# When Sparv met Superlim...
# A Sparv Plugin for Natural Language Understanding Analysis of Swedish

Felix Morger

Språkbanken Text

## 1 Introduction: What is Sparv-Superlim?

Sparv-Superlim is a Sparv plugin (Hammarstedt et al., 2022) for making a range of different corpus analyses using the reference models trained for the tasks on the Superlim benchmark (Berdicevskis et al., 2023). The tasks available in Superlim range from sentence level classification tasks, such as automatic stance detection, to word-level relatedness and similarity judgements to inference and sentence similarity. These make it possible to integrate additional annotations and features into the Sparv Pipeline. The code and more documentation of the Sparv-Superlim plugin are available on GitHub.[1]

This technical report serves as a documentation of the plugin: how to install and use it for meaningful corpus analysis. A complementary goal of this report is to also show the how integration of Superlim into Sparv allows for evaluating the ecological validity (Andrade, 2018) of the benchmark itself on novel data. The use cases presented in this report demonstrate how important it is to provide further analysis to a benchmark on other corpora and how integrating reference models into a text analysis tool like Sparv helps in this effort.

The rest of the report is structured as follows: Section 2 gives a background into Sparv and Superlim. Section 3 shows how to install the plugin and how to use it to annotate a corpus. In Section 4, the plugin is used on a couple of use cases to see if the annotations outputted by the reference models align with the expected political stances of Swedish political parties. I also discuss these results and their broader implications. Section 5 concludes the report.

## 2 Background

**Sparv Pipeline** (Hammarstedt et al., 2022), Språkbanken's official analysis platform, is a command-line tool developed by Språkbanken Text for annotating

[1] https://github.com/spraakbanken/sparv-sbx-superlim

text corpora. Its pipeline architecture enables the seamless integration of different corpus analyses on both the sentence, token and document-level. It comes with built-in token and sentence segmentation, part-of-speech and morphological tagging as well as automatically generated corpus statistics. These analyses can be done on raw data or on existing annotated corpora through its configuration framework using `.yaml` files. The Sparv Pipeline is highly customizable, thanks to its flexible plugin system, which is what the Sparv-Superlim plugin presented in this report relies on.

**Superlim** (Berdicevskis et al., 2023) is a multi-task benchmark for Swedish. The name is a Swedish translation of the English SuperGLUE benchmark, which the project was inspired by (*lim* is Swedish for 'glue'). The current 2.0 version consists of 14 datasets and 15 tasks. These include text-level, word-level and inference tasks as well as one generation, summarization task (SweDN). Superlim is one of the first benchmarks of its kind for NLU benchmarking on Swedish specifically, which is an especially important piece of infrastructure for Swedish NLP today as large language models (LLMs) are useful for a large variety of downstream tasks. Apart from the multi-task benchmark itself, the Superlim project released a reference implementation using mostly Swedish-based LLMs. These were for example BERT-based models (Devlin et al., 2019), such as KB-BERT (Malmsten et al., 2020) for the text-level classification tasks and GPT-based models (Radford et al., 2018), such as the GPT-SW3 series (Ekgren et al., 2023) for the word-level tasks.[2] The benchmark also includes an online platform where progress is tracked on a public leaderboard and where users can submit results.[3]

# 3 Installing and running the plugin

The most important prerequisite is Sparv, which can be installed using `pipx`.

```
python3 -m pip install --user pipx
python3 -m pipx ensurepath
pipx install sparv-pipeline
```

Once Sparv is installed, the Sparv-Superlim plugin can be added by injecting it into the isolated environment of the `sparv-pipeline` Python package (`sparv-sbx-superlim` is also a pip package):

```
pipx inject sparv-pipeline sparv-sbx-superlim
```

In order to run Sparv, a corpus has to be configured. This is done in a specific folder, where the source corpora is stored and a `config.yaml` file, which gives Sparv instructions on what to do with the corpora. A configuration can for example look like this:

[2] https://github.com/aidotse/superlim-baselines
[3] https://lab.kb.se/leaderboard/

```
metadata:
  id: corpora
  name:
    eng: corpora
    swe: korpusar
  language: swe
  description:
    eng: Test corpora for Superlim project
    swe: Test corpora for Superlim project
import:
  source_dir: source
  importer: text_import:parse
  text_annotation: text
export:
  default:
    - xml_export:pretty
  annotations:
    - <sentence>
    - <token>
    - <sentence>:sbx_superlim.migration_stance
    - <sentence>:sbx_superlim.nuclear_stance
```

This configuration file instructs `sparv` to segment the text into sentences and tokens and to annotate the sentences with `migration_stance` and `nuclear_stance`. Table 1 shows the currently available Sparv-Superlim annotations, the Superlim tasks they are based on, the annotation label and the type of text segment they process.

| **Superlim task** | **Sparv-Superlim Annotation** | **Annotation** | **Label** | **Segment** |
|---|---|---|---|---|
| absabank-imm | migration_stance | Attitude towards immigration | float between 1-5 | sentence |
| argumentation-sentences | [topic_stance] | Stance to a given topic | pro, con or neutral | sentence |
| dalaj-ged | correct_swedish | Correct Swedish | correct or incorrect | sentence |
| swenli | previous_entailment | The logical relationship of two sentences | entailment, contradiction or neutral | sentence pair |
| sweparaphrase | similarity | Similarity between two sentences | float between 1-5 | sentence pair |

Table 1: Available annotations in Sparv-Superlim at the time of writing.

Once the configuration is defined, running Sparv on the corpora is as simple as:

```
sparv run
```

After a successful run, Sparv saves the output files in an `export` folder where the annotations are formatted in `.xml` files. The folder should then look like this:

```
valmanifest/
├── config.yaml
└── xml_export.pretty
    ├── energy
    │   ├── fp-2014-v_export.xml
    │   ├── l-2018-v_export.xml
    │   ├── l-2022-v_export.xml
    │   ├── mp-2014-v_export.xml
    │   ├── mp-2018-v_export.xml
    │   └── mp-2022-v_export.xml
    └── migration
        ├── all-2014-v-migration_export.xml
        ├── m-2018-v-migration_export.xml
        ├── m-2022-v-migration_export.xml
        ├── sd-2014-v-migration_export.xml
        ├── sd-2018-v-migration_export.xml
        └── sd-2022-v-migration_export.xml
├── source
    ├── energy
    │   ├── fp-2014-v.txt
    │   ├── l-2018-v.txt
    │   ├── l-2022-v.txt
    │   ├── mp-2014-v.txt
    │   ├── mp-2018-v.txt
    │   └── mp-2022-v.txt
    └── migration
        ├── all-2014-v-migration.txt
        ├── m-2018-v-migration.txt
        ├── m-2022-v-migration.txt
        ├── sd-2014-v-migration.txt
        ├── sd-2018-v-migration.txt
        └── sd-2022-v-migration.txt
```

The files and annotations mentioned in this section are the ones used in the next section. These examples can be found on the same project page on GitHub.[4]

# 4 Analyzing Swedish politics using Sparv-Superlim

We will now see how the tool can be used in two different use cases. In these we analyze the political positions of different Swedish political parties using

[4] https://github.com/spraakbanken/sparv-sbx-superlim/examples

Swedish election manifestos (Språkbanken Text, 2024). These corpora are not in the Superlim datasets,[5] which allows us to assess the broader generalization capabilities of the reference models.

## Use case 1: The Moderates and change in stance on migration

In the last decade, Swedish politics has become more critical towards immigration, especially asylum-related migration. Many parties have advocated more restrictive migration policies, including parties that have previously shown a positive attitude towards migration. A notable example of this is the Moderate Party (Moderaterna), whose then party leader, Fredrik Reinfeldt, famously said "open up your hearts" to refugees before the 2014 election. Today, however, the Moderate Party (which at the time of writing is in government) wants to reduce refugee migration to as low as the EU minimum[6] and has boasted having the lowest number of asylum seekers in decades.[7]

To see if this shift in stance towards migration can be identified using the Sparv-Superlim plugin, we analyze the `migration_stance` annotations. We limit the analysis to sections dealing with migration specifically of the Moderate election manifestos from the national elections in 2014, 2018 and 2022. As a sanity check, we also include the Sweden Democrats (Sverigedemokraterna), who have always been strongly against immigration. For these annotations, we let Sparv-Superlim use the `KB/bert-base-swedish-cased` model, which got 0.529 Krippendorff's alpha on absabank-imm, the highest of all reference models.

Table 2 shows the average score and scoring distribution. While not a big difference, there is a slight change in score of -0.013 difference from 2014 to 2018 and -0.49 difference from 2014 to 2022 in the Moderate Party. Although not as pronounced, a similar trend can be seen in the Sweden Democrats, which has always been anti-immigration.

Perhaps more indicative of the broader political change is seen in the number of positive or very positive scores between 4 and 5. The Moderates in 2014 have the highest number of these and as seen from the examples in Table 3, these indeed are very positive towards migration. In 2022, however, there is only one score above 4.

## Use case 2: The stance on nuclear power of the Greens and the Liberals

Nuclear energy has been a contentious topic in Swedish politics, especially in the last few years as energy prices have soared. Two smaller parties have taken very different views when it comes this issue: The Green Party (Miljöpartiet),

[5] The LLMs might, however, have seen this corpus during pretraining.

[6] https://moderaterna.se/var-politik/migrationspolitik/

[7] https://www.government.se/press-releases/2024/08/sweden-has-more-emigrants-than-immigrants-for-the-first-time-in-half-a-century/

| | | | Score distribution | | | |
|---|---|---|---|---|---|---|
| Corpus | Year | Avg. score | 1-2 | 2-3 | 3-4 | 4-5 |
| M | 2014 | 3.4 | 0 (0.0) | 16 (20.51) | 52 (66.67) | 10 (12.82) |
| M | 2018 | 3.27 | 0 (0.0) | 19 (21.11) | 65 (72.22) | 6 (6.67) |
| M | 2022 | 2.91 | 1 (2.7) | 22 (59.46) | 13 (35.14) | 1 (2.7) |
| SD | 2014 | 3.26 | 0 (0.0) | 5 (35.71) | 8 (57.14) | 1 (7.14) |
| SD | 2018 | 3.16 | 0 (0.0) | 8 (34.78) | 15 (65.22) | 0 (0.0) |
| SD | 2022 | 3 | 0 (0.0) | 14 (53.85) | 12 (46.15) | 0 (0.0) |

Table 2: Summary of `migration_stance` annotations on selected section of the election manifesto corpora of the Moderate Party (M) and Sweden Democrats (SD). Score distribution shows the distribution of scores in frequency and in percentage % in parentheses. Note, though not indicated in the table, the Moderate Party wrote a joint manifesto with three other coalition parties called the Alliance (Alliansen) in 2014.

who historically have been against the reliance on as well as the expansion of nuclear energy, and the Liberals (Liberalerna) who have wanted to expand it.

To see if these positions can be identified by Sparv-Superlim, we use the `nuclear_stance` annotator to classify sentences in the respective party manifestos of the Green Party and the Liberals from the national elections in 2014, 2018 and 2022. Unlike use case 1, these manifestos do not have have specific sections on nuclear energy, so we apply it instead to the whole corpus. We let Sparv again use the `KB/bert-based-swedish-cased` for classification, which got a Krippendorff's alpha score of 0.555 on the argumentation_sentences task, the third highest score of all the reference models in Superlim.[8]

Table 4 shows the results. Unlike use case 1, the summary of the results are opposite to the expected results one would expect given the parties well known positions. Although both parties both have a higher proportion of `Pro` labels, the Liberals have an even higher proportion of `Con` labels than the Green Party, which on the contrary has the higher proportions of `Pro` labels. Furthermore, considering that the proportion of `Non` labels are close to 50% in all manifestos and that nuclear energy is not the most frequent topic discussed in these corpora, there seems to be in general too many `Pro` and `Con` false positives.

Table 5 further illustrates this problem with false positives. Here, five selected examples are picked with a `Con` label from Liberals manifesto in year 2014. As we can see, none of these are about the topic of nuclear energy and should therefore have been classified as `Non`. Unlike the task used for use case 1 (absabank-imm), the task for argumentation_sentences is to predict the right stance given a certain topic. It might be the case that the model trained on argumentation_sentences ignores the topic and just tries to classify a general

[8]Because of reasons relating to model size and lack of access to GPUs, we did not use the best scoring model `KBLab/megatron-bert-large-swedish-cased-165k`, which got 0.628 Krippendorff's alpha.

| Sentence | migration_stance |
|---|---|
| *Vi lever i en orolig tid, där människor tvingas fly för sina liv från krig och förtryck.* ("We live in troubled times, where people are forced to flee for their lives from war and oppression.") | 5.06 |
| *Istället för att sluta oss mot omvärlden har vi sagt att Sverige ska föra en human asylpolitik och vara en fristad för dem som flyr undan förföljelse och förtryck.* ("Instead of closing ourselves off to the outside world, we have said that Sweden should pursue a humane asylum policy and be a haven for those fleeing persecution and oppression.") | 4.44 |
| *Vi har visat att det är möjligt att sätta medmänskligheten i första hand och att öppna dörren för dem som behöver skydd.* ("We have shown that it is possible to put humanity first and to open the door to those who need protection.") | 4.16 |
| *På samma sätt har de som kommer till Sverige idag kunskaper och erfarenheter som är värdefulla för oss.* ("Similarly, those who come to Sweden today have knowledge and experience that are valuable to us") | 4.00 |
| *Vi är förberedda för att kunna ge de människor som kommer idag samma möjlighet att bidra med arbete och nyföretagande.* ("We are prepared to give the people who arrive today the same opportunity to contribute with work and new businesses.") | 4.23 |

Table 3: Five examples from sentences the Alliance manifesto in 2014 (which the Moderates co-wrote), where the score is equal to or above 4 (i.e. positive).

| Label distribution | | | | |
|---|---|---|---|---|
| Party | Year | Con | Non | Pro |
| L | 2014 | 62 (14.94) | 178 (42.89) | 175 (42.17) |
| | 2018 | 83 (13.07) | 311 (48.98) | 241 (37.95) |
| | 2022 | 63 (11.8) | 287 (53.75) | 184 (34.46) |
| MP | 2014 | 38 (8.28) | 244 (53.16) | 177 (38.56) |
| | 2018 | 21 (4.56) | 226 (49.02) | 214 (46.42) |
| | 2022 | 19 (7.6) | 134 (53.6) | 97 (38.8) |

Table 4: Summary of `nuclear_stance` annotations on the election manifesto corpora of the Liberals (L) and the Green Party (MP). Avg. score is the mean of all scores across all annotations. Label distribution shows the distribution of labels in frequency and in percentage % in parentheses. Note, though not shown in the table, the Liberals were called "Folkpartiet" in 2014.

"sentiment". This is only preliminary speculation, however, and would have to be investigated further in future work.

## 4.1 Discussion

For both use case 1 and use case 2, we have defined a non-formal objective to measure the external validity of the models, where we rely on real-world knowledge about the positions of political parties and what is written in their political manifestos. Intuitively, the predictions of these models should correspond to this knowledge. That is, the average `migration_stance` score should be lower for the manifesto of the Moderates in 2014 than in 2018-2022 and there should be a higher proportion of `Pro` labels in support of nuclear energy in the Liberal party manifestos than in those of the Green Party. For use case 1, we saw that the scores corresponded to the last years' changes regarding the Moderate party's stance on migration, while for use case 2 the model's predictions were contrary to what we know about the political positions on nuclear energy of the Liberals and the Green Party.

As with any corpus statistic, the proportion of sentences of specific stance is not a perfect metric for the complete political stance of a text: Many times, sentences form a compositional meaning or stance (i.e. against migration) which cannot be captured by the sentences without context and some meanings can carry more weight on the whole than others. Nonetheless, the proportion of one or the other stances could be an indication of overall sentiment, just like a certain number of words in a text can be an indication of topics and sentiments in a text.

For use case 1, we have seen evidence that `migration_stance` can be useful for automatically identifying sentences with positive or negative sentiments towards migration. This can be useful to find and analyze migration sentiments in a text with a corpus explorer tool like Språkbanken's word platform Korp (Borin et al., 2012) or a document analysis tool like Språkbanken's data platform Mink.

| Sentence | `nuclear_stance` |
|---|---|
| *Situationen i Mellanöstern skapar ett stort humanitärt lidande.* ("The situation in the Middle East is causing great humanitarian suffering.") | con |
| *Bristande tillgänglighet ses numera som diskriminering.* ("Lack of access is now seen as discrimination.") | con |
| *Det är lätt att glömma hur det var förut.* ("It's easy to forget how it used to be.") | con |
| *Då var det den mobbade – inte mobbaren – som fick flytta och byta skola.* ("Back then, it was the bullied - not the bully - who had to move and change schools.") | con |
| *Det vi uppnått genom ansvarstagande för ekonomin och arbetslinjen kan snabbt förskingras.* ("What we have achieved by taking responsibility for the economy and the work line can quickly be squandered.") | con |

Table 5: Five selected examples from sentences the in the Liberals' manifesto in 2014. These are falsely classified as `Con`.

[9]

Although these use cases are not formal evaluations, they nonetheless ground the models in the real world and give further insights into the models' performance and what they learn from the data they have been trained on. This in turn provides an external validation to the Superlim benchmark. The two uses cases show how additional corpus analysis can provide important, additional context to NLU benchmarks and the models that they are supposed to evaluate. Since these predictions are available in Sparv with the Sparv-Superlim plugin, such analyses can easily be complemented with other features available in the Sparv Pipeline.

# 5 Conclusions

In this technical report, I have presented Sparv-Superlim, a Sparv plugin which uses the reference models and benchmarks from Superlim to provide corpus analysis for Swedish. So far, the plugin supports annotations from five different Superlim tasks, resulting in ten different sentence-level annotations (six annotations per topic in argumentation-sentences and four for the other available Superlim tasks, see Table 1). Further additions to the plugin based on the Superlim reference models are encouraged from the broader community.

Two available annotations have been showcased in this this technical report:

[9] https://spraakbanken.gu.se/mink/

`migration_stance` (from absabank-imm) and `nuclear_stance` (from argumentation_sentences). I have shown how to use them in Sparv to analyze Swedish election manifestos. In these cases, I have shown that both these annotations are not perfect, but `migration_stance` seem better for downstream analysis than `nuclear_stance`. Integrating a multi-task benchmark like Superlim into Sparv makes it possible to analyze the predictions of models trained on the benchmark on novel data and as such can serve as an external evaluation of its ecological validity.

# References


Hammarstedt, M., Schumacher, A., Borin, L., & Forsberg, M. (2022). *Sparv 5 user manual* (tech. rep.). Göteborg, Institutionen för svenska, flerspråkighet och språkteknologi.

Berdicevskis, A., Bouma, G., Kurtz, R., Morger, F., Öhman, J., Adesam, Y., Borin, L., Dannélls, D., Forsberg, M., Isbister, T., Lindahl, A., Malmsten, M., Rekathati, F., Sahlgren, M., Volodina, E., Börjeson, L., Hengchen, S., & Tahmasebi, N. Superlim: A Swedish language understanding evaluation benchmark (H. Bouamor, J. Pino, & K. Bali, Eds.). In: *Proceedings of the 2023 conference on empirical methods in natural language processing* (H. Bouamor, J. Pino, & K. Bali, Eds.). Ed. by Bouamor, H., Pino, J., & Bali, K. Singapore: Association for Computational Linguistics, 2023, December, 8137–8153. https://doi.org/10.18653/v1/2023.emnlp-main.506

Andrade, C. (2018). Internal, external, and ecological validity in research design, conduct, and evaluation. *Indian journal of psychological medicine*, *40*(5), 498–499.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding (J. Burstein, C. Doran, & T. Solorio, Eds.). In: *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (J. Burstein, C. Doran, & T. Solorio, Eds.). Ed. by Burstein, J., Doran, C., & Solorio, T. Minneapolis, Minnesota: Association for Computational Linguistics, 2019, June, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Malmsten, M., Börjeson, L., & Haffenden, C. (2020). Playing with words at the national library of sweden – making a swedish bert.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. (2018). Improving language understanding by generative pre-training.

Ekgren, A., Gyllensten, A. C., Stollenwerk, F., Öhman, J., Isbister, T., Gogoulou, E., Carlsson, F., Heiman, A., Casademont, J., & Sahlgren, M. (2023). Gpt-sw3: An autoregressive language model for the nordic languages. https://arxiv.org/abs/2305.12987

Språkbanken Text. (2024). Svenska partiprogram och valmanifest. https://doi.org/10.23695/NC55-GD27

Borin, L., Forsberg, M., & Roxendal, J. Korp — the corpus infrastructure of språkbanken (N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis, Eds.). In: *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)* (N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis, Eds.). Ed. by Calzolari, N., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., & Piperidis, S. Istanbul, Turkey: European Language Resources Association (ELRA),

2012, May, 474–478. http://www.lrec-conf.org/proceedings/lrec2012/pdf/248_Paper.pdf