

The Polish COBUILD and its Influence on Polish Lexicography

Mirosław Bańko

Abstract

The paper deals with the origin of *Inny słownik języka polskiego* (ISJP, Bańko 2000), an innovative general-purpose dictionary of Polish which was modelled on British dictionaries of the COBUILD series. The influence of ISJP on other Polish dictionaries is also addressed. The question the present paper provokes is whether dictionaries should be adjusted to the expectations of the users, or should the users be expected to adjust themselves to new forms of lexicographic description. No simple answer is offered, but the case of ISJP suggests that innovations need time to be absorbed and the form in which they are accepted may be different from that in which they were originally introduced.

The paper begins with some autobiographical notes which explain how the idea of compiling a COBUILD-like dictionary of Polish was born and put into practice. Next other Polish dictionaries which were based on ISJP or followed its defining style are presented. Unlike the COBUILD series, all Polish dictionaries referred to in this paper have been designed for native speakers of the language. This makes them an interesting example of how ideas migrate from one country to another and how they are applied in a different domain than originally envisaged.

1. Life's mysterious coincidences

Some new developments in Polish lexicography have their origin in a series of coincidences, the first of which happened in 1987 in Norway. I stayed there as a

seasonal worker (a university teacher from Poland picking up strawberries on a Norwegian farm was not an exception at the time). One day I visited a provincial bookshop and turned my attention to an English language dictionary which had an original layout and pleasant-to-read, full-sentence definitions. It was the *Collins COBUILD English Language Dictionary*, newly published under the editorship of John Sinclair and Patrick Hanks. It cost 195 Norwegian crowns – twice my Polish salary. The temptation, however, was too strong. ‘At least I’ll have something to read in the evenings,’ I thought.

A few months later I wrote a letter to Collins Publishers, asking them to send me a copy of *Looking Up* (Sinclair 1987), an account of how the COBUILD project was carried out. All I could offer in exchange was a promise to write a review of the COBUILD dictionary for Polish readers. To my surprise, an envelope from the publisher arrived with the book inside as well as some promotional leaflets. ‘Another coincidence,’ I could say. I had no choice but to live up to my promise and publish an article on COBUILD (it was not a typical review since the dictionary was unavailable on the Polish market (nor, in fact, anywhere behind the Iron Curtain). In addition, I prepared a review of some of COBUILD’s teaching materials, and both pieces appeared in the same volume (Bańko 1989b, Bańko 1989a).

Up to that time I had had a very limited notion of lexicography. Dictionaries seemed interesting to me only to the extent that they dealt with grammar. I thought of dictionary users as of machines whose job was to process grammatical information in a way strictly determined by lexicographers. My whole doctoral dissertation was devoted to devising algorithms for such humanoid machines (a part of it was later published in IJL, Bańko 1992). It was only COBUILD that opened my eyes to other aspects of lexicographic description as well as other aspects of the lexicographer’s work. In particular, Patrick’s papers on definitions made me aware of some simple truths: that ‘lexicography is an art,’ that defining ‘is a literary, not a scientific activity,’ that ‘when theory comes into lexicography, all too often common sense goes out’ (Hanks 1979: 36, 37, 33, respectively), and that dictionary definitions may be both precise and readable, anchored in lexicographic tradition as well as current linguistic theory (Hanks 1987, whose stance towards theoretical linguistics was luckily not purely dismissive).

However, I would not have become a lexicographer if it had not been for one more happy coincidence. It was the year 1991, the early days of market economy in Poland, which could be introduced, or rather re-introduced, thanks to political changes in Central and Eastern Europe. A lot of state-owned firms were privatised, and one of the first and successful privatisations was that of

Polish Scientific Publishers PWN (*Państwowe Wydawnictwo Naukowe*). PWN were the leading producer of monolingual dictionaries of Polish as well as general encyclopaedias, but their reference books were outdated at the time and politically biased, because in the preceding decades they had to reflect the official ideology (see, e.g., Wierzbicka 1995 on political distortions in dictionary definitions). In addition, the head of the dictionaries department at PWN was about to retire and a successor was badly needed. The happy coincidence that made me take up the post was – to simplify things a little – that PWN's editor-in-chief, professor Jan Kofman, a historian, and a friend of mine, professor Zygmunt Saloni, a linguist and active metalexigrapher, used to travel on the same train from Warsaw to Białystok, a city in north-east Poland, where they were giving lectures at a branch of the University of Warsaw.

For both of them, Białystok was a place of academic exile at the time. In the 1980s, Jan Kofman was a political dissident, engaged in anti-communist activity. Zygmunt Saloni, though less active politically, sympathised with the illegal opposition in Poland and was regarded as potentially dangerous to the ancien régime. When the situation in Poland changed and PWN were transformed into a private company, Saloni became their scientific consultant and was asked to recommend someone who would take care of Polish language dictionaries. I was not number one on Saloni's list of candidates, but a better-suited friend of mine preferred to go to a Swedish university as a teacher of Polish (that might have been quite a reasonable decision, I think, in the preceding decade but more questionable in the 1990s).

Thus, not entirely by chance but certainly with the help of good fortune, I became responsible for PWN dictionaries, those existing at the time and those that were yet to be compiled.

2. Why COBUILD?

Many of the dictionaries that were in PWN's offer at the beginning of the 1990s were updated and modernised in the very same decade. Among those were, for example, a major spelling dictionary (incidentally, spelling dictionaries are the most popular type in Poland), a dictionary of foreign words (foreign-word dictionaries are the second with respect to popularity), and a dictionary of correct usage. Also, work started on a new edition of a three-volume general-purpose dictionary and ended with the publication of a four-volume edition at the beginning of the next decade. However, PWN managers expected a completely new dictionary to be compiled as well, which would be a symbol of the changes that

came. The expectation made me think of the COBUILD dictionary of 1987 as a potential model for a new dictionary of Polish.

The first COBUILD dictionary was, as is well known, a highly innovative reference book which broke with many conventions of traditional lexicography. It was, however, a dictionary for foreign learners of English, which made it less attractive as a model for a dictionary I was supposed to design for the Polish market. As a consequence, necessary adjustments had to be introduced in the 'Polish COBUILD' project. Firstly, the number of entries (including subentries) was extended to about 50.000. Secondly, pronunciation was not included with the exception of those few cases where it could pose problems for Polish users. Thirdly, the syntactic patterns had to be adjusted to Polish grammar. Fourthly, and perhaps most importantly, a new system for providing inflectional information had to be developed since the simple idea of listing all inflected forms next to the headword would not work for a highly inflecting language such as Polish. I overcame the last problem by devising an original system of coding inflectional information which was both simple and effective in that it enabled the users to derive any word-form of any word included in the dictionary with the help of just three tables, one drawing and a modest number of inflected forms listed in the entry (to this end my 'over-technical' doctoral dissertation, devoted to inflectional information in dictionaries, proved useful).

The proposal for *Nowy słownik języka polskiego* 'A New Polish Dictionary,' which was the original title of the 'Polish COBUILD' (unfortunately later changed to *Inny słownik języka polskiego*, literally 'An Alternative Polish Dictionary,' or worse still 'A Different Polish Dictionary'), was approved by PWN authorities in early 1992. The decision was preceded by a public presentation to which a number of linguists were invited and asked for their opinion. Regrettably, no publisher with experience in dictionary editing or the dictionary market was present, and the discussion was a bit too academic.

It took me more than a year to complete the proposal for *Inny słownik języka polskiego* (ISJP) and to turn it into a dictionary style guide. That time might have been used to build a corpus of contemporary Polish as well, but work on the corpus began only later and ISJP entries had to be compiled first on the basis of paper citation slips that had been collected at PWN in the previous years. Half way through the project both the slips and corpus data were used, while towards the end all editors worked with the corpus only (it had some 50 million words by then and has now grown to about 100 million – for a sample, see <http://korpus.pwn.pl>). Some of the early entries were then compared with the corpus data in the revisions. In effect, not all ISJP entries are equally corpus-based (verbs and function words were prepared entirely with the use of the cor-

pus, while nouns were still mostly based on paper slips, with adjectives falling in-between the two).

After seven years of editorial work, the dictionary appeared in print in May 2000 (Bańko 2000). The project cost nearly one million US dollars (including production costs), only a bit less than a film version of *Pan Tadeusz*, a Polish national epic, which was released a few months before, directed by Andrzej Wajda. When in August 2000 I met Patrick at the Euralex congress in Stuttgart and gave him a sixteen-page sample copy of ISJP, he smiled and shook his head: 'I wish you all the best,' he said, 'but I don't think you are going to make a commercial success.' He was consistent in his scepticism. The first time we talked in 1992, in Oxford, where I was a guest of Oxford University Press, my plan to prepare a Polish version of the COBUILD dictionary met with both his interest and disbelief in the prospects of the undertaking.

Despite the difficulty of turning a learners' dictionary into a dictionary for the general public, and the doubts whether such an undertaking could succeed, I decided to go ahead with the plan. My contention was that even if the dictionary itself would sell below expectations, the work put in it would not be in vain. What I hoped to achieve was to define new standards of lexicographic description which would be more in line with linguistic theory, more adequate to linguistic data, and more consistent than those formerly applied. Such improved standards, I thought, could be used in other PWN dictionaries, thus making the investment in ISJP more worthwhile.

What in particular did I want to improve in the dictionaries of Polish? The goals were numerous and the advantage of COBUILD as the model was that it allowed – or promised to allow – to achieve them all. First, the proper objects of lexicographic description – the linguistic units – had to be identified: in earlier dictionaries, all too often accidental phrases were defined, whether lexicalised or not (see Bogusławski 1978 for the importance of the problem). Second, the definitions had to be enhanced since they lacked the necessary rigour, too often being just lists of quasi-synonyms, and they were too encyclopaedic. Third, the interplay between the meaning of linguistic units and their context had to be brought to light, and the COBUILD-style full-sentence definitions were perfectly suited to this end. Fourth, the system of coding syntactic information had to be developed because in former dictionaries grammar notes were restricted to those features of words which were relevant to their inflectional properties. Fifth, the corpus had to be introduced as a standard source of lexicographic data (earlier dictionaries of Polish were based on paper slips).

In a sense, ISJP became a laboratory for testing new descriptive techniques, and no wonder that the work on the dictionary was summarised later in a mono-

graph (Bańko 2001). A short passage from the English summary explains the goal of the book:

The book is partly a recapitulation of the editorial work [on ISJP], but above all it provides an analysis of the methodological foundations of the dictionary. However, this is not so much a book on *An Alternative Polish Dictionary* itself, as on lexicography as seen from the latter's perspective, as well as on certain linguistic problems considered from the standpoint of practical lexicography. (Bańko 2001: 335)

3. What has been gained?

Patrick was right: ISJP did not become a commercial success. One may wonder whether mistakes had not been made while introducing it to the market. The title of the new dictionary might have been discouraging (how many of us want to have a 'different' dictionary or even an 'alternative' one?). The price was uninviting: the two-volume dictionary was 50 percent more expensive than an older three-volume PWN dictionary that was on sale at the time. The promotion was weak and the promotional budget was disproportionately low when compared to the costs of producing the dictionary. This time I could complain about ill fortune because the sales of a six-volume encyclopaedia, the most profitable PWN book of the time, were falling down rapidly, and the situation in the company was generally bad.

It took seven years to sell the first edition of ISJP. The time would have been even longer if a different distributor, Świat Książki, had not bought a considerable number of copies to sell them by mail-order. Apparently the dictionary was not well-suited to the needs of an average dictionary user in Poland. I received recognition from some reputable scholars in Poland who admired the style of ISJP definitions, and with time the dictionary became a standard reference work in the descriptions of Polish, whether oriented to morphology or syntax, semantics or pragmatics. It did not become, however, a standard dictionary for the general user.

On the other hand, ISJP gained some popularity in its new roles in which it was cast. In 2006, without the extra column (a step also seen in the recent COBUILD dictionaries) and without obscene words, it was published under the title *Wielki słownik ucznia* 'A Comprehensive Junior Dictionary' (Bańko 2006b). This dictionary was intended for the school market (a two-volume dictionary for schoolchildren was obviously too large, but PWN did not decide to invest in its abridgement). In 2007, the same dictionary, now divided into six

volumes, appeared again (Bańko 2007), this time in a series of dictionaries accompanying the Wednesday editions of the biggest Polish daily newspaper, *Gazeta Wyborcza*. While in the preface to the school dictionary, ISJP was mentioned as its basis, neither of these two was mentioned in the preface to the newspaper edition.

In 2005, a CD-ROM edition of ISJP was published by a PWN daughter company and its new version is currently on sale (Bańko 2006a). Both have the title *Multimedialny słownik ucznia* 'The Junior Multimedia Dictionary' and both are deprived of obscene words. Fortunately, the notes on grammar, synonyms, antonyms, etc. from ISJP's extra column were retained, now integrated into the body of the relevant entries.

A number of other Polish dictionaries have to be mentioned here as well, because they have adopted the full-sentence definitions of ISJP. These are, first of all, *Wieża Babel. Słownik wyrazów obcych nie tylko dla gimnazjalisty* 'Tower of Babel. A Dictionary of Borrowed Words, Not Only for Secondary School Pupils' by Radosław Pawelec (1999), and *W kilku słowach. Słownik frazeologiczny języka polskiego* 'In a Few Words. A Dictionary of Polish Idioms' by Katarzyna Mosiołek-Kłosińska and Anna Ciesielska (2001). Both are intended for young users and distributed by another PWN daughter company which operates on the school market. In addition, they both belong to the same dictionary series edited by Maria Krajewska, a former PWN editor who worked on ISJP before. Krajewska herself compiled one more dictionary for the same series, *Mój pierwszy prawdziwy słownik* 'My First Real Dictionary' (Krajewska 2000), which is lavishly illustrated and intended for even younger users. Another dictionary with definitions in the COBUILD style was prepared by a friend of mine, but eventually the publisher lost interest in it and the book has not been completed. Perhaps the most interesting example of using full-sentence definitions in Polish lexicography is a dictionary of dialectal idioms by Maciej Rak (2005). Its author points to ISJP as a model and explains why contextual definitions best suited his needs.

The influence of ISJP on Polish lexicography turns out to be still greater if one takes into account the dictionaries which followed ISJP in a less direct way, without making reference to it. Two such examples are known to me. The first is the four-volume *Uniwersalny słownik języka polskiego* 'The Universal Dictionary of Polish,' edited by Stanisław Dubisz, which was published by PWN in 2003 as a successor to the earlier three-volume *Słownik języka polskiego* 'The Polish Language Dictionary,' edited by Mieczysław Szymczak. The structure of definitions in both dictionaries is alike, but their contents are quite different: *Słownik języka polskiego* has encyclopaedic definitions, sometimes incompati-

ble with the real usage of a word, while in *Uniwersalny słownik języka polskiego*, largely under the influence of ISJP, definitions are less scientific and reflect the viewpoint of an average user of Polish. Another example is *Wielki słownik angielsko-polski* ‘A Comprehensive English-Polish Dictionary’ and *Wielki słownik polsko-angielski* ‘A Comprehensive Polish-English Dictionary’ that appeared under the editorship of Jadwiga Linde-Usiekniewicz in 2002 and 2004, respectively. Both volumes were compiled by PWN editors who made everyday use of ISJP entries even before the latter was published (ISJP helped them to decide which expressions should be printed in boldface and paired with their English equivalents). Because both the English-Polish and the Polish-English part of the dictionary were prepared in cooperation with Oxford University Press and because OUP sells them on the British market (Linde-Usiekniewicz 2005), I hope the debt I owe to British lexicographers, and Patrick in particular, has at least partially been repaid.

4. Conclusions

Lexicography changes so slowly that most present-day dictionaries are similar to those published a few hundred years ago (unless one looks at them carefully enough). Generally, tradition in lexicography is not less important than innovation, and they are two ends of a scale on which the development of lexicography takes place (cf. the title of Béjoint’s 1994 book). However, even small changes accumulate and they may influence the future course of events. The case of ISJP shows, firstly, a great influence of one dictionary on another, and secondly, a much weaker but also much wider influence of the other dictionary on the lexicography in the same country. In addition, the case of ISJP reminds us that when working on a dictionary, one should not lose sight of its primary users, their habits, and their needs. After all, it is the users that dictionaries are compiled for.

Lexicographic innovations, when first introduced, may not catch on or they may be absorbed in a different form than originally intended. The effects of our work may fall short of our expectations, but even small effects are important, especially in lexicography, where the process of compiling a dictionary consists of a great number of small steps. I hope the story of the ‘Polish COBUILD’ has not ended yet, but let me end it here with an anecdote that stresses the importance of detail in lexicographers’ work. In *The New Oxford Dictionary of English* (NODE), edited by Judy Pearsall and Patrick Hanks, I once saw the entry for *Częstochowa*, which referred to a pilgrim town in Poland, allegedly ‘famous for the statue of the black Madonna in its church.’ The information was wrong:

Częstochowa has a painting of the black Madonna, not a sculpture. I wrote a letter to Patrick, who was chief editor of current English dictionaries at OUP at the time, and pointed out the mistake. He replied with a letter of thanks, but still more rewarding for me was the fact that in a later edition of NODE the mistake was corrected.

References

A. Dictionaries

- Bańko, M.** (ed.). 2000. *Inny słownik języka polskiego* (2 volumes). Warsaw: Wydawnictwo Naukowe PWN.
- Bańko, M.** (ed.). 2006a. *Multimedialny słownik szkolny języka polskiego* (CD-ROM). Warsaw: Wydawnictwo Naukowe PWN.
- Bańko, M.** (ed.). 2006b. *Wielki słownik ucznia* (2 volumes). Warsaw: Wydawnictwo Naukowe PWN.
- Bańko, M.** (ed.). 2007. *Słownik języka polskiego* (6 volumes). Warsaw: Biblioteka Gazety Wyborczej & Wydawnictwo Naukowe PWN.
- Dubisz, S.** (ed.). 2003. *Uniwersalny słownik języka polskiego* (4 volumes). Warsaw: Wydawnictwo Naukowe PWN.
- Krajewska, M.** 2000. *Mój pierwszy prawdziwy słownik*. Warsaw: Wydawnictwo Szkolne PWN.
- Linde-Usiekiewicz, J.** (ed.). 2002. *Wielki słownik angielsko-polski. English-Polish Dictionary*. Warsaw: Wydawnictwo Naukowe PWN.
- Linde-Usiekiewicz, J.** (ed.). 2004. *Wielki słownik polsko-angielski. Polish-English Dictionary*. Warsaw: Wydawnictwo Naukowe PWN.
- Linde-Usiekiewicz, J.** (ed.). 2005. *Oxford-PWN Polish-English English-Polish Dictionary* (2 volumes). Oxford: Oxford University Press.
- Mosiółek-Kłosińska, K. & A. Ciesielska.** 2001. *W kilku słowach. Słownik frazeologiczny języka polskiego*. Warsaw: Wydawnictwo Szkolne PWN.
- Pawelec, R.** 1999. *Wieża Babel. Słownik wyrazów obcych nie tylko dla gimnazjalisty*. Warsaw: Wydawnictwo Szkolne PWN.
- Pearsall, J.** (ed.), **P. Hanks** (chief ed., current English dictionaries), *et al.* 1998. *The New Oxford Dictionary of English*. (First edition.) New York: Oxford University Press.
- Rak, M.** 2005. *Słownik frazeologiczny gwary Dębna w Górach Świętokrzyskich*. Cracow: Scriptum.
- Sinclair, J. M.** (chief ed.), **P. Hanks** (managing ed.), *et al.* 1987. *Collins COBUILD English Language Dictionary*. (First edition.) London: William Collins Sons & Co. Ltd.

Szymczak, M. (ed.). 1978-1981. *Słownik języka polskiego* (3 volumes). Warsaw: Wydawnictwo Naukowe PWN.

B. Other literature

Bańko, M. 1989a. 'Book Review: G. Fox & D. Kirby (eds.). 1987. Learning Real English with Collins COBUILD English Language Dictionary (Collins ELT)' in Z. Saloni (ed.). 1989. *Studia z polskiej leksykografii współczesnej*. Białystok: Dział Wydawnictw Filii Uniwersytetu Warszawskiego, Vol. III, 341-346.

Bańko, M. 1989b. 'Book Review: J. M. Sinclair, P. Hanks, *et al.* (eds.). 1987. Collins COBUILD English Language Dictionary (Collins), i jego definicje' in Z. Saloni (ed.). 1989. *Studia z polskiej leksykografii współczesnej*. Białystok: Dział Wydawnictw Filii Uniwersytetu Warszawskiego, Vol. III, 325-340.

Bańko, M. 1992. 'Lemmatization Algorithms for Dictionary Users. A Case Study'. *International Journal of Lexicography* 5.3: 199-220.

Bańko, M. 2001. *Z pogranicza leksykografii i językoznawstwa. Studia o słowniku jednojęzycznym*. Warsaw: Wydział Polonistyki Uniwersytetu Warszawskiego.

Béjoint, H. 1994. *Tradition and Innovation in Modern English Dictionaries*. Oxford: Clarendon Press.

Bogusławski, A. 1978. 'Towards an Operational Grammar'. *Studia Semiotyczne* 8: 29-90.

Hanks, P. 1979. 'To What Extent Does a Dictionary Definition Define?' in R. K. Hartmann (ed.). 1979. *Dictionaries and their Users: Papers from the 1978 B.A.A.L. Seminar on Lexicography* (Exeter Linguistic Studies 4 & ITL Review of Applied Linguistics 45-46). Exeter: University of Exeter Press, 32-38.

Hanks, P. 1987. 'Definitions and Explanations' in J. M. Sinclair (ed.). 1987, 116-136.

Sinclair, J. M. (ed.). 1987. *Looking Up: An account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English Language Dictionary*. London: Collins ELT.

Wierzbicka, A. 1995. 'Dictionaries and Ideologies: Three Examples from Eastern Europe' in B. B. Kachru & H. Kahane (eds.). 1995. *Cultures, Ideologies, and the Dictionary. Studies in Honor of Ladislav Zgusta* (Lexicographica Series Maior 64). Tübingen: Max Niemeyer, 181-195.