Semantic Role Labelling without Deep Syntactic Parsing

Konrad Gołuchowski[1,2] and Adam Przepiórkowski[2,1]

[1] University of Warsaw
[2] Institute of Computer Science, Polish Academy of Sciences

kodie@mimuw.edu.pl
adamp@ipipan.waw.pl

Abstract. This article proposes a method of Semantic Role Labelling for languages with no reliable deep syntactic parser and with limited corpora annotated with semantic roles. Reasonable results may be achieved with the help of shallow parsing, provided that features used for training such shallow parsers include both lexical semantic information (here: hypernymy) and syntactic information.

Keywords: argument identification, semantic role classification, shallow parsing, chunking

1 Introduction

Semantic Role Labelling (SRL) is a well-known task within semantic analysis. The idea is to annotate predicate arguments in the sentence with special labels to indicate the semantic relation between the argument and the verb. In the sentence below, three semantic roles (i.e. `Buyer`, `Goods` and `Time`) are indicated.

[$_{\texttt{Buyer}}$ Frank] **bought** [$_{\texttt{Goods}}$ a new car] [$_{\texttt{Time}}$ yesterday].

While much SRL work has been done for English, hardly any is reported for Polish, a language with only prototype-quality parsers and without large or balanced corpora annotated at the level of semantic labels. Moreover, unlike English, Polish is a language with rich inflection and relatively free word order. The experiments reported here were conducted on a small (*83,000-word*) corpus of transcribed phone conversations concerning public transportation in Warsaw, the so-called LUNA corpus [10]. Therefore, annotated situations and semantic roles are limited. This paper proposes a method of semantic labelling in just such setups: without deep syntactic parsing and with a very limited manually annotated corpus.

2 Related work

In [20], the SRL task is divided into three subtasks: argument identification, semantic role classification and joint annotation. The current paper focuses on the first two: finding phrase boundaries and assigning roles to argument phrases.

Argument identification typically consists of syntactic parsing, followed by binary classification of parse nodes according to whether they represent arguments of predicates [3], perhaps with additional heuristics [21]. Supervised machine learning is also used in the task of semantic role classification, with much work devoted to feature selection [3, 2, 19, 14].

Few papers propose methods not involving deep parsing. [6] uses SVM to assign to syntactic chunks IOB tags derived from semantic roles. [18] uses shallow parsing for Chinese SRL. [15] compares systems that use deep and shallow parsing and shows implications of the lack of complete trees for argument identification.

3 SRL without complete syntactic parsing

Most approaches mentioned above make significant use of the syntactic parse tree. Many features for classification and argument identification heuristics are based on spans of tree nodes and relations between them. Such features are designed to reflect some relation between the syntactic realization of arguments and their semantics. The lack of a syntactic parse also makes argument identification task much harder: since semantic arguments are normally realised as syntactic constituents, obtaining the tree of a sentence is paramount to constraining the set of candidates for arguments. Finally, information contained in the syntactic tree helps to decide which of possibly several predicates in the sentence governs a given argument.

To alleviate the problem of no complete syntactic parse information, shallow parsing is applied here to extract basic syntactic (nominal, prepositional, etc.) groups (roughly: chunks [1]), using a shallow grammar (a cascade of regular grammars) manually developed within a different project [5, 4]. Initial experiments indicate that a certain level of correlation exists between such groups and predicate arguments. Treating these groups as arguments yields acceptable recall but very low precision:

Arguments match criteria	**Precision**	**Recall**	**F-measure**
Exact match	0.15	0.44	0.22
Overlap	0.29	0.86	0.43

Given these unsatisfactory results, an *additional* approach to shallow parsing was implemented, based on IOB tagging [16] and taking advantage of both syntactic and semantic features. The tagger trained here implements the linear-chain Conditional Random Fields model.

The key problem was the selection of features relevant for identifying argument boundaries. The following features have been used to train the model: 1) word shape (e.g. *Ul* for "Desk", where *U* stands for a sequence of upper-case letters and *l* stands for a sequence of lower-case letters), 2) the most general hypernym of the word (based on Polish WordNet, [13, 9]) if available and base form of the word otherwise, 3) the word's part of speech, 4) the word's case (if relevant), 5) the syntactic group (if the word is contained inside one) identified by the shallow grammar, 6) all above features for the two immediately adjacent words.

4 Semantic role classifier

After argument identification, the next step is to assign a semantic role to each argument. The MaxEnt-based classifier implemented in the SciKit package [12] was trained for this purpose.

As in the argument identification task, many commonly used features in this task could not be used due to the lack of a parse tree. In particular, it is not known which of possibly many predicates actually governs a given argument. Instead, the closest potential predicates to the left and to the right of the argument are identified, and the following features are used to decide what semantic role the predicate assigns to the argument: 1) the base form of each predicate, 2) their parts of speech (PoS), 3) whether predicates are negated, 4) the type of the syntactic group (if overlapping with argument boundaries), 5) the case of the noun in the argument (if applicable), 6) the left-most preposition in the argument, 7) PoS of the first and last words in the argument, 8) the most general hypernyms of the words from argument available in the WordNet, and the base form of the word otherwise, 9) the words' prefixes and suffixes of length three.

Originally, the LUNA corpus was annotated with 64 FrameNet-like roles [8]. However, more than half of these roles occurred less than 15 times in the corpus. Therefore, FrameNet-like roles were semi-automatically projected to 19 more general VerbNet-based thematic roles [17] utilizing Semlink resources [11].

5 Evaluation

When determining whether a given argument has been identified correctly, one may require the complete identity of spans or loosen this requirement to mere overlap. Both approaches have been used in the past. A compromise but still relatively strict approach is proposed here: the potential argument is judged as correctly identified if it differs from the gold standard argument at most with respect to initial or final potential modifiers (i.e., particles, adjectives, etc.), as in the English NPs *books about Bali* vs. *even these books about Bali* (containing the focus particle *even* and the demonstrative *these*).

The usual 10-fold cross-validation was performed. For the classification of semantic roles, two sets of results are given: for role classification as a separate task assuming prior gold-standard argument identification (Table 1) and as a joint task with argument identification (Tables 2 and 3). Note that the results in Table 1 are optimistic, as they assume not only prior identification of arguments, but also – in the last row – the correct identification of the predicate governing the argument. This table also contains a comparison of semantic role classification using FrameNet-like roles and thematic VerbNet-like roles.

In order to measure the importance of particular syntactic and semantic features introduced in Section 4, experiments were repeated with different feature

Table 1. Results of semantic role labelling on gold standard arguments

Task	**Accuracy**
Semantic role labelling (FrameNet roles)	0.65
Semantic role labelling (VerbNet roles)	0.74
Semantic role labelling (VerbNet roles) + correct predicate	0.76

Table 2. Results of argument identification task and semantic role classification task with proposed compromise solution for arguments' agreement.

Task	**Precision**	**Recall**	**F-measure**
Argument identification	0.71	0.68	0.70
Arg. identification + semantic role classification	0.61	0.57	0.59

Table 3. Results of argument identification task and semantic role classification task when identity of argument spans is required.

Task	**Precision**	**Recall**	**F-measure**
Argument identification	0.69	0.64	0.67
Arg. identification + semantic role classification	0.58	0.54	0.56

Table 4. Results of semantic role classification on gold standard arguments with different set of features

Features	**Accuracy**
Predicate features(1,2,3) + Case(5)	0.55
Predicate features(1,2,3) + Preposition(6)	0.57
Predicate features(1,2,3) + Syntactic group(4) + Case(5)	0.60
Predicate features(1,2,3) + Case(5) + Preposition(6)	0.62
Predicate features(1,2,3) + All syntactic features(4,5,6,7)	0.66
Predicate features(1,2,3) + Hypernyms(8)	0.63
Predicate features(1,2,3) + Case(5) + Preposition(6) + Hypernyms(8)	0.71
Syntactic group(4) + Case(5) + Preposition(6) + Hypernyms(8)	0.67
Predicate's PoS(2) + Syntactic group(4) + Case(5) + Preposition(6) + Hypernyms(8)	0.67
Predicate's lemma(1) + Syntactic group(4) + Case(5) + Preposition(6) + Hypernyms(8)	0.73
All features (1-9)	0.74

sets. The most significant results are presented in Table 4, which expands the 2nd row of Table 1.

Due to the fact that arguments are usually adjective, noun or prepositional phrases, syntactic features are crucial in the task of argument identification as they are of great help in extracting such phrases. Moreover, predicates very often impose certain restrictions on syntactic features of semantic roles. In fact, case (5) and preposition (6) presents high correlation with semantic role occurrences and are very useful in semantic role classification. Only the base forms of predicates seem important in case of predicate features. Also, one can observe that the use of hypernyms (from the Polish WordNet) greatly improves the results.

6 Summary and future work

This article presents initial experiments with semantic role labelling for Polish. It proposes a method that does not require a syntactic parse tree to identify arguments and instead relies on the output of the shallow parser both for argument identification and for semantic role classification. Evaluation includes the impact of different features on the accuracy of semantic role classification.

Because the IOB tagger produces arguments that do not overlap, there is no need for joint annotation, used for example in [20]. However, future work should examine the possibility of increasing argument identification recall by introducing overlapping arguments. The step of joint annotation also gives the opportunity to take advantage of probabilities of various semantic roles for each argument.

During the experiments it turned out that the corpus employed here is rather noisy, apart from being small and unbalanced. In order to build a robust SRL system, a bigger and cleaner corpus is needed. One possibility to build such a corpus is to exploit the existence of parallel corpora and SRL tools for English (see [7]).

References

1. Abney, S.: Parsing by chunks. In: Berwick, R., Abney, S., Tenny, C. (eds.) Principle-Based Parsing, pp. 257–278. Kluwer (1991)
2. Fleischman, M., Kwon, N., Hovy, E.: Maximum entropy models for framenet classification. In: In Proceedings of the Conference on Empirical Methods in Natural Language Processing (2003)
3. Gildea, D., Jurafsky, D.: Automatic labeling of semantic roles. Computational Linguistics 28(3), 245–288 (2002)
4. Głowińska, K.: Anotacja składniowa NKJP. In: Przepiórkowski, A., Bańko, M., Górski, R.L., Lewandowska-Tomaszczyk, B. (eds.) Narodowy Korpus Języka Polskiego. Wydawnictwo Naukowe PWN, Warsaw (2012)
5. Głowińska, K., Przepiórkowski, A.: The design of syntactic annotation levels in the National Corpus of Polish. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC 2010. ELRA, Valletta, Malta (2010)
6. Hacioglu, K., Pradhan, S., Ward, W., Martin, J.H., Jurafsky, D.: Semantic Role Labeling by Tagging Syntactic Chunks. In: Proceedings of CoNLL-2004. pp. 110–113 (2004)
7. Johansson, R., Nugues, P.: A FrameNet-based semantic role labeler for Swedish. In: Proceedings of the COLING/ACL on Main conference poster sessions. pp. 436–443. COLING-ACL '06, Association for Computational Linguistics, Stroudsburg, PA, USA (2006)
8. Johnson, C.R., Fillmore, C.J., Petruck, M.R., Baker, C.F., Ellsworth, M.J., Ruppenhofer, J., Wood, E.J.: FrameNet: Theory and Practice (2002)

9. Maziarz, M., Piasecki, M., Szpakowicz, S.: Approaching plWordNet 2.0. In: Proceedings of the 6th Global Wordnet Conference. Matsue, Japan (2012)
10. Mykowiecka, A., Marasek, K., Marciniak, M., Rabiega-Wiśniewska, J., Gubrynowicz, R.: Annotated corpus of Polish spoken dialogues. In: Vetulani, Z., Uszkoreit, H. (eds.) Human Language Technology. Challenges of the Information Society, Lecture Notes in Computer Science, vol. 5603, pp. 50–62. Springer Berlin / Heidelberg (2009)
11. Palmer, M.: SemLink—linking PropBank, VerbNet, FrameNet and WordNet. In: Proceedings of the Generative Lexicon Conference. Pisa, Italy (2009)
12. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830 (2011)
13. Piasecki, M., Szpakowicz, S., Broda, B.: A Wordnet from the Ground Up. Oficyna Wydawnicza Politechniki Wroclawskiej, Wrocław (2009)
14. Pradhan, S., Hacioglu, K., Krugler, V., Ward, W., Martin, J.H., Jurafsky, D.: Support vector learning for semantic argument classification. Machine Learning 60(1–3), 11–39 (2005)
15. Punyakanok, V., Roth, D., Yih, W.t.: The importance of syntactic parsing and inference in semantic role labeling. Computational Linguistics 34(2), 257–287 (2008)
16. Ramshaw, L.A., Marcus, M.P.: Text chunking using transformation-based learning. In: Proceedings of the Third Workshop on Very Large Corpora. pp. 82–94. ACL, Cambridge, MA (1995)
17. Schuler, K.K.: VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon. Ph.D. thesis, University of Pennsylvania (2006)
18. Sun, W., Sui, Z., Wang, M., Wang, X.: Chinese semantic role labeling with shallow parsing. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3. pp. 1475–1483. EMNLP '09, Association for Computational Linguistics, Stroudsburg, PA, USA (2009)
19. Surdeanu, M., Harabagiu, S., Williams, J., Aarseth, P.: Using predicate-argument structures for information extraction. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics – Volume 1. pp. 8–15. ACL '03, Association for Computational Linguistics, Stroudsburg, PA, USA (2003)
20. Toutanova, K., Haghighi, A., Manning, C.D.: A global joint model for semantic role labeling. Computational Linguistics 34(2), 161–191 (2008)
21. Xue, N.: Calibrating features for semantic role labeling. In: In Proceedings of EMNLP 2004. pp. 88–94 (2004)