

Krzysztof Szafran
Institute of Informatics, Warsaw University
kszafran@mimuw.edu.pl

STRUCTURE OF MEANING IN THE DICTIONARY OF POLISH OF THE 16th CENTURY

ABSTRACT: The dictionary of Polish of the 16th century is one of the three great dictionaries of historical Polish. It is being compiled and published with great effort mainly in traditional form, i.e. on paper. Structure of its entries is very complicated and varied. In the paper we present a short analysis of the structure of the semantic part of entries. As a result of this analysis, a formal description of the structure in XML is shown.

KEYWORDS: Polish of the 16th century; formal description; XML.

1.1. Preliminary notes

This paper will be concerned with a dictionary of Polish of the 16th century: „Słownik polszczyzny XVI wieku” (MAYENOWA, PEPEŁOWSKI(1966-); henceforth SPXVIw). A short introduction to it can be found in PIOTROWSKI, SZAFRAN(2005).

SPXVIw is one of the largest and most important dictionaries being prepared in Poland. The dictionary, which has been in preparation since the 1940s (the first volume was published in 1966), is still published mainly in traditional form, i.e. on paper.

A few years ago some efforts were undertaken (see SZAFRAN(2003)) to prepare a formal description of the entries of the dictionary. The paper shows some results that were obtained. It discusses briefly the structure of the main part of an entry in the dictionary, i.e. the semantic part, and presents possible formal description of this structure in a form of a DTD¹. More exhaustive description of an entry and its structure can be found in SZAFRAN(2007).

The semantic part of an entry in *SPXVIw* is the most important, largest, and most complicated part of an entry.

The basic structure of meaning is described by a multilevel hierarchy of so-called numbered semantic sections. There are separate sections for each semantic shade or for each difference in scope. Each of them can be divided according to the same rule. It leads to a great differentiation of types of semantic sections. Additionally there exists a rule which allows an editor to create a new

¹ We assume that the reader is familiar with XML and DTD formalisms to a degree, which allow her/him understand very simple examples presented in the paper.

type of the section when necessary. These sections are numbered with Arabic numbers, small Latin letters, small Greek letters, and so on.

Sometimes such hierarchy is recognized as insufficient and in such a case additional levels of hierarchy are introduced (numbered by capital Latin letters). There are entries in which the semantic part consists of eight numbered levels of description.

1.2. Structure of meaning

1.2.1. Basic structure

A basic notion used for the description of the semantic part of an entry in SPXVIw is the semantic section (in Polish: *rubryka znaczeniowa*). Such a section includes language evidence and can consist of several parts:

1. general part, immediately following the definition of meaning, containing material which does not belong to any of the following (sub)parts,
2. syntactic structures part²,
3. grammatical sections part³,
4. part containing connections with specific types of words⁴,
5. phraseological part,
6. part describing ironical, facetious, offensive, contemptuous usages,
7. part describing metaphorical usages,
8. **subsections for semantic shadows and differences in scope** (so-called numbered subsections).

Additionally, this structure is defined in an iterative way. It means that each numbered subsection and part describing metaphorical usages can be divided in the same way as above.

In practice, there exist entries with rather deep structures. For example, in the entry **PRZECIWI** there are eight levels of description, numbered: **I, 1., A., AA, a., α., αα., ααα.**

Some of the above mentioned types of sections can be divided and described much more precisely. For example, sections created on the basis of connection can have (among others) one of the following headers:

- In connection with a name or a family name,
- In connection with a locality name,
- In connection with an adjective derived from a locality name,
- Possibly other: *In connection with (...)*,
- In connections in series,
- In pleonastic connections,
- In intensifying repetition,
- In opposition,
- In comparison,
- In formula (...),
- Titles,
- Etymological explanations.

² See e.g. *przechod przez co (...), przechod dokąd (...)* in the entry **PRZECHOD**.

³ See e.g. *N sg w funkcji orzecznika: (...)* in the entry **PRZEDZIWNY** or *G sg w funkcji przydawki do tytułu utworu: (...)* in the entry **PRZEKŁADANIE**.

⁴ See e.g. *W połączeniu z przymiotnikiem od nazwy organu: (...)* or *W charakterystycznych połączeniach: (...)* in the entry **PRZECHOD**.

Also the phraseological part can be divided for:

- proverbs,
- phrases,
- idioms,
- expressions,
- comparisons,
- series,
- adverbial expressions,
- connections⁵,
- repetitions.

Semantic (sub)sections are established for distinct semantic shades, on the basis of meaning (see e.g. **KAZAĆ**), scope (see e.g. **KOŁO**) and grammatical characteristics (e.g. adjective in the position of a noun). They are usually numbered with Arabic numbers. Successive levels of the structure are numbered with small Latin letters, single small Greek letters, double small Greek letters, and so on.

In specific situations, numbers on some levels can be omitted. It means that on this level there is only one numbered semantic section. Such a case can be seen in the entry **PRZE** shown in the next paragraph (there is no number immediately before 'Bo rozumiała').

1.2.2. Multi-stage partitioning

Editorial instructions anticipate that for some more complicated situations the schema shown in the previous paragraph is insufficient. In such cases, additional levels of the semantic structure are introduced.

The most typical situations when additional partitioning is necessary are as follows.

1. Language evidence described in the entry belongs to two different grammatical categories.

PRZECIW (4752) *praep i av*

(...)

I. *Przyimek* (4750):

Cum D (4749), *cum G* (w znaczu **1.**) (1)

1. *W funkcji przestrzennej;* (...)

(...)

II. *Przystówek: po przeciwnej stronie* (...)

In the entry **PRZECIW** there are two parts. The first one contains material described as a pronoun (*praep, przyimek*), and the second one contains material categorized as an adverb (*av, przystówek*).

Both parts are numbered with Roman numbers.

2. The entry describes a pronoun which has two (or more) different meanings.

PRZE (2780) *praep*

(...)

I. *Cum A* (2778):

1. *Wskazuje* } (...)

(...)

⁵ This section is different from the sections *In connection with* (...)

II. Cum G: bez (2): Bo rozumiała (...)

Additionally, in this entry attention should be drawn to the different ways of presenting semantic discrimination of the adverb in the entries **PRZECIW** and **PRZE**.

3. The entry describes an adverbial noun.

PRZECHODZENIE (12) *sb n*

(...)

I. Rzeczownik od „przechodzić“: (...)

(...)

II. Rzeczownik od „przechodzić się“: (...)

(...)

III.1. [Miejsce służące do (...)]

(...)

4. The entry concerns some other situations.

PRZEGRANY (25) *part praet pass pf*

(...)

I. W funkcji właściwej (22):

1. (...)

(...)

II. W funkcji czynnej: który przegrał (...)

(...)

III. W funkcji rzeczownika: >>przegrana<<

Such partitioning can be found also in entries that describe perfect and imperfect forms of verbs (i.e. with grammatical information of the form: *vb pf i impf*).

1.2.3. Additional levels in the semantic structure

In some entries we can observe extensions of the semantic structure as compared to the one presented earlier. There are additional levels of structure, which can be visible below the basic level of the semantic section (the one numbered with Arabic numbers). Those additional levels are numbered by capital Latin letters, double capital Latin letters and so on, and can be used at any level of the structure. As one can see, semantic sections numbered by letter A or letters AA in the example below are positioned in different places in section 1. and section 5. As an example, we present a small extract from the entry **PRZED**.

I. Cum I (10175) (...)

(...)

1. Przy określaniu stosunków przestrzennych (...)

A. Punktem odniesienia jest obiekt nieożywiony (...)

(...)

a. W pobliżu (...)

(...)

B. Punktem odniesienia w zasadzie jest (...)

AA. W stosunku do (...)

a. O kolejności zajmowanego (...)

(...)

BB. W stosunku do (...)

- (...)
- 5. *O kolejności czasowej (...)*
 - a. *Wskazuje na odstęp czasowy wstecz (...)*
 - (...)
 - b. *Wskazuje na moment odniesienia (...)*
 - A. *Momentem odniesienia (...)*
 - AA. *W stosunku do pory dnia (...)*

1.3. XML representation of the structure of meaning

Despite the great variety of possible types of the semantic sections, most of them are organized in a rather uncomplicated way. Of course, many of them have their own specific subsections, and each of them has its own specific details. If we wanted to describe all these sections, we would have to define many different XML-elements, but this job would be more laborious than difficult. For this reason, in the DTD presented in SZAFRAN(2007), we have defined only the elements which were necessary for showing the principle on which the description of the structure is based, i.e. the elements describing the numbered sections. From the great variety of the unnumbered sections, we have defined only those elements which were necessary to describe examples of the entries from SPXVIw presented in the book.

1.3.1. Multi-stage partitioning

The most external or the highest level of description in the semantic part of the entry is the level corresponding to the multi-stage partitioning, which was described in p. 1.2.2 (page 3). This level of description is represented by the XML element `<poziom-zewnetrzny>`. This element was defined as obligatory and the only possible subelement of the `<semantyka>` element, which describes the semantic part of the entry (see line 825 in Figure 1). If multi-stage partitioning doesn't exist in the entry, this level of description has to be "hidden" (in presentation) by an optional attribute `jedyny` (see line 832 in Figure 1). In turn, the `<poziom-zewnetrzny>` element always refers to the `<poziom-1>` element, which always exists. This element describes the first level of the numbered sections in the semantic structure.

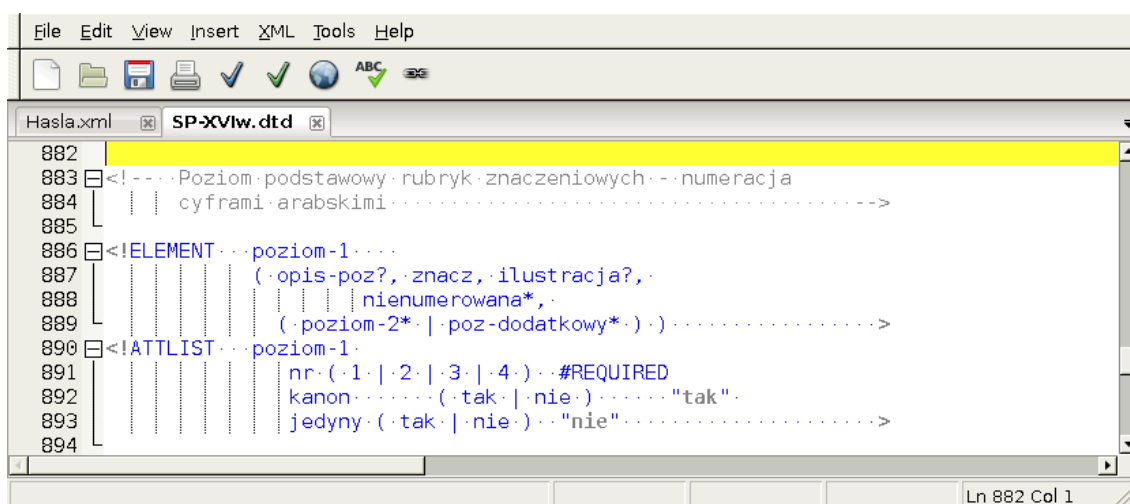
Figure 1. Definition of the `<poziom-zewnetrzny>` element.

1.3.2. Basic structure of meaning

Because the numbered semantic sections create a backbone for the meaning structure of the entry, we have defined them in XML in a way which guarantees the correct sequence of levels. We have obtained it by defining several elements of the form shown in Figure 2.

The matter consists in defining a sequence of XML-elements in such a way that element `<poziom-1>` requires⁶ `<poziom-2>`, which in turn requires `<poziom-3>`, and so on. Each of the above mentioned elements has the obligatory attribute `nr`. However, the set of values for each of the elements is different: 1, 2, 3, ... for `<poziom-1>`, a, b, c, ... for `<poziom-2>`, and so on (see e.g. line 891 in Figure 2). Optional attribute `jedyny` makes it possible to omit the number when it is necessary.

More attentive readers probably noticed that in the definition of `<poziom-zewnetrzny>` in line 828, next to `<poziom-1>` and in the definition of `<poziom-1>` in line 889, next to `<poziom-2>` the XML-element `<poz-dodatkowy>` (as an optional alternative) can be seen. Such a solution guarantees a possibility of placing elements describing additional levels of the meaning structure mentioned in paragraph 1.2.3. (page 4) in any place. Similar constructions can be seen in many other definitions.



```
File Edit View Insert XML Tools Help
Hasla.xml SP-XV/w.dtd
882 <!-- Poziom podstawowy rubryk znaczeniowych --> numeracja
883 <!-- cyframi arabskimi -->
884 <!--
885 <!--
886 <!-- ELEMENT --> poziom-1 -->
887 <!-- (. opis-poz?, .znacz, .ilustracja?, .
888 <!-- nienumerowana*, .
889 <!-- (. poziom-2* | .poz-dodatkowy*) -->
890 <!-- ATTLIST --> poziom-1
891 <!-- nr (.1 | .2 | .3 | .4) --> #REQUIRED
892 <!-- kanon (.tak | .nie) --> "tak"
893 <!-- jedyny (.tak | .nie) --> "nie"
894 <!--
```

Figure 2. Definition of the `<poziom-1>` element.

Attention should be drawn to line number 887 in the definition of the `<poziom-1>` element. In this line one can see only one obligatory element, i.e. `<znacz>`. This element marks the meaning defined by this level of description. Because the `<poziom-i>` elements only define the structure of the meaning part of the entry, the meaning as such has to be defined separately. That is why the same element `<znacz>` has to be present on every level of the structure.

⁶ Because the depth of the structure is different for different entries all `<level-i>` elements (starting from `i=2`) have to be defined as optional.

1.3.3. Additional levels in the semantic structure

```
846 847 <!ELEMENT poz-dodatkowy ( opis-poz-dod?,
848 ( poz-dodatkowy* |
849 poziom-1 |
850 poziom-2 |
851 poziom-3 ) ) >
852 <!ATTLIST poz-dodatkowy
853 nr ( A | AA | AAA | B | BB | BBB |
854 C | CC | CCC ) #REQUIRED
855 kanon ( tak | nie ) "tak"
856
857 <!ELEMENT opis-poz-dod ( #PCDATA | liczba-wystapień ) * >
858
```

Figure 3. Definition of the <poz-dodatkowy> element.

Because this type of the semantic section can be found virtually in any place (on an arbitrary level of the structure), we propose describing them with a single XML-element <poz-dodatkowy>. Our definition can be seen in Figure 3.

1.3.4. Filling the structure

The semantic part of a typical entry looks as follows:

PRZA (67) *sb f*

(...)

1. *Konflikt, spór; Lis Mącz, PolAnt; causa PolAnt, Modrz; iudicium PolAnt; controversia Mącz (43): (...)*

(...)

2. *Skarga sądowa, pozew, powództwo; treść skargi; controversia JanStat (21): (...)*

(...)

3. *Sprawa, problem; causa, iudicium PolAnt (3): (...).*

Apart from the numbers, which belong to the structure definitions, first parts (sometimes called headers) of the three presented semantic sections consist of the same elements: an explanation of the meaning, Latin equivalents, information about the source and information about the number of appearances. All of them are components of the <ZNACZ> element mentioned earlier. Simple XML definition of all these elements may look as in Figure 4.

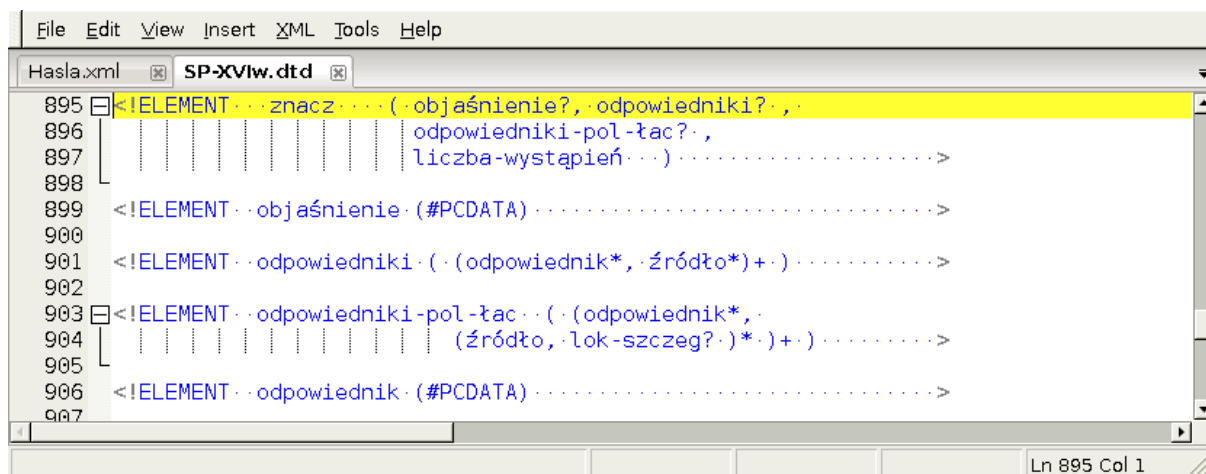


Figure 4. Definition of some elements filling the structure.

1.4. Sample unnumbered sections

Only by way of example we present here two types of unnumbered sections, quite frequently seen in SPXVIw. The first are the so-called syntactic structures.

prza czyja [= kogo dotycząca] [w tym: pron poss (1), G sb (1)] (2): (...)
 (...)
 przebierać komu [zawsze: sobie] (2): (...)

As one can see, the phrase "syntactic structure" does not appear in the text of the entry. The reader has to know that the description which looks as shown in the example describes a syntactic structure (if such categorization is imported for her/him). But in many cases unnumbered sections in the dictionary have their own explicit headers, for example phrases, idioms, (adverbial) expressions, and so on.

The structure of the header of this section is similar to the headers of the numbered sections. The header shows the schema of the syntactic structure. Besides, it contains some additional elements, which explain the meaning of the schema. The XML-representation can be seen below.

```

<nienumerowana>
  <składniowa>
    <schemat>
      <opis-sch>
        <opis-sch-podst>
          przą czyja
        </opis-sch-podst>
        <os-czyli>kogo dotycząca</os-czyli>
        <rn-w-tym>
          <rn-w-czym>pron poss</rn-w-czym><liczba-wystąpień>1</liczba-wystąpień>
          <rn-w-czym>G sb</rn-w-czym><liczba-wystąpień>1</liczba-wystąpień>
        </rn-w-tym>
      </opis-sch><liczba-wystąpień>2</liczba-wystąpień>
    </schemat>
    <ilustracja> (...) </ilustracja>
  </składniowa>
</nienumerowana>

```

Here's another example of the unnumbered section. This section belongs to the category "on the basis of connection" and subcategory "in opposition".

W przeciwstawieniach: >>kwaśny (3), kwaszony (...) przasny<< (4):

This section informs the reader that there are two different forms of the opposition: more frequent *kwaśny - przasny* and rarer *kwaszony - przasny*. In XML, this section can be described in the following way.

```
<nienumerowana>
  <na-zas-połączeń typ-poł="w-przeciwstawieniach">
    <zasada-poł>
      <opis-zasady>
        <el-lewy>kwaśny</el-lewy><liczba-wystąpien>3</liczba-wystąpien>
        <el-lewy>kwaszony</el-lewy>
        <el-prawy>przasny</el-prawy>
      </opis-zasady><liczba-wystąpien>4</liczba-wystąpien>
    </zasada-poł>
    <ilustracja> (...) </ilustracja>
  </na-zas-połączeń>
</nienumerowana>
```

1.5. Few remarks about metaphoric meanings

And finally we want to present an example of the section which describes metaphorical usages of a word. This section starts from the word *Przen* and contains few⁷ other unnumbered sections

1. PRZEBADAĆ SIĘ (2) *vb impf*

(...)

Przekłuwać swoje ciało, przebijać się zadając sobie rany (2):

(...)

Przen (1):

W porównaniu: (1): SarnUzn D3v cf »swym mieczem się przebadać«.

Zwroty: »swym mieczem się przebadać« = przytaczać argumenty sprzeczne z własnymi twierdzeniami (1): Pirwey [adwersarze] powiadałże Sin był (...)

»sam się przebadać« (1): SarnUzn D3v cf »swym mieczem się przebadać«.

Szereg: »pobijać się i przebadać« (1): SarnUzn D3v cf »swym mieczem się przebadać«.

Formally, this section is categorized as the unnumbered section, but its structure corresponds to the structure of the numbered semantic section of the first level, in spite of the fact that the general part of it is empty.

An interesting question is whether there exist entries for which the metaphorical usages section has multilevel structure.

Acknowledgment

The text was proofread by Radosław Moszczyński.

⁷ In the entry PRZEBADAĆ SIĘ, in section describing metaphorical usages there exist three unnumbered subsections. It's interesting, that in such a way is described only **one example** from sermons of Sarnicki.

References

- MAYENOWA M.R., PEPEŁOWSKI F.** (editors-in-chief) (1966—), *Słownik polszczyzny XVI wieku* [Dictionary of Polish in the 16th century], Wrocław: Ossolineum, (Since 1995 published by Instytut Badań Literackich, PAN, Warszawa.).
- SZAFRAN K.** (2003) *Ku elektronicznej wersji Słownika polszczyzny XVI wieku.*(Tekst referatu wygłoszonego na Zjeździe Polskiego Towarzystwa Językoznawczego w Tarnowie, wrzesień 2003) *Biuletyn PTJ*, vol. LX, ISSN 0032-3802, ss. 89 -- 97, 2004.
- PIOTROWSKI, T., SZAFRAN K.** (2005), The dictionary of Polish of the 16th century and the computer: from paper to (structured) file, in: Kiefer K., Kiss G., Pajzs J. (editors), *Papers in CL Complex 2005*, Budapest, pp 171-179.
- SZAFRAN K.** (2007), *Analiza i formalny opis struktury "Słownika polszczyzny XVI wieku"*, Wydawnictwa UW, ISBN 978-83-235-0358-3, 218s., Warszawa 2007.