Semantic Role Labeling

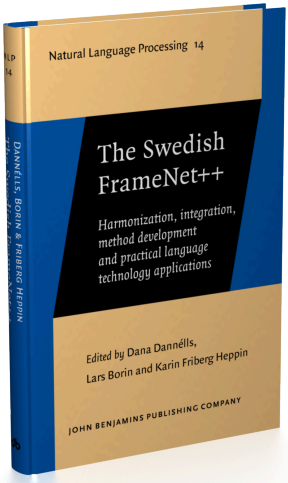
Natural Language Processing 14

CHAPTER 10

Semantic role labeling*

Richard Johansson | University of Gothenburg
Karin Friberg Heppin | HeppiLing AB
Dimitrios Kokkinakis | University of Gothenburg

https://doi.org/10.1075/nlp.14.10joh



Pages 263–280 of
The Swedish FrameNet++: Harmonization, integration, method development and practical language technology applications
Edited by Dana Dannélls, Lars Borin and Karin Friberg Heppin
[**Natural Language Processing**, 14] 2021. xiv, 333 pp.

CHAPTER 10

Semantic role labeling[1]

Richard Johansson[1], Karin Friberg Heppin[2]
and Dimitrios Kokkinakis[1]
[1]University of Gothenburg / [2]HeppiLing AB

We investigate the feasibility of automatic *semantic role labeling* (SRL) using Swedish FrameNet (SweFN). In the first part of the chapter, we describe a baseline system using a traditional division into segmentation and labeling steps. These subsystems are implemented as separate machine learning models, and we explore a wide range of syntactic and lexical features for these models. In the second part, we turn to the question of how the frame-to-frame relations defined in FrameNet allow us to use the annotated examples more effectively. The cross-frame generalization methods reduce the number of errors made by the labeling classifier by 27%. For previously unseen frames, the reduction is even more significant: 50%.

1. Introduction

The release of the FrameNet lexical database and annotated corpora (Baker et al. 1998; Fillmore & Baker 2009), as well as other role-semantic resources such as PropBank (Palmer et al. 2005), have allowed the development of automatic systems that extract and label semantic relationships between a target word and its semantic arguments within a sentence. This general task is referred to as *semantic role labeling* (SRL; Gildea & Jurafsky 2002) and has been proposed as a useful component technology for applications such as information extraction, summarization, and machine translation (Palmer et al. 2010).

For instance, given the sentence *Yesterday, Jane bought an orange at the cafe*, we are told that the target word *bought* evokes the frame `Commerce_buy` and we want to extract its semantic roles. In this case, we should identify *Yesterday* as the TIME, *Jane*

1. Parts of this chapter build on and elaborate content previously presented in Johansson et al. (2012) and Johansson (2012).

Linguistic examples in this chapter are glossed using the *Leipzig Glossing Rules* https://www.eva.mpg.de/lingua/resources/glossing-rules.php, with the following addition(s): IP: punctuation; PA: preposition complement; RA: adverbial of location; SS: subject.

https://doi.org/10.1075/nlp.14.10joh

as the BUYER, *an orange* as the GOODS, and *at the cafe* as the PLACE. The labels used for the slot fillers, correspond to the *frame elements* of the frame `Commerce_buy`.

Most research in semantic role labeling has focused on English (Gildea & Jurafsky 2002; Johansson & Nugues 2007a; Johansson 2008; Màrquez et al. 2008; Das et al. 2010), the development of a Swedish framenet (SweFN; Borin et al. 2010), discussed in Chapter 2 in this volume, allows us to investigate the feasibility of using this resource in constructing automatic role-semantic analyzers for Swedish.

In this chapter, after a brief introduction to SweFN, we describe the design and implementation of a number of SRL systems based on it. Our first system is a baseline approach using a conventional design as a series of classifiers operating on a syntactic tree. However, since SweFN comes with fewer annotated examples than its English counterpart, its performance leaves room for improvement. To mitigate this scarcity of annotated data, we then extend the baseline system by investigating several approaches to reusing data across frames, and we see that these cross-frame generalization methods result in clear improvements in SRL performance.

2. The Swedish FrameNet

SweFN is a lexical resource under development (Friberg Heppin & Toporowska Gronostaj 2012), based on the original English language FrameNet constructed by the Berkeley research group (Baker et al. 1998; Fillmore & Baker 2009). It is found on the SweFN++ website,[2] 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 indicate semantic types.

The experiments described in this chapter were carried out using the version of SweFN of October, 2011. This version of SweFN covered 519 frames with around 18,000 lexical units. The lexical units in SweFN are defined in relation to Saldo, a free Swedish electronic association lexicon (Borin et al. 2013; see also Chapter 3 in this volume). Each entry in Saldo corresponds to one sense of a lemma and can only be associated with one frame. As discussed in Chapter 8 of this volume, this

2. https://spraakbanken.gu.se/en/projects/swedish-framenet-swefn

can be problematic as different aspects of a sense may fit into different frames. The solution is to propose a new Saldo entry, or simply to determine which of the possible frames the Saldo entry should be associated with.[3] In this version of SweFN, there were 31 frames in SweFN which do not match a frame in the Berkeley FrameNet. Of these, there were 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 in SweFN as of October 2011 was 2,974. The most well-annotated frames were `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. Semantic role labeling with SweFN

As discussed in Section 1, in FrameNet-based SRL we are given a *target word* and the *frame* it evokes, and our task is to extract and label its *semantic roles*: the parts of the sentence that fill a slot in the frame evoked by the target word.[4] More generally, an automatic system may also find frame-evoking target words and determine their frames, and this task is sometimes referred to as *frame-semantic parsing* (Johansson & Nugues 2007a; Das et al. 2010). In this work, we assume that the target words and their frames are given as input to the system.

Based on SweFN, we implemented a system that extracts the semantic roles for a given target word with a given SweFN frame. Following Gildea & Jurafsky (2002), it is common to split the SRL task into two stages:

1. In the first *segmentation* step, the system identifies the boundaries of the instantiations of the semantic roles for the current target word. For instance, in the sentence *She sent me a letter*, with the target word *sent*, the instantiations are *She, me*, and *a letter*.
2. In the second *labeling* step, semantic role labels are assigned to the unlabeled semantic role spans identified in the previous step. In the previous example, the labels would be Sender, Recipient and Theme, respectively.

---

3. Alternatively, there is now a "Guest_LU" category in SweFN; see Chapter 8.

4. There is some terminological variation: in FrameNet literature, the term *frame-evoking element* is used for target words, while the semantic roles are called *frame elements*. In PropBank contexts the corresponding terms are *predicate* and *arguments*, respectively.

As in most previous implementations, we used a syntactic parse tree as the basis of the semantic role extraction; we assumed that the boundaries of every semantic role coincide with the projection of a subtree in the syntactic tree. The tasks of segmentation and labeling then reduce to two classification problems on nodes in a syntactic tree. Most early work in SRL used a constituency-based syntactic representation (Gildea & Jurafsky 2002; Màrquez et al. 2008) while newer work often uses dependency syntax (Johansson & Nugues 2007a; Johansson 2008). We chose the latter approach due to the availability of a Swedish treebank that uses a dependency representation (Nilsson et al. 2006).

Figure 1 shows an example of how syntactic subtrees correspond to semantic roles in a a sentence. In the figure, an example sentence is annotated with a dependency tree (drawn above the sentence) and a semantic role structure (below). In this example, the target word *promenerar* 'walks' evokes the frame `Self_motion`, and the four semantic roles correspond to the subtrees of the dependency tree that are syntactically connected to this target word: *vi* 'we' is a syntactic subject (SS in the dependency tree) and fills the Self_mover role; *söderut* 'south' is an adverbial of location (RA) and fills the role Direction; *från Lindholmen* 'from Lindholmen' another adverbial of location that is a Source; and *längs stranden* 'along the shore' is yet another adverbial of location, this time filling the semantic role Path.

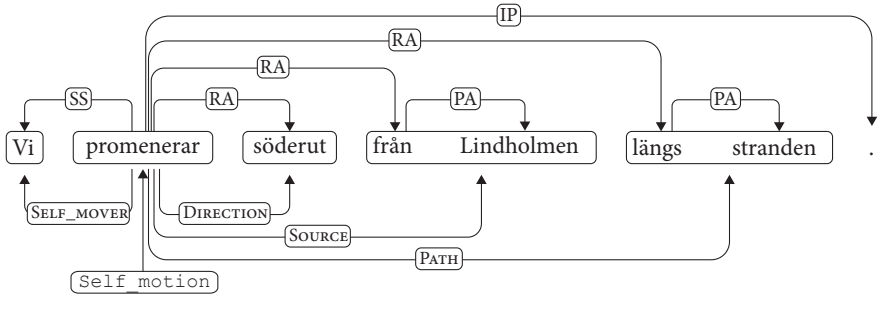


Figure 1. A sentence with dependency syntax (above) and semantic role structure (below)

3.1 Segmentation and labeling classifiers

As discussed above, the segmentation classifier extracts subtrees from a syntactic tree corresponding to semantic roles. In a dependency tree, the tree nodes correspond to words, and the segmentation classifier finds the *heads* of the relevant subtrees. For instance, in Figure 1, *Vi* 'We', *söderut* 'south', *från* 'from', and *längs* 'along' would be classified positively as semantic role heads, while all other tree nodes would be classified negatively.

We implemented the segmentation classifier as a linear support vector machine using the LIBLINEAR library (Fan et al. 2008). In addition, there is a challenge posed by the imbalance between the positive and negative classes: for a target word in a reasonably long sentence, the vast majority of the words are *not* heads of semantic roles. To deal with this issue, we used a rule-based pruning algorithm proposed by Xue & Palmer (2004) to discard a large number of obvious negatives before applying the segmentation classifier. In the dependency-based version of this pruning algorithm (Johansson & Nugues 2007a), we walk recursively up the dependency tree, collecting the direct dependents along the way, and discard all other tree nodes.

The labeling classifier was implemented as a linear multiclass classifier with an output space depending on the frame of the given target word: we only allow the classifier to output semantic role labels in the frame evoked by the target word. We trained this classifier using the online Passive–Aggressive learning algorithm (Crammer et al. 2006). In addition, we imposed a uniqueness constraint on the semantic role labels output by the classifier, so that every role is allowed to appear only once for a given target word.

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

4. Experiments

In this section, we carry out several experiments evaluating aspects of the baseline SweFN-based semantic role labeler.

4.1 Experimental data and preprocessing

In these experiments, we used the example sentences provided for the frames in SweFN. Each sentence was part-of-speech-tagged by the HunPos tagger (Halácsy et al. 2007) trained on the Stockholm–Umeå corpus (Ejerhed et al. 1992), and then parsed by the LTH dependency parser (Johansson & Nugues 2008a), a second-order search-based parser. We trained the parser on the dependency-converted version of the Talbanken treebank (Nilsson et al. 2006).

Table 1. List of features considered for the segmentation and labeling classifiers

Frame	Arg. left child word
Frame elements	Arg. left child POS
Target word	Arg. right child word
Target POS	Arg. right child POS
Target lemma	Arg. left sibling word
Target saldo ID	Arg. left sibling POS
Voice	Arg. right sibling word
Position	Arg. right sibling POS
Target parent word	Subcat. frame
Target parent POS	Verb chain has subj.
Target 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

4.2 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 0.706 and a recall of 0.648, and the labeling accuracy was 0.643.

This accuracy can be compared to a baseline accuracy of 0.304 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: 0.719 compared to 0.472.

4.3 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 0.525, an drop by 0.118 in absolute numbers 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). In the latter part of this chapter, we will investigate different methods to take the structure of FrameNet into account.

4.4 Analysis of features

Tables 2 and 3 show 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 names of features used by both classifiers are set in boldface. The Frame and Frame elements features were added manually to the labeling classifier feature set.

Table 2. Features selected for the segmentation classifier

POS path	Arg. l. child POS (Coarse)
Arg. rel to parent	Arg. l. sibl. POS (Coarse)
Dep. relation path	Controller has obj.
Frame elements	Target parent POS (Coarse)
Arg. POS (Fine)	**Arg. POS (Coarse)**
Verb chain has subj.	

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.

4.5 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

Table 3. Features selected for the labeling classifier

Frame	**Dep. relation path**
Frame elements	Position
Arg. saldo ID	Arg. lemma
Arg. POS (Fine)	Target POS (Fine)
Arg. POS (Coarse)	Arg. r. child POS (Coarse)
Arg. word	

under such circumstances, we made a cross-validation over the set of frames. As opposed to the evaluation in Section 4.2, 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, the segmentation quality is almost unchanged compared to the previous experiment: the precision and recall values were 0.706 and 0.639, respectively. However, the labeling classifier performs much worse in this experiment and the accuracy drops to 0.494.

Compared to the experiment in Section 4.2, 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 Section 4.4: features for segmentation are structural in nature while labeling features rely on lexical information.

The poor performance of the labeling classifier in this context shows that we need to approach cross-frame generalization more carefully. Note that the label generalization approach investigated in Section 4.3 is necessary in this case – in other case, no label prediction could be made for frame elements in new frames.

4.6 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 previously 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 0.643 to 0.650.

4.7 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 & Palmer (2002) and Johansson & Nugues (2008b), but there are also SRL systems that try to bypass the syntactic step (Collobert & 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 0.642 and 0.540 respectively, and the role classification accuracy was 0.612. This should be contrasted with the results we get when using the LTH parser: a segmentation precision and recall of 0.706 and 0.648 and a labeling accuracy of 0.643. This shows that the choice of the parser has a very large impact on the frame element extraction performance. The reason for the difference in SRL performance is likely due to quality differences in the underlying parse trees, as previously investigated for English for the same parsers (Johansson 2008).

4.8 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 SweFN annotation. In this evaluation, the segmentation had a precision of 0.673 and a recall of 0.611, and the labeling had an accuracy of 0.526. 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.

Table 4. Summary of frame element segmentation results (CV = cross-validation)

	P	*R*
Segmentation	0.706	0.648
CV over sentences	0.706	0.648
CV over frames	0.706	0.639
LTH parser	0.706	0.648
MaltParser	0.642	0.540
Medical texts	0.673	0.611

Table 5. Summary of frame element classification results (CV = cross-validation)

	Accuracy
Baseline majority classifier	0.304
Classification accuracy	0.643
Core frame elements	0.719
Noncore frame elements	0.472
CV over sentences	0.643
CV over frames	0.494
LTH parser	0.643
MaltParser	0.612
Word features	0.643
Word and cluster features	0.650
Medical texts	0.526

4.9 Summary of results for the baseline systems

The results of the experiments described in the previous sections are summarized in Tables 4 and 5. As a general observation, the performance figures are lower than what we typically see for English (Johansson & Nugues 2007b; Das et al. 2010), which is unsurprising since the amount of annotated examples in SweFN is much smaller than in FrameNet.

5. Using the FrameNet relational structure to improve the semantic role labeler

As we have seen, SweFN is a fairly small resource and not much annotated material is available, and this limits the performance attainable by automatic classifiers trained on this dataset. In particular, the scarce amount of data makes it very hard for the machine learning methods to discern general linguistic facts concerning the syntactic–semantic linking patterns, such as the relation between the voice of a verb, the syntactic functions of its arguments, and the semantic roles of the arguments (Dowty 1991).

In the latter part of this chapter, we will investigate how the frame-to-frame relations described in the FrameNet ontology can allow us to generalize across frames. This allows the classifier to learn general linguistic facts, and it also leads to more efficient use of the annotated data. To allow this kind of generalization, we formulate the semantic role selection problem as a classification task with *non-atomic labels*: that is, a set of labels instead of a single label. This cross-frame generalization method reduces the number of errors made by the classifier by 27%, improving the accuracy from 0.544 to 0.665. When evaluating on frames for which the classifier has not been trained, the accuracy improves from 0.072 (random performance) to 0.534, a 50% error reduction.

We first describe the implementation of the improved semantic role labeling system that uses the frame-to-frame relations to improve its accuracy, and we discuss several feature representation approaches that take these relations into account in order to improve the generalization capabilities of the system. We then carry out an evaluation of the different representation approaches and finally discuss the results.

5.1 A classifier using non-atomic semantic role labels

The labeling classifier described in Section 3.1 is a quite typical example of how semantic role classifiers are normally implemented: in this model, each frame is independent of all other frames. However, in our case, when the amount of training data is quite small, the limitations of this standard approach become apparent:

- Since there are many frames, the amount of training data for each frame is very limited.
- Basic linguistic facts, such as which roles are likely to appear in subject position, have to be relearned for each frame.

To remedy these problems, we developed a classifier using non-atomic labels: instead of just a simple label `Ingestion`:INGESTOR, the classifier can predict several labels, using some sort of decomposition into meaningful parts. For instance, we can make use of the frame-to-frame inheritance relations defined in FrameNet, and analyze the `Ingestion`:INGESTOR label as { `Ingestion`:INGESTOR, `Ingest_-substance`:INGESTOR, `Manipulation`:AGENT, `Intentionally_affect`: AGENT, `Intentionally_act`:AGENT, `Transitive_action`:AGENT }. This decomposition and several alternatives will be described in detail in Section 5.2.

As described in Section 3.1, our baseline classifier is a linear classifier trained using an online learning algorithm (Crammer et al. 2006). Assume that the frame F lists a set of semantic role labels $l_1, \ldots, l_n$, then the classifier predicts a semantic role label l^* for a given argument a using this model:

$$l^* = \arg\max_{l \in F} w \cdot \Phi(a, l)$$

Here Φ is a feature function describing features of the argument a taking the semantic role label l, and w is a weight vector produced by some training algorithm.

This classifier model can easily be generalized to the non-atomic case. We then assume that each role label l can be decomposed using a decomposition function D, which returns a set of labels. We now apply the feature function to each sub-label λ instead of the main label l.

$$l^* = \arg\max_{l \in F} \sum_{\lambda \in D(l)} w \cdot \Phi(a, l)$$

Non-atomic classification has been described in a number of publications. It is fairly common in text categorization, where *hierarchical* classification (Moschitti et al. 2012) is probably the most common type. One of the most similar to ours is the action classifier by Roth & Tu (2009), which handled a large label set by decomposing the labels into meaningful parts.

5.2 Generalization methods

We investigated several ways of analyzing the labels, and most of them were based on the properties of the frames defined in the FrameNet ontology, including several different frame-to-frame relations as well as semantic types of the frames. SweFN does not explicitly define such properties, but since the frames and frame elements are for the most part based on their English counterparts, we used the English ontology and assumed that properties of frames including frame-to-frame relations would carry over to the Swedish frames (see also Chapters 2 and 8 in this volume). In case a frame was not found in the English FrameNet ontology, we did not decompose the label.

The first decomposition method we tried was based on *frame-to-frame relations*. For instance, the frame `Ingestion`, evoked by items such as *eat* and *drink*, labels the active participant (who eats or drinks) as INGESTOR. FrameNet defines an *inheritance* relation between this frame and a more general frame, `Ingest_substance`, which itself inherits from `Manipulation`, and so on. Furthermore, in a frame-to-frame relation, FrameNet defines correspondences between frame elements: for instance, the INGESTOR in `Ingestion` corresponds to INGESTOR in `Ingest_substance` and AGENT in `Manipulation`.

In our model, we used the following frame-to-frame relations:

Inheritance: specific to general, e.g. COMMUNICATION_NOISE to COMMUNICATION.
Subframe: from component to complex, e.g. SETTING_OUT to TRAVEL.
Causative-of: causative to inchoative, e.g. CAUSE_TEMPERATURE_CHANGE to INCHOATIVE_CHANGE_OF_TEMPERATURE.
Inchoative-of: inchoative to stative, e.g. INCHOATIVE_CHANGE_OF_TEMPERATURE to TEMPERATURE.
Using: child to parent, e.g. COMMUNICATION_NOISE to MAKE_NOISE.
PERSPECTIVE-ON: perspectivized to neutral, e.g. RIDE_VEHICLE to USE_VEHICLE.

To analyze a label in terms of frame-to-frame relations, we applied the transitive closure of each relation and returned the resulting set. For instance, when applying the Inheritance relation to the `Ingestion`:INGESTOR label, we get the following set: { `Ingestion`:INGESTOR, `Ingest_substance`:INGESTOR, `Manipulation`:AGENT, `Intentional_affect`:AGENT, `Intentional_act`:AGENT, `Transitive_action`:AGENT }.

The second decomposition method was based on the *semantic type* of the semantic role. For instance, the `Ingestion`:INGESTOR role needs to be filled by an entity of the semantic type SENTIENT. The decomposition of this role then simply becomes { `Ingestion`:INGESTOR, SENTIENT }.

The third decomposition method was based on the simple notion of *label generalization*: if two semantic roles in two different frames have the same name, then we use the same label. For instance, we change `Ingestion`:INGESTOR and `Ingest_substance`:INGESTOR to INGESTOR. We normalized the spelling, punctuation, and capitalization of the labels before generalizing.

6. Experiments in cross-frame generalization

We evaluated the classifier on the example sentences in SweFN. The frame and the argument were given to the classifier, which then had to predict the semantic role. We evaluated in two different ways: *In-frame* evaluation, where a 5-fold cross-validation was carried out over the set of sentences, and *Out-frame* evaluation, where the cross-validation was done over the set of frames. The out-frame setting simulates the situation where a new frame has been defined, but no training data have been annotated. Without any sort of cross-frame generalization, the classification in the out-frame setting becomes a random baseline.

Table 6 shows the results of using the frame-to-frame relations for analyzing the semantic role labels. We see that decomposition based on Inheritance is by far the most effective of these, although the highest performance is obtained when combining all types of relation-based decompositions.

Table 6. Classification results with generalization based on frame-to-frame relations

Classifier	In-frame	Out-frame
Baseline	0.544	0.072
Inheritance	0.587	0.281
Using	0.558	0.205
Subframe	0.548	0.115
Causative-of	0.545	0.097
Perspective-on	0.545	0.081
Inchoative-of	0.544	0.080
All except Inheritance	0.560	0.240
All relations	0.596	0.369

The effect of analyzing labels in terms of semantic type is similar. The in-frame performance is higher than that of relation-based decomposition, while the out-frame performance is a bit lower. The two generalization methods seem to complement each other, since we get a higher performance by combining them. Table 7 shows the results.

Finally, Table 8 shows the effect of using label generalization. This is by far the most effective method. However, we get even higher performance by combining it with the other two methods.

Table 7. Adding semantic type generalization

Classifier	In-frame	Out-frame
Semantic type	0.617	0.317
Semantic type + relations	0.635	0.426

Table 8. Results with label generalization

Classifier	In-frame	Out-frame
Label generalization	0.659	0.515
LG + ST + relations	0.665	0.534

7. Conclusion

We have presented the first results of experiments on semantic role labeling using 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.

The performance figure are lower than what we typically see for English (Johansson & Nugues 2007a; Das et al. 2010), reflecting the size of the training material, which is fairly limited so far. 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: by generalizing role labels across frames and by adding cluster-based features. Both methods result in clear improvements.

When developing NLP systems for a language where resources are scarce, it is crucial to make effective use of the available data, and generalizing roles across frames is one way to utilize the annotated data more effectively. This also makes sense from a theoretical point of view, since predicting multiple labels allows the machine learner to learn general facts as well as specifics. For the task of FrameNet-based semantic role classification, a similar approach was used by Matsubayashi et al. (2009), who defined a classifier making use of *role groups*; the effect of the role groups turns out to be similar to our non-atomic classification approach.

Our experiments showed very significant error reductions when we used the cross-frame generalization methods. This was especially notable in the case of out-frame evaluation, which is to be expected since the baseline in this case was a random selection. The best classifier used all three types of label decomposition, and achieved a 26% in-frame and a 50% out-frame error reduction.

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 semi-supervised approaches (Fürstenau & Lapata 2009) and the inclusion of automatically produced training data (Johansson & Nugues 2006). It could possibly also be relevant to investigate approaches that integrate the syntactic and semantic analysis steps more tightly (Johansson 2009).

References

Baker, Collin F., Charles J. Fillmore & John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of Coling 1998*, 86–90. Montréal: ACL.

Borin, Lars, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj & Dimitrios Kokkinakis. 2010. The past meets the present in Swedish FrameNet++. In *Proceedings of EURALEX 2010*, 269–281. Ljouwert/ Leeuwarden: Fryske Akademy.

Borin, Lars, Markus Forsberg & Lennart Lönngren. 2013. SALDO: A touch of yin to WordNet's yang. *Language Resources and Evaluation* 47(4): 1191–1211. https://doi.org/10.1007/s10579-013-9233-4

Brown, Peter F., Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra & Jenifer C. Lai. 1992. Class-based *n*-gram models of natural language. *Computational Linguistics* 18(4): 467–479.

Collobert, Ronan & Jason Weston. 2007. Fast semantic extraction using a novel neural network architecture. In *Proceedings of ACL 2007*, 560–567. Prague: ACL.

Crammer, Koby, Ofer Dekel, Joseph Keshet, Shai Shalev-Schwartz & Yoram Singer. 2006. Online passive-aggressive algorithms. *Journal of Machine Learning Research* 7: 551–585.

Das, Dipanjan, Nathan Schneider, Desai Chen & Noah A. Smith. 2010. Probabilistic frame-semantic parsing. In *Proceedings of NAACL/HLT 2010*, 948–956. Los Angeles.

Dowty, David R. 1991. Thematic proto-roles and argument selections. *Language* 67(3): 574–619. https://doi.org/10.1353/lan.1991.0021

Ejerhed, Eva, Gunnel Källgren, Ola Wennstedt & Magnus Åström. 1992. *The linguistic annotation system of the Stockholm–Umeå corpus project – description and guidelines*. Research report. Umeå: Umeå University, Dept. of Linguistics.

Fan, Rong-En, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang & Chih-Jen Lin. 2008. Liblinear: A library for large linear classification. *Journal of Machine Learning Research* 9: 1871–1874.

Fillmore, Charles J. & Collin Baker. 2009. A frames approach to semantic analysis. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 313–339. Oxford: Oxford University Press.

Friberg Heppin, Karin & Maria Toporowska Gronostaj. 2012. The rocky road towards a Swedish FrameNet – creating SweFN. In *Proceedings of LREC 2012*, 256–261. Istanbul: ELRA.

Fürstenau, Hagen & Mirella Lapata. 2009. Semi-supervised semantic role labeling. In *Proceedings of EACL 2009*, 220–228. Athens. https://doi.org/10.3115/1609067.1609091

Gildea, Daniel & Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics* 28(3): 245–288. https://doi.org/10.1162/089120102760275983

Gildea, Daniel & Martha Palmer. 2002. The necessity of syntactic parsing for predicate argument recognition. In *Proceedings of ACL 2002*, 239–246. Philadelphia: ACL.

Halácsy, Péter, András Kornai & Csaba Oravecz. 2007. HunPos – an open source trigram tagger. In *Proceedings of ACL 2007: Demo and poster sessions*, 209–212. Prague: ACL.

Johansson, Richard. 2008. Dependency-based semantic analysis of natural-language text. Lund University. (PhD thesis).

Johansson, Richard. 2009. Statistical bistratal dependency parsing. In *Proceedings of EMNLP 2009*, 561–569. Singapore: ACL. https://doi.org/10.3115/1699571.1699586

Johansson, Richard. 2012. Non-atomic classification to improve a semantic role labeler for a low-resource language. In *Proceedings of *SEM 2012*, 95–99. Montréal: ACL.

Johansson, Richard, Karin Friberg Heppin & Dimitrios Kokkinakis. 2012. Semantic role labeling with the Swedish FrameNet. In *Proceedings of LREC 2012*, 3697–3700. Istanbul: ELRA.

Johansson, Richard & Pierre Nugues. 2006. A FrameNet-based semantic role labeler for Swedish. In *Proceedings of Coling/ACL 2006*, 436–443. Sydney: ACL. https://doi.org/10.3115/1273073.1273130

Johansson, Richard & Pierre Nugues. 2007a. LTH: Semantic structure extraction using nonprojective dependency trees. In *Proceedings of SemEval 2007*, 227–230. Prague: ACL. https://doi.org/10.3115/1621474.1621522

Johansson, Richard & Pierre Nugues. 2007b. Using WordNet to extend FrameNet coverage. In *Proceedings of the Nodalida workshop FRAME 2007: Building frame semantics resources for Scandinavian and Baltic languages*, 27–30. Lund: Lund University.

Johansson, Richard & Pierre Nugues. 2008a. Dependency-based syntactic–semantic analysis with PropBank and NomBank. In *Proceedings of the CoNLL 2008 shared task*, 183–187. Manchester: ACL. https://doi.org/10.3115/1596324.1596355

Johansson, Richard & Pierre Nugues. 2008b. The effect of syntactic representation on semantic role labeling. In *Proceedings of Coling 2008*, 393–400. Manchester: ACL. https://doi.org/10.3115/1599081.1599131

Koo, Terry, Xavier Carreras & Michael Collins. 2008. Simple semi-supervised dependency parsing. In *Proceedings of ACL/HLT 2008*, 595–603. Columbus: ACL.

Màrquez, Lluís, Xavier Carreras, Ken Litkowski & Suzanne Stevenson. 2008. Semantic role labeling: An introduction to the special issue. *Computational Linguistics* 34(2): 145–159. https://doi.org/10.1162/coli.2008.34.2.145

Matsubayashi, Yuichiroh, Naoaki Okazaki & Jun'ichi Tsujii. 2009. A comparative study on generalization of semantic roles in framenet. In *Proceedings of ACL/IJCNLP 2009*, 19–27. Singapore: ACL. https://doi.org/10.3115/1687878.1687883

Moschitti, Alessandro, Qi Ju & Richard Johansson. 2012. Modeling topic dependencies in hierarchical text categorization. In *Proceedings of ACL 2012*, 759–767. Jeju: ACL.

Nilsson, Jens, Johan Hall & Joakim Nivre. 2006. MAMBA meets TIGER: Reconstructing a Swedish treebank from antiquity. In *Treebanking for discourse and speech: Proceedings of the NODALIDA 2005 special session on treebanks for spoken language and discourse*, 119–132. Copenhagen: Samfundslitteratur.

Nivre, Joakim, Johan Hall, Jens Nilsson, Atanas Chanev, Gülşen Eryığıt, Sandra Kübler, Svetoslav Marinov & Erwin Marsi. 2007. MaltParser: A languageindependent system for data-driven dependency parsing. *Natural Language Engineering* 13(2): 95–135. https://doi.org/10.1017/S1351324906004505

Palmer, Martha, Dan Gildea & Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics* 31(1): 71–105. https://doi.org/10.1162/0891201053630264

Palmer, Martha, Daniel Gildea & Nianwen Xue. 2010. Semantic role labeling. *Synthesis Lectures on Human Language Technologies* 3(1): 1–103. https://doi.org/10.2200/S00239ED1V01Y200912HLT006

Roth, Dan & Yuancheng Tu. 2009. Aspect guided text categorization with unobserved labels. In *Proceedings of ICDM 2009*, 962–967. Miami Beach: IEEE. https://doi.org/10.1109/ICDM.2009.129

Xue, Nianwen & Martha Palmer. 2004. Calibrating features for semantic role labeling. In *Proceedings of EMNLP 2004*, 88–94. Barcelona: ACL.