

Semantic Role Labeling with the Swedish FrameNet

Richard Johansson, Karin Friberg Heppin, Dimitrios Kokkinakis

Språkbanken, Department of Swedish, University of Gothenburg
Box 100, SE-40530 Gothenburg, Sweden
{richard.johansson, karin.friberg.heppin, dimitrios.kokkinakis}@svenska.gu.se

Abstract

We present the first results on semantic role labeling using the Swedish FrameNet, which is a lexical resource currently in development. Several aspects of the task are investigated, including the selection of machine learning features, the effect of choice of syntactic parser, and the ability of the system to generalize to new frames and new genres. In addition, we evaluate two methods to make the role label classifier more robust: cross-frame generalization and cluster-based features. Although the small amount of training data limits the performance achievable at the moment, we reach promising results. In particular, the classifier that extracts the boundaries of arguments works well for new frames, which suggests that it already at this stage can be useful in a semi-automatic setting.

Keywords: Frame semantics, semantic role labeling, Swedish

1. Introduction

The FrameNet lexical database and annotated corpora, based on the theory of semantic frames (Fillmore et al., 2003), have allowed the implementation of automatic systems for the extraction of *semantic roles* (Gildea and Jurafsky, 2002; Johansson and Nugues, 2007; Márquez et al., 2008; Das et al., 2010). While most research has focused on English, the ongoing development of a Swedish FrameNet (Borin et al., 2010) allows us to investigate the feasibility of using this resource in constructing an automatic role-semantic analyzer for Swedish.

2. The Swedish FrameNet

The Swedish FrameNet, SweFN, is a lexical resource under development, based on the English version of FrameNet constructed by the Berkeley research group. It is found on the SweFN website¹, and is available as a free resource. All lexical resources used for constructing SweFN are freely available for downloading.

The SweFN frames and frame names correspond to the English ones, with some exceptions, as do the selection of frame elements including definitions and internal relations. The meta-information about the frames, such as semantic relations between frames, is also transferred from the Berkeley FrameNet. Compared to the Berkeley FrameNet, SweFN is expanded with information about the domain of the frames, at present: general language, the medical and the art domain. The frames also contain notation about semantic types.

As of October 2011, SweFN covered 519 frames with around 18,000 lexical units. The lexical units are gathered from SALDO, a free Swedish electronic association lexicon (Borin, 2010). A lexical unit from SALDO cannot populate more than one frame. This can be problematic as different aspects of one lexical unit may fit into different frames. The solution is to propose a new SALDO entry, or simply to determine which of the possible frames should be populated by the unit in question. At present there are

31 frames in SweFN which do not match a frame in the Berkeley FrameNet. Of these, there are eight completely new frames while the others have been modified in some way.

Crucially for the work presented in this paper, every frame is exemplified with at least one example sentence. The total number of sentences is currently 2,974. The most well-annotated frames are EXPERIENCER_OBJ with 38 sentences, CAUSE_MOTION with 21, and CAUSE_HARM with 19. These sentences form the training material used in the following sections.

3. Automatic Extraction of Semantic Roles

We implemented a system that extracts the semantic roles for a given predicate with a given frame. We split the task into two stages: 1) *segmentation*, identifying the span of a semantic argument, and 2) *labeling*, assigning a semantic role label to a given argument span. Following most previous implementations, we used a syntactic parse tree as the basis of the semantic role extraction; we assumed that every semantic role span coincides with the projection of a subtree in the syntactic tree. The tasks of segmentation and labeling then reduce to a classification problem on syntactic tree nodes. Each sentence was parsed by the LTH dependency parser (Johansson and Nugues, 2008a), a second-order search-based parser. We trained the parser on a Swedish treebank (Nilsson et al., 2005). Figure 1 shows an example of a sentence annotated with a dependency tree and semantic role structure.

The segmentation classifier was implemented as a linear support vector machine that classifies syntactic nodes as filling a semantic role or not. In addition, we used the pruning approach by Xue and Palmer (2004) before classification. The labeling classifier was implemented as a linear multiclass classifier with a flexible output space depending on the frame of the given predicate; we trained this classifier using an online learning algorithm. In addition, we imposed a uniqueness constraint on the semantic role labels output by the classifier, so that every role may appear only once for a given predicate.

¹<http://spraakbanken.gu.se/eng/swefn>

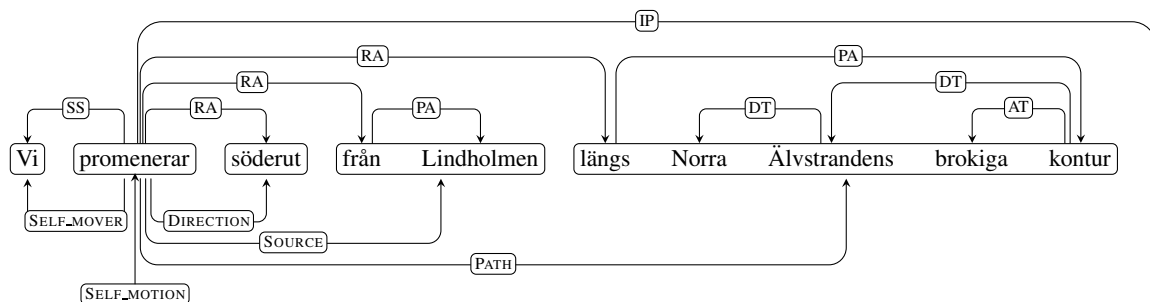


Figure 1: Example of a sentence annotated with a syntactic dependency tree (above) and frame element structure (below).

We considered a large number of features for the classifiers, shown in Table 1. Most of these are commonly used features; for their definition, see for instance Johansson (2008). We then applied a standard greedy forward feature selection procedure to determine which of them to use. The features containing SALDO ID refer to the entry identifiers in the SALDO lexicon. Note that the POS tags have coarse and fine variants, such as VERB and VERB-FINITE-PRESENT-ACTIVE respectively, and we considered both of them.

FRAME	ARG. LEFT CHILD WORD
FRAME ELEMENTS	ARG. LEFT CHILD POS
PRED. WORD	ARG. RIGHT CHILD WORD
PRED. POS	ARG. RIGHT CHILD POS
PRED. LEMMA	ARG. LEFT SIBLING WORD
PRED. SALDO ID	ARG. LEFT SIBLING POS
VOICE	ARG. RIGHT SIBLING WORD
POSITION	ARG. RIGHT SIBLING POS
PRED. PARENT WORD	SUBCAT. FRAME
PRED. PARENT POS	VERB CHAIN HAS SUBJ.
PRED. REL. TO PARENT	CONTROLLER HAS OBJ.
ARG. WORD	ARG. REL TO PARENT
ARG. POS	ARG. SET OF CHILD DEP. RELS
ARG. LEMMA	ARG. SET OF CHILD POS
ARG. SALDO ID	ARG. SET OF CHILD WORDS
DEP. RELATION PATH	POS PATH

Table 1: List of features considered.

4. Analysis

4.1. Cross-validation over Sentences

To estimate the performance of our system, we carried out a 5-fold cross-validation over the set of example sentences since the material is too small to be split into a training and a test set.

We evaluated segmentation and labeling separately. To measure the segmentation performance, we used standard precision and recall measures: A span was counted as correctly extracted if its boundaries coincided exactly with one in the gold standard. Before the evaluation, we normalized the spans by removing punctuation at their boundaries. The segmentation procedure had a precision of 70.6 and a recall of 64.8, and the labeling accuracy was 64.3.

This accuracy can be compared to a baseline accuracy of 30.4 for a classifier that always selects the most common

label for a given frame. The accuracy was higher for core frame elements than for non-core: 71.9 compared to 47.2.

4.2. Cross-frame Role Label Generalization

The machine learning model used for the labeling stage treats identical role labels in different frames as equivalent; for instance, the label TIME in the SELF_MOTION frame is regarded as equivalent to the TIME label in COMMERCE_BUY. While this may be seen as a simplification of the FrameNet model, it allows training data for role labels to be shared across frames, which may be particularly important in our setting since the training set is small.

To evaluate the effect of generalization, we trained a classifier using frame-specific labels such as SELF_MOTION:TIME and COMMERCE_BUY:TIME. In this setting, the labeling accuracy was 52.5, a drop by 11.8 points from the accuracy achieved previously. This result shows the effectiveness of a simple label-based generalization; in the future we may compare it to more theoretically well-founded methods, such as using the frame and role hierarchies defined in FrameNet (Matsubayashi et al., 2009).

4.3. Analysis of Features

Tables 2 and 3 shows the results of the feature selection procedure for the segmentation and labeling classifiers, respectively. The features are listed in the order returned by the feature selection (top to down, left to right), which may serve as an approximation of their relative importance. The FRAME and FRAME ELEMENTS features were added manually to the labeling classifier feature set.

POS PATH	ARG. L. CHILD POS (COARSE)
ARG. REL TO PARENT	ARG. L. SIBL. POS (COARSE)
DEP. RELATION PATH	CONTROLLER HAS OBJ.
FRAME ELEMENTS	PRED. PARENT POS (COARSE)
ARG. POS (FINE)	ARG. POS (COARSE)
VERB CHAIN HAS SUBJ.	

Table 2: List of segmentation features.

We note that the resulting feature sets for the two tasks are fairly different: For the segmentation task, only *structural* features have been selected, while the labeling task needs structural and *lexical* features. This result may of course change for larger training sets, where lexical features may be expected to have a more measurable impact.

FRAME	DEP. RELATION PATH
FRAME ELEMENTS	POSITION
ARG. SALDO ID	ARG. LEMMA
ARG. POS (FINE)	PRED. POS (FINE)
ARG. POS (COARSE)	ARG. R. CHILD POS (COARSE)
ARG. WORD	

Table 3: List of labeling features.

4.4. Cross-validation over Frames

One potential application of a semantic role labeler is to assist lexicographers annotating examples in a newly created frame where the frame elements are known but no training data are available. To have an idea of the performance of the system under such circumstances, we made a cross-validation over the set of frames. As opposed to the evaluation in §4.1., there would be no train/test split for the example sentences associated with a frame; instead, they would be put as a whole either into the test set or the training set. In this evaluation, our system achieved a segmentation precision and recall of 70.6 and 63.9, and a labeling accuracy of 49.4.

Compared to the experiment in §4.1., we see that the segmentation system is remarkably robust in new frames: The recall dropped by 1 percent point but the precision was unchanged. These figures also make sense in light of the feature selection results from §4.3.: features for segmentation are structural in nature while labeling features rely on lexical information. Note that the cross-frame role label generalization approach investigated in §4.2. is necessary in this case – in other case, no label prediction could be made for frame elements belonging to new frames.

4.5. Increasing Classifier Robustness by Adding Cluster Features

Since the label classifier uses word features, lexical sparsity may be a problem due to the small size of our training set. To reduce lexical sparsity, we added features based on hierarchical clusters constructed using the Brown algorithm (Brown et al., 1992). Similar features have previous been used in applications such as dependency parsing (Koo et al., 2008). The Brown algorithm clusters word into hierarchies represented as bit strings; for instance, the cluster 10110 is divided into two clusters 101100 and 101101.

We added a feature to the role label classifier representing the cluster of the argument head word. Based on tuning on a development set, we found that it was best not to use the full bit string, but only a prefix if the string was longer than 12 bits. We evaluated the new classifier using the same evaluation protocol, and we found that the cluster feature improved the classification performance from 64.3 to 65.0.

4.6. The Effect of Syntactic Parser Choice

The syntactic parser serves as a backbone of the semantic role analyzer: First of all its output directly affects which parts of the text are considered for classification, and secondly the most useful features are directly derived from the syntactic trees. This has been noted repeatedly, for instance by Gildea and Palmer (2002) and Johansson and Nugues

(2008b), but there are also SRL systems that try to bypass the syntactic step (Collobert and Weston, 2007). In any case, in a syntax-based semantic role extraction system the quality of the automatically produced syntactic trees is crucial.

Since our previous experiments were carried out using parse trees from the LTH parser, we developed a new system that used MaltParser (Nivre et al., 2007) instead. This parser was trained on the same treebank. With MaltParser, the segmentation precision and recall values were 64.2 and 54.0 respectively, and the role classification accuracy was 61.2. This should be contrasted with the results we get when using the LTH parser: a segmentation precision and recall of 70.6 and 64.8 and a labeling accuracy of 64.3. This shows that the quality of parse trees has a very large impact on the frame element extraction performance.

4.7. Evaluation in the Medical Domain

We finally evaluated the system on a new domain: medical text. 211 sentences were extracted from Swedish medical corpora, and were manually annotated with frames and their arguments by two annotators. We used a 5-fold cross-validation procedure where each training part was concatenated with the example material from the FrameNet annotation. In this evaluation, the segmentation had a precision of 67.3 and a recall of 61.1, and the labeling had an accuracy of 52.6. The fact that these results are clearly lower than the previous figures shows that there is significant domain sensitivity for the segmenter as well as the labeling classifier and a need for domain adaptation to the medical domain.

	<i>P</i>	<i>R</i>
Segmentation	70.6	64.8
CV over sentences	70.6	64.8
CV over frames	70.6	63.9
LTH parser	70.6	64.8
MaltParser	64.2	54.0
Medical texts	67.3	61.1

Table 4: Summary of frame element segmentation results.

	Accuracy
Baseline majority classifier	30.4
Classification accuracy	64.3
Core frame elements	71.9
Noncore frame elements	47.2
CV over sentences	64.3
CV over frames	49.4
LTH parser	64.3
MaltParser	61.2
Word features	64.3
Word and cluster features	65.0
Medical texts	52.6

Table 5: Summary of frame element classification results.

5. Conclusion

We have presented the first results of experiments on semantic role labeling using the new Swedish FrameNet. In addition, we have investigated the efficacy of machine learning features for the segmentation and labeling tasks, and studied how well the system performs on frames without training material and on a specialized textual domain. Due to the small training set size, lexical features cause the role labeling classifier to suffer from feature sparsity. We tried to address this problem in two ways: cross-frame label generalization and adding cluster-based features. Both methods result in clear improvements. Our figures are summarized in Tables 4 and 5.

The magnitude of the figures reflects the size of the training material, which is fairly limited so far. Annotating sentences is very time-consuming and we will thus have to live with small training sets for the foreseeable future. However, this constraint may force us to abandon the traditional, heavily lexicalized, brute-force approaches and instead lead us into a territory of interesting research opportunities. Possible directions include semisupervised approaches (Fürstenau and Lapata, 2009) and the inclusion of automatically produced training data (Johansson and Nugues, 2006).

6. Acknowledgements

We are grateful to Percy Liang for the implementation of the Brown clustering software. This work was partly funded by the Centre for Language Technology at Gothenburg University.

7. References

- Lars Borin, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj, and Dimitrios Kokkinakis. 2010. The past meets the present in the Swedish FrameNet++. In *Proceedings of EURALEX*.
- Lars Borin. 2010. Med Zipf mot framtiden - en integrerad lexikonresurs för svensk språkteknologi. *LexicoNordica*, 17.
- Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, and Jenifer C. Lai. 1992. Class-based n -gram models of natural language. *Computational Linguistics*, 18(4):467–479.
- Ronan Collobert and Jason Weston. 2007. Fast semantic extraction using a novel neural network architecture. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 560–567, Prague, Czech Republic.
- Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A. Smith. 2010. Probabilistic frame-semantic parsing. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 948–956, Los Angeles, United States, June.
- Charles Fillmore, Christopher Johnson, and Miriam Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16(3):235–250.
- Hagen Fürstenau and Mirella Lapata. 2009. Semi-supervised semantic role labeling. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 220–228, Athens, Greece, March.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- Daniel Gildea and Martha Palmer. 2002. The necessity of syntactic parsing for predicate argument recognition. In *Proceedings of the 40th Annual Conference of the Association for Computational Linguistics*, pages 239–246, Philadelphia, United States.
- Richard Johansson and Pierre Nugues. 2006. A FrameNet-based semantic role labeler for Swedish. In *Proceedings of Coling/ACL*, pages 436–443, Sydney, Australia, July 17–21.
- Richard Johansson and Pierre Nugues. 2007. Semantic structure extraction using nonprojective dependency trees. In *Proceedings of SemEval-2007*, pages 227–230, Prague, Czech Republic, June 23–24.
- Richard Johansson and Pierre Nugues. 2008a. Dependency-based syntactic–semantic analysis with PropBank and NomBank. In *Proceedings of the CoNLL Shared Task*, pages 183–187, Manchester, United Kingdom.
- Richard Johansson and Pierre Nugues. 2008b. The effect of syntactic representation on semantic role labeling. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 393–400, Manchester, United Kingdom.
- Richard Johansson. 2008. *Dependency-based Semantic Analysis of Natural-language Text*. Ph.D. thesis, Lund University.
- Terry Koo, Xavier Carreras, and Michael Collins. 2008. Simple semi-supervised dependency parsing. In *Proceedings of ACL-08: HLT*, pages 595–603, Columbus, United States.
- Lluís Màrquez, Xavier Carreras, Ken Litkowski, and Suzanne Stevenson. 2008. Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2):145–159.
- Yuichiroh Matsubayashi, Naoaki Okazaki, and Jun’ichi Tsujii. 2009. A comparative study on generalization of semantic roles in FrameNet. In *Proceedings of ACL/IJCNLP*, pages 19–27, Singapore.
- Jens Nilsson, Johan Hall, and Joakim Nivre. 2005. MAMBA meets TIGER: Reconstructing a Swedish treebank from antiquity. In *Proceedings of NODALIDA Special Session on Treebanks*.
- Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chaney, Gülşen Eryiğit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.
- Nianwen Xue and Martha Palmer. 2004. Calibrating features for semantic role labeling. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 88–94, Barcelona, Spain.