly [ˈkreli] *adj.* **1.** przejrzysty **2.** przeźroczysty
cullen [ˈkʌlən] *n.* trzeźwość
dress [dres] *n.* **1.** sukienka **2.** ubiór
feltous [ˈfeltəs] *adj.* **1.** brzemienny, ciężarny, w ciąży **2.** zadłużony
frighten [ˈfraɪtən] *v.* **1.** przestraszyć
gerd [gɜːd] *n.* porost
hat [hæt] *n.* czapka
helk [helk] *adj.* **1.** cichy **2.** delikatny
introduce [ˈɪntrədjuːs] *v.* **1.** wprowadzić **2.** włożyć, wsadzić **3.** przedstawić
legalize [ˈliːgəlaɪz] *v.* zalegalizować
lenate [ləˈneɪt] *v.* niedowierzać
luton [ˈluːtən] *n.* kominiarka
marbish [ˈmɑːbɪʃ] *adj.* szybki
mirk [mɜːk] *v.* przypominać
punish [ˈpʌnɪʃ] *v.* karać
reaff [riːf] *v.* chrapać
remestian [rəˈmestɪən] *n.* **1.** *geogr.* państwo ościenne **2.** sąsiad **3.** *bud.* ścianka działowa
scolb [skəʊlb] *n.* przyspieszenie, zryw, zrywność
spating [ˈspeɪtɪŋ] *n.* **1.** fundament **2.** deska, listwa
strod [strɒd] *v.* zabronić
sturton [ˈstɜːtən] *v.* przybierać
sulk [sʌlk] *v.* boczyć się, dąsać się
tentrophilia [ˌtentrəˈfɪlɪə] *n.* **1.** *med.* katar **2.** *astron.* planeta zewnętrzna
vab [væb] *v.* **1.** kąsać **2.** krytykować
vase [vɑːz] *n.* wazon
woozy [ˈwuːzi] *adj.* **1.** podchmielony **2.** zdezorientowany

Version 2

English dictionary

amazing [əˈmeɪzɪŋ] *adj.* **1.** very surprising or unexpected **2.** very difficult to believe **3.** of very good quality

ban [bæn] *v.* **1.** to forbid *(something)* officially

batenate [ˈbætəneɪt] *v.* **1.** to fry *(something)* lightly on a medium-heat pan **2.** to give *(someone)* a better, more responsible job

blouse [blaʊz] *n.* a shirt for a woman or girl

boss [bɒs] *n.* a person in charge of workers

cover [ˈkʌvə] *n.* **1.** something that is put onto or over something else to protect it **2.** something that is put onto or over something else to keep dirt out

crelly [ˈkreli] *adj.* **1.** allowing a person to see through **2.** allowing light to pass through

cullen [ˈkʌlən] *n.* ability to think clearly and quickly

dress [dres] *n.* **1.** a piece of clothing for a woman or girl that covers her body from shoulders to legs **2.** the way someone dresses

feltous [ˈfeltəs] *adj.* **1.** having an unborn baby growing inside the body **2.** having unpaid debt

frighten [ˈfraɪtən] *v.* **1.** to cause *(someone)* to be afraid

gerd [gɜːd] *n.* primitive plant of yellow or grey colour growing on rocks

hat [hæt] *n.* any piece of clothing that people wear on the head

helk [helk] *adj.* **1.** speaking in a soft voice and not a lot **2.** avoiding saying things which might hurt other people

introduce [ˈɪntrədjuːs] *v.* **1.** to put *(something)* into use, operation or a place for the first time **2.** to put *(something)* into an opening or hole **3.** to give the name of *(someone)* to another person when they first meet

legalize [ˈliːgəlaɪz] *v.* to make a law which allows people to do *(sth that was not allowed before)*

lenate [ləˈneɪt] *v.* to find it difficult to believe *(someone)*

luton [ˈluːtən] *n.* a piece of clothing worn on the head and usually over most of the face

marbish [ˈmɑːbɪʃ] *adj.* capable of fast movement

mirk [mɜːk] *v.* to look similar to *(something else)*

punish [ˈpʌnɪʃ] *v.* to make *(someone)* suffer because they have done something wrong or broken the law

reaff [riːf] *v.* to breathe noisily when sleeping

remestian [rəˈmestɪən] *n.* **1.** *geogr.* a country directly bordering on another country **2.** a person who lives next to you or near you **3.** *bud.* a thin wall separating two rooms

scolb [skəʊlb] *n.* the ability of a car or other vehicle to increase its speed rapidly

spating [ˈspeɪtɪŋ] *n.* **1.** the solid layer of cement, bricks, or stones that is under a building to support it **2.** a long narrow strip of wood used for construction

strod [strɒd] *v.* no longer allow*(something)*

sturton [ˈstɜːtən] *v.* *(of a river)* to swell with water threatening to flood its banks

sulk [sʌlk] *v.* to show that you are annoyed by being silent and having an unhappy expression on your face

tentrophilia [ˌtentrəˈfɪlɪə] *n.* **1.** *med.* a slight illness that makes it difficult to breathe through the nose and causes liquid to flow from the nose **2.** *astron.* a planet whose orbit lies beyond the orbit of Jupiter

vab [væb] *v.* **1.** *(of an insect or animal)* to bite painfully and quickly **2.** to express disapproval of *(someone or something)* by talking about their faults

vase [vɑːz] *n.* a tall container used for putting flowers in or for decoration

woozy [ˈwuːzi] *adj.* **1.** slightly drunk **2.** having confused or unclear thoughts

Version 3

English dictionary for Polish learners

amazing [əˈmeɪzɪŋ] *adj.* **1.** zdumiewający <=very surprising or unexpected> **2.** niewiarygodny <=very difficult to believe> **3.** doskonały <=of very good quality>

ban [bæn] *v.* **1.** zdelegalizować <=to forbid *(something)* officially>

batenate [ˈbætəneɪt] *v.* **1.** podsmażyć <=to fry *(something)* lightly on a medium-heat pan> **2.** awansować <=to give *(someone)* a better, more responsible job>

blouse [blaʊz] *n.* bluzka <=a shirt for a woman or girl>

boss [bɒs] *n.* szef, szefowa <=a person in charge of workers>

cover [ˈkʌvə] *n.* **1.** osłona <=something that is put onto or over something else to protect it> **2.** pokrywka, pokrywa <=something that is put onto or over something else to keep dirt out>

crelly [ˈkreli] *adj.* **1.** przejrzysty <=allowing a person to see through> **2.** przeźroczysty <=allowing light to pass through>

cullen [ˈkʌlən] *n.* trzeźwość <=ability to think clearly and quickly>

dress [dres] *n.* **1.** sukienka <=a piece of clothing for a woman or girl that covers her body from shoulders to legs> **2.** ubiór <=the way someone dresses>

feltous [ˈfeltəs] *adj.* **1.** brzemienny, ciężarny, w ciąży <=having an unborn baby growing inside the body> **2.** zadłużony <=having unpaid debt>

frighten [ˈfraɪtən] *v.* **1.** przestraszyć <=to cause *(someone)* to be afraid>

gerd [gɜːd] *n.* porost <=primitive plant of yellow or grey colour growing on rocks>

hat [hæt] *n.* czapka <=any piece of clothing that people wear on the head>

helk [helk] *adj.* **1.** cichy <=speaking in a soft voice and not a lot> **2.** delikatny <=avoiding saying things which might hurt other people>

introduce [ˌɪntrəˈdjuːs] *v.* **1.** wprowadzić <=to put *(something)* into use, operation or a place for the first time> **2.** włożyć, wsadzić <=to put *(something)* into an opening or hole> **3.** przedstawić <=to give the name of *(someone)* to another person when they first meet>

legalize [ˈliːgəlaɪz] *v.* zalegalizować <=to make a law which allows people to do *(sth that was not allowed before)*>

lenate [ləˈneɪt] *v.* niedowierzać <=to find it difficult to believe *(someone)*>

luton [ˈluːtən] *n.* kominiarka <=a piece of clothing worn on the head and usually over most of the face>

marbish [ˈmɑːbɪʃ] *adj.* szybki <=capable of fast movement>

mirk [mɜːk] *v.* przypominać <=to look similar to *(something else)*>

punish [ˈpʌnɪʃ] *v.* karać <=to make *(someone)* suffer because they have done something wrong or broken the law>

reaff [riːf] *v.* chrapać <=to breathe noisily when sleeping>

remestian [rəˈmestɪən] *n.* **1.** *geogr.* państwo ościenne <=a country directly bordering on another country> **2.** sąsiad <=a person who lives next to you or near you> **3.** *bud.* ścianka działowa <=a thin wall separating two rooms>

scolb [skəʊlb] *n.* przyspieszenie, zryw, zrywność <=the ability of a car or other vehicle to increase its speed rapidly>

spating [ˈspeɪtɪŋ] *n.* **1.** fundament <=the solid layer of cement, bricks, or stones that is under a building to support it> **2.** deska, listwa <=a long narrow strip of wood used for construction>

strod [strɒd] *v.* zabronić <=no longer allow*(something)*>

sturton [ˈstɜːtən] *v.* przybierać <=*(of a river)* to swell with water threatening to flood its banks>

sulk [sʌlk] *v.* boczyć się, dąsać się <=to show that you are annoyed by being silent and having an unhappy expression on your face>

tentrophilia [ˌtentrəˈfɪlɪə] *n.* **1.** *med.* katar <=a slight illness that makes it difficult to breathe through the nose and causes liquid to flow from the nose> **2.** *astron.* planeta zewnętrzna <=a planet whose orbit lies beyond the orbit of Jupiter>

vab [væb] *v.* **1.** kąsać <=*(of an insect or animal)* to bite painfully and quickly> **2.** krytykować <=to express disapproval of *(someone or something)* by talking about their faults>

vase [vɑːz] *n.* wazon <=a tall container used for putting flowers in or for decoration>

woozy [ˈwuːzi] *adj.* **1.** podchmielony <=slightly drunk> **2.** zdezorientowany <=having confused or unclear thoughts>

Version 4

Słownik angielsko-polski

amazing [əˈmeɪzɪŋ] *adj.* **1.** zdumiewający <=bardzo zaskakujący lub nieoczekiwany> **2.** niewiarygodny <=bardzo trudny do uwierzenia> **3.** doskonały <=bardzo dobrej jakości>

ban [bæn] *v.* **1.** zdelegalizować <=oficjalnie zabronić *(czegoś)*>

batenate [ˈbætəneɪt] *v.* **1.** podsmażyć <=lekko *(coś)* usmażyć na umiarkowanie nagrzanej patelni> **2.** awansować <=dać *(komuś)* atrakcyjniejsze, bardziej odpowiedzialne stanowisko>

blouse [blaʊz] *n.* bluzka <=górna część ubrania kobiet lub dziewcząt; odpowiednik koszuli u mężczyzn>

boss [bɒs] *n.* szef, szefowa <=osoba na kierowniczym stanowisku w pracy>

cover [ˈkʌvə] *n.* **1.** osłona <=coś, co się nakłada na coś innego dla ochrony przed zniszczeniem> **2.** pokrywka, pokrywa <=coś, co się nakłada na coś innego dla ochrony przed zakurzeniem>

crelly [ˈkreli] *adj.* **1.** przejrzysty <=dający się przeniknąć wzrokiem> **2.** przeźroczysty <=przepuszczający światło>

cullen [ˈkʌlən] *n.* trzeźwość <=zdolność do jasnego i szybkiego myślenia>

dress [dres] *n.* **1.** sukienka <=artykuł odzieży noszony przez kobiety i dziewczynki zakrywający ciało od ramion do nóg> **2.** ubiór <=sposób ubierania się>

feltous [ˈfeltəs] *adj.* **1.** brzemienny, ciężarny, w ciąży <=noszący w sobie jeszcze nie narodzone dziecko> **2.** zadłużony <=posiadający zaległe zobowiązania finansowe>

frighten [ˈfraɪtən] *v.* **1.** przestraszyć <=wywoływać u *(kogoś)* uczucie strachu>

gerd [gɜːd] *n.* porost <=prymitywna roślina koloru zielonkawego lub szarego rosnąca na skałach>

hat [hæt] *n.* czapka <=artykuł odzieży, który nosi się na głowie>

helk [helk] *adj.* **1.** cichy <=mówiący cichym głosem i niewiele> **2.** delikatny <=starający się nie mówić o rzeczach, które mogłyby urazić innych ludzi>

introduce [ˈɪntrədjuːs] *v.* **1.** wprowadzić <=spowodować, że *(coś)* zacznie być obecne lub używane w nowym miejscu> **2.** włożyć, wsadzić <=umieścić *(coś)* w otworze> **3.** przedstawić <=podać imię lub nazwisko *(kogoś)* innej osobie na początku spotkania>

legalize [ˈliːgəlaɪz] *v.* zalegalizować <=zezwolić aktem prawnym *(coś, co było uprzednio zabronione)*>

lenate [ləˈneɪt] *v.* niedowierzać <=mieć wątpliwości, czy *(ktoś)* mówi prawdę>

luton [ˈluːtən] *n.* kominiarka <=artykuł odzieży noszony na głowie i zakrywający zwykle większość twarzy>

marbish [ˈmɑːbɪʃ] *adj.* szybki <=mogący się szybko przemieszczać>

mirk [mɜːk] *v.* przypominać <=wyglądać podobnie do *(czegoś innego)*>

punish [ˈpʌnɪʃ] *v.* karać <=zadawać cierpienie *(komuś, kto czynił źle lub złamał prawo)*>

reaff [riːf] *v.* chrapać <=oddychać w hałaśliwy sposób w czasie snu>

remestian [rəˈmestɪən] *n.* **1.** *geogr.* państwo ościenne <=państwo sąsiadujące bezpośrednio z danym państwem> **2.** sąsiad <=osoba mieszkająca obok lub niedaleko> **3.** *bud.* ścianka działowa <=cienka ściana niekonstrukcyjna pomiędzy pomieszczeniami>

scolb [skəʊlb] *n.* przyspieszenie, zryw, zrywność <=zdolność samochodu lub innego pojazdu do gwałtownego zwiększania szybkości jazdy>

spating [ˈspeɪtɪŋ] *n.* **1.** fundament <=twarda podstawa budynku wykonana z betonu, cegieł lub kamieni> **2.** deska, listwa <=długi, wąski kawałek drewna używany w budownictwie>

strod [strɒd] *v.* zabronić <=przestać na *(coś)* zezwalać>

sturton [ˈstɜːtən] *v.* przybierać <=*(o rzece)* wypełniać się wodą grożąc wylaniem>

sulk [sʌlk] *v.* boczyć się, dąsać się <=okazywać niezadowolenie małomównością i nieszczęśliwym wyrazem twarzy>

tentrophilia [ˌtentrəˈfɪlɪə] *n.* **1.** *med.* katar <=stan chorobowy charakteryzujący się niedrożnością nosa i cieknięciem z nosa> **2.** *astron.* planeta zewnętrzna <=planeta, której orbita leży poza orbitą Jowisza>

vab [væb] *v.* **1.** kąsać <=*(o zwierzęciu, owadzie)* ugryźć dotkliwie i nieoczekiwanie> **2.** krytykować <=wyrażać niezadowolenie z *(czegoś lub kogoś)* przez wskazywanie na ich wady>

vase [vɑːz] *n.* wazon <=wydłużone naczynie na kwiaty cięte lub do dekoracji>

woozy [ˈwuːzi] *adj.* **1.** podchmielony <=lekko pijany> **2.** zdezorientowany <=mający pomieszane lub niejasne myśli>

Version 5

English dictionary for Polish learners

amazing [əˈmeɪzɪŋ] *adj.* **1.** <=very surprising or unexpected> zdumiewający **2.** <=very difficult to believe> niewiarygodny **3.** <=of very good quality> doskonały

ban [bæn] *v.* **1.** <=to forbid *(something)* officially> zdelegalizować

batenate [ˈbætəneɪt] *v.* **1.** <=to fry *(something)* lightly on a medium-heat pan> podsmażyć **2.** <=to give *(someone)* a better, more responsible job> awansować

blouse [blaʊz] *n.* <=a shirt for a woman or girl> bluzka

boss [bɒs] *n.* <=a person in charge of workers> szef, szefowa

cover [ˈkʌvə] *n.* **1.** <=something that is put onto or over something else to protect it> osłona **2.** <=something that is put onto or over something else to keep dirt out> pokrywka, pokrywa

crelly [ˈkreli] *adj.* **1.** <=allowing a person to see through> przejrzysty **2.** <=allowing light to pass through> przeźroczysty

cullen [ˈkʌlən] *n.* <=ability to think clearly and quickly> trzeźwość

dress [dres] *n.* **1.** <=a piece of clothing for a woman or girl that covers her body from shoulders to legs> sukienka **2.** <=the way someone dresses> ubiór

feltous [ˈfeltəs] *adj.* **1.** <=having an unborn baby growing inside the body> brzemienny, ciężarny, w ciąży **2.** <=having unpaid debt> zadłużony

frighten [ˈfraɪtən] *v.* **1.** <=to cause *(someone)* to be afraid> przestraszyć

gerd [gɜːd] *n.* <=primitive plant of yellow or grey colour growing on rocks> porost

hat [hæt] *n.* <=any piece of clothing that people wear on the head> czapka

helk [helk] *adj.* **1.** <=speaking in a soft voice and not a lot> cichy **2.** <=avoiding saying things which might hurt other people> delikatny

introduce [ˌɪntrəˈdjuːs] *v.* **1.** <=to put *(something)* into use, operation or a place for the first time> wprowadzić **2.** <=to put *(something)* into an opening or hole> włożyć, wsadzić **3.** <=to give the name of *(someone)* to another person when they first meet> przedstawić

legalize [ˈliːgəlaɪz] *v.* <=to make a law which allows people to do *(sth that was not allowed before)*> zalegalizować

lenate [ləˈneɪt] *v.* <=to find it difficult to believe *(someone)*> niedowierzać

luton [ˈluːtən] *n.* <=a piece of clothing worn on the head and usually over most of the face> kominiarka

marbish [ˈmɑːbɪʃ] *adj.* <=capable of fast movement> szybki

mirk [mɜːk] *v.* <=to look similar to *(something else)*> przypominać

punish [ˈpʌnɪʃ] *v.* <=to make *(someone)* suffer because they have done something wrong or broken the law> karać

reaff [riːf] *v.* <=to breathe noisily when sleeping> chrapać

remestian [rəˈmestɪən] *n.* **1.** *geogr.*<=a country directly bordering on another country> państwo ościenne **2.** <=a person who lives next to you or near you> sąsiad **3.** *bud.*<=a thin wall separating two rooms> ścianka działowa

scolb [skəʊlb] *n.* <=the ability of a car or other vehicle to increase its speed rapidly> przyspieszenie, zryw, zrywność

spating [ˈspeɪtɪŋ] *n.* **1.** <=the solid layer of cement, bricks, or stones that is under a building to support it> fundament **2.** <=a long narrow strip of wood used for construction> deska, listwa

strod [strɒd] *v.* <=no longer allow*(something)*> zabronić

sturton [ˈstɜːtən] *v.* <=*(of a river)* to swell with water threatening to flood its banks> przybierać

sulk [sʌlk] *v.* <=to show that you are annoyed by being silent and having an unhappy expression on your face> boczyć się, dąsać się

tentrophilia [ˌtentrəˈfɪlɪə] *n.* **1.** *med.*<=a slight illness that makes it difficult to breathe through the nose and causes liquid to flow from the nose> katar **2.** *astron.* <=a planet whose orbit lies beyond the orbit of Jupiter> planeta zewnętrzna

vab [væb] *v.* **1.** <=*(of an insect or animal)* to bite painfully and quickly> kąsać **2.** <=to express disapproval of *(someone or something)* by talking about their faults> krytykować

vase [vɑːz] *n.* <=a tall container used for putting flowers in or for decoration> wazon

woozy [ˈwuːzi] *adj.* **1.** <=slightly drunk> podchmielony **2.** <=having confused or unclear thoughts> zdezorientowany

Version 6

Słownik angielsko-polski

amazing [əˈmeɪzɪŋ] *adj.* **1.** <=bardzo zaskakujący lub nieoczekiwany> zdumiewający **2.** <=bardzo trudny do uwierzenia> niewiarygodny **3.** <=bardzo dobrej jakości> doskonały

ban [bæn] *v.* **1.** <=oficjalnie zabronić *(czegoś)*> zdelegalizować

batenate [ˈbætəneɪt] *v.* **1.** <=lekko *(coś)* usmażyć na umiarkowanie nagrzanej patelni> podsmażyć **2.** <=dać *(komuś)* atrakcyjniejsze, bardziej odpowiedzialne stanowisko> awansować

blouse [blaʊz] *n.* <=górna część ubrania kobiet lub dziewcząt; odpowiednik koszuli u mężczyzn> bluzka

boss [bɒs] *n.* <=osoba na kierowniczym stanowisku w pracy> szef, szefowa

cover [ˈkʌvə] *n.* **1.** <=coś, co się nakłada na coś innego dla ochrony przed zniszczeniem> osłona **2.** <=coś, co się nakłada na coś innego dla ochrony przed zakurzeniem> pokrywka, pokrywa

crelly [ˈkreli] *adj.* **1.** <=dający się przeniknąć wzrokiem> przejrzysty **2.** <=przepuszczający światło> przeźroczysty

cullen [ˈkʌlən] *n.* <=zdolność do jasnego i szybkiego myślenia> trzeźwość

dress [dres] *n.* **1.** <=artykuł odzieży noszony przez kobiety i dziewczynki zakrywający ciało od ramion do nóg> sukienka **2.** <=sposób ubierania się> ubiór

feltous [ˈfeltəs] *adj.* **1.** <=noszący w sobie jeszcze nie narodzone dziecko> brzemienny, ciężarny, w ciąży **2.** <=posiadający zaległe zobowiązania finansowe> zadłużony

frighten [ˈfraɪtən] *v.* **1.** <=wywoływać u *(kogoś)* uczucie strachu> przestraszyć

gerd [gɜːd] *n.* <=prymitywna roślina koloru zielonkawego lub szarego rosnąca na skałach> porost

hat [hæt] *n.* <=artykuł odzieży, który nosi się na głowie> czapka

helk [helk] *adj.* **1.** <=mówiący cichym głosem i niewiele> cichy **2.** <=starający się nie mówić o rzeczach, które mogłyby urazić innych ludzi> delikatny

introduce [ˌɪntrəˈdjuːs] *v.* **1.** <=spowodować, że *(coś)* zacznie być obecne lub używane w nowym miejscu> wprowadzić **2.** <=umieścić *(coś)* w otworze> włożyć, wsadzić **3.** <=podać imię lub nazwisko *(kogoś)* innej osobie na początku spotkania> przedstawić

legalize [ˈliːgəlaɪz] *v.* <=zezwolić aktem prawnym *(coś, co było uprzednio zabronione)*> zalegalizować

lenate [laˈneɪt] *v.* <=mieć wątpliwości, czy *(ktoś)* mówi prawdę> niedowierzać

luton [ˈluːtən] *n.* <=artykuł odzieży noszony na głowie i zakrywający zwykle większość twarzy> kominiarka

marbish [ˈmɑːbɪʃ] *adj.* <=mogący się szybko przemieszczać> szybki

mirk [mɜːk] *v.* <=wyglądać podobnie do *(czegoś innego)*> przypominać

punish [ˈpʌnɪʃ] *v.* <=zadawać cierpienie *(komuś, kto czynił źle lub złamał prawo)*> karać

reaff [riːf] *v.* <=oddychać w hałaśliwy sposób w czasie snu> chrapać

remestian [rəˈmestɪən] *n.* **1.** *geogr.*<=państwo sąsiadujące bezpośrednio z danym państwem> państwo ościenne **2.** <=osoba mieszkająca obok lub niedaleko> sąsiad **3.** *bud.*<=cienka ściana niekonstrukcyjna pomiędzy pomieszczeniami> ścianka działowa

scolb [skəʊlb] *n.* <=zdolność samochodu lub innego pojazdu do gwałtownego zwiększania szybkości jazdy> przyspieszenie, zryw, zrywność

spating [ˈspeɪtɪŋ] *n.* **1.** <=twarda podstawa budynku wykonana z betonu, cegieł lub kamieni> fundament **2.** <=długi, wąski kawałek drewna używany w budownictwie> deska, listwa

strod [strɒd] *v.* <=przestać na *(coś)* zezwalać> zabronić

sturton [ˈstɜːtən] *v.* <=*(o rzece)* wypełniać się wodą grożąc wylaniem> przybierać

sulk [sʌlk] *v.* <=okazywać niezadowolenie małomównością i nieszczęśliwym wyrazem twarzy> boczyć się, dąsać się

tentrophilia [ˌtentrəˈfɪlɪə] *n.* **1.** *med.*<=stan chorobowy charakteryzujący się niedrożnością nosa i cieknięciem z nosa> katar **2.** *astron.* <=planeta, której orbita leży poza orbitą Jowisza> planeta zewnętrzna

vab [væb] *v.* **1.** <=*(o zwierzęciu, owadzie)* ugryźć dotkliwie i nieoczekiwanie> kąsać **2.** <=wyrażać niezadowolenie z *(czegoś lub kogoś)* przez wskazywanie na ich wady> krytykować

vase [vɑːz] *n.* <=wydłużone naczynie na kwiaty cięte lub do dekoracji> wazon

woozy [ˈwuːzi] *adj.* **1.** <=lekko pijany> podchmielony **2.** <=mający pomieszane lub niejasne myśli> zdezorientowany

Appendix 8. Educational institutions participating in the study

Educational Institution	Subjects	Percent
Gimnazjum w Uniecku	20	2.8
Gimnazjum w Raciążu	24	3.4
Gimnazjum 2 w Murowanej Goślinie	42	5.9
LO w Pniewach	46	6.5
LO w Sulechowie	30	4.2
LO, ZSZ w Kościanie	22	3.1
LO w Puszczykowie	45	6.3
IX LO w Poznaniu	54	7.6
XV LO w Poznaniu	25	3.5
LO w Kole	62	8.7
Liceum Ekonomiczne w Poznaniu	27	3.8
ZSZ 1 we Wrześni, Liceum Administracyjno-Biurowe	56	7.9
NKJO w Poznaniu	14	2.0
SJO Europa Szczecin	40	5.6
SP w Iłówcu	35	4.9
Gminna Szkoła Podstawowa w Czerniejewie	22	3.1
Wyższa Szkoła Oficerska w Poznaniu	26	3.7
Akademia Sztuk Wizualnych w Poznaniu	26	3.7
ZSZ w Kole, Technikum Chemiczne	17	2.4
ZSZ w Kole, Technikum Ochrony Środowiska	13	1.8
ZSZ 1 we Wrześni, Technikum Elektroniczne	32	4.5
UAM	34	4.8

Appendix 9. Database dictionary codes

Dictionary Code	Type	Long Description
AHD	m	American Heritage Dictionary
ALD	m	Oxford Advanced Learners Dictionary
BBC	m	BBC English Dictionary
BBI	m	BBI Combinatory Dictionary
BGW	b	Collins-BGW, Fisiak J.
BGWCD	b	YDP; Fisiak J., Collins-BGW on CD-ROM
BGWEP	b	Collins-BGW, Fisiak J., English-Polish
BGWPE	b	Collins-BGW, Fisiak J., Polish-English
Bi	b	Other unspecified bilingual
BiEP	b	Other unspecified bilingual English-Polish
BiPE	b	Other unspecified bilingual Polish-English
CIDE	m	Cambridge International Dictionary of English
Cobuild	m	COBUILD (any)
CobuildCD	b	Cobuild on CD
Coll	m	Collocation dictionary (any)
Collins		Collins unspecified
ElBi	b	Electronic bilingual
Electronic		Electronic
ElEP	b	Electronic English-Polish
ElPE	b	Electronic Polish-English
Etranslator	b	English Translator
Grzebieniowski	b	Grzebieniowski T.
Idioms	m	Dictionary of idioms
Idiomy	b	Słownik idiomów i zwrotów
Jaworska	b	Sł. ang-pol.WNT, Jaworska, T.
Kałuża	b	Kałuża, J., Exlibris
Kiesz	b	Kieszonkowy, Jaślan J. & Stanisławski J.
Langenscheidt	b	Langenscheidt's Pocket English Dictionary, Grzebieniowski T.
LASD	b	Longman Active Study Dictionary
LDLC	m	Longman Dictionary of English Language and Culture
LDOCE	m	Longman Dictionary of Contemporary English
LLA	m	Longman Language Activator
LongMono	m	Longman monolingual
LongPocket	m	Longman Pocket
LongPodr	b	Longman Podręczny, Fisiak J. et al.
LPD		Longman Pronunciation Dictionary, Wells, J.C.
LTT	b	Language Teacher Translator (Electronic)

Dictionary Code	Type	Long Description
Mini	b	Minimum/Mini
Mizgalski	b	Mizgalski, E.
Mono	m	Other unspecified monolingual
NewHot	b	New Hotline Elementary (coursebook glossary)
Other		Other unspecified
OxCon	m	Oxford Concise Dictionary
Oxford		Oxford unspecified
OxMono	m	Oxford monolingual
OxPocket	m	Oxford Pocket Dictionary
OxQuick	m	Oxford Quick Reference and Thesaurus
OxStud	m	Oxford Student's Dictionary
OxWord	b	Oxford Wordpower English-Polish
Penguin	m	New Penguin English Dictionary
Piotrowski	b	Piotrowski T. Saloni, Z., WILGA
Podr	b	Podręczny unspecified
RHC	m	Random House College Dictionary
Saloni	b	Piotrowski T., Saloni Z., Mini PWN czerwony
SOED	m	Shorter Oxford English Dictionary
STAG	b	Wielki Słownik PA i PA Stanisławski J.
StanP	b	Podręczny Słownik AP i PA Stanisławski J.
Tematyczny	b	Słownik tematyczny
Translator	b	Translator
Turystyczny	b	Słownik turystyczny
Uniwersalny	b	Uniwersalny, Grzebieniowski T. & Kaznowski A.
Webster	m	Webster unspecified
Wtranslator	b	Word Translator

Note: The table reflects dictionary identification information as provided by subjects. Therefore, it is not meant to give consistent, detailed and complete bibliographical data. The specificity of the descriptions varies widely, with, at best, specific titles (though never editions) identified, but more often just the publisher or the author, or, at worst, just catch-all categories such as *Dictionary of idioms*, *Webster unspecified* or simply *Other unspecified*.