

Grzegorz Zarzeczny (ORCID 0000-0002-5259-6753)
Uniwersytet Wrocławski, Poland

Vocabulary in selected textbooks for teaching Polish as a Foreign Language (A diachronic perspective)

Abstract

This study examines the vocabulary used in textbooks for Teaching Polish as a Foreign Language from a diachronic perspective. The analysis is based on the premise that contemporary vocabulary lists, which aim to reflect the lexicon essential for everyday communication, may serve as an indicator of the pedagogical priorities and assumptions underlying textbook design.

The corpus comprises selected TPFL textbooks published in the United States and the United Kingdom over a 60-year period following World War II. For the purposes of analysis, only those segments of the textbooks explicitly intended to expose learners to the target language were included, yielding a total of 112,246 word forms. The vocabulary of each textbook was compared against the list of word families proposed by Seretny (2011).

The findings are: (1) The proportion of high-frequency vocabulary in a textbook is largely dependent on its total textual volume. (2) No distinct groupings of textbooks can be identified based on the ratio of high-utility to low-utility vocabulary. (3) Three main lexical strategies can be observed: (a) extensive use of core vocabulary accompanied by a substantial amount of peripheral lexicon, (b) a limited selection from the core vocabulary, which nevertheless represents a significant proportion of the textbook's overall lexicon, (c) a balanced approach, with moderate representation of both high- and low-frequency vocabulary.

Keywords: Polish as a Foreign Language, foreign language textbook, vocabulary.

1. Introduction

The subject of this analysis is the vocabulary used in textbooks for teaching Polish as a Foreign Language (TPFL) from a diachronic perspective.¹ The very premise that the vocabulary taught should not be selected randomly is by no means a novel idea, although the criteria applied by those responsible for selection may vary. Today, vocabulary selection is most commonly associated with communicative needs, which necessitates the development of word lists comprising the most useful items for everyday communication – primarily those with the highest frequency of occurrence. This assumption is often linked to the rise of the communicative approach in foreign language teaching; nonetheless, it was not entirely absent from earlier stages of glottodidactic reflection. What fundamentally distinguishes more recent approaches is the systematic (scientifically grounded) selection of vocabulary, as opposed to a more intuitive method.

While frequency-motivated and pedagogically oriented word lists have existed for many languages for several decades, in the case of Polish, the first (grounded in research) compilation of this kind – by Halina Zgółkowa – dates back only to the late 20th century (Zgółkowa 1992; see also a critical discussion of earlier compilations therein). The scholar based her list, among other sources, on data from Polish word frequency lists (Kurcz et al. 1974–1979) and on her own research on spoken Polish (Zgółkowa 1983), supplemented with the so called “criterion of essential needs” (*‘kryterium potrzeb życiowych’*). Her compilation of a minimum vocabulary and two lists comprising a basic vocabulary of the Polish language remained fundamental to TPFL for an extended period. Only in recent years has a newer model emerged – largely building on the previous one – in the form of word families² lists proposed by Anna Seretny (2011).

Naturally, from a historical perspective, there have also been instances – particularly within the grammar-translation approach – where vocabulary was treated as a secondary concern relative to other instructional priorities. As Michael Swan observes: “My first German textbook (a reprint of a very old course) made a reasonable job of sequencing the structures of

¹ The analysis is a part of a broader research conducted for the doctoral dissertation “Dyskurs glottodydaktyczny w podręcznikach do nauczania języka polskiego jako obcego wydanych w USA i Wielkiej Brytanii po 1945 r.” written at the University of Wrocław under the supervision of Anna Dąbrowska.

² “Word family” (Bauer/Nation 1993). The Polish term used by Seretny is “grono wyrazowe”.

the language, but vocabulary selection was subordinated to the exigencies of a plot-line based on German mythology: one of the first nouns I learnt was *der Greif*: ‘the gryphon’. Vocabulary selection based on frequency counts came in later.” (Swan 2009:129).

Moreover, even after research-based vocabulary lists became available, they have often been used only to a limited extent in the creation of TPFL materials. This point is aptly emphasized by Seretny in her discussion of Zgółkowa’s pioneering compilation: “The authors of teaching materials relied on the lists only sporadically, and many teachers were unaware of their existence. Thus, the recommendations issued by the Expert Committee concerning vocabulary instruction found very limited application in everyday teaching practice, in contrast to the guidelines pertaining to grammatical issues (e.g., the sequence in which grammatical cases are introduced) or language functions (Seretny 2011:175)”.³

The foundation of the present analysis is therefore the following assumption: if contemporary vocabulary lists most accurately reflect the lexicon necessary for everyday communication, they may serve to a considerable extent as a litmus test of the underlying assumptions adopted by a textbook’s author. In other words, a high degree of correspondence between the vocabulary included in such a list and that found in a textbook would suggest an effort to reflect everyday communication in the material. Conversely, a low level of correspondence would indicate either the prioritization of other aspects (which, importantly, should not be judged negatively *per se*) or – though only secondarily – a lack of skill in selecting the lexical input appropriately.

The present study is exploratory in nature – above all, there is virtually no existing knowledge regarding the selection of vocabulary used in older TPFL textbooks. The aim, therefore, is to examine whether any patterns or regularities emerge in this respect that could be interpreted as (conscious or unconscious) lexical strategies, for example:

- depending on the period of publication of the textbook,
- depending on the language teaching approach employed.

It is important to keep in mind that, in a diachronic comparison, minor differences in results are not significant, as they may simply reflect the time gap between a given textbook and the contemporary vocabulary list. In other words, older textbooks will inevitably lack words such as *television*

³ All quotations from Polish texts are in my translation.

or *computer*. On the other hand, such items cannot be excluded from the current list, as they may be relevant for more recent textbooks. The decision to use a single, unmodified list for all texts is arbitrary but appears to be the most reasonable for comparative purposes. This is primarily due to the fact that the stem of the most frequently used vocabulary has remained essentially stable throughout history – the differences mentioned above concern only a small percentage of vocabulary.

2. Theoretical context

A TPFL textbook constitutes one of the (macro)genres within Foreign/Second Language Teaching (FLT/SLT) discourse, that is, a specific variety of educational discourse. The latter is defined, following Jolanta Nocoń, as: “a characteristic interaction of the *specialist–novice* type (with the archetypal roles of teacher and student) and an educational function – a specific communicative aim that may be described as one party’s (the teacher’s) intention to bring about a particular personality-related change in the interaction partner” (Nocoń 2011:190).

Following Basil Bernstein, this type of discourse is understood here to consist of two main dimensions: the “instructional” and the “regulative”. The former encompasses the content of instruction (concepts, skills, etc.), while the latter – a superior one – is responsible for reproducing the norms and values upheld by a given society, or more precisely, by that segment of society which shapes the character of the regulative discourse (Bernstein 1990, Bielecka-Prus 2010). Within this framework, the vocabulary selected by the authors of individual textbooks is primarily an element of the instructional dimension of the discourse, although – as already mentioned – it is to some extent conditioned by the regulative one.

3. Material

The empirical foundation for the analysis presented in this study consists of a selection of materials published in the United States and the United Kingdom within a 60-year period following the end of the Second World War. This scope covers the majority of textbooks (including all the most significant ones) for TPFL aimed at English speakers and published outside Poland. Omitted are several pre-war textbooks that had fallen out of use after 1945 and had previously been employed primarily in local contexts. The selection criteria included genre prototypicality (general textbooks) and scope uniformity (texts intended for beginners).

Due to the wide time span over which the selected textbooks were produced, they reflect different FLT/SLT approaches, including the direct, grammar-translation, audiolingual, and communicative approaches. Sometimes the teaching approach is explicitly signalled in the very title of the publication: for instance, “Polish Grammar” by Bolanowski or Teslar reflects a focus on language teaching understood primarily as the teaching of grammar, in contrast to e.g. Mazur’s “Colloquial Polish” which emphasizes language as a tool for communication.⁴ Nevertheless, these are always general, introductory textbooks representative of their respective periods.⁵ A list of the analyzed textbooks, along with their assigned abbreviations, is provided at the end of the text.

For the purposes of comparison, a corpus was compiled consisting of those parts of the textbooks intended to expose the learner to the target language.⁶ The structure of these materials depends on various factors, most importantly on the FLT/SLT approach adopted by the textbook author. These may include, to name just a few key oppositions: narrative or dialogic texts, original or adapted texts, coherent texts or unrelated individual sentences, and so forth. In each case, however, the essential function of this component is the presentation of the target language (and usually the target language alone), either in its form or in action.

This choice is motivated, first, by the relative comparability of such components in terms of word frequency, which is not always the case with other parts of a foreign language textbook. Second, it may be assumed that it is within these texts that vocabulary of greater relevance to language instruction is likely to appear, whereas in later components of a lesson unit – for example, in exercises – the same vocabulary is typically repeated (for reinforcement), and new items that do appear are often of lesser importance. Third, these texts are, at least in principle, intended to represent actual examples of the target language in use.

The corpus comprises approximately 100,000 words (i.e., units counted from space to space) and may be classified as a small corpus according to Sinclair’s (2001) definition. This dataset was processed automatically using the TaKIPI tagger (Piasecki 2007) and the Morfeusz morphologi-

⁴ The title of Mazur’s textbook also reflects, independently of its content, a publishing decision – it was released as part of Routledge’s “Colloquial...” series.

⁵ A separate type of publication are proper grammars intended for students – these are not included in the scope of the present analysis.

⁶ From the textbook’s genre composition perspective, this constitutes *introductory texts* (Nocoń 2009:67-69) specific for FLT/SLT.

cal analyzer (Woliński 2006). As a result, each word form was assigned a grammatical interpretation in accordance with the IPI PAN tagset (i.e., the set of morphosyntactic tags developed by the Institute of Computer Science of the Polish Academy of Sciences). The entire corpus contains over 143,000 tokens; after excluding nearly 30,000 punctuation marks, this yields 112,246 word forms.

Naturally, automatic grammatical annotation is not entirely error-free. For the TaKIPI tagger, the accuracy rate is around 90%, which implies a relatively significant – but still acceptable – margin of error. This must be kept in mind when interpreting the results of further analyses.

4. Analysis

As already noted, the most current and robust compilation of vocabulary necessary at the initial stages of TPFL is the alphabetical list of lexical families for Range I (1-1000), compiled by Seretny (2011, Annex 2: 114-122). This list includes just over 1,200 words and corresponds to the vocabulary required for A1 level.⁷ For the purposes of this study, the list was adjusted to match the corpus data.⁸ The frequency lists generated for each subcorpus were then automatically compared with the contents of Seretny's list of word families.

The table below presents the comparison results for the individual textbooks. The first column shows the year of publication, the next two columns indicate the number of word tokens (W) and word types (i.e., lexemes; L), respectively. The fourth column (LS) provides the number of lexemes shared by the given textbook and Seretny's list.

The first step was to determine what proportion of the vocabulary from the list (S) could be found in each textbook ($T = \%S$). This percentage is reported in the final column.

⁷ Research by Nation (2001:15) shows that in English, it provides coverage of about $\frac{3}{4}$ of a typical text. For the Polish language, the data look similar (Zgółkowska 1992). How much this matters for actual understanding of such a text is a separate issue – for English, in this case, knowledge of 95% of the words in the text is required (Seretny 2004:271, Milton 2009:44-70).

⁸ For example: *on, ona, ono* → *on*; *nad//nade* → *nad*.

	year	W	L	LS	T = %S
BF	1978	9944	1883	674	56%
BJ	1946	8904	2676	708	59%
CM	1964	6166	998	553	46%
FZ	1948	7809	1891	615	51%
GM	2003	3820	1073	522	44%
JA	1996	4591	1369	548	46%
MB1	1983	2137	737	415	35%
MB2	1997	5756	1390	627	52%
MJ	1988	5022	683	374	31%
RW	1959	4515	873	398	33%
SA	1973	21221	2222	843	70%
SD	1995	1974	668	356	30%
SG	1992	4110	714	419	35%
SO	1983	4699	1230	550	46%
TJ	1944	14640	2514	780	65%
WE	2001	1229	499	260	22%
WJ	1987	5432	1576	568	47%

Table 1: Number of words common to the textbooks and Seretny's list, along with the percentage of vocabulary from the list present in the textbooks

The chart below ranks all the textbooks according to the proportion of vocabulary from Seretny's list.

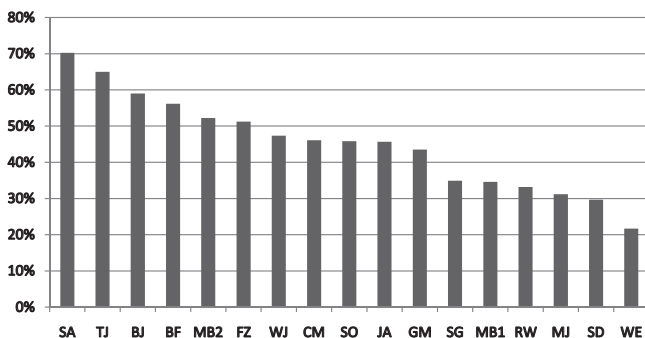


Figure 1: Percentage of vocabulary from Seretny's list in individual textbooks

As can be seen, the stock of vocabulary most essential for everyday communication varies significantly depending on the textbook. The lowest percentage is found in the publication by Wanasz (22%), while the highest

occurs in Schenker's textbook (70%). However, even in the latter case, the student is not exposed to 30% of the aforementioned vocabulary. It is also worth noting the difference between two different editions of Mazur's book – the proportion of words from Seretny's list increases by as much as 17%. This difference arises from the fact that the two editions clearly differ in their methodological approach: while the first edition is closer to the grammar-translation approach, the second – significantly revised – implements a communicative one. In effect, they are largely two distinct textbooks, which is also reflected in the vocabulary employed.

The chart also reveals at least one notable boundary between groups of the analyzed textbooks. In six textbooks, the share of vocabulary from the lexical family list does not exceed 35%, whereas in the remaining ones, it does not fall below 44%.

Above all, the data show an expected relationship between the dictionary size (and indirectly the volume of the textbook) and the percentage of words overlapping with the compared list. The second group primarily comprises shorter textbooks (except the first edition of Mazur's textbook). For this reason, a more interesting parameter is the ratio of words common to the Seretny's list relative to the dictionary size. In other words, what percentage of a given textbook's lexemes (%T) belongs to the set of vocabulary most necessary for communication (S).

	Year	W	L	LS	%T = S
BF	1978	9944	1883	674	36%
BJ	1946	8904	2676	708	26%
CM	1964	6166	998	553	55%
FZ	1948	7809	1891	615	33%
GM	2003	3820	1073	522	49%
JA	1996	4591	1369	548	40%
MB1	1983	2137	737	415	56%
MB2	1997	5756	1390	627	45%
MJ	1988	5022	683	374	55%
RW	1959	4515	873	398	46%
SA	1973	21221	2222	843	38%
SD	1995	1974	668	356	53%
SG	1992	4110	714	419	59%
SO	1983	4699	1230	550	45%
TJ	1944	14640	2514	780	31%
WE	2001	1229	499	260	52%
WJ	1987	5432	1576	568	36%

Table 2: Number of words common to the textbooks and Seretny's list, along with the percentage of the textbook's vocabulary overlapping with the list

In this case, ranking the results does not yield any particularly significant outcome. Although the range is considerable (from 26% in Bolanowski's textbook to 59% in Stone's one), the individual textbooks differ from their nearest neighbors only to a small extent.

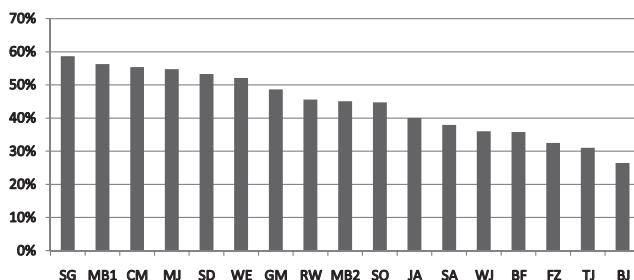


Figure 2: Percentage of the textbook vocabulary overlapping with Seretny's list

The most interesting results seem to emerge when both discussed parameters are considered together. The chart below maintains the descending order of textbooks according to the percentage of vocabulary from Seretny's list present in each text. Simultaneously, it also indicates what percentage of the entire textbook vocabulary this measure represents.

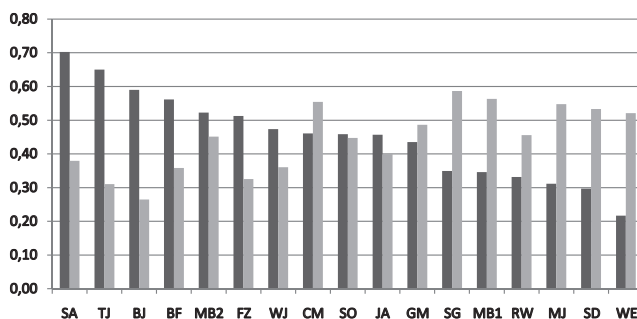


Figure 3: Combination of both parameters

After this ordering, three clear groups of texts emerge:

- Those achieving a high rate of overlap with the core vocabulary list at the expense of a significant amount of less communicatively necessary vocabulary. The best example is the audiolingual textbook by Schenker, which contains as much as 70% of the vocabu-

lary recommended by Seretny, but this constitutes only 38% of all lexemes in the publication. Other textbooks of this type include works by Teslar, Bolanowski, Birkenmayer/Folejewski, Frenkel, and Wira.

- Those providing a relatively small range of vocabulary from the core word families but making up a large proportion of all lexemes in the textbook. This group mainly includes small elementary textbooks, foremost among them Stone's booklet, which covers only 35% of the vocabulary recommended by Seretny but simultaneously represents as much as 59% of the vocabulary offered to the student. This category also includes works by Wanasz, Stok, Miska, Rozmarek, and the first edition of Mazur's textbook.
- The moderate ones, both in terms of coverage of the core vocabulary list and the use of less communicatively important lexemes. A central position is occupied by Swan's textbook (46% and 45%, respectively), accompanied by texts by Corbridge-Patkaniowska and by Gotteri/Michalak-Grey (better vocabulary coverage) as well as by Juszczak and the newer edition of Mazur's book (better core vocabulary list coverage).

5. Conclusions

The entire discussion can be summarized in a few concise conclusions:

1. The proportion of vocabulary necessary for everyday communication offered in a textbook primarily depends on its textual volume; more extensive textbooks provide 44-70%, while more modest ones cover only 22-35% of such vocabulary.
2. Despite the considerable range of results in the ratio of most necessary to less useful words at the initial learning stage (26-59%), no objectively distinct groups of texts can be clearly identified.
3. Three main lexical strategies can be identified among groups of textbooks:
 - achieving a high rate of overlap with the core vocabulary list at the expense of a significant amount of less communicatively necessary vocabulary,
 - providing a relatively small range of vocabulary from the core word clusters but constituting a significant proportion of all lexemes in the textbook,
 - moderate both in terms of coverage of the core vocabulary list and the use of lexemes that are less important in communication.

4. None of the above results can be unequivocally correlated with either the date of publication of a textbook or the author's preferred teaching approach, although certain tendencies can be observed:
 - in general, the lexicon of more recent textbooks reflects everyday communication more accurately and more effectively. However, the most notable positive outlier in this respect is Corbridge-Patkaniowska's textbook (1964), whereas the most negative outlier is Wira's one (1987).
 - similarly, as a general tendency, the vocabulary of textbooks following the grammar-translation approach tends to reflect everyday communication both less and less effectively. Yet again, Corbridge-Patkaniowska's textbook stands out as a positive exception.

6. Limitations and further research

The analysis presented above is primarily exploratory in nature. This entails a number of limitations, which merit reiteration here, as they also point to potential directions for future research.

1. First and foremost, the decision to apply a contemporary list of vocabulary deemed most essential for communication to textbooks produced over several decades has, alongside its advantages, some obvious drawbacks stemming from potential diachronic differences in this domain. While these differences were treated as negligible for the purposes of the present analysis, the ideal solution would involve using frequency lists (with a FLT/SLT orientation) that are representative of each specific period. Although such lists are currently unavailable, the dynamic development of historical corpus linguistics for Polish suggests that their compilation may become feasible in the near future. Comparing the vocabulary of textbooks from particular decades with the corresponding frequency lists would undoubtedly yield a more accurate picture of the strategies employed by their authors.
2. Although both the date of publication and the author's preferred teaching approach revealed certain tendencies – albeit by no means without exceptions – there are numerous additional variables that were not taken into account in this study. These include, for instance, the place of publication (USA vs. UK) and the nationality of the authors. Examining the lexical material from these perspectives would certainly be worthwhile.

3. Finally, an obvious additional point of reference – closely related to the factors mentioned above – would be the vocabulary of Polish language textbooks for English speakers published in Poland. Unfortunately, to date, such material has not been quantitatively analyzed. In the future, however, it could provide the most valuable reference point for further investigations.

Textbooks analyzed

- BF = Birkenmayer Sigmund S. / Folejewski Zbigniew, 1978, Introduction to the Polish Language, New York: Kościuszko Foundation.
- BJ = Bolanowski Jerome E., 1946, A New Polish Grammar, Milwaukee: Polonia Publishing Company.
- CM = Corbridge-Patkaniowska Mary, 1964, Polish: A Simplified Course for Beginners, Sevenoaks: Hodder and Stoughton.
- FZ = Frenkiel Zygmunt, 1948, The Easy Way to Speak Polish: A New and Simple Method of Learning the Polish Language, London: F. Mildner.
- GM = Gotteri Nigel / Michalak-Gray Joanna, 2003, Teach Yourself Polish, London: Teach Yourself.
- JA = Juszczak Albert, 1996, Mastering Polish, New York: Hippocrene.
- MB1 = Mazur Bolesław W., 1983, Colloquial Polish: The Complete Course for Beginners, London: Routledge.
- MB2 = Mazur Bolesław W., 1997, Colloquial Polish: The Complete Course for Beginners, 2nd Edition, London: Routledge.
- MJ = Miska Jan K., 1988, Polska Mowa: Polish Language for Beginners, Vol. 1, Chicago: Polish Teachers Association in America.
- RW = Rozmarek Wanda, 1959, Polish for Americans, Vol. 1-2, Chicago: Polish National Alliance Educational Department.
- SA = Schenker Alexander M., 1973, Beginning Polish, Vol.1, New Haven: Yale University Press.
- SD = Stok Danusia, 1995, Polish in Three Months, Woodbridge: DK Publishing.
- SG = Stone Gerald, 1992, Introduction to Polish, London: Oxford University Press.
- SO = Swan Oscar E., 1983, First Year Polish, Columbus: Slavica Publishers.
- TJ = Teslar Joseph Andrew, 1944, A New Polish Grammar, Edinburgh: Oliver & Boyd.
- WE = Wanasz Białasiewicz Ewa, 2001, Beginner's Polish, New York: Hippocrene.
- WJ = Wira Joseph, 1987, Polish. Textbook of the Polish Language for English-Speaking People, Livonia: Belweder Publishing.

References

- Bauer Laurie / Nation Paul, 1993, Word families, in: *International Journal of Lexicography*, 6(3), pp. 253-279.
- Bernstein Basil, 1990, *Odtwarzanie kultury*, Warszawa: Państwowy Instytut Wydawniczy.
- Bielecka-Prus Joanna, 2010, *Transmisja kultury w rodzinie i w szkole. Teoria Basila Bernsteina*, Warszawa: Wydawnictwo Naukowe PWN.
- Kurcz Ida / Lewicki Andrzej / Sambor Jadwiga / Woronczak Jerzy / Masłowski Władysław, 1974-1979, *Słownictwo współczesnego języka polskiego. Listy frekwencyjne*, Kraków: Polska Akademia Nauk. Instytut Języka Polskiego.
- Milton James, 2009, *Measuring Second Language Vocabulary Acquisition*, Bristol: Multilingual Matters.
- Nation Ian Stephen Paul, 2001, *Learning Vocabulary in Another Language*, Cambridge: Cambridge University Press.
- Nocoń Jolanta, 2009, *Podręcznik szkolny w dyskursie dydaktycznym – tradycja i zmiana*, Opole: Wydawnictwo Uniwersytetu Opolskiego.
- Nocoń Jolanta, 2011, Stylistyczne aspekty dyskursu edukacyjnego, in: Sokolska U. (ed.), *Odmiany stylowe polszczyzny dawniej i dziś*, Białystok: Wydawnictwo Uniwersytetu w Białymstoku, pp. 187-200.
- Piasecki Maciej, 2007, Polish tagger TaKIPI: rule based construction and optimization, in: *Task Quarterly* 11, No 1 - 2, pp. 151-167.
- Seretny Anna, 2004, *Słownictwo poziomu progowego – listy frekwencyjne a zawartość inwentarza funkcjonalno-pojęciowego*, in: Dąbrowska A. (ed.), *Wrocławska dyskusja o języku polskim jako obcym*, Wrocław: Wydawnictwo Wrocławskiego Towarzystwa Naukowego, pp. 269-276.
- Seretny Anna, 2011, *Kompetencja leksykalna uczących się języka polskiego jako obcego w świetle badań ilościowych*, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Sinclair John Houston, 2001, Preface, in: Ghadessy M./Henry A./Roseberry R.L. (eds.), *Small corpus studies and ELT. Theory and practice*, Amsterdam: John Benjamins, pp. VII-XV.
- Swan Michael, 2009, We do need methods, in: Wei L./Cook V. (eds.), *Contemporary Applied Linguistics*, vol. 1, Language Teaching and Learning, London: Continuum, pp. 117-136.
- Woliński Marcin, 2006, Morfeusz – a practical tool for the morphological analysis of Polish, in: Kłopotek M.A./Wierżchoń S.T./Trojanowski K. (eds.), *Intelligent Information Processing and Web Mining. Advances in Soft Computing*, vol 35. Berlin/Heidelberg: Springer, pp. 511-520.

- Zgólkowa Halina, 1983, Słownictwo współczesnej polszczyzny mówionej. Lista frekwencyjna i rangowa, Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza w Poznaniu.
- Zgólkowa Halina, 1992, Dobór słownictwa do nauczania języka polskiego jako obcego, in: Miodunka W. (ed.), Język polski jako obcy. Programy nauczania na tle badań współczesnej polszczyzny, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, pp. 37-104.