

A Frame-Semantic Parsing Plugin for Swedish Research Infrastructure*

Dana Dannélls^{1,*}, Shafqat Virk[†]

Språkbanken Text, University of Gothenburg, 405 30 Gothenburg, Sweden

Abstract

We present the development of a frame-semantic parsing plugin for Sparv – an annotation pipeline for Swedish. The plugin integrates a frame-semantic parser into the annotation pipeline, enabling the automatic identification of frames evoked by lexical units and the assignment of frame elements to their corresponding arguments in text. Designed to operate seamlessly within the infrastructure, the plugin takes raw text as input, and outputs semantic role information in a standardized format compatible with other annotation layers. This implementation demonstrates how frame-semantic analysis can be made available as an additional corpus annotation layer, enriching texts with structured semantic representations that go beyond syntactic or lexical features. By providing access to semantic role information, the plugin can support a wide range of research applications, including semantic search, discourse analysis, and investigation of meaning variation in language use.

Keywords

Digital humanities, Frame semantic parsing, Swedish infrastructure

1. Introduction

Språkbanken Text (SBX), the Swedish research infrastructure for language technology, develops freely available digital research platforms. One of these platforms called Strix [1] is a document-centric platform designed for researchers in the humanities, enabling them to create, annotate, and analyze documents in meaningful and semantically informed ways. The platform, along with other SBX platforms, is built on the annotation pipeline, Sparv [2], which makes it possible to analyze digital material in depth and to formulate nuanced research questions about the data. For example, semantic analysis is performed with the help of lexical-semantic databases such as Saldo [3] and Swedish FrameNet [4]. This process produces semantic representations that make it possible to identify words according to their senses within a conceptual network. As a result, it is possible to highlight words according to the semantics of LEADERSHIP (Figure 1a) or display the frequencies of three major cities in Sweden mentioned in the same collection over time (Figure 1b), or visualize all the geographical locations referenced in the collection (Figure 1c).

Currently, the semantic analysis operates at the word level, that is, it relies on a simple word-matching approach against lexical resources, and does not provide deeper sentence-level interpretation. As a result, questions about which titles or roles are mentioned in specific places and periods, and how these occurrences might relate to historical developments, cannot be systematically explored. To move beyond word-level analysis and capture the underlying meaning within sentences, semantic role labeling (SRL) – the task of automatically assigning semantic roles – can be applied. Recent advances in transformer-based language models have made it possible to train an SRL model for Swedish [5]. In this paper, we describe how this model was integrated into Sparv to enhance semantic analysis. We highlight both the technical implementation and the potential applications for digital humanities research, demonstrating how SRL

Huminfra Conference 2025, Stockholm, 12–13 November 2025.

*You can use this document as the template for preparing your publication.

*Corresponding author.

†These authors contributed equally.

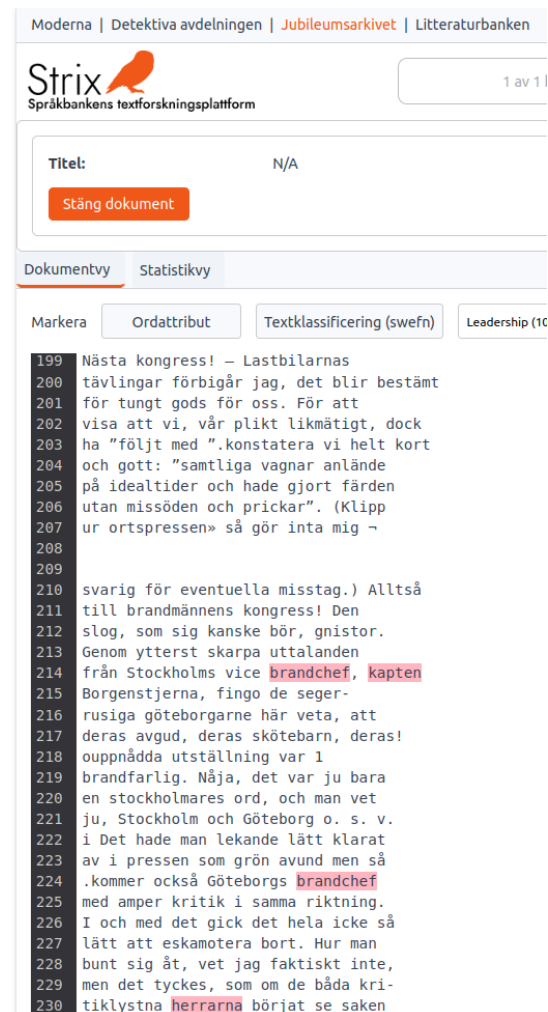
✉ dana.dannells@svenska.gu.se (D. Dannélls); shafqat.virk@svenska.gu.se (S. Virk)

ORCID 0000-0002-3338-2979 (D. Dannélls); 0000-0002-5030-9191 (S. Virk)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

integration can facilitate new forms of text analysis, improve interpretability, and enable more nuanced exploration of linguistic and cultural phenomena.



(a) Analysis according to the LEADERSHIP frame in Swedish FrameNet



(b) Frequencies of Stockholm, Gothenburg, and Malmö in the collection over time



(c) Geographic locations mentioned in the collection

Figure 1: Different analyses of the Jubileumsarkivet collection in Strix

2. Sparv

Språkbanken Text's analysis platform, Sparv, is a command-line tool for annotating text corpora [6]. The modular architecture of Sparv enables the integration of different corpus analyses on the token and sentence levels. Current analyses include tokenization, morphological, part-of-speech, compound, name entity recognition, sentiment, sentence segmentation, dependency parsing, and word sense disambiguation (WSD) [7], just to name a few.

Sparv is highly customizable due to its flexible plugin system. All modules (importers, annotators, exporters, model builders) are replaceable, making it easy to extend the system with custom plugins.¹

¹<https://github.com/spraakbanken/sparv>

3. Swedish frame semantic parser

Frame semantic parsing, the automatic processing of enhancing written text with semantic information, aims at identifying and classifying words and phrases according to their semantics following FrameNet principles [8] that are based on frame semantics theory [9, 10]. Training a frame semantic model requires access to large amounts of semantically annotated sentences, but with the advantages of recent transformer techniques, it was possible to train a language-specific parser on Swedish FrameNet [4], a resource with a relatively small amount of semantically annotated training data compared to English [5].

The parser is currently available in two model sizes—small and base [11]—both of which we retrained. Their evaluation results on the Swedish dataset are relatively low (see Table 1). Our goal is to train the large model, which is expected to yield significantly better performance.

Model	Task	Dataset	Precision	Recall	F1
Base	Args Extraction	Validation	0.501	0.450	0.474
	Frame Classification	Validation	0.468	0.468	0.468
	Trigger Identification	Validation	0.513	0.510	0.511
	Args Extraction	Test	0.521	0.456	0.487
	Frame Classification	Test	0.452	0.452	0.452
	Trigger Identification	Test	0.515	0.515	0.515
Small	Args Extraction	Validation	0.374	0.339	0.356
	Frame Classification	Validation	0.437	0.437	0.437
	Trigger Identification	Validation	0.454	0.451	0.452
	Args Extraction	Test	0.393	0.350	0.371
	Frame Classification	Test	0.434	0.434	0.434
	Trigger Identification	Test	0.419	0.408	0.413

Table 1

Comparison of Base and Small models on validation and test datasets for argument extraction, frame classification, and trigger identification.

Both models achieve moderate performance on the three evaluated tasks: argument extraction, frame classification, and trigger identification. In the validation set, the scores range from approximately 0.36 to 0.47 F1, with precision and recall showing balanced values. In the test set, the models achieve slightly higher performance in some cases, with F1 scores around 0.45–0.52. These results indicate that the models are able to capture relevant semantic and structural patterns in the data, though there is still room for improvement in terms of precision and recall, particularly for argument extraction. The larger model is expected to perform even better with competitive evaluation scores to English counterpart parsers.

4. Swedish Frame Semantic parser plugin

As part of the recent restructuring of Sparv, a plugin-based architecture was introduced that enables new annotation schemes and tools to be flexibly integrated into the processing pipeline. This modular design allows researchers to extend the system with minimal overhead, facilitating the addition of task-specific layers of linguistic analysis beyond the core annotations such as tokenization, POS tagging, and lemmatization.

Within this framework, we developed a plugin for frame-semantic parsing, aimed at providing semantic role information as an additional layer of text analysis. The plugin integrates the developed frame-semantic parser into the annotation pipeline, automatically identifying frames evoked by lexical units, and assigning frame elements to their corresponding arguments in the text. This enriches the corpus with structured semantic representations that go beyond surface-level syntactic analysis, enabling a more nuanced exploration of meaning in context. An example of how the plug-in operates is illustrated in the following.

Input

Ordförande för förhandlingarna har varit fröken M. Nordénfeldt, Göteborg.

Output

Activity: förhandlingarna

Leader: fröken M. Nordénfeldt

Place: Göteborg

In this example, the input sentence *Ordförande för förhandlingarna har varit fröken M. Nordénfeldt, Göteborg.* ‘The chairman of the negotiations has been Miss M. Nordénfeldt, Gothenburg.’ was parsed according to the semantic frame LEADERSHIP, triggered by the lexical unit *Ordförande* ‘chairman’.

From a technical perspective, the plugin is designed to operate seamlessly within the standardized data exchange format of the infrastructure. It consumes pre-processed linguistic information (e.g., lemmas, POS tags, dependency relations) already available in the pipeline and outputs semantic annotations in a format consistent with other modules. This ensures interoperability with downstream components and facilitates combined analyses. The modular design further allows the plugin to be independently maintained or upgraded, ensuring its adaptability to future parser improvements or alternative semantic frameworks. The plugin is available as a github repository at <https://github.com/shafqatvirk/sparv-plugin-semparse-swe.git>, and the instructions to install and use a plugin can be found at <https://spraakbanken.gu.se/sparv/user-manual/intro/>.

5. Use case

The following use case illustrates a hypothetical example of how the frame-semantic parsing plugin could be applied to analyze diachronic Swedish texts and trace how the concept of *marknad* (‘market’) has shifted over time. With appropriate training or fine-tuning, the parser could distinguish between a Physical Market Frame, where *marknad* denotes a concrete site of exchange (involving elements such as Location, Seller, Buyer, Goods, Transaction, Time, and Atmosphere), and an Abstract Market Frame, where it refers to an economic or systemic domain (with elements such as Domain, Participants, Goods/Services, Mechanism, Condition, Competition, Outcome, and Influence). By applying these frame distinctions across historical corpora, we can quantify when and how references to *marknad* shift from physical to abstract contexts, thereby providing empirical evidence of semantic change in Swedish economic discourse.

6. Conclusion and future work

Sparv-SRL is a key plugin enabling advanced natural language understanding (NLU) for Swedish. Integrated into Språkbanken Text’s annotation pipeline, it enriches sentences with semantic role information. The integration of the frame-semantic plugin demonstrates both the extensibility of the new infrastructure and the benefits of incorporating deeper linguistic analysis. By augmenting annotated corpora with semantic roles, the plugin opens new possibilities for linguistic research and downstream applications, including semantic search, discourse analysis, and studies of meaning variation over time.

In future work, we plan to enhance the search functionality in Strix, to enable researchers not only to analyze but also to visualize search results in ways that align with their specific research questions.

Acknowledgments

This work has been funded and supported by Språkbanken – jointly funded by its 10 partner institutions and the Swedish Research Council (2025–2028; project id 2023-00161) as well as (2018–2024; dnr 2017-00626). We would like to acknowledge the Swedish national research infrastructure Huminfra,

funded for the years 2022-2024 and 2005-2028, contracts 2021-00176 and 2023-00171 respectively, and the participating partner institutions.

References

- [1] M. Forsberg, Y. Ali Mohammed, E. Sköldbberg, M. Öhrman, SO in Strix: a lexicographic case study of entry vectors, in: *Sixty years of Swedish computational lexicography / Dana Dannélls, Kristian Blensenius and Lars Borin (eds.)*, De Gruyter Brill, Berlin, 2025, pp. 289–303.
- [2] M. Hammarstedt, A. Schumacher Olsson, L. Borin, M. Forsberg, Sparv 5.3.0: Språkbanken's Analysis Platform – Technical Report, Technical Report, University of Gothenburg, Göteborg, 2025.
- [3] L. Borin, M. Forsberg, Saldo: the hub of Språkbanken's lexical research infrastructure, in: *Sixty years of Swedish computational lexicography / Dana Dannélls, Kristian Blensenius and Lars Borin (eds.)*, De Gruyter, Berlin, 2025, p. 97–111.
- [4] D. Dannélls, L. Borin, M. Forsberg, K. F. Heppin, M. T. Gronostaj, Swedish FrameNet, in: D. Dannélls, L. Borin, K. F. Heppin (Eds.), *The Swedish FrameNet++: Harmonization, integration, method development and practical language technology applications*, volume 14 of *Natural Language Processing*, John Benjamins, Amsterdam, 2021, pp. 37–66.
- [5] D. Dannélls, R. Johansson, L. Y. Buhr, Transformer-based Swedish semantic role labeling through transfer learning, in: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, ELRA and ICCL, Turin, Italy, 2024.
- [6] M. Hammarstedt, A. Schumacher, L. Borin, M. Forsberg, Sparv 5 Developer's Guide, Technical Report, University of Gothenburg, 2022.
- [7] R. Johansson, L. Nieto Piña, Combining relational and distributional knowledge for word sense disambiguation, in: *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA)*, volume 109, Linköping Electronic Conference Proceedings, Linköping University Electronic Press, Vilnius, Lithuania, 2015, pp. 69–78.
- [8] J. Ruppenhofer, M. Ellsworth, M. R. Petruck, C. R. Johnson, C. F. Baker, J. Scheffczyk, FrameNet II: Extended Theory and Practice, Technical Report, 2016. URL: <https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf>, berkeley: ICSI.
- [9] C. J. Fillmore, Frame semantics and the nature of language, *Annals of the New York Academy of Sciences* 280 (1976) 20–32. URL: <https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.1976.tb25467.x>. doi:<https://doi.org/10.1111/j.1749-6632.1976.tb25467.x>.
- [10] C. J. Fillmore, Frame semantics, in: *Linguistic Society of Korea (Ed.)*, *Linguistics in the Morning Calm*, Hanshin Publishing Co., Seoul, 1982, pp. 111–137.
- [11] D. Chanin, Open-source frame semantic parsing, arXiv preprint arXiv:2303.12788 (2023).