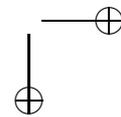Adam Przepiórkowski

**On the Computational Usability of
Valence Dictionaries for Polish**

Nr 971

Warszawa, December 2003

Abstract

Valence dictionaries, i.e., dictionaries specifying possible arguments of words, usually verbs, play an important role in natural language processing (NLP): they are used, e.g., in syntactic parsing, part-of-speech tagging, information extraction and in word sense disambiguation. The aim of this report is to critically evaluate the design of existing valence dictionaries for Polish and to identify their strengths and weaknesses. This is a preliminary step towards the design of a reusable valence dictionaries for Polish, useful in NLP applications.

Keywords: NLP, valence, MRDs, syntactic dictionaries

Streszczenie

O użyteczności słowników walencyjnych języka polskiego w przetwarzaniu języka naturalnego

Słowniki walencyjne, tj. słowniki podające informację o możliwych argumentach wyrazów, przede wszystkim czasowników, grają ważną rolę w przetwarzaniu języków naturalnych (ang. *natural language processing*; NLP): są one używane w analizatorach składniowach, w anotacji korpusów częściami mowy, w ekstrakcji informacji oraz w semantycznej dezambiguacji form wyrazowych. Celem niniejszego raportu jest krytyczna ocena istniejących słowników walencyjnych języka polskiego. Jest to pierwszy krok zmierzający do zaprojektowania uniwersalnych słownika walencyjnego dla języka polskiego, użytecznego w zastosowaniach NLP.

Słowa kluczowe: przetwarzanie języka naturalnego, walencja, słowniki elektroniczne, słowniki składniowe
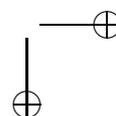
The aim of this article is to discuss the kinds of information provided by the currently available valence dictionaries for Polish, from the perspective of their usability in Natural Language Processing (NLP) tasks. First, §1 describes the scope and purpose of valence dictionaries, and §2 presents existing valence dictionaries for Polish, then §3, the core section of this article, contains the discussion of various design flaws of those dictionaries, and, finally, §4 contains final remarks. Conclusions of this discussion are taken into account in the design of a machine-readable valence dictionary for Polish, currently developed at the Institute of Computer Science, Polish Academy of Sciences (Przepiórkowski, 2004).

1   Introduction

Valence dictionaries are lexica which contain information about predicates', usually verbs', argument structures (so-called valence frames). For example, valence dictionaries for Polish should contain the information that the Polish verb LUBIĆ 'to like' always takes a (nominative) subject and either an (accusative) object (*Janek lubi Marysię* 'Janek.NOM likes Marysia.ACC'), or an infinitival phrase (*Janek lubi patrzeć na Marysię* 'Janek.NOM likes [to look.INF at Marysia]'), or a subordinate clause introduced by one of a class of complementisers and question words (e.g., *Janek lubi, kiedy Marysia na niego patrzy* 'Janek.NOM likes [when Marysia looks at him]'), possibly among other valence frames. Sometimes valence dictionaries contain additional semantic information, e.g., which valence frames correspond to which meanings of the verb, and what semantic restrictions apply to what arguments of the verb (e.g., the subject of LUBIĆ is assumed to be human, or at least sentient.

Valence dictionaries are extremely useful resources, not only within NLP. First of all, and most obviously, they may be used as general references for learners and users of a given language. They are also very valuable for linguistic research on types of predicate structures and on morphosyntactic realisations of semantic arguments.

From the NLP perspective, valence dictionaries are of pivotal importance for the Computational Linguistic task of constructing so-called deep parsers, i.e., parsers which find the full syntactic and possibly some semantic structure for natural language sentences; cf. Przepiórkowski *et al.* 2002 on constructing such a deep parser for Polish. Deep parsers are used, e.g., in some Machine Translation (e.g., Verbmobil, Wahlster 2000) and Question-Answering (e.g., Javelin, Durme *et al.* 2003) systems. Valence information is also useful in shallow parsing, where only certain aspects of the structure of a sentence
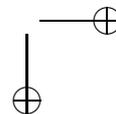
are taken into account, e.g., only noun phrases or only predicates and their arguments. Information Extraction is one of typical application areas of such shallow parsers, cf., e.g., Piskorski *et al.* 2003. Finally, and perhaps surprisingly, such valence information is useful for the task of part of speech (POS) disambiguation, e.g., for the purpose of corpus annotation or speech recognition. A high profile example of a rule-based tagger making use of valence information is ENGCG (Voutilainen, 1995).

This brief overview of some of the NLP applications of valence dictionaries shows that a well-designed large-scale machine-readable valence dictionary for a given language may substantially increase the quality of various NLP systems for that language. The following section takes this perspective in examining the currently available valence dictionaries for Polish.
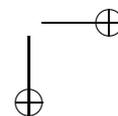
2 Valence Dictionaries for Polish

There is currently no publicly available electronic valence dictionary for Polish. There are three publicly available traditional dictionaries containing valence information:

- *Słownik syntaktyczno-generatywny czasowników polskich*, henceforth SS-GCP, Polański 1992, published in 5 volumes which appeared between 1980 and 1992; this is probably the most extensive existing source of valence information on Polish verbs;

- *Inny słownik języka polskiego*, ISJP, Bańko 2000, a 2-volume general dictionary of Polish which contains various grammatical characteristics of lexemes and their meanings, including valence information;

- *Słownik walencyjny czasowników niemieckich i polskich*, SWCNiP, Morciniec *et al.* 1995, a valence dictionary of German verbs and their Polish counterparts, rather modest both with respect to the number of lexemes and the exhaustiveness of valence information.

Since such dictionaries usually use non-trivial typesetting conventions, they cannot be easily converted to the electronic form using existing OCR software. At the time of writing this article, there are at least two projects (one academic, cf. Grund 2000, and one commercial) aiming at converting Polański 1992 to an electronic form, but the results of these efforts are currently not publicly available.

Another electronic valence dictionary, also not publicly available at the time of writing this article, is:

- *Syntactic Dictionary of Polish Verbs*, SDPV, Świdziński 1994, an unpublished list of valences of some Polish verbs.

Let us illustrate the kinds and format of information that these dictionaries provide, taking the verb ZACZĄĆ 'start, begin' as an example.

2.1 Morciniec *et al.* 1995

SWCNiP offers the following information about the verb ZACZĄĆ:[1]

(1)   *Actants*: nom prp:*od*+gen/instr/adv
     Meaning: nom   HUM INST OBJ ABSTR
                prp   –
                instr OBJ: *text*, *phrase*, etc.
                adv   QUAL

According to (1), ZACZĄĆ takes two arguments: a nominative ('nom') subject and a complement, which may be realised as a PP headed by the genitive-taking preposition *od* (cf. 'prp:*od*+gen'), as an instrumental phrase (cf. 'instr'), or as an adverbial phrase ('adv'). Semantically, the nominative subject may refer to humans ('HUM'), institutions ('INST'), countable objects ('OBJ') or abstract objects ('ABSTR') such as theories, opinions, activities, states and qualities. Moreover, semantic reference of instrumental complements is limited to such countable objects as texts or phrases (presumably, as in 'starting a speech'), while adverbial complements express quality ('QUAL'), i.e., in this case, manner (presumably, as in 'starting in some way').

This information is both more modest and more controversial than in the case of the other valence dictionaries for Polish, discussed below.

2.2 Świdziński 1994

The following two entries are provided for ZACZĄĆ in SDPV:

```
zacząć 1 [Vp 10c]: 11 B; 12 od-D$/$z-D;
                   15 Ps; 16 V; 22 B+od-D; @1
zacząć 2 [Qp 10c]: 14 Q
```

Thus, according to this dictionary, there are two verbs ZACZĄĆ: they are both perfective ('p' in '[Vp 10c]') and they belong to the same inflectional class ('10c'), but ZACZĄĆ 1 is a regular subject-taking verb ('V' in '[Vp 10c]'),

---

[1]Here and below, valence information is presented in formats close to that of the valence dictionaries under discussion.
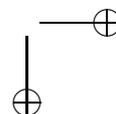
while ZACZĄĆ 2 is a quasi-verb, *i.e.*, a subjectless verb. The latter verb's only argument is an infinitival phrase headed by a quasi-verb ('14 Q'[2]), as in *Zaczęło padać* 'It started to rain', lit. 'started rain.INF'.

The regular verb ZACZĄĆ 1, on the other hand, has a number of valence frames: apart from the subject, it make take either an accusative nominal phrase ('11 B'; e.g., *Janek zaczął książkę* 'Janek started a book'), or prepositional phrases headed by *od* ('12 od-D\$'; e.g., *Zaczął od sprawozdania* 'He started with the report', lit. 'Started.3.SG.MASC from report'; the following dollar sign indicates possible idiomatic realisations of this frame) or by *z* ('12 \$/z-D'; the preceding dollar sign indicates that all realisations are idiomatic), or an adverbial phrase ('15 Ps'), or an infinitival phrase headed by a regular verb ('16 V'; e.g., *Janek zaczął śpiewać* 'Janek started to sing'), or two complements: an accusative nominal phrase and a PP headed by *do* ('22 B+od-D'; e.g., *Janek zaczyna dzień od kawy* 'Janek starts a day with a coffee'), or, finally, an indirect speech complement ('@1'; e.g., *„Pada", zaczął Janek* '"It's raining," started Janek').

2.3 Polański 1992

Apart from a number of frames realised only in idiomatic expressions involving ZACZĄĆ, SSGCP specifies three meanings of this verb, given below as I., II. and III., with three valence frames corresponding to the first meaning:

I. 'start to act; start, enter an initial period of some state'

1. $\mathbf{NP_N} - \left\{ \begin{array}{c} \mathbf{IP} \\ \mathbf{NA} \end{array} \right\}$
   $\mathrm{NP_N} \longrightarrow \left[\begin{array}{c} +\mathrm{Hum} \end{array} \right]$

2. $\mathbf{NP_N} - \mathbf{IP}$
   $\mathrm{NP_N} \longrightarrow \left[\begin{array}{c} +\mathrm{Anim} \\ -\mathrm{Hum} \end{array} \right] \left[\begin{array}{c} -\mathrm{Abstr} \\ -\mathrm{Anim} \end{array} \right]$

3. $\qquad - \mathbf{IP}$

II. 'take, make use of something'

$\mathbf{NP_N} - \mathbf{NP_{Acc}}$
$\mathrm{NP_N} \longrightarrow \left[\begin{array}{c} +\mathrm{Hum} \end{array} \right]$
$\mathrm{NP_{Acc}} \longrightarrow \left[\begin{array}{c} -\mathrm{Abstr} \\ -\mathrm{Anim} \end{array} \right]$

---

[2]The numbers 14, 11, etc., refer to valence *schemes*, e.g., in case of regular verbs, 11 refers to the 'subject + nominal complement' scheme, 12 — to the 'subject + prepositional complement' scheme, etc.
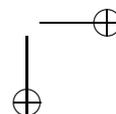
III. 'start talking'

$$\mathbf{NP_N} - \left\{ \begin{array}{l} \mathbf{OR} \\ \mathbf{NP_{Acc}} \end{array} \right\}$$
$$\mathrm{NP_N} \longrightarrow [\ +\mathrm{Hum}\ ]$$
$$\mathrm{NP_{Acc}} \longrightarrow [\ \mathrm{Inf}\ ]$$

Apart from the third frame of the first meaning, all frames mention the nominative nominal phrase subject ('$\mathrm{NP_N}$'). In case the subject is human ('+Hum'), the complement may either be an infinitival phrase ('IP'; e.g., *Janek zaczął czytać* 'Janek started to read') or an action nominal ('NA'; e.g., *Janek zaczął pranie* 'Janek started washing up'). Only an infinitival complement is possible in case the subject is non-human (i.e., in case it is non-human animate or a non-animate concrete object). The third frame of the first meaning corresponds to the quasi-verb ZACZĄĆ 2 of SDPV (but without the information that the infinitival complement must be headed by a quasi-verb).

In case of the second meaning, the nominative subject must be semantically human, while the accusative NP complement must be a concrete object, as in *Janek zaczął butelkę wina* 'Janek started a bottle of wine'.

Finally, in case of the third meaning, the complement is either indirect speech (*oratio recta*) or an accusative NP expressing information, as in *Zaczął swoje przemówienie* 'He started his speech'.

2.4 Bańko 2000

Last but not least, ISJP is a general dictionary of Polish with valence information given on the margin, next to each meaning within a lexical entry. For (the non-reflexive) ZACZĄĆ, 6 related non-idiomatic meanings are provided with 6 different valence frames. The following table provides and briefly explains those valence frames, without citing the corresponding meanings, most of which are related to the meanings exemplified in previous subsections.

(2)

[(BEZOK)]	an optional infinitival phrase
[BEZOK]	an obligatory infinitival phrase
[B]	an accusative phrase
[B+(*od*-D/N)]	two arguments: an accusative phrase and either an optional PP headed by the genitive-taking *od* or an instrumental phrase
[(CYT)]	optional indirect speech
[(*w*-Ms)]	optional PP headed by the locative-taking *w*; e.g., *Zaczęła w handlu* 'She started in trade'

5
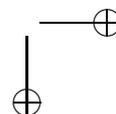
3 Design Deficiencies

The following subsections discuss various general design problems exhibited by Polish valence dictionaries, ignoring particular omissions or errors.

3.1 Arguments vs. Adjuncts

One fundamental question that valence dictionaries crucially rely on and consistently fail to address is what criteria are assumed for deciding whether a given type of dependents should be mentioned in valence frames. For example, should the phrases emphasised below be explicitly given in valence frames of the respective verbs?

(3)   Łódź leży *niedaleko Warszawy*.
      Łódź lies near   Warsaw

      'Łódź is situated not far from Warsaw.'

(4)   Pamiętam   go   *miłym*.
   remember.1.SG him.ACC *nice*.INS

      'I remember him as nice.'

The Polish-German valence dictionary SWCNiP does not seem to record the verbs LEŻEĆ 'lay, be situated' and PAMIĘTAĆ 'remember' at all.

Świdziński's SDPV gives two relevant valence frames for LEŻEĆ, namely, with prepositional phrases (PPs) headed by the locative-taking preposition *na*,[3] and with adverbial phrases. The latter valence frame subsumes the use of LEŻEĆ in (3), on the assumption that *niedaleko* 'near' in that example is an adverb, and not a genitive-taking preposition. Moreover, SDPV also specifies a valence frame for PAMIĘTAĆ 'remember' corresponding to (4). Similar frames specifying an accusative object and an instrumental adjective are given for a few other verbs, e.g., WIDZIEĆ 'see', cf. (5), but not, e.g., for ZASTAĆ 'find', cf. (6).

(5)   Widziałem   go   pijanym.
      saw.1.SG.MASC him.ACC drunk.INS

      'I saw him drunk.'

(6)   Zastałem   go   pijanym.
      found.1.SG.MASC him.ACC drunk.INS

      'I found him drunk.'

---

[3]However, other possible locative prepositional phrases, such as [w + locative], are not included in that entry.
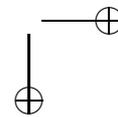
According to the valence frame for LEŻEĆ as 'be situated' in Polański's SSGCP, this verb may occur with an NP_{Loc}, i.e., a 'locative nominal phrase' (Polański, 1980, p. 9), but — as is made clear in the preface — such nominal phrases subsume locative adverbs and — as suggested by examples illustrating this and similar valence frames — they also subsume locative prepositional phrases, so sentences such as (3) are accounted for. On the other hand, SSGCP apparently does not treat instrumental dependents of PAMIĘTAĆ as worth recording — there is no valence frame matching (4).[4]

Finally, Bańko's ISJP does not include any explicit information about instrumental adjective phrases in the lexical entry for PAMIĘTAĆ, but it does mention the dependent of LEŻEĆ, specifying it as an adjunct ('OK', i.e., Pol. *okolicznik*).

These differing treatments of the locative dependent of LEŻEĆ and the instrumental predicative dependent of PAMIĘTAĆ reflect the confusion about the argument/adjunct distinction. As discussed in Przepiórkowski 1999a,b, 2002, the theory of valence has been flawed ever since its conception in Tesnière 1959, where three *pairwise incompatible* criteria were given for distinguishing complements and adjuncts (here summarised after Vater 1978, p. 22):

- the *morphological-syntactic criterion*: complements are noun phrases, adjuncts are prepositional phrases;

- the *semantic criterion*: complements express the persons or things participating in the process in a special way, whereas adjuncts express the time, the place, the manner, etc. connected with that process;

- the *functional criterion*: complements, unlike adjuncts, are indispensable to complete the meaning of the verb; hence the number of complements, but not adjuncts, is limited for every verb.

Subsequently, a number of further tests for distinguishing complements and adjuncts have been proposed (e.g., substitution by verbal proforms, iterability, extraction across clausal boundaries, etc.), again resulting in different partitions of dependents into arguments and adjuncts.

Given the multitude of views on what constitutes an argument and what information should be provided in valence frames, any precise definition of argument will be controversial and to some extent arbitrary. Nevertheless, such a definition, translatable into a formal test for argumenthood, should be

---

[4]Such an instrumental phrase is, however, included in a valence frame for WIDZIEĆ 'see', as in (5), so the lack of such information in the lexical entry of PAMIĘTAĆ might be an accidental omission or it might reflect different views of the argument/adjunct dichotomy of different SSGCP editors.
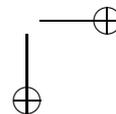
provided with any valance dictionary, if this dictionary is meant to be used in NLP applications.

3.2 Syntax-Semantics Interface

Two of the valence dictionaries for Polish, SSGCP and SWCNiP, provide some semantic information, namely, semantic restrictions on arguments. Of the other two dictionaries, SDPV lacks any semantic information, while ISJP provides semantic information about the arguments and about the lexical semantics of the predicate, but only in free text, in bodies of lexical entries.

It should be noted that the kind of semantic information provided by SSGCP and SWCNiP is imprecise and defeasible. For example, 'human' should almost always mean 'sentient' or 'volitional agent', information about animateness is defeasible not only in the context of fairy stories and magic, but also as soon as one starts talking about computers and cars, etc. Given the human ability to use language creatively and metaphorically (Lakoff and Johnson, 1980), examples violating such semantic restrictions are easy to find in everyday contexts. This questions the ultimate usability of such information.

On the other hand, this is not to say that there are no hard semantic restrictions on arguments. For example, it is well-known that so-called embedded interrogatives express semantic questions when used as complements of some verbs, and propositions, when used as complements of other verbs (e.g., 'to *ask* who came' is to ask *a question*, while 'to *know* who came' is to know *an answer to a question*; cf. Ginzburg and Sag 2000 and references therein). This, however, is a restriction at a more abstract level than '±human'.

More importantly, the dictionaries listed above do not provide any information about the correspondence between the predicate's semantic arguments and its syntactic arguments. Making explicit which syntactic arguments correspond to which semantic roles is important, e.g., in such tasks as Information Extraction and Machine Translation. Two examples illustrating that such correspondence is not trivial and should be stated in valence dictionaries are so-called *psych verbs* and the *raising/control* distinction.

Consider two psychological predicates, PRZESTRASZYĆ 'to frighten' and PRZESTRASZYĆ SIĘ 'to get scared of':

(7) Burza   przestraszyła Janka.
     storm.NOM frightened   Janek.ACC

     'The storm frightened Janek.'

(8) Janek   przestraszył się burzy.
     Janek.NOM got scared of   storm.GEN

     'Janek got scared of the storm.'
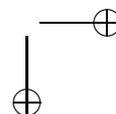
At a certain level of granularity, they express the same semantic relation, namely about x frightening y. However, the two semantic arguments of this relation, x and y, are realised in two syntactically different ways: in case of PRZESTRASZYĆ, (7), x is realised as the subject and y — as the accusative object, while in case of PRZESTRASZYĆ SIĘ, (8), x is realised a genitive object, while y — as the subject.

Conversely, verbs such as ZACZĄĆ 'start' and PRÓBOWAĆ 'try', with similar syntactic arguments (a subject and an infinitival complement), have different semantic arguments.

(9)   a.   Janek   zaczyna czytać.
                Janek.NOM starts   read.INF
                'John is starting to read.'

        b.   Janek   próbuje czytać.
                Janek.NOM tries   read.INF
                'John is trying to read.'

More precisely, verbs such as ZACZĄĆ, called raising verbs, are usually treated as semantically mono-valent, with the semantic argument corresponding to the proposition expressed by the infinitival complement, while verbs such as PRÓBOWAĆ, called control verbs or equi verbs, have two semantic arguments, directly corresponding to the syntactic arguments. This difference in mapping between semantic arguments and syntactic arguments is, e.g., responsible for the following differences in the behaviour of the two verbs:

(10)   a.   Zaczęło padać.
                started rain.INF
                'It started to rain.'

        b.   Gra   zaczyna być   warta świeczki.
                game starts   be.INF worth candle
                'It starts being worth the effort.' (idiomatic)

(11)   a.   *Próbowało padać.
                tried   rain.INF

        b.   *Gra   próbuje być   warta świeczki.
                game tries   be.INF worth candle

Without an explicit statement of the relationship between a predicate's semantic argument structure and its syntactic argument structure, the usefulness of a valence dictionary in text understanding or generating will be limited.

9
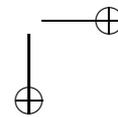
3.3 Insufficient Formal Apparatus

As shown in Świdziński 1993 and further discussed in Przepiórkowski 1997, raising verbs, such as ZACZĄĆ, illustrate the inherent weakness of the division of verbs into regular verbs, taking nominative subjects, and quasi-verbs, which do not occur with nominative subjects. As (12) from Świdziński 1993 shows, DZIWIĆ 'surprise, make wonder' may occur with a sentential subject — this is clear because the clause *że Maria woli Piotra* is coordinated with an uncontroversial nominative subject, *jej brak gustu*:

(12)   Jana   dziwi,   że Maria woli   Piotra, i   jej   brak
        Jan.GEN surprises [that Maria prefers Piotr] and [her lack
        gustu.
        good taste.GEN].NOM

        'That Maria prefers Piotr, and her lack of good taste, surprise Jan.'

Similarly, in (13) below, it can be shown that the embedded *że*-clause is the subject of *zaczęło*, a form of ZACZĄĆ 'start'.

(13)   Jana   zaczęło dziwić,   że Maria woli   Piotra.
        Jan.GEN started surprise.INF [that Maria prefers Piotr]

        'It started surprising Jan that Maria prefers Piotr.'

Thus ZACZĄĆ, as used in (13), does not fit into either group of verbs. Moreover, it would be wrong to say that ZACZĄĆ takes sentential subjects — it does, but only when its infinitival complement takes sentential subjects. This is the essence of raising: a raising verb such as ZACZĄĆ 'raises' the subject of its infinitival complement, whatever its categorial status, into its own subject position. To give another example, in (10a), the form *zaczęło* does not combine with an overt subject precisely because the verb *padać* 'to rain' does not expect an overt subject.

The dictionaries ISJP and SDPV, which assume the simple bifurcation of verbs into regular verbs and quasi-verbs, cannot easily handle such raising dependencies. Interestingly, although raising has been one of the most important and best understood topics of research in generative linguistics, the syntactic-generative dictionary of Polish verbs SSGCP does not handle raising verbs properly either. As discussed in §2.3, ZACZĄĆ, in its relevant meaning, is assigned roughly two valence frames: I.1.–2., according to which it takes the nominative subject and an infinitival complement, and I.3., according to which its a subjectless verb subcategorising only for an infinitival complement. That
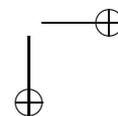
is, even the crude approximation of raising present in SDPV, namely, that ZA-CZAĆ as a subjectless quasi-verb occurs with infinitival complements headed by similarly subjectless verbs, is lost in SSGCP.

This discussion shows that valence dictionaries should take better advantage of the results of linguistic theorising of the past decades and, in particular, should embrace at least some of the formal apparatus assumed within modern linguistic theories.

3.4 Overly Specific Information

Finally, although valence dictionaries are dictionaries of lexemes and the information they provide should be true for all forms of a given lexeme, they usually contain case information which is true only for some forms of the lexeme. For example, the direct object of verbs is usually specified as accusative, even though it is realised as genitive in case of gerundial forms in *-nie/-cie*, often assumed to belong to the verbal lexeme, as nominative in case of passive participles, and as genitive in the scope of verbal negation (roughly speaking):[5]

(14) a. Janek zobaczył Paryż.
          Janek saw   Paris.ACC
          'Janek saw Paris.'

      b. zobaczenie Paryża
          seeing   Paris.GEN
          'seeing Paris'

      c. Janek nie chciał zabaczyć Paryża.
          Janek not wanted see.INF   Paris.GEN
          'Janek didn't want to see Paris.'

Thus, such valence dictionaries implicitly rely on the users' knowledge about morphosyntactically induced case variations of the lexeme's arguments, instead of specifying such information in an explicit manner. Although this problem can be alleviated at further levels of linguistic processing, a valence dictionary which specifies information true for all forms of a given lexeme should be preferred to a dictionary which gives information true to only some of the forms.

---

[5]For a detailed analysis of such case variations in Polish see, e.g., Przepiórkowski 1999a and references therein.
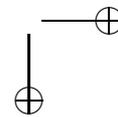
4 Conclusions

Valence dictionaries are very valuable for many different applications, including various NLP applications, and the construction of a valence dictionary is a substantial effort, so it is crucial that valence dictionaries be designed in a maximally transparent way that maximises their versatility and usefulness. In this article, I have discussed various common design flaws of existing valence dictionaries for Polish, including: the lack of any clear definition of what kinds of dependents are included in valence frames, the lack of any non-defeasible lexical semantic information, the lack of any information linking syntactic arguments to semantic roles, formal language incapable of the proper treatment of certain classes of predicates, such as raising predicates, and overly specific information, true for only some forms of a given lexeme.

I hope that this discussion will help build a well-designed and reusable valence dictionary for Polish. An initial attempt in that direction is described in Przepiórkowski 2004.

References

Bańko, M., editor (2000). *Inny słownik języka polskiego*. Wydawnictwo Naukowe PWN, Warsaw.

Durme, B. V., Huang, Y., Kupść, A., and Nyberg, E. (2003). Towards light semantic processing for Question Answering. In *Proceedings of the HLT-NAACL 2003 Workshop on Text Meaning*, pages 54–61, Edmonton.

Ginzburg, J. and Sag, I. A. (2000). *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*. CSLI Publications, Stanford, CA.

Grund, D. (2000). Komputerowa implementacja słownika syntaktyczno-generatywnego czasowników polskich. *Studia Informatica*, **3**(41), 243–256.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. Chicago University Press, Chicago, IL.

Morciniec, N., Cirko, L., and Ziobro, R. (1995). *Słownik walencyjny czasowników niemieckich i polskich / Wörterbuch zur Valenz Deutscher und Polnischer Verben*. Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław.

Piskorski, J., Homola, P., Marciniak, M., Mykowiecka, A., Przepiórkowski, A., and Woliński, M. (2003). Information extraction for Polish using the SProUT platform. In *Proceedings of IIS:IIPWM 2004*.

Polański, K., editor (1980). *Słownik syntaktyczno-generatywny czasowników polskich*, volume I (A–M) of Polański 1992. Zakład Narodowy im. Ossolińskich / Wydawnictwo Polskiej Akademii Nauk, Wrocław.
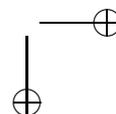
Polański, K., editor (1980–1992). *Słownik syntaktyczno-generatywny czasowników polskich*. Zakład Narodowy im. Ossolińskich / Instytut Języka Polskiego PAN, Wrocław / Kraków.

Przepiórkowski, A. (1997). Transmisja wymagań składniowych. *Polonica*, **XVIII**, 29–50.

Przepiórkowski, A. (1999a). *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. Ph. D. dissertation, Universität Tübingen, Germany.

Przepiórkowski, A. (1999b). On complements and adjuncts in Polish. In R. D. Borsley and A. Przepiórkowski, editors, *Slavic in Head-Driven Phrase Structure Grammar*, pages 183–210. CSLI Publications, Stanford, CA.

Przepiórkowski, A. (2002). Verbal proforms and the structural complement-adjunct distinction in Polish. In P. Kosta and J. Frasek, editors, *Current Approaches to Formal Slavic Linguistics*, pages 405–414, Frankfurt am Main. Peter Lang.

Przepiórkowski, A. (2004). Towards the design of a syntactico-semantic lexicon for Polish. In *Proceedings of IIS:IIPWM 2004*.

Przepiórkowski, A., Kupść, A., Marciniak, M., and Mykowiecka, A. (2002). *Formalny opis języka polskiego: Teoria i implementacja*. Akademicka Oficyna Wydawnicza EXIT, Warsaw.

Świdziński, M. (1993). Dalsze kłopoty z bezokolicznikiem. In J. Sambor, J. Linde-Usiekniewicz, and R. Huszcza, editors, *Językoznawstwo synchroniczne i diachroniczne*, pages 303–314. Wydawnictwa Uniwersytetu Warszawskiego, Warsaw.

Świdziński, M. (1994). Syntactic dictionary of Polish verbs. Ms., University of Warsaw and Universiteit van Amsterdam.

Tesnière, L. (1959). *Éléments de Syntaxe Structurale*. Klincksieck, Paris.

Vater, H. (1978). On the possibility of distinguishing between complements and adjuncts. In W. Abraham, editor, *Valence, Semantic Case and Grammatical Relations*, volume 1 of *Studies in Language Companion Series (SLCS)*, pages 21–45. Benjamins, Amsterdam.

Voutilainen, A. (1995). Morphological disambiguation. In F. Karlsson, A. Voutilainen, J. Heikkilä, and A. Anttila, editors, *Constraint Grammar: A Language-Independent Systsem for Parsing Unrestricted Text*, pages 165–284. Mouton de Gruyter, Berlin.

Wahlster, W., editor (2000). *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin.
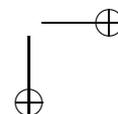
Adres autora: Adam Przepiórkowski
                 Instytut Podstaw Informatyki PAN
                 ul. Ordona 21
                 01-237 Warszawa
                 Polska
E-mail: `adamp@ipipan.waw.pl`