

# Multisłownik: Linking plWordNet-based Lexical Data for Lexicography and Educational Purposes

**Maciej Ogrodniczuk**

Institute of Computer Science  
Polish Academy of Sciences  
maciej.ogrodniczuk@ipipan.waw.pl

**Joanna Bilińska**

University of Warsaw  
j.bilinska@uw.edu.pl

**Zbigniew Bronk**

Institute of Computer Science  
Polish Academy of Sciences  
zbigniew.bronk@ai.pl

**Witold Kieras**

Institute of Computer Science  
Polish Academy of Sciences  
wkieras@ipipan.waw.pl

## Abstract

Multisłownik is an automated integrator of Polish lexical data retrieved from multiple available online sources intended to be used in various scenarios requiring access to such data, most prominently dictionary creation, linguistic studies and education. In contrast to many available internet dictionaries Multisłownik is WordNet-centric, capturing the core definitions from Słowosieć, the Polish WordNet, and linking external resources to particular synsets. The paper provides details of construction of the resource, discussed the difficulties related to linking different logical structures of underlying data and investigates two sample scenarios for using the resulting platform.

## 1 Introduction

*Multisłownik* (Pol. *multidictionary*) is a linguistic integration platform for Polish lexical data retrieved from multiple available online sources intended to be used in various research and educational scenarios. The difficulty of such setting is clear: lexical data is created for different purposes resulting in various underlying structures and representation formats, tailored to specific requirements of each subfield of linguistics. For instance, morphological dictionaries may not differentiate word senses when inflectional patterns of each sense is the same; in turn, when they are different, senses can be assigned properly but at the same time usage examples from corpora restricted to a given sense may be difficult to retrieve.

The paper presents an attempt of creating such linked resource for Polish using computational

methods. Section 2 presents similar attempts for other languages, Section 3 describes the data sources used, Section 4 documents the decisions made during the process of data linking, Section 5 provides two sample scenarios based on the integrated data and Section 6 summarizes the paper and presents the work in progress.

## 2 Related Work

In contemporary lexicography there can be seen a tendency to integrate dictionaries into portals<sup>1</sup> mainly provided as a source of information for the ordinary users rather than linguists and researchers. Usually the idea of such portals is to give maximum data big publicity as possible with a minimal effort.

As compared to FRAN, a Slovenian dictionary of a similar type<sup>2</sup>, gathering in-house lexical resources available to the Fran Ramovš Institute of the Slovenian Language ZRC SAZU, the initial assumption was that external resources will be used as well. The reason for such a decision was a desire to present Polish vocabulary in an extensive way which seemed to be impossible while using only open resources or those published by a single unit. Unlike in Slovenia, the main Polish linguistic sources were prepared by various publishing houses and research centres. However, because of the authors' rights, not all the dictionaries could be used in the same way. Therefore some of the dictionary data is only presented as references and information whether the searched word can be found in a given dictionary. By default FRAN presents results 'dictionary by dictionary'

<sup>1</sup>See e.g.: <https://en.oxforddictionaries.com/>, <http://www.termania.net/>, <http://dictionaryportal.eu/>.

<sup>2</sup>See <http://fran.si>.

ordering them from the general one (with definitions) through etymological and historical to more specialised ones (e.g. spelling dictionary, medical lexicon or the dictionary of climber’s language).

Online dictionary of the PWN publishing house<sup>3</sup> offers a similar approach to Polish: entries from dictionaries of several types are presented “as is” on a single Web page together with language use comments, encyclopaedia entries and corpus-based examples. Even less used-friendly Dictionary Portal of such type<sup>4</sup> mainly facilitates searches in various dictionaries providing references to source entries.

Multisłownik combines the concepts of a dictionary portal and a general dictionary trying emulate a traditional dictionary. Therefore the query results are presented in a form of an automatically generated dictionary-like entry.

### 3 Sources of Lexical Data

Multisłownik integrates three different kinds of lexical resources:

1. traditional dictionaries created by philologists and meant for human readers only, either web-based or digitalized
2. electronic datasets created by computational linguists for both human users and automatic processing in NLP implementations
3. community-based lexical collections developed online.

The main two sources of lexical entries, forming the core of Multisłownik, are plWordNet (Piasecki et al., 2009)<sup>5</sup> and Grammatical Dictionary of Polish (Saloni et al., 2012; Saloni et al., 2015; Woliński and Kieraś, 2016)<sup>6</sup>. Several others contributing to its content are: Polish language version of Wikipedia and Wikisource, Walenty valency dictionary (Przepiórkowski et al., 2014) and National Corpus of Polish (Przepiórkowski et al., 2012, NKJP)<sup>7</sup>. Various other lexical datasets are linked to each entry.

We briefly characterize these sources below showing their lexical potential and pointing out

<sup>3</sup>See <http://sjp.pwn.pl>.

<sup>4</sup>See <http://dictionaryportal.eu/>.

<sup>5</sup>See <http://plwordnet.pwr.wroc.pl/wordnet/>.

<sup>6</sup>Pol. Słownik gramatyczny języka polskiego, SGJP, see <http://sgjp.pl>.

<sup>7</sup>Pol. Narodowy Korpus Języka Polskiego, see <http://nkjp.pl>.

their most important features hindering integration.

#### 3.1 plWordNet

plWordNet (Piasecki et al., 2009) is a lexico-semantic network reflecting the lexical system of Polish inspired by Princeton WordNet (Miller, 1995)<sup>8</sup>. It contains sets of synonymous lexical units (synsets) interconnected with lexico-semantic and derivational relations such as synonymy, hypo-/hypernymy or mero-/holonymy. plWordNet is currently the largest wordnet in the world and contains 178K synsets, 259K word senses and over 600K relations.

Apart from a very rough assignment of part-of-speech category (one of: noun, verb, adjective, adverb) to each lexical unit, plWordNet does not cover any other grammatical information such as grammatical gender for nouns or aspect for verbs. Some of this information may be derived from relations such as verb–noun *mpar\_VN* relation linking verbs and derived gerunds. Currently plWordNet does not cover numerals and uninflected parts of speech.

#### 3.2 SJP.pl

SJP.pl is a Web-based dictionary created by Polish enthusiasts of word games (mainly Scrabble). It aggregates vocabulary from various contemporary printed dictionaries, including spelling and foreign words dictionaries, and classifies them as permitted or non-permitted in word games. Currently it contains ca. 200,000 lexemes. SJP.pl is being developed by the community of its users. As the list of forms noted in SJP.pl is distributed under the terms of open source license it is also used as a data source for spell-checkers. Apart from inflectional forms SJP.pl entries usually also contain short definitions. For Multisłownik it serves mainly as a supplementary source of lexical and grammatical data, especially when the word searched by the user is not present in SGJP.

#### 3.3 Grammatical Dictionary of Polish

Inflectional information is based on The Grammatical Dictionary of Polish (Pol. *Słownik gramatyczny języka polskiego*, SGJP) (Saloni et al., 2012; Woliński and Kieraś, 2016). SGJP is the largest existing linguistically elaborated data set of Polish inflectional morphology, from the very

<sup>8</sup>See <http://wordnet.princeton.edu>.

beginning developed as an electronic dictionary, now in its third edition turned into Web-based linguistic resource. SGJP serves as a main source of grammatical information for widely used morphological analyzer Morfeusz (Woliński, 2006; Woliński, 2014), as well as for the new general dictionary known as The Great Dictionary of Polish (Pol. *Wielki słownik języka polskiego*), currently under development (Żmigrodzki, 2007).

The integration of morphological data with plWordNet senses is hindered by high inflectional variation of Polish lexemes.

### 3.4 National Corpus of Polish

The National Corpus of Polish (Przepiórkowski et al., 2012) is the most prominent corpus of general Polish, providing a balanced representation of contemporary Polish. For Multisłownik it offers real usage examples. To ensure that they represent extensive variety of possible usage of the word it looks for corpus examples for all the possible non-syncretic forms from the inflectional paradigm of the word. For each such form a corpus frequency is also provided.

The corpus data is limited only to NKJP as the largest and most representative corpus of Polish available. Still, closing the dataset in 2010 makes it less and less up to date each year. As a consequence, NKJP does not reflect the newest Polish vocabulary such as the word *prekariat* ‘prekariat’ which appears in 1-billion-word data set only twice while its actual frequency in daily and weekly newspapers is much higher in the recent years.

### 3.5 Wiktionary and Wikipedia

Wiktionary<sup>9</sup> and Wikipedia<sup>10</sup> are open-source, multilingual, community-developed dictionary and encyclopaedia fully available to download in XML format. For Multisłownik they are used as additional sources of lexemes, inflection forms, definitions, examples, collocations, information on pronunciation and etymology.

### 3.6 Other Linked Sources

Multisłownik also provides information about the presence of a search word in various other lexical resources unable to integrate directly due to licence or format constraints. The list of

<sup>9</sup>See <https://pl.wiktionary.org/wiki>.

<sup>10</sup>See <https://pl.wikipedia.org/wiki>. Note: due to its character, Wikipedia covers mostly nominal entries.

such resources is extremely heterogeneous. It contains both specialized linguistic dictionaries, both digitalized versions or paper dictionaries and Web-based developments as well as community-based lexical databases. The list of linked sources varies from well known general dictionaries such as PWN dictionaries (Słownik Języka Polskiego PWN, Słownik Wyrazów Obcych PWN, Doroszewski’s classical dictionary, available as scanned pages<sup>11</sup>, through the electronic Dictionary of 17th & 18th Century Polish (Instytut Języka Polskiego PAN, 2010) to various resources capturing the newest vocabulary, both academia-based (such as the entries from the Language Observatory of the University of Warsaw<sup>12</sup>) and community-based, e.g. urban slang dictionaries<sup>13</sup>. Other sources include the Great Dictionary of Polish (Żmigrodzki, 2007), dictionaries of Polish personal and place names<sup>14</sup> and dictionaries of synonyms, antonyms and crossword definitions<sup>15</sup>.

Their integration was motivated by practical reasons put forward by lexicographers: it saves user’s time and effort used for searching the word in all these sources separately.

## 4 Integration

Integration of multiple dictionary resources, heterogeneous by nature, poses various problems due to diverse representation and scope of lexical properties, different levels of detail and incompleteness of coverage of lexical entries. For online resources this situation gets additionally hindered by their constant change: new entries are added to lexicons, models are getting restructured and new data sources appear regularly. Based on all these assumptions we believe that the close integration of resources in such setting (such as combining them into a common LMF<sup>16</sup> resource) is a myth — the complexity of such resource would need to exceed the complexity of its parts, already very high for most of the resources. Our approach is differ-

<sup>11</sup>See <http://doroszewski.pwn.pl/>.

<sup>12</sup>See <http://nowewyrazy.uw.edu.pl/>.

<sup>13</sup>See e.g. Słownik miejski, <http://www.miejski.pl/>.

<sup>14</sup>See <http://nlp.actaforte.pl:8080/Nomina/Nazwiska> and <http://nlp.actaforte.pl:8080/Nomina/Miejscowosci>.

<sup>15</sup><http://synonimy.net>, <http://antonimy.net>, <http://krzyzowka.net>.

<sup>16</sup>Lexical Markup Framework, an ISO 24613:2008 standard for machine-readable dictionary lexicons (Francopoulo, 2013).

ent and assumes interfacing related sources rather than absorbing them into a single common 'super-resource'.

At the same time a common point of reference is needed to serve as the core of the integration; for Multisłownik we decided it to be Słowosieć, the Polish WordNet (Piasecki et al., 2009), further referred to as plWordNet, the most extensive freely available semantic resource offering lexeme-to-sense mapping. plWordNet contains extensive description of lexical-semantic relations for Polish with interlinked synsets and short definitions, currently featuring over 300K lexical relations, 320K synsets and 1.2M inter-synset relations. In Multisłownik it serves as the main source of lexemes and semantic information.

Since plWordNet and SGJP make the most prominent resources covering respectively semantic and grammatical layers, comparison of these resources was of vital importance. As for the data set, SGJP contains 150K entries which do not have their counterparts in plWordNet (not taking into account negated adjectives, representing in SGJP as separate entries). On the other hand, plWordNet contains 20K entries absent from SGJP. plWordNet contains many multiword lexical units (over 30% of the total number) while SGJP does not cover any multiword entries apart from hyphenated entries such as *vis-a-vis* or *ping-pong* and a small sample of words functioning today only as parts of fixed phraseological expressions. Homonymy is the main problem of linking plWordNet data to SGJP; the set of homonyms contains 3450 nouns, 926 adjectives and 586 verbs.

The integration process starts with plWordNet taking over its semantic domains, lexical relation and synset relation types. SGJP is the main source of grammatical data and other resources are used to populate the entry.

Figure 1 presents a simple Web application interfacing Multisłownik platform. Sections provide information about pronunciation and etymology of the entry, its plWordNet senses with SGJP inflection variants assigned properly, related words retrieved from Wikidictionary, concordance from NKJP and information on presence of the lexeme in available online sources.

Information on pronunciation is presented in two formats: IPA and AS. For each sense its domain, definition, example and selected semantic

relations as well as English translation are presented. Grammatical information covers grammatical class, selective categories and inflection pattern symbol. Inflection section presents selected inflectional forms:

- for nouns – singular genitive and locative and plural nominative and genitive
- for adjectives – singular nominative feminine and neutral and plural nominative masculine
- for verbs – selected personal forms.

Syntax information is presented according to Walenty model and annotation. Frequency data and NKJP-based quotations are currently dynamically retrieved using PELCRA search engine.

## 5 Possible Usage Scenarios

The aggregation platform is intended to reflect a standard dictionary, therefore the results are presented in a form similar to a dictionary entry and reflect its microstructure. Each entry provides a number of slots for information: headword, pronunciation, etymology, senses/definitions, grammar information (inflectional patterns), translations into English, derived words and collocations, concordances with quantitative data from the NKJP. An important part are links to online dictionaries of surnames, geographical names, antonyms, synonyms, city slang vocabulary and new vocabulary which makes getting information about the contents of other sources, popularity or importance of lemmata very straightforward.

### 5.1 Lexicographic Scenario

Multisłownik is by its nature a highly heterogeneous resource on many levels: it integrates synchronic and diachronic dictionaries, specialist and general purpose dictionaries, scientific-driven and crowd sourced lexical databases. Thus it does not provide a sound lexicographic description but it can serve as an instant support for a professional lexicographer working in the field of extending a specific dictionary or a linguistic text annotation.

Since Polish is a highly inflectional language, morphological resources are crucial to almost any natural language processing task. For this reason grammatical data sets need constant development especially in reference to new vocabulary. A lexicographer working on this task needs to determine both grammatical features of the lexical entry (such as gender for nouns and aspect for verbs)

Multisłownik

🔍

---

**PLÝWAK**

**Wymowa**  
IPA: ˈpwivak, AS: pɥvɨk Źródło: Wikisłownik

**Pochodzenie**  
pol. pływać Źródło: Wikisłownik

**Znaczenia, charakterystyka gramatyczna**

◊ Rzeczownik męskoosobowy, wzór odmiany: **B3k**  
Charakterystyczne formy fleksyjne:  
lp D.: *pływaka*  
lm M.: *pływacy*  
depr.: *pływaki* Źródło (z tabelą odmiany): SGJP

**pływak<sub>1</sub>** – (osoba) «mężczyzna uprawiający pływanie» Źródło: Słowosiec

*Mikołaj był najszybszym pływakiem w polskiej reprezentacji olimpijskiej.*

Synonimy: -

Hiperonimy: [wodniak<sub>1</sub>](#)

Hiponimy: [nurek<sub>2</sub>](#), [kraulista<sub>1</sub>](#), [żabkarz<sub>1</sub>](#), [grzbiecista<sub>1</sub>](#), [delfinista<sub>1</sub>](#), [zmiennista<sub>1</sub>](#) [mniej ▲]

Tłum. ang.: *swimmer<sub>1</sub>*

Agens/subiekt: [pływać<sub>2</sub>](#)

Żeńskaść: [pływaczka<sub>1</sub>](#)

Derywacyjność: [pływactwo<sub>1</sub>](#)

Synonimia międzyparadygmatica dla relacyjnych: [pływacki<sub>1</sub>](#)

◊ Rzeczownik męskozwierzotny, wzór odmiany: **B3k**  
Charakterystyczne formy fleksyjne:  
lp D.: *pływaka*  
lm M.: *pływaki* Źródło (z tabelą odmiany): SGJP

**pływak<sub>2</sub>** – (wytwór) «boja» Źródło: Słowosiec [więcej ▼]

**pływak<sub>3</sub>** – (wytwór) «splawik wędkarski» Źródło: Słowosiec [więcej ▼]

**pływak<sub>4</sub>** – (wytwór) «szczelny zbiornik, wypełniony powietrzem lub gazem, służący do utrzymania samolotu na powierzchni wody» Źródło: Słowosiec [więcej ▼]

**pływak<sub>6</sub>** – (wytwór) «element sieci rybackiej utrzymujący ją w pozycji pionowej i na zamierzonej głębokości» Źródło: Słowosiec [więcej ▼]

**pływak<sub>7</sub>** – (wytwór) «zakończenie przyrządu służącego do oznaczania poziomu cieczy w jakimś zbiorniku» Źródło: Słowosiec [więcej ▼]

◊ Rzeczownik męskozwierzotny, wzór odmiany: **B3k**  
Charakterystyczne formy fleksyjne:  
lp D.: *pływaka*  
lm M.: *pływaki* Źródło (z tabelą odmiany): SGJP

**pływak<sub>5</sub>** – (zwierzę) «drapieżny chrząszcz z rodziny pływakowatych» Źródło: Słowosiec [więcej ▼]

**Wyrazy pokrewne**

- rzecz. pływanie (n), pływalnia (f), pływaczka (f)
- czas. pływać, płynąć
- przym. pływacki, pływakowy

Źródło: Wikisłownik

**Związki wyrazowe**

[pływak żółtobrzeżek](#), [pływak lapoński](#), [pływak szerokobrzeżek](#), [zatapiacz pływaka](#), [świąd pływaków](#)

**Frekwencja form i cytaty z NKJP (zrównoważonego)**

**pływak:** 307

... rok, czas po Październiku, wydawał się **świt**. Zanurzyłam się jak **pływak** w morze nadziei. Nie spostrzegłam, że na dźwięk słowa "odwilż" moje...

... Geralt wyrwał się, rzucił w przód rozgarniając śmieci piersią jak **pływak** wodę, rąbnął z całej siły, z góry, z mocą naparł na ostrze...

... zewnętrznego, „z przestrzeni” powiedzielibyśmy my. Pomarańczowy **pływak** od sieci rybackiej. Ciężki, z polistyrenu, o dwu wielkich uszach....

... serca. Daleko, daleko jarzył się czerwienią jeszcze jeden czepek, **pływak** podniósł rękę w geście pozdrowienia. Poznałem Sergiusza Michalkowa....

... jej kolana. Kobieta popłynęła na dół, walcząc z wichrem niby **pływak** z rozchukanymi balwanami. Brązowy papier pomknął dalej i stara...

... wykonywał równocześnie gwałtowne ruchy rękami i nogami, jak **pływak**, ale w niczym nie poprawiało to jego rozpaczliwej sytuacji....

**Występowanie w słownikach i encyklopediach**

Źródło	Zawartość	Linki
Słowosiec	semantyka, powiązania, tłumaczenia	<a href="#">pływak</a>
SGJP	gramatyka, fleksja	<a href="#">pływak</a>
SJP.PL	znaczenia, formy odmiany	<a href="#">pływak</a>
Wikisłownik	znaczenia, wymowa, etymologia, fleksja, tłumaczenia	<a href="#">pływak</a>
Wikipedia	znaczenia, opis encyklopedyczny	<a href="#">Pływak</a>
Walenty	składnia	-
WSJP	znaczenia, odmiana, etymologia...	-
Sł. nazwisk	odmiana nazwisk i ich występowanie w Polsce	<a href="#">Pływak</a>
Sł. miejscowości	występowanie i odmiana nazw polskich miejscowości	-
PWN (SJP+SO+SJPDor.)	znaczenia, odmiana, ortografia, cytaty	<a href="#">pływak</a>
SJPDor.	znaczenia, cytaty, odmiana	<a href="#">pływak</a>
Slang miejski	znaczenia slangowe	-
sv7ii.pl	słownictwo polskie XVII i XVIII w.	-
Oberwatorium UW	najnowsze słownictwo polskie	-
Synonimy.NET	synonimy	-
Antonimy.NET	antonimy	-
Krzyżówka.NET	definicje do krzyżówek	-

Figure 1: Test front-end of Multisłownik

and some specific word endings. Consider for example a noun PARKOUR ‘a training discipline’, which does not appear in the *Grammatical Dictionary of Polish*. Since she is dealing with an obvious loanword the lexicographer needs to determine, whether the noun declines or it has all its forms homonymous. If it declines, some alternative word endings need to be determined, such as *-u* or *-a* in genitive singular (both are possible). Also a grammatical gender needs to be assigned (could be either neuter or masculine inanimate). Since the word refers to a rather niche sport activity, a regular lexicographer cannot rely on her own experience and needs to consult some external lexical resources. By simply typing the word *parkour* in Multisłownik’s search bar the lexicographer gains access to

1. basic definition (provided by plWordnet)
2. characteristic inflectional forms and hypothetical gender value (provided by Multisłownik’s own heuristic algorithms)
3. usage examples for four different inflectional forms including their frequencies (found in the National Corpus of Polish).

Based on these informations a proper grammatical description of the word can be formulated and included in the dictionary.

On the other hand a human annotator conducting a morphological, syntactic or semantic text annotation needs a constant access to large lexical data sets supporting her work. Text samples often do not provide a sufficiently large context to determine the proper meaning of a text token or the annotator simply does not have enough specialist knowledge to determine i.e. a lemma of a word. Consider a locative phrase *w Sycowie* (“in Syców/Sycowo”) in which a proper name can be lemmatized either as SYCÓW or SYCOWO. Both endings (*-ów* and *-owo*) are correct and both are very common in Polish names of settlements, both form a locative case form ending with *-owie* but only one of the resulting base forms actually exists and refers to a small town Syców in southwestern Poland. The proper lemma can be easily determined in Multisłownik in which a proper names’ declension dictionary is integrated.

## 5.2 Educational Scenario

Although the platform is aimed at the linguistically- and lexicographically-aware user, it can also be an attractive source of informa-

tion for wider audience, for instance high school pupils. Searching for random words can be a good start point to teach the students what is the dictionary microstructure and how it can differ between dictionaries. After this stage we plan to present the dictionary by looking up the words. We would suggest following queries for teaching purposes, aiming to present the platform to the young people:

1. Check the word KAFAR and PROMULGOWAĆ in Google and in Multisłownik — what are the differences, information given, which source gives you more information on the lemma in the first hit (without further clicking)?
2. What is GEN.PL of MECZ or DAT.SG of MUCHA? (results from the grammatical dictionary)
3. What are the possible lemmata for the word form “danie” (the grammatical dictionary)
4. Which animals groups are called STADO? (the National Corpus of Polish)
5. Who is KALETNIK (plWordNet)
6. What are the other words derived from SEKRET (plWordNet)
7. What are the antonyms of the word SEKRET? (the dictionary of antonyms)
8. Is the form “Dania” in “Dania jest piękna” and “Dania hiszpańskie są smaczne” pronounced in the same way? (Wikisłownik)
9. What is the difference in meaning of NYGUS in general Polish and in the city slang? (plWordNet, slang dictionary)
10. Is the word form ŁABADŹ always incorrect? (dictionary of surnames and 16–17 century dictionary)
11. What is the origin of the words KSIEŻYC and ŁABEDŹ? (Wikisłownik)
12. Is there a place (city, town, village) called “Łabędź” in Poland? (dictionary of surnames)
13. What does the word TRZECIOTEŚCIK mean? (language observatory)
14. What are the synonyms for the DOM? (plWordNet)
15. Which case is “tysiącpięćsetletniemu”? (grammatical dictionary)

The classes on using the dictionary portal would be even more attractive to students when cross-words or other word games (e.g. Scrabble) are used as search targets. One of such activities could be deciphering a coded information with the usage of Multisłownik conducted in a following way:

- Formulating a question that needs to be answered.
- Providing the coded answer with some or all characters replaced with numbers connected to the questions that lead to decoding the secret characters.
- Possible types of questions:
  - “The last letter of the synonym of the word SEKRET that ends with letter T”.
  - “What is the origin of the word KUŚNIERZ? The first letter of the original language name is the secret character number X”.
  - “Is there a surname Łabądz in Polish? If yes, the secret letter is N, if no, the secret letter is C”.

## 6 Conclusions and Further Steps

Multisłownik already proved useful in many scenarios related to combining lexical information by offering a simple yet practical method of referring to multiple sources at the same time.

The most obvious further direction for extension of Multisłownik is adding more data; it occurs that even resources less relevant to the current task, e.g. numerous historical corpora can help lexicographers retrieve usage examples from historical texts to trace back the change of word meanings.

Another type of interesting functionality of Multisłownik would be searching for so called “cultural traces” of a given word. Apart from offering the user extensive dictionary-based grammatical and semantic information also references of a given word or phrase to important artwork (e.g. its presence novel and movie titles, lyrics of popular song or famous quotes) could be tracked. This would require building much larger datasets based on library catalogues, movie databases and Wikiquote, integrated and sorted according to its impact on both high and popular culture.

## Acknowledgments

The work reported here was carried out within the research project financed by the Polish National Science Centre (contract number 2014/15/B/HS2/00182) and was partially financed as part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education.

## References

- Gil Francopoulo. 2013. *LMF. Lexical Markup Framework*. ISTE - Wiley.
- Instytut Języka Polskiego PAN. 2010. *Słownik języka polskiego XVII i 1. połowy XVIII w. [En. Dictionary of 17 century and 1st half of 18 century Polish]*. Warszawa.
- Piotr Żmigrodzki. 2007. O projekcie Wielkiego słownika języka polskiego. *Język Polski*, 5(LXXXVII):265—267.
- George A. Miller. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41.
- Maciej Piasecki, Stanisław Szpakowicz, and Bartosz Broda. 2009. *A Wordnet from the Ground Up*. Oficyna Wydawnicza Politechniki Wrocławskiej.
- Adam Przepiórkowski, Mirosław Bańko, Rafał L. Górski, and Barbara Lewandowska-Tomaszczyk, editors. 2012. *Narodowy Korpus Języka Polskiego [Eng.: National Corpus of Polish]*. Wydawnictwo Naukowe PWN, Warsaw.
- Adam Przepiórkowski, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski, and Marek Świdziński. 2014. Walenty: Towards a comprehensive valence dictionary of Polish. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, pages 2785–2792, Reykjavík, Iceland. ELRA.
- Zygmunt Saloni, Marcin Woliński, Robert Wołosz, Włodzimierz Gruszczyński, and Danuta Skowrońska. 2012. *Słownik gramatyczny języka polskiego*. Warszawa, 2. edition.
- Zygmunt Saloni, Marcin Woliński, Robert Wołosz, Włodzimierz Gruszczyński, and Danuta Skowrońska. 2015. *Słownik gramatyczny języka polskiego*. 3. edition, online publication.
- Marcin Woliński and Witold Kieraś. 2016. The online version of Grammatical Dictionary of Polish.

In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC 2016*, pages 2589–2594, Portorož, Slovenia. ELRA, European Language Resources Association (ELRA).

Marcin Woliński. 2006. Morfeusz – a practical tool for the morphological analysis of Polish. In Mieczysław A. Kłopotek, Sławomir T. Wierzchoń, and Krzysztof Trojanowski, editors, *Proceedings of the International Intelligent Information Systems: Intelligent Information Processing and Web Mining 2006 Conference*, pages 511–520, Wisła, Poland, June.

Marcin Woliński. 2014. Morfeusz Reloaded. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, pages 1106–1111, Reykjavík. European Language Resources Association.