

Piotr Żmigrodzki

Polish Academy of Sciences, Institute of Polish Language

Polish Academy of Sciences Great Dictionary of Polish History, presence, prospects¹

Abstract

Polish Academy of Sciences Great Dictionary of Polish – history, presence, prospects

The paper presents a lexicographical project involving the development of the newest general dictionary of the Polish language: the Polish Academy of Sciences Great Dictionary of Polish [*Wielki słownik języka polskiego PAN*]. The project is coordinated by the Institute of Polish Language at the Polish Academy of Sciences and carried out in collaboration with linguists and lexicographers from several other Polish academic centres. The paper offers a concise discussion of the genesis of the project and the range of information included in the dictionary under construction, as well as the organisation of work necessary in the case of an online dictionary gradually made available on the Internet as its development progresses.

Key words:

Polish language, lexicography, general dictionary of Polish, online dictionary

Streszczenie

Wielki słownik języka polskiego PAN. Historia, stan aktualny, perspektywy

W artykule przedstawiono projekt leksykograficzny dotyczący opracowania najnowszego ogólnego słownika języka polskiego: *Wielkiego słownika języka polskiego PAN*. Projekt jest koordynowany przez Instytut Języka Polskiego Polskiej Akademii Nauk, realizowany przy współpracy językoznawców i leksykografów z kilku innych polskich ośrodków akademickich. Artykuł przedstawia skrótowe omówienie początków projektu i zakres informacji, które znajdują się w przygotowywanym słowniku, jak również pewne aspekty organizacji pracy nad słownikiem elektronicznym oraz perspektywy pracy nad nim w przyszłości.

Słowa kluczowe:

język polski, leksykografia, słownik ogólny języka polskiego, słownik internetowy

The following paper presents a work-in-progress: *Wielki słownik języka polskiego PAN* [the PAN – Polska Akademia Nauk, Polish Academy of Sciences – Great Dictionary of Polish; the abbreviation of the Polish title, WSJP, will be used throughout the text]. Aiming at a possibly full outline of the whole undertaking, we will begin with a brief

¹ Academic project financed 2007–2012 from the academic and scientific fund as a development project (R 17 004 03).

description of the existing repertoire of general dictionaries of Polish, then move on to sketch the background of the project, and finally present the dictionary itself, focusing on its content and selected technical aspects.

1. Polish lexicography: recent history

In Poland, there were two major multi-volume dictionaries of the Polish language published the 20th century:

A) The Warsaw Dictionary (*Słownik warszawski, SW*), edited by three linguists: Jan Karłowicz, Antoni Kryński and Władysław Niedźwiecki. It comprised 7 volumes, published between 1900 and 1927. With its estimated 280,000 entries, the Warsaw Dictionary is considered to be the largest inventory of Polish vocabulary. Due to the fact that its conceptual origins date back to the 19th century, the Warsaw Dictionary had long been underestimated and even sharply criticised, not always justifiably. In recent years, there appeared a monograph of the dictionary (Majdak 2009), and the lexicographical work itself was made available in a digitalised form (<http://ebuw.uw.edu.pl>).

B) The 11-volume PAN Dictionary of Polish [*Słownik języka polskiego PAN, SJPD*] edited by Witold Doroszewski – known as the “Doroszewski dictionary”, published 1958–1969. Although its approx. 125,000 entries make less than a half of the SW entries, the description in SJPD is much richer, including in particular authentic examples of word usage. Separate sections in the entry description were devoted to phraseologisms and proverbs. Inflection tables and a system of reference markers provided detailed information on inflexion. The dictionary played a major role in the Polish lexicography of the second half of the 20th century, becoming the source of material (and the theoretical basis) for many smaller popular dictionaries, especially the 3-volume PWN Dictionary of Polish [*Słownik języka polskiego PWN*, called “the Szymczak dictionary” – SJPSz], which sold in two million copies between 1978 and 2004, and the Little Dictionary of Polish [*Mały słownik języka polskiego, MSJP*], first published 1968 and re-issued a number of times, in its original form as well as in various modified versions. Even the 2003 Universal Dictionary of Polish [*Uniwersalny słownik języka polskiego*] directly draws from the tradition of SJPD and the lexicographical framework developed by Witold Doroszewski.

Since the late 1970s, however, the SJPD framework had been criticised by lexicographers of the younger generation. In mid-1980s efforts were made to create a new great dictionary of Polish but due to a number of unfortunate circumstances and also because of the political changes in Poland, this attempt was unsuccessful. After the breakthrough of 1989, it seemed that the emergence of private publishing houses would prompt new lexicographical works. Indeed, there appeared many popular dictionaries (e.g. the Dictionary of Contemporary Polish, edited by Bogusław Dunaj [*Słownik współczesnego języka polskiego – SJP-Dun*]), yet the need for a comprehensive academic lexicographical description of the Polish language remained unfulfilled².

² For more information about the latest history of Polish lexicography, see e.g. Żmigrodzki 2009a.

2. *The WJSP project: general description*

The history of the present project dates back to the autumn of 2004. During a conference entitled “Polish communication and language policy vis-à-vis the challenges of the 21st century”, linguists representing different academic centres passed a resolution which included the following statements:

[...] in order to strengthen the status of the Polish language, one should
[...]

2) in the shortest time possible commence the preliminary works on a great dictionary of 21st-century Polish, which would be a concerted endeavour of Polish humanists, and especially the whole linguistic branch of Polish studies (the project should be initiated by the Committee on Linguistics of the Polish Academy of Sciences).

(Gajda et al., eds., 2005: 416)

Acting in accordance with this point of the resolution, the Chairman of the PAN Committee on Linguistics announced a contest for the best dictionary project. The first presentation of proposed frameworks took place at the Committee’s session on October 3rd, 2005; the PAN Institute of Polish Language [PAN IPL] was represented by Bogusław Dunaj, Renata Przybylska, and Piotr Żmigrodzki (for a published version of the authorial team’s presentation, see Dunaj, Przybylska, Żmigrodzki 2006). The PAN IPL had been systematically developing the initial concept since January 1st, 2006; financing the work of the team from the statutory funds. A detailed dictionary project framework ensued. It was presented at the meeting of the PAN Committee on Linguistics on 3rd December 2006. (see Żmigrodzki et al., 2007). The presentation met with a favourable reception, which was reflected in the Committee’s resolution. In order to secure additional funding sources, a grant application was submitted to the Ministry of Science and Higher Education. Efforts to win the Ministry’s support continued for the whole year and the funding agreement was eventually signed only in December 2007. Since 13th December 2007 the WSJP project enjoys the status of a development project (registered as R 17 004 03). It is entitled “The Great Dictionary of Polish: the basic lexical inventory of the Polish language” [*Wielki słownik języka polskiego – podstawowy zasób leksykalny polszczyzny*], with the author of the present paper acting as the project leader.

The most important principles governing the 5-year-long project can be summed up as follows:

- objective: creating an exhaustive lexicographical description for 15,000 most frequently used lexemes of the Polish language, together with discontinuous units (idioms) containing these lexemes, and selected derivatives;
- mode of presentation: the dictionary is developed in an electronic version (the project team members work via the Internet) and will be available online for free (the introduction of paid access to more sophisticated functions, such as advanced search, is being considered);
- there will be no printed version of the whole dictionary; in the future, however, the WSJP database may serve as a source for derived dictionaries, which could be published in the printed form;

As regards the characteristics of the dictionary, we should emphasise that it is going to be:

- in principle synchronic: although the year 1945 was accepted as the beginning of the time span covered, due to the nature of the sources, to which we shall return later on, the overwhelming majority of the material will belong to the last decades of the 20th and the beginning of the 21st century.
- in principle descriptive: the authors are not going to eliminate from description any lexicographical facts deemed incorrect or – for whatever reasons – unworthy of being noted in a dictionary, as long as these facts are well attested in the sources. The authors will only point out the normative unacceptability of a given fact, basing on the Normative Dictionary of Polish [Słownik poprawnej polszczyzny] (for a more detailed overview, cf. Żmigrodzki 2008b), and mark the stylistic qualification of sub-standard units.
- an academic dictionary in which the authors aim to employ wherever possible the achievements of Polish 20th-century linguistics, especially in the field of semantic, inflexional and syntactic description of lexical units, at the same time keeping in mind that the description must be accessible to a very broad group of Polish language users.

The main source database of the dictionary is the National Corpus of Polish [Narodowy Korpus Języka Polskiego, NKJP], a collective undertaking of several academic units (including PAN IPL), carried out as a development project parallel to WSJP and available for free on the Internet (<http://nkjp.pl>). The second most important source inventory is an auxiliary corpus created at the PAN IPL specifically to serve the needs of the emerging dictionary; it comprises texts which for various reasons were not (and are not going to be) included in the NKJP. Polish Internet sites constitute the third source. Finally, the authors of particular entries may rely on their own excerption. Although we are quite aware that this set of sources is not perfect and might be criticised especially by philologists and lexicographers representing more traditional approaches, we believe that a better corpus of sources for WSJP would not be feasible within the foreseeable time. The NKJP corpus, or, to be precise, one of its earlier trial versions, served as the basis for the list of entries to be included in the dictionary at the present stage of its development. Approx. 15,000 entry words were selected, mainly according to the frequency of given units in the corpus; it turned out, however, that less frequent lexemes need to be added to the list in order to fill in the gaps in certain word paradigms (e.g. names of the days of the week, names of months, Zodiac signs, numerals).

3. Project team

Fulfilling the expectations expressed in the above-quoted resolution, the WSJP is a kind of a linguistic joint venture. As of January 2011, the team of its authors (not counting former or present temporary collaborators) included:

Academic centre	Number of persons
PAN Institute of Polish Language	13
Jagiellonian University (Kraków)	10
University of Warsaw	8
University of Silesia (Katowice)	4
Nicolaus Copernicus University (Toruń)	2
University of Warmia and Mazury (Olsztyn)	1

There are also four IT specialists and one graphic designer involved in the project. When it comes to the inflexional description of WSJP entries, the editors of the Grammatical Dictionary of Polish [*Słownik gramatyczny języka polskiego*] (Saloni et al. 2007), who allowed us to use their descriptions in WSJP, should also be counted among the WSJP authors³.

The key role in the WSJP project is played by the WSJP Workroom – a ten-member unit established within the structures of the PAN Institute of Polish Language in January 2008 and responsible in particular for the coordination of work and solving more significant problems arising in the process of entry construction. The WSJP Workroom is headed by Piotr Żmigrodzki; responsible for major issues related to the conceptual framework are also Renata Przybylska and Katarzyna Węgrzynek. Since 2008, there also exists a WSJP unit at the Institute of Polish Language, University of Warsaw; its head is Mirosław Bańko. The Katowice and Toruń teams do not form separate official units; the former works under the supervision of Piotr Żmigrodzki and Magdalena Pastuch, while the Toruń-Olsztyn group (which is responsible for WSJP function lexemes) is headed by Maciej Grochowski. The majority of team members are young academics: assistant professors, Ph.D. students and even Polish Language students of the last years. Such a combination of the experience of renowned researchers and lexicographers and the fresh outlook of young editors seems beneficial for the project, also in the light of its assumed prospectiveness: the dictionary is to be further developed after 2012, and due to the open nature of the project, the work should continue without end. Only ten members of the team are regular employees of the PAN IPL, the rest work per assignment. This situation has its advantages and disadvantages. On the one hand, we were able to launch the dictionary project without creating extra employment positions at the Institute, and the freelance contractors are financially motivated to work more efficiently; on the other hand, at the initial stages, the team turnover was relatively high, there were almost 60 people altogether involved in the project at various points of time.

³ The full list of authors and collaborators can be found at: <http://wsjp.pl> in the tab "Autorzy".

4. Technical aspects

As has already been said, the dictionary exists in an online version. It consists of three components:

- a relational database (MySQL) on a computer server;
- an edition panel (interface), by means of which the editors enter lexicographical data in the database, filling in respective forms reflecting the microstructure of specific types of entries;
- a presentation panel, by means of which the completed dictionary entries are presented to the user.

The unquestionable advantage of the WSJP IT solution, designed by Mateusz Żółtak, Paweł Fronczak and Tomasz Żółtak, consists in the fact that the dictionary entries can be edited without any specialist software; the edition panel is an electronic form, which (after logging in) can be opened in any web browser. Since the text corpus is also accessible online, the dictionary can be developed anywhere and anytime, the only technical requirement being a computer with an Internet connection. All documents, such as editorial manual and guidelines, are uploaded onto a special protected website, so that practically all information is exchanged between the dictionary authors via the Internet. This solution has proven extremely useful in the light of the geographical dispersion of the co-workers and the workspace limitations of the PAN IPL.

The picture below (Fig. 1) shows the initial view of the edition panel after the

creation of a new entry. By clicking <+> and the yellow triangle signs, the editor opens each field, and can either type in a text (if the field is a text field) or select one of the listed options (list field). List fields are especially useful where the coherence of the description is important, e.g. in the case of labels or thematic classification (see Fig. 2).

The screenshot shows a web-based form for editing dictionary entries. At the top, there is a text input field labeled 'Forma hasłowa' and a dropdown menu with 'do opracowania'. Below this are several sections, each with a yellow triangle icon and a '+' sign to expand it. The sections include: 'Typ hasła: zwykłe', 'Podtyp' (expanded), 'Warianty (0) +', 'Wymowa' (expanded), 'Pochodzenie' (expanded), 'Chronologizacja' (expanded), 'Podhasła (0) +', 'Frazeologizmy (0) +', 'Skrzydlate słowa (0) +', and 'Przysłowia (0) +'. The 'Autor' section is expanded, showing 'Autor: PZ' and 'Data warsz.: 20 04 2009'. Below these are three sections for 'Uwagi redaktora', 'Uwagi nadredaktora', and 'Uwagi superkoordynatora'. At the bottom, there are buttons for 'zapisz', 'zapisz i wyjdź', 'usuń hasło', 'podgląd', and 'powrót', along with two buttons labeled 'mżewn wszystko' and 'zwn wszystko'.

Fig. 1. Entry view (general) in the WSJP edition panel

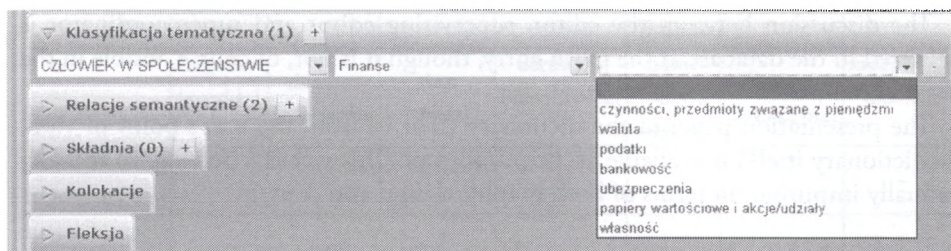


Fig. 2. WSJP list field with an open list of selectable options [Thematic classification: “MAN IN SOCIETY” → “Finance” → “things and actions related to handling money” / “currency” / “taxes” / “banking” / “insurances” / etc.]

All members of the WSJP team have their own user accounts and a passwords with which their log in to the system. There are three main categories of WSJP authors and one additional one:

- editor: the basic category; editors can create their own entries and edit them, they can also view the entries created by other editors but cannot modify them.
- supervising editor: the first proofreader of the entries created by editors. The supervising editor can view all entries but can only modify the entries created by editors who were assigned to his or her supervision by the system.
- supercoordinator: the person who does the final proofreading of the entry before it is accepted for presentation. The supercoordinator can view and modify all entries.

The additional category is the:

- specialist: this person fills in only one specified field, but in all dictionary entries. At present, only three fields are under the charge of specialists: origin, thematic classification and chronology. Due to the specific character of these fields, efficiency is maximised (and the risk of errors and inconsistencies diminished) when one person is responsible for all entries.

The general procedure of entry creation is following:

- the editor creates an entry in accordance with guidelines for a given entry type, filling in all fields except those reserved for specialists, and passes the entry on to the supervising editor;
- the supervising editor checks if the guidelines were followed properly and if the description is adequate; all remarks are entered in a special field;
- the editor modifies the entry, taking into account the supervising editor’s remarks; when the number of sub-entries is settled, also the specialists begin their work;
- the supervising editor controls the entry again and accepts it;
- the supercoordinator controls the entry; any remarks are entered in a respective field;
- the editor (after a discussion, if it ensues) modifies the entry;
- the supercoordinator makes sure that the description is complete (the specialists filled in their fields) and accepts the entry for presentation. The entry is then visible in the presentation panel.

The discussion between the editor, supervising editor and supercoordinator is registered in the database at the given entry, though it is not, of course, visible in the presentation panel.

The presentation panel of the dictionary (that is, from the user's point of view, the dictionary itself) is available at: <http://wsjp.pl>. This website (see Fig. 3) is being gradually improved, in terms of both graphic design and content.

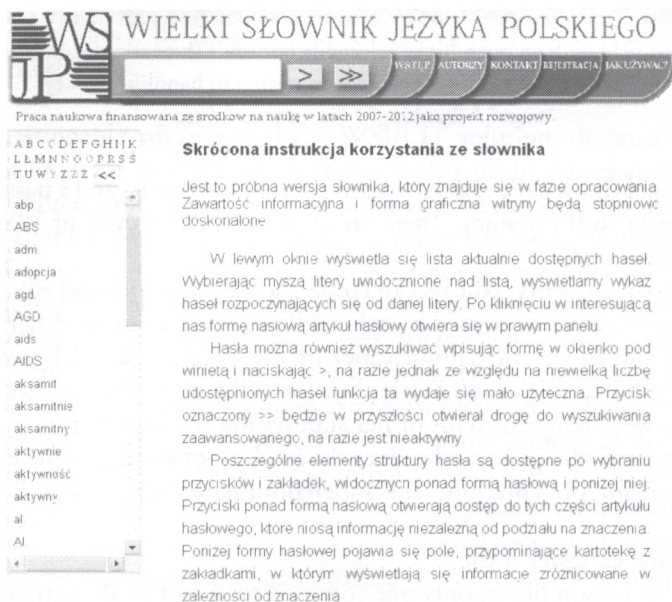


Fig. 3. The front page of the Great Dictionary of Polish – January 2011

5. Dictionary information range

The microstructure of a given entry, as well as the range of information included, depend on the type of the lexicographical object described. We distinguished seven entry types:

- regular (single words);
- discontinuous (idioms, proverbs, winged words);
- abbreviation;
- acronym;
- proper name;
- functional lexeme;
- morpheme.

The following table shows a general list of entry sections depending on entry types⁴:

⁴ [?] For easier orientation in the (untranslated) illustration material included in this paper (e.g. Fig. 4 and 5), Polish names of entry types and dictionary fields are given in square brackets [translator's note].

	regular [zwykłe]	discon- tinuous [nieciągłe]	abbrevia- tion [skrót]	acronym [skrótowiec]	proper name [nazwa własna]	functional lexeme [funkcyjne]	morpheme [morfem]
headword form [forma hasłowa]	+	+	+	+	+	+	+
entry sub-type [podtyp hasła]	+	+	–	–	–	–	–
variant(s) [wariant(y)]	+	+	+	+	+	+	+
chronology [chronologi- zacja]	+	–	+	+	+	+	–
origin [pochodzenie]	+	–	+	+	+	+	+
semantic description: [opis znaczenia:] — guideword [identyfikator] — definition [definicja]	+	+	–	+	+	+	+
labels [kwalifikatory]	*	*	*	*	–	*	*
thematic classification [klasyfikacja tematyczna]	+	+	+	+	–	–	–
semantic relations [relacje znaczeniowe]	+	+	–	–	–	+	–
inflection [fleksja]	+	+	–	+	+	*	–
syntax [składnia]	+	+	–	+	–	+	–
collocations [kolokacje]	+	+	–	+	+	–	–
quotations [cytaty]	+	+	+	+	+	+	–
abbreviation [skrót]	+	–	–	–	+	+	–

informacja normatywna [normative information]	*	*	*	*	*	*	*
notes on usage [noty o użyciu]	*	*	*	*	*	*	*
derivatives [pochodne]	-	-	-	-	+	-	-
lexemes [leksemy]	-	-	-	-	-	-	+
expansion [rozwinięcie]	-	-	+	+	-	-	-

Tab. 1. Components of the WSJP microstructure depending on the entry type. Components marked with <+> are always present in a given entry type, those marked with a <-> are not, and an asterisk <*> indicates that the use of the component in a given entry type is facultative and depends on the characteristics of the specific entry

Particular sections of the above table reflect the fields of the dictionary database; their internal structure may vary according to the entry type. The subsequent part of the paper gives a brief overview of particular fields⁵.

- A) Headword form. We follow conventions well-grounded in the history of Polish lexicography: with nouns, the headword form is the nominative singular (or plural, in the case of plurale tantum), with adjectives and numerals – the nominative singular masculine, with verbs – the infinitive form. As regards “discontinuous” units, we chose a non-traditional solution: for idiomatic expressions of the sup-type “verbal phrase”, the headword is – following Grochowski 1982 – a sequence with the verb in third person singular, together with variable pronouns, eg. *ktoś* [‘someone’, nominative] *upadł* [‘fall’, 3rd person singular masculine, past tense] *na* [on] *głowę* [‘head’ accusative singular] [literally “someone fell on the head”, Polish idiom meaning “someone is acting oddly, someone is crazy”, “someone lost their mind”].
- B) Entry sub-type. This is a technical field, i.e. it is not visible for the dictionary user. For “regular” entries, the sub-type is related to the lexical category of a given item (noun, verb, adjective, etc.), for “discontinuous” ones – with the structure (clause, verb phrase, noun phrase). The choice of the sub-type determines which forms will be added to other fields to be filled in (e.g. Syntax, Inflexion, Collocations).
- C) Variants. The notion of variance can be understood in a number of ways, and so the information included in the field Variants refers to different phenomena, depending on entry type. For “regular” entries, phonetic-orthographic variants are noted here, that is such cases where a change in spelling is accompanied by

⁵ The following description does not apply to „functional” entries; in the case of this entry type, the WSJP general entry structure is considerably modified.

a change in pronunciation, e.g. *pośpieszny* and *pospieszny* [adjective; ‘hurried, hasty, fast’]. In the case of idiomatic expressions, variants are for example *ktoś chwyta kogoś za słowa* and *ktoś łapie kogoś za słowa* [literally ‘someone grasps/catches someone by the words’ – someone is picking on someone’s words, deliberately paying attention to the form of a statement and not to its meaning] or *zielone papiery* and *żółte papiery* [‘green/yellow papers’ – a document issued by a doctor, stating that a given person is mentally ill], that is – in a nutshell – sequences differing in one item while having the same overall meaning. The problem of variance, or rather the instability of the form of idiomatic expressions in Polish, has yet to wait for a proper theoretical analysis; it was the more difficult to invent a system of variant dictionary notation which would enable an automatic entry search, whichever variant the user types in. We managed to solve that problem thanks to the support of our IT specialists; the phrase most frequently appearing in the NJKP is treated as the basic form, and all variants are “linked up” to it with empty reference entries.

- D) Chronology. The name of the field might be a little misleading. Initially, the authors of the dictionary wanted to include here the exact time of the first appearance of a given headword in Polish texts, yet in the present circumstances, this plan proved unfeasible. Thus, as is the practice of many other dictionaries, we give information about the appearance of a particular word (or rather its graphic form) in an older dictionary of Polish. Although this compromise has been criticised by some Polish researchers, we believe that even this kind of information on chronology may be of some help to the dictionary user (and it is of course possible to complete the chronology data in the future).
- E) Origin. Here, too, the information included in the dictionary is at the moment rather provisional in character. We offer etymological information on lexemes of foreign origins, drawing from available dictionaries of foreign terms and etymological dictionaries. Further work on a better verification of etymological information is planned for the future.
- F) Semantic description. Statistical survey (cf. Żmigrodzki, Ulitzka, Nowak 2005) confirms that it is the semantic description that the average user is looking for when consulting a dictionary. Therefore, we try to treat it with due attention. There exist countless critical analyses of word definitions found in 20th-century Polish dictionaries (in fact, opening a text on semantics with some critical remarks on dictionary definitions has turned into a veritable tradition). This results from the fact that dictionary definitions clearly fall behind the significant developments in the methodology of semantic description which took place in the last decades. Thus, already at the initial stages of conceptual planning of the dictionary, it was our objective to make the semantic description reflect the achievements of contemporary semantics to the greatest extent possible (for details, see e.g. Żmigrodzki 2009b). Simply copying into dictionary definitions the explications drawn up by semanticists is of course out of the question, since definitions fashioned in that way would be unclear even for exceedingly well-educated dictionary users. What we try to do, however, is draw inspiration from these explications and adapt the metalanguage of the description to the perception

capabilities of ordinary language users. Thanks to the fact that the entries are not created in the alphabetical order but according to a thematic classification, the work on semantic description is made easier in that the authors can start by identifying the problems and strategies of defining lexemes which belong to a given semantic field.

Apart from the definition, in the case of entries with more than one meaning there is one more component in the semantic description field, namely the guideword (or, as we call it, the semantic identifier). Guidewords are single words or short phrases indicating the meaning of the lexeme explained in the particular sub-entry. The idea of guidewords was borrowed from Western lexicography (e.g. LDOCE or Elexiko web platform). On opening an entry, the user finds a list of guidewords which refer to the sub-entries; selecting one of the guidewords, the user opens the respective sub-entry. This is illustrated below.

The screenshot shows the WSJP website interface. At the top, the title 'WIELKI SŁOWNIK JEZYKA POLSKIEGO' is displayed. Below the title is a search bar and navigation buttons: 'WSTĘP', 'AUTOREZY', 'KONTAKT', 'REJESTRACJA', and 'JAK UŻYWAĆ'. A secondary navigation bar contains 'CIBONOŁO GŁOSIĄCA', 'FRAZYOLOGIZMI', and 'FOKUS WSZYSTKO'. A vertical alphabetical index is on the left, with 'spodenki' selected. The main content area shows the entry for 'spodenki' with five numbered guidewords: 1. krótkie, 2. majtki, 3. sportowe, 4. dziecięce, and 5. ekspresywnie. A sub-entry for the first meaning (1. krótkie) is open, showing the definition: 'część ubrania, przypominająca spodnie, ale z krótkimi nogawkami, kończącymi się przed kolanem'. To the right of the definition is a vertical menu with buttons: 'DEFINICJA', 'WARIANTY', 'KWALIFIKACJA TEMATYCZNA', 'RELACJE ZNACZENIOWE', 'POŁĄCZENIA', 'CYTATY', and 'ODMIANA'.

Fig. 4. Entry for the entry *spodenki* [shorts], with the guidewords for 5 meanings and an open sub-entry for meaning 1

The guidewords are included in the sub-entries after the entry is divided into separate meanings and their definitions are developed. Unlike definition creation, the invention of guidewords is not governed by any strict rules; the editor must propose such guidewords which will allow the user to differentiate between the particular meanings easily.

- G) Labels. In existing dictionaries, it became a standard to employ labels – abbreviations indicating that a given lexical unit is stylistically marked, specialist or dated. In the WSJP we employ a system of labels developed on the basis of a critical

analysis of the choice and use of labels in other Polish dictionaries. As a rule, the label(s) can be found before the semantic definition; in some specified cases labels are also used to mark inflexion forms (see below).

- H) Thematic classification. The WSJP is the first general dictionary of Polish which makes use of a thematic classification of the vocabulary. We employ a three-tier classification scheme (about 80 categories altogether). In older dictionaries, labels marking the lexical units as specialist could partly serve as a classifying system, yet in this way the stylistic marking of the unit (specialist versus non-specialist) was not kept distinct from the reference of the lexeme to the real world. In our classification, every separate meaning of an entry is of course categorised independently. The description of the WSJP classification can be found in an article by B. Batko-Tokarz (2008).
- I) Semantic relations. The sub-entries can include lexical units exhibiting the relations of synonymy, antonymy, hyperonymy or incompatibility to the headword (for “function” entries, we additionally include quasi-synonyms and quasi-antonyms). These relations are understood narrowly; we follow e.g. M. Grochowski (1982), who in turn adapted J. Lyons’s classification (1968) to the Polish language. Synonymy is thus bilateral implication, hyperonymy – unilateral implication, etc. Hence, the WSJP should not be treated as a practical dictionary of synonyms, whose aim is to offer the user various suggestions for the substitution of a given lexeme in a text.
- J) Inflexion. The WSJP is the first general dictionary of Polish providing direct and exhaustive information about inflexion. We include full inflexion paradigms for all inflected lexemes, as well as the indication of gender, aspect of verbs, comparison, etc. This information is provided courtesy of the authors of the Grammatical Dictionary of Polish [Słownik gramatyczny języka polskiego] (cf. Saloni et al. 2007), who, to our delight, agreed to cooperate with the WSJP project. Inflexion facts are also noted with regards to „discontinuous” entries; in this case, however, it is manually typed in by the editors. The inflexion database is stored on the server, which enables the importation of the inflexion paradigm(s) to the entry as soon as the editor begins working on a given lexeme. The editor can select one of the suggested paradigms (due to inflexion homonymy, there may be more than one) or mark particular paradigm items with chronology, frequency and style labels.
- K) Syntax. The valence of the units is indicated; this information is made available to the user in the shape of symbolic syntactic schema. These might be accompanied – if need be – by a note that a given unit takes an unusual syntactic sequence or that there apply rules of semantic selection (that is the meaning of words determines their ability to form a sequence with a given lexical unit).
- L) Collocations. Collocations, here understood as statistically frequent combinations of the entry with other lexemes, form the main bulk of the WSJP exemplification material. Recent lexicographical publications clearly state that collocations illustrate usage better than full-sentence quotations. The arrangement and structuring of collocations in the WSJP varies according to entry type and sub-type (it is different for verbs, nouns, adjectives etc.). An advanced collocation search of the dictionary database will be possible in the future, offering substantial assistance to research-

ers studying the lexical collocation of Polish words. The current possibilities of acquiring collocations for the dictionary leave a lot to wish for. The editors use the resources of the NKJP corpus. Searching for collocations, they employ either of the two corpus browsers, Poliqaip or PELCRA. Unfortunately, the collected data must be entered to the WSJP forms manually.

- M) Quotations. We also include a small number (max. five per one headword meaning) of authentic quotations, comprising single full sentences or longer. Both quotations and collocations are taken mainly from the NKJP, some come also from other sources previously mentioned in the present paper.
- N) Abbreviation. The WSJP also notes frequently used abbreviations of a given lexeme, e.g. *dr* from *doktor* [doctor], *zob.* from *zobacz* [see, as in “see above/below” etc.]. These abbreviations are also described in separate entries.
- O) Normative information. As has already been said, the WSJP is a descriptive dictionary; thus, we do not eliminate linguistic facts considered to violate norms or for whatever reasons deemed unfit for a dictionary. If such controversial facts are sufficiently common, we include them in respective fields of the dictionary, noting in the field Normative Information that a given form or usage of the entry deviates from the linguistic norm (as contained in the latest edition of the PWN Press Great Normative Dictionary of Polish). We do not verify the correctness ourselves, nor do we decide on the correctness of units not included in the normative dictionary.
- P) Notes on usage. This field includes any additional information that could not be entered in the previous fields, for example:
- that the unit is often used as a component of a proper name;
 - that the unit is sometimes confused with another one (paronymy of the kind *adaptować* – *adoptować* [to adapt – to adopt]);
 - that the unit is often used in a specific semantic sense (e.g. *samochód* [motor car] meaning often *samochód osobowy* [automobile, passenger motor car]);
 - that there are deviations from the established graphic form of the unit.
- The fields listed below are included only in selected entry types.
- R) Derivatives – this field is completed for „proper name” entries, in the description of the names of towns and states. The sub-fields include:
- the name of the male inhabitant
 - the name of the female inhabitant
 - the derivative adjective,
- which are subsequently described in separate entries.
- S) Expansion. This information is offered for entry-types „abbreviation” and „acronym”. An abbreviation is (e.g. *nr* = numer [number], *prof.* = profesor [professor], *cdn.* = ciąg dalszy nastąpi [to be continued]) is not a lexical unit in itself but rather a graphic representation of a lexeme or phrase. Consequently, abbreviations are not defined in the dictionary; the expansion is provided instead, referring the user to the entry which describes a particular lexical unit.

Acronyms, on the other hand, have both expansions and definitions. The expansion of an acronym is the sequence of phrases it refers to, whereas its definition is the semantic interpretation of that sequence. The expansion of the acronym *PIT*,

for example, is “Personal Income Tax”, but its [Polish] definition can be divided into at least two meanings: 1. ‘podatek od dochodów osobistych, płacony przez osoby prywatne w Polsce’ [personal income tax, paid in Poland by natural persons]; 2. ‘formularz związany z rozliczeniem podatkowym, składany w Urzędzie Skarbowym’ [the form including the tax statement, submitted to the Tax Office].

T) Lexemes. This field is used for the entry-type “morpheme” and contains several examples of words in which the headword morpheme is found.

6. The mode of presentation of lexicographical information

The WSJP is – to use the term coined in Żmigrodzki 2008a – a primarily online dictionary, which means that it has been developed to be presented on the computer screen. As a result, the basic entry view that presents itself to the user is a structured “tab view”.

On selecting an entry, the user first sees only the description components shared by all meanings (above the headword) and the label. A label must be selected to access a folder with tabs containing the data for a given sub-entry, i.e. meaning (see Fig. 4 above). The user opens tabs by clicking them and their content becomes visible. Respecting the habits of some users, though, we also offer them the possibility to view entries in a linear, “show-all” mode, in which all sections of the description are presented one after another on a single page (see Fig. 5). An entry viewed in this arrangement can be also printed out (although the dictionary is designed to be used on a PC).

spodenki

CHRONOLOGIZACJA: SJPWAR: spodenki
SJPDr
SJPSz
SJPDun
ISJP
PSWP
USJP

ODMIANA: część mowy: rzeczownik
rodzaj gramatyczny: p2
liczba pojedyncza / liczba mnoga

M: spodenki
D: spodenek
C: spodenkom
B: spodenki
N: spodenkami
Ms: spodenkach
W: spodenki

1. krótkie

DEFINICJA: część ubrania, przypominająca spodnie, ale z krótkimi nogawkami, kończącymi się przed kolanem

WARIANTY: • rzadziej spodenki

KWALIFIKACJA TEMATYCZNA: CODZIENNE ŻYCIE CZŁOWIEKA
> Ubranie
> rodzaje ubrań

RELACJE ZNACZENIOWE: • hiperonim: ubiór

POŁĄCZENIA: • damskie, męskie; białe, czarne, czerwone, granatowe, niebieskie, dzinsowe, skórzane; krótkie, króciutkie, kuse, przkrótkie, dopasowane, obcisłe, opięte; zmięte spodenki
• spodenki dla chłopca, męzczyzny, spodenki do kolan
• spodenki na szelkach
• para spodenek
• rozdarcie na spodenkach
• mieć (na sobie), nosić, zakładać; ściągnąć, zdjąć; opuścić, podciągnąć spodenki

CYTATY: • Kobiety miały na sobie tylko wzorzyste chusty, upięte w kształt spódniczek [...]. Mężczyźni nosili krótkie spodenki i coś w rodzaju orzewiowych kamizelek z przątkowanego perkalu.
źródło: NK: P. Marian Brandys: Z panem Bieganiem w Abisyni, 1962
• Modna kobieta pozostawi w swojej szafie spodenki do kolan oraz rybaczki.
źródło: Katarzyna Kulńska: Trendy wiosna-lato... 2007, Echo Dnia, 8.08.2006
• Pośrodku smukłych pni palmowych, u stóp podryfego pagórka stał dżip, policjanci w krótkich spodenkach i czerwonych turbanach kwilił nerucho [...].
źródło: NK: P. Włoczek Żukowski: Kamienne tablice, 1994
• Z tak brutalnie potraktowanego nosa kapie krew kap, kap, [...] kładła kaia na kocu: leżac sztywno, wciąga z kieszeni

Fig. 5. Entry *spodenki* [shorts/short trousers] in the „Show all” view (excerpt)

The dictionary entries can be accessed in different ways.

- the headword can be selected from a list on the left-hand panel (see Fig. 3 above); to find the headword on the list, the user can type its first letter(s);
- word forms can be entered in the search window (simple search); the program then lists all entries containing the given form (Fig. 6).

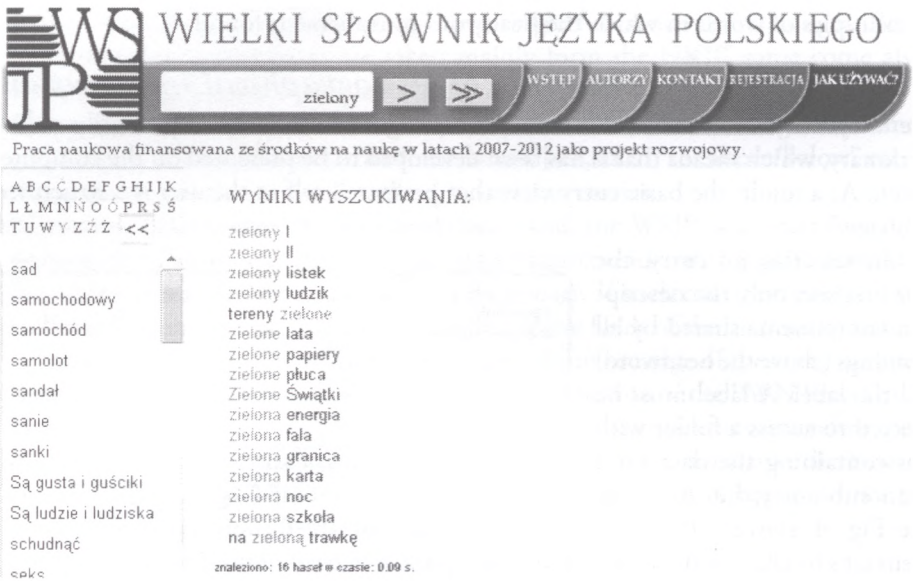


Fig. 6. Simple search in the Great Dictionary of Polish

- the advanced search option can be used, allowing the user to choose various search criteria, e.g. thematic classification, lexical category, origin or chronology. The highly complex structure of the dictionary database, which is only partially suggested by the presentation panel, enables the implementation of very sophisticated search options; this, however, will probably require the employment of additional IT solutions. We are considering the introduction of an access fee in the case of more advanced search options, or at least the obligatory registration of the users in the dictionary users database.

What needs to be pointed out is also the fact that the entry view in the presentation panel is each time generated in response to the user's query from the dictionary database in its current state. In this way, every change made to an existing entry by its editor is almost immediately visible in the end form of the dictionary. New entries are made available to the users without delay. The current address of the dictionary is: <http://wsjp.pl>.

7. Dictionary meta-information

So far, we have been focusing on the lexicographical data included directly in the entry section of the dictionary. Yet, as we know from metalexicographical publications (e.g. Hartmann & James 1998), apart from the entry section a lexicographical work contains also a number of other parts, variously situated in different types of lexicons. Depending on their place in relation to the entry section, they are often called front or back matters respectively.

In large dictionaries, the most important of these non-entry parts is the introduction, sketching, on the one hand, the genesis of the work and its place in lexicography in general, and, on the other hand, presenting the theoretical basis of the description, explaining lexicographical conventions, principles of entry creation etc. All this information will of course be present in the WSJP as well, yet due to the online form of the dictionary it is going to be organized in a different way.

By “metadata” we understand all information included in the WSJP which does not constitute the entries proper. Planning the organisation of metadata in the WSJP, we generally wanted to follow the practices of already existing online/electronic dictionaries, in order to provide the user with an easy and relatively intuitive access to the information, and, above all, to make the organisation of information meet the needs of the user, who, viewing a particular entry, is looking for particular data. Users should be allowed to access the information they are looking for without first having to get through a long introduction or loading the information they do not need at the moment.

According to our initial conceptual framework, the WSJP metadata will be structured into at least three levels:

1. Basic contextual information – in the form of „balloons” appearing when a given object is indicated with the cursor. The objects in question are especially:
 - a. buttons and tabs – a brief description of what happens when the button/tab is selected
 - b. abbreviations of dictionary titles in the field Chronology – a shortened bibliographical description of the dictionary
 - c. other abbreviations used in the dictionary entries, in particular :
 - i. labels
 - ii. all abbreviations and symbols in the tab Inflexion
 - iii. abbreviations and symbols in the tab Symbols
 - iv. abbreviations of names of languages in the tab Origin

This solution is often employed in many Windows applications and websites. In the current version of our dictionary, it can be found in the field Chronology (Fig. 7):

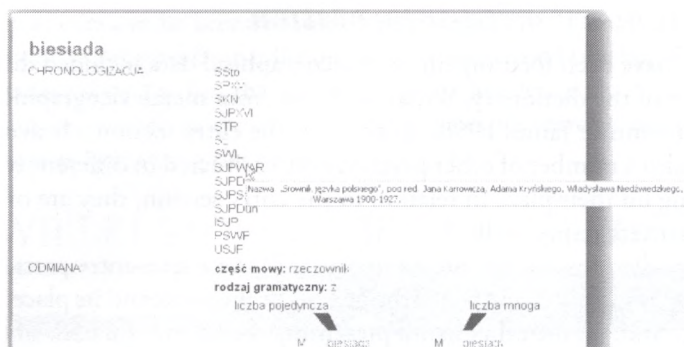


Fig. 7. WSJP, entry: *biesiada* [feast (noun)], view: „Show all”
(excerpt with metadata for the field Chronology)

1. Expanded contextual information could be accessed in an open tab by right-clicking a button (or in some other way) and would include a short description related to the field and entry (sub)type, or some cross-reference to a longer text dealing with the topic.
2. General information, with each thematic section available after selecting a button on the front page of the dictionary (shown in Fig. 6) or clicking a reference link in the texts of level two. General information would include:
 - a. a full description of principles governing the creation of entries (of course structured as a hypertext);
 - b. technical guidelines for dictionary users;
 - c. information of the history of the project and a list of its authors;
 - d. bibliographical data of scientific and other publications related to the dictionary.

As for now, only fragmentary pieces of metadata are available; just as many other dictionary functions, the metadata section still needs to wait for its further development and the implementation of proper IT-solutions.

3. The present state of the project

The present stage of our work on the dictionary will continue till the end of 2012; by that time, we plan to create entries for 15,000 most frequently use lexemes of the Polish language (including idioms and some derivative lexemes). The list of these units was developed basing on an analysis of Polish computer corpora and frequency dictionaries. As of January 2011, about 10,000 have been created but not all of them have yet been subject to the final content verification and accepted for presentation to the external user. During the realisation of the project, various difficulties were revealed, slowing down our progress. These are above all technical, IT-related matters. Although it had first seemed that the structure of the dictionary database was perfectly predictable and could be planned beforehand, with time it turned out that that certain changes are inevitable. Other problems are related with our resource database. The main source of material for the dictionary, the NKJP corpus, is being developed parallel to

our project and still [early 2011] has not been completed. Finally, the most important conclusion arising from our experiences is that some of the contemporary theoretical approaches which we wanted to employ in our lexicographical description prove ineffective, as they were developed on the basis of a limited number of examples and have little explanatory potential when confronted with a large bulk of linguistic material. This accounts for example for the above-mentioned concept of semantic relations, particularly with regards to synonymy, and also for strict definition-writing rules: it is sometimes quite impossible to fulfill the prerequisite of limiting the vocabulary of definitions to irreducible or even just semantically simple units. As is always the case in lexicography, the time factor plays an important role here as well: the duration of the project being strictly determined, one needs to reach a compromise between the pace and the manner of entry creation.

8. *The prospective future of the dictionary*

The work we plan to complete by the end of 2012 is of course just the beginning. The number of entries should be further expanded until practically all lexical units of 21st-century Polish language are described. Due to the electronic form of our lexicographical work – a form open by its nature – the development of the WSJP can continue without end; on the one hand, new entries can be always added, and on the other hand, the existing descriptions can be extended, new fields included, entries improved.

Concluding this very brief overview of issues concerning the PAN Great Dictionary of Polish, on my own behalf and on behalf of the team of authors I would like to express the hope that our project will successfully reach its planned conclusion and then will be further developed, that the dictionary will become visibly present in the Polish lexicography of the 21st century and – in the form we developed for this purpose – will prove helpful to many users.

References

Dictionaries:

- LDOCE: *Longman Dictionary of Contemporary English*. Fourth Edition. — Harlow 2003: Pearson Ltd.
- MSJP: Stanisław SKORUPKA, Halina AUDERSKA, Zofia ŁEMPICKA (eds.): *Mały słownik języka polskiego*. — Warszawa: Państwowe Wydawnictwo Naukowe, 1968. New edition: Elżbieta SOBOL (ed.). — Warszawa: Wydawnictwo Naukowe PWN, 1993.
- SJPD: Witold DOROSZEWSKI (ed.): *Słownik języka polskiego PAN*, vol. 1–11. — Warszawa: Państwowe Wydawnictwo Naukowe, 1958–1969.
- SJPDun: Bogusław DUNAJ (ed.): *Słownik współczesnego języka polskiego*. — Warszawa: Wilga, 1996.
- SJPSz: Mieczysław SZYMCZAK (ed.): *Słownik języka polskiego PWN*, vol. 1–3, Warszawa: Wydawnictwo Naukowe PWN, 1978–1981; *Suplement*, Mirosław BAŃKO, Maria KRAJEWSKA, Elżbieta SOBOL (eds.). — Warszawa: Wydawnictwo Naukowe PWN, 1992.
- SW: Jan KARŁOWICZ, Adam KRYŃSKI, Władysław NIEDŹWIEDZKI (eds.): *Słownik języka polskiego*, vol. 1–8. — Warszawa 1900–1927; ebuw.uw.edu.pl.
- USJP: Stanisław DUBISZ (ed.): *Uniwersalny słownik języka polskiego PWN*, vol. 1–4. Special edition, vol. 1–6. — Warszawa: Wydawnictwo Naukowe PWN, 2003. Electronic edition: 2004.

Other references:

- BATKO-TOKARZ Barbara (2008): Tematyczny podział słownictwa w *Wielkim słowniku języka polskiego*. — [In:] ŻMIGRODZKI, PRZYBYLSKA, 31–48.
- DUNAJ Bogusław, PRZYBYLSKA Renata, ŻMIGRODZKI Piotr (2006): Zarys koncepcji wielkiego słownika języka polskiego. — *Polonica XXVI–XXVII*, 5–16.
- GAJDA Stanisław, MARKOWSKI Andrzej, PORAYSKI-POMSTA Józef, eds. (2005): *Polska polityka komunikacyjno-językowa wobec wyzwań XXI wieku*. — Warszawa: Elipsa.
- GROCHOWSKI Maciej (1982): *Zarys leksykologii i leksykografii. Zagadnienia synchroniczne*. — Toruń: Wydawnictwo UMK.
- LYONS John (1968): *Introduction to theoretical Linguistics*. — Cambridge: Cambridge University Press.
- MAJDAK Magdalena (2009): *Słownik warszawski: koncepcja — realizacja — recepcja*. — Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego.
- SALONI Zygmunt, GRUSZCZYŃSKI Włodzimierz, WOLIŃSKI Marcin, WOŁOSZ Robert (2007a): *Słownik gramatyczny języka polskiego (CD-ROM)*. — Warszawa: Wiedza Powszechna.
- SALONI Zygmunt, GRUSZCZYŃSKI Włodzimierz, WOLIŃSKI Marcin, WOŁOSZ Robert (2007b): Grammatical Dictionary of Polish. Presentation by The Authors. — *Studies in Polish Linguistics* 4, 5–26.
- ŻMIGRODZKI Piotr (2008a): *Słowo — słownik — rzeczywistość. Z zagadnień leksykografii i metaleksykografii*. — Kraków: Lexis.
- ŻMIGRODZKI Piotr (2008b): Nowy *Wielki słownik języka polskiego* a problemy poprawności językowej. — [In:] Małgorzata ŚWIĘCICKA (ed.): *Siła słów i ludzi*; Bydgoszcz: Wydawnictwo Uniwersytetu Kazimierza Wielkiego, 88–100.
- ŻMIGRODZKI Piotr (2009a): *Wprowadzenie do leksykografii polskiej*, 3rd expanded edition. — Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- ŻMIGRODZKI Piotr (2009b): *Najważniejsze zasady opisu semantycznego* w *Wielkim słowniku języka polskiego*. — *Linguistica Copernicana* 1, 183–197.
- ŻMIGRODZKI Piotr, BAŃKO Mirosław, DUNAJ Bogusław, PRZYBYLSKA Renata (2007): *Koncepcja wielkiego słownika języka polskiego — przybliżenie drugie*. — [In:] ŻMIGRODZKI, PRZYBYLSKA, 9–21.
- ŻMIGRODZKI Piotr, PRZYBYLSKA Renata, eds. (2007): *Nowe studia leksykograficzne*. — Kraków: Lexis.
- ŻMIGRODZKI Piotr, PRZYBYLSKA Renata, eds. (2008): *Nowe studia leksykograficzne 2*. — Kraków: Lexis.

Translated by Zofia Ziemann