

# Reusing Swedish FrameNet for training semantic roles

Ildikó Pilán, Elena Volodina

Språkbanken, University of Gothenburg  
Box 200, Göteborg 405 30, Sweden

Email: ildiko.pilan@svenska.gu.se, elena.volodina@svenska.gu.se

## Abstract

In this article we present the first experiences of reusing the Swedish FrameNet (SweFN) as a resource for training semantic roles. We give an account of the procedure we used to adapt SweFN to the needs of students of Linguistics in the form of an automatically generated exercise. During this adaptation, the mapping of the fine-grained distinction of roles from SweFN into learner-friendlier coarse-grained roles presented a major challenge. Besides discussing the details of this mapping, we describe the resulting multiple-choice exercise and its graphical user interface. The exercise was made available through Lärka, an online platform for students of Linguistics and learners of Swedish as a second language. We outline also aspects underlying the selection of the incorrect answer options which include semantic as well as frequency-based criteria. Finally, we present our own observations and initial user feedback about the applicability of such a resource in the pedagogical domain. Students' answers indicated an overall positive experience, the majority found the exercise useful for learning semantic roles.

**Keywords:** Swedish FrameNet, semantic roles, exercise generator

## 1. Introduction

The main objective of the implemented exercise is to help students raise their awareness of semantic (thematic) roles, which is a challenging task partly due to the lack of available example sentences for each role in the printed literature. Luckily, SweFN has proven to be a rich source comprising 5954 example sentences. However, SweFN is a resource whose practical application remains little explored within the pedagogical domain so far. For example, it would be unreasonable to expect students to differentiate between the hundreds of different roles that SweFN contains, thus, a number of roles needed to be grouped under more general headings.

Our goal with the exercise for training semantic roles is, thus, two-fold. On the one hand, we aim at reusing, and through this, attempt to evaluate SweFN and other relevant language resources for Swedish. On the other hand, we intend to offer practice material for students of Linguistics, in this particular case for the association of a semantic role to a word or a group of words, which can be a challenging task not only for students of Linguistics, but sometimes also for experienced linguists. Below, we give an overview of the resources used (section 2), present the exercise itself (section 3), and summarize some feedback about users' initial impressions and our experience when working with the relevant resources (section 4).

## 2. Resources

FrameNet aims at describing specific situations (*frames*) representing canonical events and situations, and the semantic role of the participants they involve (*frame elements*, FEs) (Baker et al., 1998). SweFN<sup>1</sup> is a freely available resource currently under development (Friberg Heppin and Gronostaj, 2012; Johansson et al., 2012). The

frames are analogous with those of the English FrameNet with only a few exceptions. The information for each frame includes also example sentences which have been carefully selected from non-adapted real-life corpora and then equipped with manual annotations. Below, an example sentence (1a) for the *Apply\_heat* frame from SweFN is shown together with its English translation (1b). It contains a lexical unit (LU), i.e. the verb evoking the frame (*baka*, Eng. “bake”) and the frame elements *Food* and *Duration*.

### (1) *Apply\_heat* frame

- a. [Baka]<sub>LU</sub> [potatisarna]<sub>Food</sub> [ca 45 - 60 minuter]<sub>Duration</sub> beroende på storlek.  
b. [Bake]<sub>LU</sub> [the potatoes]<sub>Food</sub> [ca. 45 – 60 minutes]<sub>Duration</sub> depending on their size.

Lärka<sup>2</sup> is a language learning platform comprising an exercise generator for two target groups: students of (Swedish) Linguistics and second language learners of Swedish (Volodina et al. 2013, Volodina et al. 2012). The exercise presented in this paper has been integrated into Lärka's module for students of Linguistics.

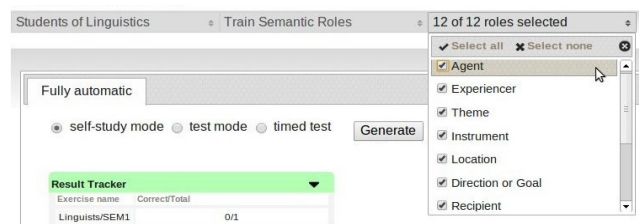


Figure 1: Lärka, general menus.

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

<sup>2</sup> <http://spraakbanken.gu.se/larka/>

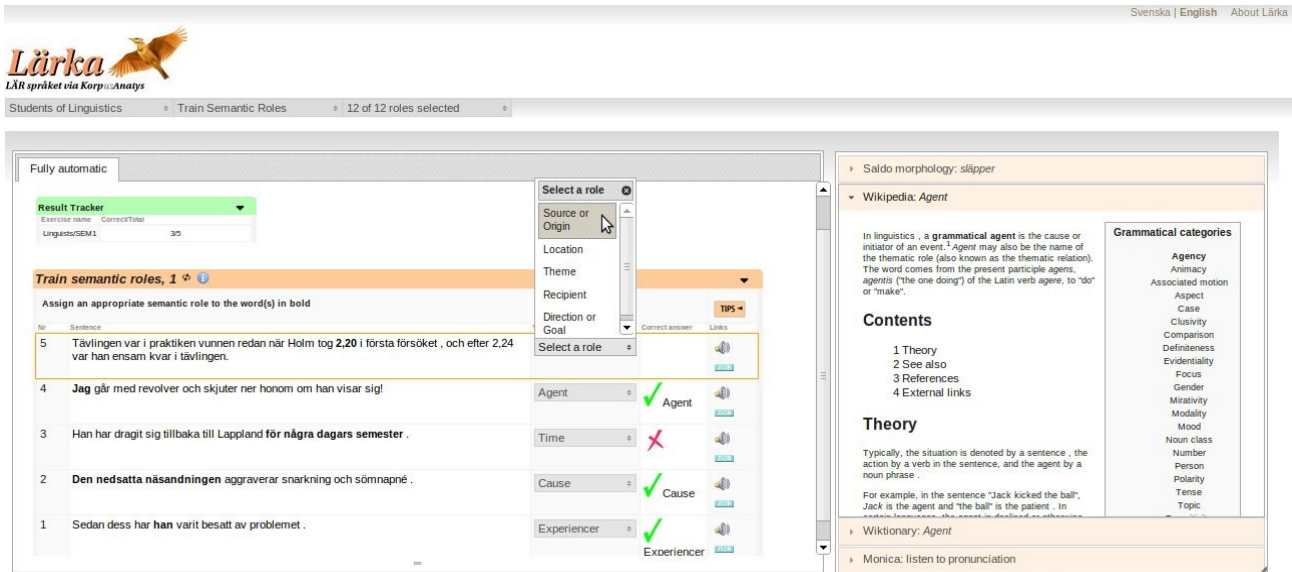


Figure 2: Exercise for training semantic roles in Lärka. User interface.

At the moment Lärka offers five exercise types, namely exercises for training *word classes*, *syntactic relations* and *semantic roles* for students of Linguistics; and exercises for training *word knowledge* and *inflectional paradigms* for second language learners. Exercises share the same *format* (multiple-choice), the same *context size* (sentence), as well as the same *reference materials* (Wikipedia, Wiktionary, morphological lexicon for Swedish SALDO morphology, and a text-to-speech module provided by SitePal<sup>3</sup>). Further, exercises can be trained in different *modes*: self-study, test or timed test (see Figure 1). Feedback is provided in the form of immediate

correct or incorrect response symbol and a result tracker. Each exercise is accompanied by information that can be opened through clicking on an *i*-icon.

### 3. Semantic role exercise

The semantic role exercise has been implemented as a multiple choice task. Users are presented with one sentence at a time in which one or more words are highlighted in bold (*target*), see Figure 2 above. For the successful completion of the task, the correct semantic role for the highlighted element(s) should be chosen from a list of five possible roles. Once the answer is provided, a new sentence is selected and displayed.

The graphical interface allows the user to click on all

Role	Nr	Role	Nr	Role	Nr	Role	Nr	Role	Nr	Role	Nr
<b>AGENT</b>		<b>968</b>	<b>EXPERIENCER</b>	<b>265</b>	<b>LOCATION</b>	<b>491</b>	<b>ORIGIN</b>	<b>145</b>	<b>MANNER</b>	<b>298</b>	
Agent	749	Grinder	1	Experiencer	175	Location	49	Source	125	Manner	298
Actor	15	Healer	3	Victim	90	Attraction	2	Origin	12	<b>PURPOSE</b>	<b>164</b>
Assailant	23	Helper	18	<b>THEME</b>	<b>376</b>	New_area	3	Heat_source	2	Purpose	153
Avenger	4	Ingestor	17	Theme	347	Seat_of_emotion	3	Hot_cold_source	1	Desired_goal	2
Creator	39	Killer	6	Whole_patient	8	Place	431	Name_source	1	Intended_goal	9
Deformer	6	Kindler	4	Co_theme	3	<b>GOAL</b>	<b>174</b>	Source_of_information	4	<b>CAUSE</b>	<b>126</b>
Destroyer	4	Leader	10	Mass_theme	9	Goal	146	<b>TIME</b>	<b>823</b>	Cause	106
Distributor	5	Obligator	1	Themes	7	Direction	28	Time	800	Internal_cause	5
Dodger	2	Operator	9	Theme_1	1	<b>RECIPIENT</b>	<b>54</b>	Destination_time	2	External_cause	2
Employer	19	Participant_1	4	Theme_2	1	Recipient	36	End_time	2	Preventing_cause	9
Evader	9	Participant_2	2	<b>INSTRUMENT</b>	<b>85</b>	Buyer	18	Interval	5	Formational_cause	4
Examiner	1	Survivor	5	Instrument	62			Relative_time	1		
Excreter	9	Teacher	3	Connector	4			Start_time	3		
				New_tool	5			Time_of_creation	10		
				Old_tool	5						
				Tools	4						
				Weapon	5						

Figure 3: Mapping of specific FE into abstract semantic roles.

<sup>3</sup> <http://www.sitepal.com/>

running words in the sentence whereby reference materials show articles that contain information about them. This reference window can be hidden if so wished. The user may also see background information about the exercise item (e.g. the semantic frame) in the form of a JSON<sup>4</sup> object.

The sentences used as exercise items are all examples extracted from SweFN and they amounted to 5954 sentences at the moment of writing. However, sentences shorter than seven tokens have been excluded to ensure a larger context which might help identifying the correct role.

Currently there are 12 different semantic roles available for training (see highlighted roles in capital letters in Figure 3). Reference materials during the selection of these roles included Fillmore (1968), Teleman et al. (1999), Jurafsky and Martin (2009) and discussions with researchers involved in the development of the SweFN project.

Domain- or verb-specific semantic roles have been mapped into broader, abstract semantic roles (*thematic relations*) mainly on the basis of the Frame Element Taxonomy<sup>5</sup> (Litkowski, 2010). Each of the selected abstract roles (e.g. *Theme*) were looked up in this taxonomy and some of their child nodes (e.g. *Themes*, *Theme\_1* etc.) were grouped together, including the parent role itself. The complete list of category groupings is presented in Figure 3, where *Nr* represents the number of sentences for the relevant role in the SweFN resource.

During the creation of the exercise, the selection of *distractors*, i.e. the incorrect answer options, presented a particular challenge. We opted for a selection based on a combination of: (a) semantic relatedness among the roles, (b) frequency information and (c) randomness. For the first aspect, certain roles have been grouped together into the generalized semantic roles *Actor* and *Undergoer* (Van Valin, 1999) which we complemented with two additional categories (*Place*, *Other*). The thematic roles for each of these categories are presented in Table 1.

Generalized semantic roles		Additional macro-categories	
ACTOR	UNDERGOER	PLACE	OTHER
Experiencer	Theme	Location	Purpose
Agent	Patient	Origin (Source)	Cause
	Recipient	Direction (Goal)	Manner
			Instrument
			Time

Table 1: Semantics-based grouping.

4 JSON – JavaScript Object Notation, a lightweight data-interchange format <<http://www.json.org/>>

5 <http://www.cires.com/db/feindex.html>

When dividing the roles into frequency bands, we distinguished three categories according to the number of times the roles appeared in SweFN: low-, middle- and high-frequency thematic roles. This grouping is shown in Table 2.

High frequency (>200)		Middle frequency (150-200)		Low frequency (< 150)	
ROLE	NR	ROLE	NR	ROLE	NR
Agent	968	Direction (Goal)	174	Cause	125
Time	822	Purpose	164	Recipient	54
Location	490	Origin (Source)	142	Instrument	85
Theme	376				
Manner	298				
Experiencer	265				

Table 2: Frequency-based grouping.

On the basis of the two groupings presented in Tables 1 and 2, the list of distractors proposed per each exercise item are selected in the following way: the first distractor is a semantically-related one (Table 1), the second and the third are from the same frequency band (Table 2), and the fourth is a random distractor chosen from the complete list of roles.

#### 4. SweFN and Lärka: initial experiences

During the creation of the exercise, we had an opportunity to gain insight into how easy-to-use the structure and the content of SweFN was. The structure and the format proved to be very convenient which ensured a simple and fast way for the extraction of the example sentences. There were some instances which were not suitable for our purposes (e.g. LUs with medical drug names), these have been excluded from the exercise. Being that the precise number and name of thematic roles differ across different reference resources in the literature, to find or create a hierarchy and macro-categories for such roles, which would be widely accepted, was challenging. We have also realized that the availability of syntactic annotation for the example sentences in the future would allow for the observation of the interaction and the correspondences among constituents at different linguistic levels.

Besides SweFN, we received feedback from the first users about this exercise as well as about our learning platform, Lärka, in general. Linguists who have tested our platform found the structure of the page clear, and the *Reference* section useful. Furthermore, they expressed preference for the generation of more than one exercise item at a time to reduce waiting times. They have suggested that the role names in Swedish should be accompanied by the English

terms (since they are more widely known) together with a short definition, which we have already added. Besides linguists, we carried out an initial evaluation with a group of students who have used different Lärka exercises during the laboratory sessions of a university course in Linguistics. The answers for the question concerning the exercise for semantic roles is presented in Figure 4.



Figure 4: Results of the evaluation question for the semantic role exercise.

Students were asked to rate how helpful the exercise was for training semantic roles on a scale from 1 (not helpful at all) to 6 (very helpful). Out of a total of 18 respondents, about 28% were very positive (scores 5 and 6) about the helpfulness of the exercise for training semantic roles, whilst 39% of students found it useful to some extent (score 4). According to the comments provided, some students perceived the exercise somewhat difficult, either due to an insufficient amount of background knowledge and lack of familiarity with the terms used, or because certain categories of roles seemed similar and were, therefore, harder to distinguish. Some students continued training with Lärka also outside the classroom as autonomous student practice. The responses for the exercise items answered by students have been logged for future analysis.

## 5. Concluding remarks

In this article we have reported on the recently added Lärka-based exercise for training semantic roles. We have described the user interface, the algorithm for exercise generation as well as explained the reasons for a number of decisions made during the development.

The exercise has been positively accepted by both linguists and a first group of student evaluators. They provided feedback, among others, about the difficulty level of the exercise and the clarity of the categories and terms used. The results indicate a potential need for variants of this exercise of different difficulty levels, which would ensure suitable practice also for students in the initial stage of their studies. In the future, larger-scale and more in-depth evaluations could further confirm the appropriateness of SweFN as a resource for this type of exercise. Moreover, we plan to analyse the collected data about students' performance to identify error-prone roles and SweFN example sentences which might be less

suitable from a pedagogical perspective. Changes to the categorization of fine-grain SweFN roles will also be considered.

Besides the aspects mentioned above, our list of future additions and improvements also contains:

- a possibility to see semantic roles of the unmarked (non-target) part of the sentence, e.g. through adding tooltips;
- adding syntactic trees to each sentence to allow for cross-level comparison (i.e. syntactic structure versus semantic analysis and word classes);
- adding a new exercise format for more advanced training where an option of analysing the whole sentence into the constituent roles will be provided

SweFN offers also potential for being used in the context of second language learning which yet needs to be explored and tested.

## 6. Acknowledgements

We want to extend our thanks to our two SweFN colleagues – Maria Toporowska Gronostaj and Richard Johansson – who have been extremely helpful during the brainstorming, implementation and testing phases, not only with their advice, but also through providing access to materials, general guidance and honest comments about the produced results.

## 7. References

- Baker, C. F; Fillmore, C. J; Lowe, J. B. (1998). The Berkeley FrameNet project. In *Proceedings of COLING/ACL*. Montreal, Canada, pp. 86–90 .
- Fillmore, C. J. (1968). The case for case. In Bach and Harms, (Eds.), *Universals in Linguistic Theory*, pp. 1–88. Holt, Rinehart, and Winston, New York.
- Friberg Heppin, K; Toporowska Gronostaj, M. (2012). The Rocky Road towards a Swedish FrameNet-Creating SweFN. In *Proceedings of LREC-2012*, pp. 256–261.
- Johansson, R.; Friberg Heppin, K; Kokkinakis, D. (2012). Semantic Role Labeling with the Swedish FrameNet. In *Proceedings of LREC*, pp. 3697–3700.
- Jurafsky, D.; Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2 ed.). Pearson International Edition.
- Litkowski, K. (2010). CLR: Linking events and their participants in discourse using a comprehensive FrameNet dictionary. In *Proceedings of the 5th international workshop on semantic evaluation*, pp. 300–303.
- Teleman, U.; Hellberg, S; Andersson, E. (1999). *SAG. Svenska Akademiens grammatik*.

- Van Valin Jr, R. D. (1999). Generalized semantic roles and the syntax-semantics interface. *Empirical issues in formal syntax and semantics*, 2, pp. 373–389.
- Volodina, E; Pijetlovic, D; Pilán, I; Johansson Kokkinakis, S. (2013). Towards a gold standard for Swedish CEFR-based ICALL. In *Proceedings of the Second Workshop on NLP for Computer-Assisted Language Learning. Nodalida 2013*, Oslo, Norway.
- Volodina, E; Borin, L; Loftsson, H; Arnbjörnsdóttir, B.; Leifsson, G. Ö. (2012). Waste not, want not: Towards a system architecture for ICALL based on NLP component re-use. Workshop on NLP in Computer-Assisted Language Learning. In *Proceedings of the SLTC 2012 workshop on NLP for CALL. Linköping Electronic Conference Proceedings*, 80, pp. 47–58.