Semantic Roles in Grammar Engineering

Wojciech Jaworski **Adam Przepiórkowski**

Institute of Computer Science, Polish Academy of Sciences
ul. Jana Kazimierza 5, 02-668 Warszawa

`wjaworski@mimuw.edu.pl`   `adamp@ipipan.waw.pl`

Abstract

The aim of this paper is to discuss difficulties involved in adopting an existing system of semantic roles in a grammar engineering task. Two typical repertoires of semantic roles are considered, namely, VerbNet and Sowa's system. We report on experiments showing the low inter-annotator agreement when using such systems and suggest that, at least in case of languages with rich morphosyntax, an approximation of semantic roles derived from syntactic (grammatical functions) and morphosyntactic (grammatical cases) features of arguments may actually be beneficial for applications such as textual entailment.

1 Introduction

The modern notion of semantic – or thematic – roles stems from the lexical semantic work of Gruber 1965 (his *thematic relations*) and Fillmore 1968 (so-called *deep cases*), and was popularised by Jackendoff 1972, but traces of this concept may already be found in the notion of *kāraka* in the writings of the Sanskrit grammarian Pāṇini (4th century BC); see, e.g., Dowty 1991 for a historical introduction. Fillmore's deep cases are Agentive, Dative, Instrumental, Factive, Locative, Objective, as well as Benefactive, Time and Comitative, but many other sets of semantic roles may be found in the literature; for example, Dalrymple 2001, p. 206, cites – after Bresnan and Kanerva 1989 – the following ranked list of thematic roles: Agent, Benefactive, Recipient/Experiencer, Instrument, Theme/Patient, Locative.

In Natural Language Processing (NLP), one of the most popular repertoires of semantic roles is that of VerbNet (Kipper et al. 2000; `http://verbs.colorado.edu/~mpalmer/projects/verbnet.html`),

a valence lexicon of English based on Levin's (1993) classification of verbs according to the diathesis phenomena they exhibit. The VerbNet webpage states that it contains 3769 lemmata divided into 5257 senses. There are 30 semantic roles used in VerbNet 3.2,[1] including such standard roles as Agent, Beneficiary and Instrument, but also more specialised roles such as Asset (for quantities), Material (for stuff things are made of) or Pivot (a theme more central to an event than the theme expressed by another argument). This resource is widely used in NLP, and it was one of the main lexical resources behind the Unified Lexicon of English (Crouch and King, 2005), a part of an LFG-based semantic parser (Crouch and King, 2006) employed in tasks such as question answering (Bobrow et al., 2007a) and textual entailment (Bobrow et al., 2007b).

Another system of semantic roles considered here is that developed by Sowa (2000; `http://www.jfsowa.com/krbook/`) for the purpose of knowledge representation in artificial intelligence. There are 18 thematic roles proposed in Sowa 2000, p. 508, including standard roles such as Agent, Recipient and Instrument, but also 4 temporal and 4 spatial roles. Unlike in case of VerbNet, there is no corpus or dictionary showing numerous examples of the acutal use of such roles – just a few examples are given (on pp. 506–510). On the other hand, principles of assigning thematic roles to arguments may be formulated as a decision tree, which should make the process of semantic role labelling more efficient.

But why should we care about semantic roles at all? From the NLP perspective, the main reason is that they are useful in tasks approximating reasoning, such as textual entailment. Take the follow-

---

[1] Table 2 on the VerbNet webpage lists 21 roles, of which Actor is not actually used; the 10 roles which are used but not listed there are Goal, Initial_Location (apart from Location), Pivot, Reflexive, Result, Trajectory and Value, as well as Co-Agent, Co-Patient and Co-Theme.

ing two Polish sentences, with their naïve meaning representations in (1a)–(2a):

(1) Anonim   napisał artykuł na *SEM.
     anonymous wrote   paper for *SEM
     'An anonymous person wrote a paper for *SEM.'

   a. $\exists a \exists p\, article(a) \wedge person(p) \wedge$
       $anonymous(p) \wedge write(p, a, starsem)$

   b. $\exists e \exists a \exists p\, article(a) \wedge person(p) \wedge$
       $anonymous(p) \wedge write(e) \wedge$
       $agent(e, p) \wedge patient(e, a) \wedge$
       $destination(e, starsem)$

(2) Anonim   napisał artykuł.
     anonymous wrote   paper
     'An anonymous person wrote a paper.'

   a. $\exists a \exists p\, article(a) \wedge person(p) \wedge$
       $anonymous(p) \wedge write(p, a)$

   b. $\exists e \exists a \exists p\, article(a) \wedge person(p) \wedge$
       $anonymous(p) \wedge write(e) \wedge$
       $agent(e, p) \wedge patient(e, a)$

While it is clear that (2) follows from (1), this inference is not obvious in (1a)–(2a); making such an inference would require an additional meaning postulate relating the two *write* predicates of different arities. In contrast, when dependents of the predicate are represented via separate semantic roles, as in the neo-Davidsonian (1b)–(2b) (cf. Parsons 1990), the inference from (1b) to (2b) is straightforward and follows from general inference rules of first-order logic; nothing special needs to be said about the writing events.

Also, building on examples from Bobrow et al. 2007b, p. 20, once we know that *flies* is a possible hyponym of *travels*, we may infer *Ed travels to Boston* from *Ed flies to Boston*. Given representations employing semantic roles, e.g., $\exists e\, fly(e) \wedge agent(e, ed) \wedge destination(e, boston)$ and $\exists e\, travel(e) \wedge agent(e, ed) \wedge destination(e, boston)$, all that is needed to make this inference is a general inference schema saying that, if P is a hypernym of Q, then $\forall e\, Q(e) \rightarrow P(e)$. A more complicated set of inference schemata would be necessary if the neo-Davidsonian approach involving semantic roles were not adopted.

2 Problems with standard repertoires of semantic roles

As noted by Bobrow et al. 2007b, p. 20, standard VerbNet semantic roles may in some cases make inference more difficult. For example, in *Ed travels to Boston*, VerbNet identifies *Ed* as a Theme, while in *Ed flies to Boston* – as an Agent. The solution adopted there was to use "a backoff strategy where fewer role names are used (by projecting down role names to the smaller set)".

In order to verify the usefulness of well-known repertoires of semantic roles, we performed a usability study of the two sets of semantic roles described above. The aim of this study was to estimate how difficult it would be to create a corpus of sentences with verbs' arguments annotated with such semantic roles. For this purpose, 37 verbs were selected more or less at random and 843 instances of arguments of these verbs (in 393 sentences, but only one verb was considered in each sentence) were identified in a corpus. In two experiments, the same 7 human annotators were asked to label these arguments with VerbNet and with Sowa's semantic roles.

In both cases interannotator agreement (IAA) was below our expectations, given the fact that VerbNet comes with short descriptions of semantic roles and a corpus of illustrative examples, and that Sowa's classification could be (and was for this experiment) formalised as a decision tree. For VerbNet roles, Fleiss's κ (called *Fleiss's Multi-π* in Artstein and Poesio 2008, as it is actually a generalisation of Scott's π rather than Cohen's κ) is equal to 0.617, and for Sowa's system it is a little higher, 0.648. According to the common wisdom (reflected in Wikipedia's entry for "Fleiss' kappa"), values between 0.41 and 0.60 reflect moderate agreement and between 0.61 and 0.80 – substantial agreement. Hence, the current results could be interpreted as moderately substantial agreement. However, Artstein and Poesio 2008, p. 591, question this received wisdom and state that "only values above 0.8 ensured an annotation of reasonable quality".

This opinion is confirmed by the more detailed analysis of the distribution of (dis)agreement provided in Tab. 1. The top table gives the number of arguments for which the most commonly assigned Sowa's role was assigned by n annotators (n ranges from 2 to 7; not from 1, as there were no arguments that would be assigned 7 different roles by the 7 annotators) and the most commonly assigned VerbNet role was assigned by m annotators (m also ranges from 2 to 7). For example, the cell in the row labelled 7 and in the column labelled

		VerbNet						
		2	**3**	**4**	**5**	**6**	**7**	
	2	6	8	3	0	0	0	17
S	**3**	8	39	39	17	25	3	131
o	**4**	2	26	49	37	20	5	139
w	**5**	4	11	48	45	11	15	134
a	**6**	1	9	18	16	35	20	99
	7	0	3	11	47	52	210	323
		21	96	168	162	143	253	843

		VerbNet						
		2	**3**	**4**	**5**	**6**	**7**	
	2	0.71%	0.95%	0.36%	0.00%	0.00%	0.00%	2.02%
S	**3**	0.95%	4.63%	4.63%	2.02%	2.97%	0.36%	15.54%
o	**4**	0.24%	3.08%	5.81%	4.39%	2.37%	0.59%	16.49%
w	**5**	0.47%	1.30%	5.69%	5.34%	1.30%	1.78%	15.90%
a	**6**	0.12%	1.07%	2.14%	1.90%	4.15%	2.37%	11.74%
	7	0.00%	0.36%	1.30%	5.58%	6.17%	24.91%	38.32%
		2.49%	11.39%	19.93%	19.22%	16.96%	30.01%	100.00%

Table 1: Interannotator agreement rate for VerbNet and Sowa role systems; the top table gives numbers of arguments, the bottom table gives normalised percentages

6 contains the information that 52 arguments were such that all annotators agreed on Sowa's role and 6 agreed on a VerbNet role. The final row and the final column contain the usual marginals; e.g., out of 843 arguments, in case of Sowa's system 253 arguments were annotated unanimously, and in case of VerbNet roles – 323 arguments. The lower table gives the same information normalised to percentages. Note that for a significant percent of examples (almost 18% in case of Sowa's system and almost 14% in case of VerbNet) there is no majority decision and that the concentration of examples around the diagonal means that the lack of consensus is largely independent of the choice of the role system.

Some of the most difficult cases were discussed with annotators and the conclusion reached was that there are two main reasons for the low IAA: numerous cases where more than one role seems to be suitable for a given argument and cases where there is no suitable role at all. (In fact, as in case of LECZYĆ 'treat, cure' discussed below, it is sometimes difficult to distinguish these two reasons: more than one role seems suitable because none is clearly right.)

The first situation is caused by the fact that a distinction between the roles is often highly subjective; for example, when *a doctor is treating a girl*, is (s)he causing a structural change? The answer to this question determines the distinction between Patient and Theme in Sowa's system. It could be "no" when the doctor only prescribes some medicines, but it could be "yes" when (s)he operates her. Furthermore, some emphasis is put on volitionality in Sowa's system: the initiatior of an action can be either Agent or Effector, depending on whether (s)he causes the action voluntarily or not – something that is often difficult to decide even when a context of a sentence is given.

On the other hand, the Agent role is extended in VerbNet to 'internally controlled subjects such as forces and machines', but it is easy to confuse this role with Theme. For example, in *The horse jumped over the fence*, the *horse* is – somewhat counterintuitively – marked as Theme, as it must bear the same role as in *Tom jumped the horse over the fence*, where the Agent role is already taken by *Tom*. Other commonly confused pairs are Stimulus and Theme, Topic and Theme, and Patient and Theme. Moreover, there are cases where more than one role genuinely (not as a result of confusion) matches a given argument. For example, in the Polish sentence *Ona ładuje się w foremkę, którą ktoś jej podsunął* 'She squeezes/loads herself into a/the mould that somebody offered her', the argument *w foremkę* 'into mould' can be rea-

sonably marked as both: a spatial Destination and a functional Result.

The other common reason for interannotator disagreement is the lack of a suitable role. For example, returning to the sentence *A doctor is treating a girl*, it seems that neither of the two systems has an obvious role for the person being cured (hence the impression of potential suitability of a number of roles). In Polish sentences involving the verb LECZYĆ 'treat, cure', the object of treatment was variously marked as Agent, Beneficiary, Patient or Source when using VerbNet roles, and as Agent, Beneficiary, Experiencer, Patient, Recipient or Result when using Sowa's system. Thus, in *Zwierzę jest leczone z tych chorób* 'An animal is treated for these diseases', in the VerbNet experiment the animal was marked as Beneficiary (by 3 annotators), as Patient (×3) and as Source (×1), and in the Sowa experiment – as Beneficiary (×2), as Patient (×2), as Recipient (×2) and as Result (×1). Similarly, for *Mąż leczył się na serce*, lit. 'Husband treated himself for his heart', the husband was annotated as Agent (×2), Beneficiary (×2), Patient (×2) and Source (×1) when using VerbNet roles and as Agent (×1), Beneficiary (×2), Experiencer (×1), Patient (×2) and Recipient (×1) when using Sowa's roles.

Another major problem with the attempt to use these sets of semantic roles was a high percentage of verb occurrences with multiple arguments assigned the same semantic role. In case of Sowa's system 4.36% of sentences had this problem on the average (the raw numbers for the 7 annotators are: 2, 5, 8, 9, 17, 31, 34 out of 347 sentences with no coordination of unlikes in argument positions;[2] note the surprisingly large deviation) and in case of VerbNet – 2.47% sentences were so affected (7, 7, 7, 8, 9, 10, 12).

On the basis of these experiments, as well as various remarks in the literature (see, e.g., the reference to Bobrow et al. 2007b at the beginning of this section), we conclude that semantic role systems such as VerbNet or Sowa's are perhaps not really well-suited for the grammar engineering task – and certainly not worth the time, effort

[2]In case of arguments realised as a coordination of unlikes, e.g., a nominal phrase and a sentential clause, annotators routinely assigned distinct semantic roles to different conjuncts, so that one argument received a number of different roles (from the same annotator) and, consequently, there were many more duplicates in the remaining $393 - 347 = 46$ sentences than in the 347 sentences free from coordination of unlikes considered here.

and money needed to construct reasonably-sized corpora annotated with them – and that other approaches must be explored.

3 Syntactic approximation of semantic roles

In Jaworski and Przepiórkowski 2014 we propose to define 'semantic roles' on the basis of morphosyntactic information, including morphological cases, following the Slavic linguistic tradition stemming from the work of Roman Jakobson (see, e.g., Jakobson 1971a,b). In particular, since the broader context of the work reported here is the development of a syntactico-semantic LFG (Lexical-Functional Grammar; Bresnan 2001; Dalrymple 2001) parser for Polish, we build on the usual LFG approach of obtaining semantic representations on the basis of f-structures, i.e., non-tree-configurational syntactic representations (as opposed to more surfacy tree-configurational c-structures) containing information about predicates, grammatical functions and morphosyntactic features; this so-called description-by-analysis (DBA) approach has been adopted for German (Frank and Erk, 2004; Frank and Semecký, 2004; Frank, 2004), English (Crouch and King, 2006) and Japanese (Umemoto, 2006).

In the usual DBA approach, semantic roles are added to the resulting representations on the basis of semantic dictionaries external to LFG grammars (Frank and Semecký, 2004; Frank, 2004; Crouch and King, 2005, 2006). When such FrameNet- or VerbNet-like dictionaries are not available, grammatical function names (subject, object, etc.) are used instead of semantic roles (Umemoto, 2006). Unfortunately, this latter approach is detrimental for tasks such as textual entailment, as LFG grammatical functions represent the surface relations, so, e.g., a passivised (deep) object bears the grammatical function of (surface) subject. Other diathesis phenomena also result in different grammatical functions assigned to arguments standing in the same semantic relation to the verb, e.g., the recipient of the verb GIVE will normally be assigned a different grammatical function depending on whether it is realised as an NP (as in *John gave Mary a book*) or as a PP (*John gave a book to Mary*).

Although currently no reasonably-sized dictionaries of Polish containing semantic role information are available, we do not resort to grammatical

functions as names of semantic roles, but rather guess approximations of semantic roles on the basis of grammatical functions and morphosyntactic features. For example, subjects of active verbs are marked as R0 (the 'semantic role' approximating the Agent), but subjects of passsive verbs, as well as objects of active verbs, are marked as R1 (roughly, the Undergoer, i.e., Patient, Theme or Product).[3] Apart from grammatical functions and the voice value of the verb, also morphosyntactic features of arguments are taken into account, especially, for PP arguments, the preposition lemma and the grammatical case it governs. So, for example, both the OBJ-TH (dative NP) arguments and certain OBL (PP) arguments, e.g., those headed by the preposition DLA 'for', are translated into the R2 'semantic role', which approximates the Beneficiary and Recipient semantic roles. This results in the same semantic representations of *Papkin upolował dla Klary krokodyla* 'Papkin.NOM hunted a crocodile.ACC for Klara', lit. 'Papkin hunted for Klara crocodile', and *Papkin upolował Klarze krokodyla*, lit. 'Papkin.NOM hunted Klara.DAT crocodile.ACC'.

The advantage of this morphosyntax-based approach is that it is fully deterministic (only one 'semantic role' may be assigned to a given argument) and that it ensures high uniqueness of any 'semantic role' in the set of arguments of any verb (only 6 of the 347 sentences considered above, i.e., 1.73%, have the same 'semantic role' asigned to a couple of arguments, compared with 2.47% and 4.36% in the experiments described in this paper; see Jaworski and Przepiórkowski 2014 for additional data). The disadvantage is that sometimes wrong decisions are made; for example, OBL arguments of type Z[inst] 'with' may have one of at least three meanings: Perlative (R7), Thematic (R1) and Co-agentive (R0); in fact, the sentence *Zrób z nim porządek*, lit. 'do with him order', is ambiguous between the last two and may mean either 'Deal with him' (R1) or 'Clean up with him' (R0). However, the procedure will always assign only one of these 'roles' to such Z[inst] arguments (currently R7).

4 Conclusions

When developing a semantic parser, it makes sense to aim at neo-Davidsonian representations with semantic roles relating arguments to events, as such representations facilitate textual entailment and similar tasks. In this paper we reported on experiments which show that the practical usability of two popular repertoires of semantic roles in grammar engineering is limited: as the IAA is low, systems trained on corpora annotated with such semantic roles are bound to be inconsistent, limiting the usefulness of resulting semantic representations in such tasks. In case of a language that does not have a resouce such as VerbNet, the question arises then whether it makes sense to invest considerable time and effort into creating it.

In this and the accompanying paper Jaworski and Przepiórkowski 2014 we suggest an answer in the negative and propose to approximate semantic roles on the basis of syntactic and morphosyntactic information. Admittedly, this proposal is currently rather programmatic, as it is supported only with anectodal evidence. It seems plausible that the usefulness of resulting representations for textual entailment should be comparable to – or maybe even better than – that of semantic representations produced by semantic role labellers trained on rather inconsistently annotated data, but this should be quantified by further experiments.[4] If this hypothesis turns out to be true, however, the method we propose has the clear advantage of being overwhelmingly cheaper: instead of many person-years of building a resource such as Verb-Net (and then training a role labeller, etc.), a couple of days of a skilled researcher are required to define and test reasonable translations from (morpho)syntax to 'semantic roles'.

References

Ron Artstein and Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics* 34(4), 555–596.

Daniel G. Bobrow, Bob Cheslow, Cleo Condoravdi, Lauri Karttunen, Tracy Holloway King,

---

[3]We use symbols such as R0 or R1 instead of more meaningful names in order to constantly remind ourselves that we are dealing with approximations of true semantic roles; this also explains scare quotes in the term 'semantic role' when used in this approximative sense.

[4]To the best of our knowledge, no testing data for such tasks are available for Polish and Polish has never been included in evaluation initiatives of this kind. The approach described here is currently employed in a large-scale grammar development effort carried out at the Institute of Computer Science, Polish Academy of Sciences, in connection with the CLARIN-PL Research Infrastructure, and we hope to further report on its usefulness in the future.

Rowan Nairn, Valeria de Paiva, Charlotte Price, and Annie Zaenen. 2007a. PARC's Bridge and Question Answering System. In *Proceedings of the GEAF07 Workshop*, pages 26–45.

Daniel G. Bobrow, Cleo Condoravdi, Richard Crouch, Valeria de Paiva, Lauri Karttunen, Tracy Holloway King, Rowan Nairn, Charlotte Price, and Annie Zaenen. 2007b. Precision-focused textual inference. In *Proceedings of the ACL–PASCAL Workshop on Textual Entailment and Paraphrasing at ACL 2007*, pages 16–21.

Joan Bresnan. 2001. *Lexical-Functional Syntax*. Blackwell Textbooks in Linguistics, Malden, MA: Blackwell.

Joan Bresnan and Jonni M. Kanerva. 1989. Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic Inquiry* 20(1), 1–50.

Dick Crouch and Tracy Holloway King. 2005. Unifying Lexical Resources. In *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*.

Dick Crouch and Tracy Holloway King. 2006. Semantics via f-structure rewriting. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG'06 Conference*, Universität Konstanz, Germany: CSLI Publications.

Mary Dalrymple. 2001. *Lexical Functional Grammar*. San Diego, CA: Academic Press.

David Dowty. 1991. Thematic Proto-roles and Argument Selection. *Language* 67(3), 547–619.

Charles J. Fillmore. 1968. The Case for Case. In Emmon Bach and Robert T. Harms (eds.), *Universals in Linguistic Theory*, pages 1–88, New York: Holt, Rinehart and Winston.

Anette Frank. 2004. Generalisations over Corpus-induced Frame Assignment Rules. In Charles Fillmore, Manfred Pinkal, Collin Baker and Katrin Erk (eds.), *Proceedings of the LREC 2004 Workshop on Building Lexical Resources from Semantically Annotated Corpora*, pages 31–38, ELRA, Lisbon.

Anette Frank and Katrin Erk. 2004. Towards an LFG Syntax-Semantics Interface for Frame Semantics Annotation. In Alexander Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing (CICLing 2004)*, volume 2945 of *Lecture Notes in Computer Science*, pages 1–12, Heidelberg: Springer.

Anette Frank and Jiří Semecký. 2004. Corpus-based Induction of an LFG Syntax-Semantics Interface for Frame Semantic Processing. In Silvia Hansen-Schirra, Stefan Oepen and Hans Uszkoreit (eds.), *Proceedings of the 5th International Workshop on Linguistically Interpreted Corpora at COLING 2004*, Geneva.

Jeffrey Gruber. 1965. *Studies in Lexical Relations*. Ph. D.thesis, Massachusetts Institute of Technology.

Ray Jackendoff. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: The MIT Press.

Roman O. Jakobson. 1971a. Beitrag zur allgemeinen Kasuslehre. Gesamtbedeutungen der russischen Kasus. In *Selected Writings II*, pages 23–71, The Hague: Mouton.

Roman O. Jakobson. 1971b. Morfologičeskie nabljudenija nad slavjanskim skloneniem. In *Selected Writings II*, pages 154–183, The Hague: Mouton.

Wojciech Jaworski and Adam Przepiórkowski. 2014. Syntactic Approximation of Semantic Roles. In Adam Przepiórkowski and Maciej Ogrodniczuk (eds.), *Advances in Natural Language Processing: Proceedings of the 9th International Conference on NLP, PolTAL 2014, Warsaw, Poland, September 17–19, 2014*, volume 8686 of *Lecture Notes in Artificial Intelligence*, Heidelberg: Springer.

Karin Kipper, Hoa Trang Dang, William Schuler, and Martha Palmer. 2000. Building a class-based verb lexicon using TAGs. In *Proceedings of TAG+5 Fifth International Workshop on Tree Adjoining Grammars and Related Formalisms*.

Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Terence Parsons. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*. Cambridge, MA: The MIT Press.

John F. Sowa. 2000. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks Cole Publishing Co.

Hiroshi Umemoto. 2006. Implementing a Japanese Semantic Parser Based on Glue Approach. In *Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation*, pages 418–425, Huazhong Normal University, Wuhan, China: Tsinghua University Press.