

CONTENTS

1	Introduction	2
2	Quick Settings	3
3	Result View	3
4	Analysis Modes	4
5	Advanced Settings	4
6	File Upload	6

1 INTRODUCTION

Sparv is an annotation tool developed by Språkbanken that is used for analysing the corpora in Korp and the texts in Strix. The Sparv web interface¹ can be used for annotating your own texts.

This user manual describes most of the functions available in the Sparv web interface. There are also exercises in Swedish available for download².

To make a Sparv analysis you can just type or paste some text into the text field (or hit one of the example buttons) and then press the green **Run** button. When the analysis is done the result is shown as a table below the text input field.

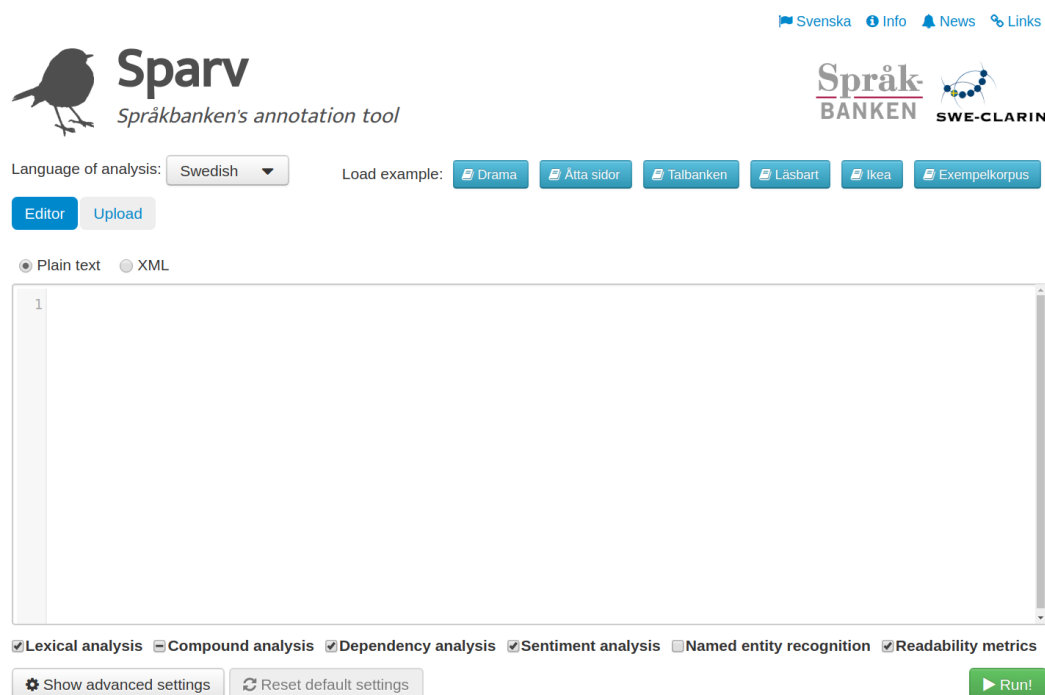


Figure 1: The Sparv web interface

¹<https://spraakbanken.gu.se/sparv>

²https://github.com/spraakbanken/sparv-docs/raw/master/exercises/sparvovningar_hw2017.pdf

2 QUICK SETTINGS

The check boxes under the text input field can be used to quickly select or deselect different types of annotations. For example, if you would like to exclude the dependency trees from the result you can deselect the dependency analysis check box.

Above the text field you can choose between plain text or XML input. If your text already has some mark-up you need to select XML, otherwise all existing annotations will be lost after the analysis.

3 RESULT VIEW

When the analysis is done the result is shown as a table below the text input field. Each column represents one type of analysis on word level. When hovering over the names in the table header a short explanation is displayed. The same goes for many abbreviations used in the analysis, e.g. the labels in the dependency analysis, the named entity categories and the morphosyntactic tags.

The dependency analysis is displayed as a dependency tree in the beginning of every sentence in addition to the dependency relations being shown in the result table. Structural attributes, i.e. analyses that can stretch across several words or the entire text (e.g. readability metrics or name tags) are shown as XML tags in the table.

By pressing the **XML** button above the result table you can download the analysis in XML format.

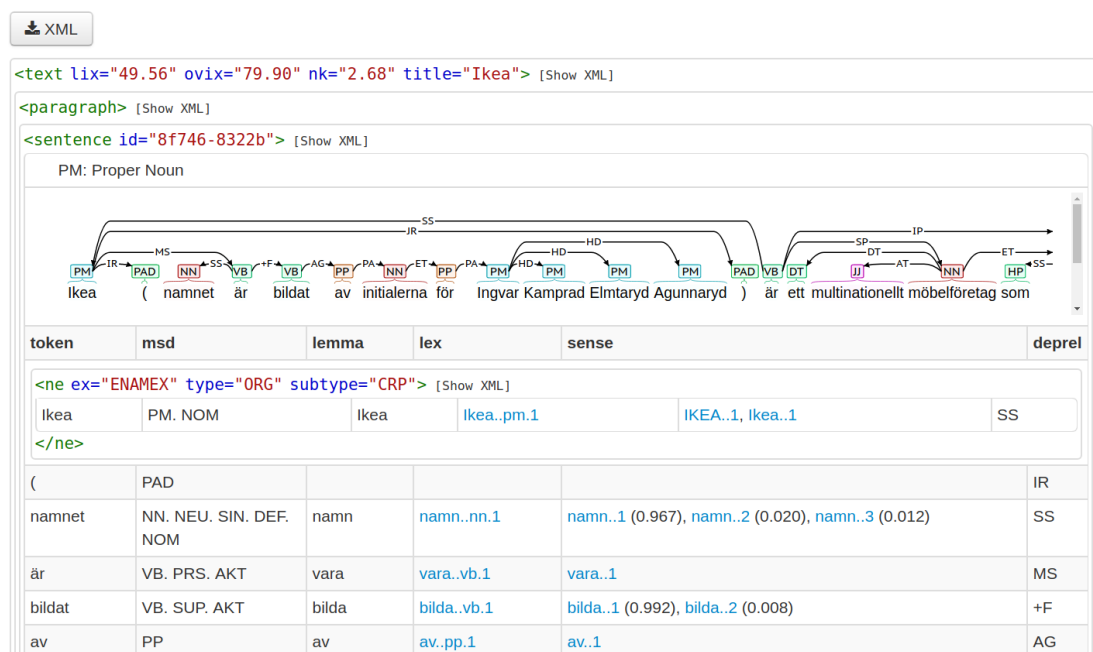


Figure 2: A completed Sparv analysis

4 ANALYSIS MODES

Below the Sparv logo you can find a dropdown menu for picking the language of analysis. This menu lets you analyse texts written in other languages than Swedish. Currently Sparv supports 20 different languages but most of them only support part-of-speech tagging and lemmatisation.

Sparv has an analysis mode for annotating Swedish texts from the 1800's. In this mode information from two older dictionaries (Dalin and Swedberg) is used in order to obtain better analyses for words with older spelling.

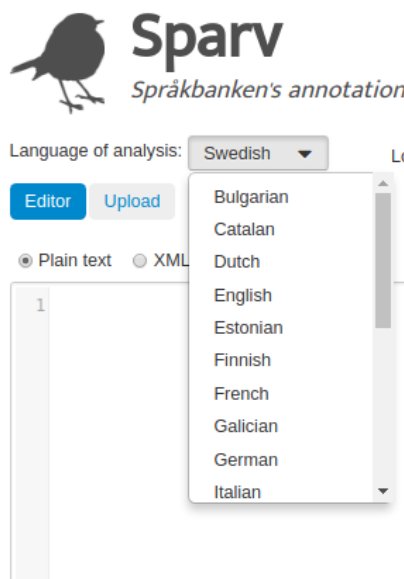


Figure 3: The language picker

5 ADVANCED SETTINGS

When analysing texts without any mark-up it is often enough to paste the text in the input field and hit the **Run** button. If the text contains XML you need to adjust the settings in order to let Sparv know what the different tags in your data represent.

You can open the settings by clicking on **Show advanced settings**. Then you can specify if the segmentation into paragraphs, sentences, and words should be done automatically or if existing tags should be used instead. You can also specify the name of the document element and enter additional tags and attributes that are available in your input data. Any tags and attributes not specified in the settings will be removed from the result.

The advanced settings let you control exactly which analyses should be included in the result. You can select and deselect analyses by clicking on the different attributes. E.g. if you are only interested in the readability metric **lix** you can deselect **ovix** and **nk** in the advanced settings. If you want to know more about a setting you can click on the question mark beside it and a short explanation will be displayed. The reset button allows you to return to the default settings.

ⓘ Corpus name:

ⓘ Word segmentation

ⓘ Sentence segmentation

ⓘ Sentence chunk

ⓘ Segment by

ⓘ Paragraph segmentation

ⓘ Paragraph chunk

ⓘ Segment by

ⓘ Document element ⓘ Tag:

ⓘ Structural attributes

ⓘ Extra structural elements

ⓘ Positional attributes ⓘ Lexical analysis

ⓘ Compound analysis

ⓘ Dependency analysis

ⓘ Sentiment analysis

ⓘ Named entity recognition

ⓘ Text attributes ⓘ Readability metrics

☒ Show XML tags

Figure 4: The advanced settings in XML mode

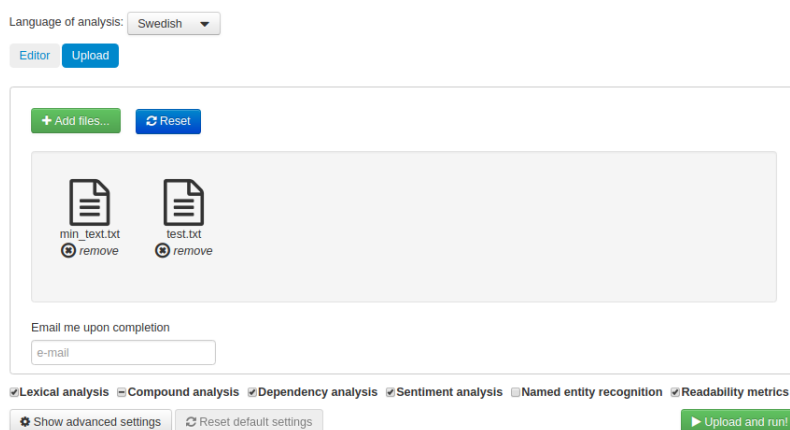
6 FILE UPLOAD

Instead of feeding text directly into the text field you can use the file upload mode by pressing the **Upload** button. Valid input formats are txt (plain text) and XML. Documents created with Microsoft Word or similar text processing software cannot be analysed.

If you choose to upload XML files you will need to adjust the settings to keep Sparv from interpreting your mark-up as text. You also need to adjust the language of analysis if you upload text which is written in another language than Swedish. Sparv allows for uploading multiple files at a time. The settings and choice of language will be applied to all files analysed in the same run.

Some analyses take quite some time to process and if you don't want to wait that long you can fill in your e-mail address and you will get an e-mail with a download link as soon as the analysis is completed. If you have supplied your e-mail address it is okay to close your browser while your text is being processed.

In file upload mode the result is not shown as a table but instead you will be able to download a zip file containing the annotated text in XML format.



The screenshot shows the Sparv file upload interface. At the top, there is a dropdown menu for 'Language of analysis' set to 'Swedish'. Below it are 'Editor' and 'Upload' buttons. The main area contains a '+ Add files...' button and a 'Reset' button. Below these, two files are listed: 'min_text.txt' and 'test.txt', each with a 'remove' button. A text input field for 'Email me upon completion' is present. At the bottom, there are checkboxes for various analysis options: 'Lexical analysis' (checked), 'Compound analysis' (unchecked), 'Dependency analysis' (checked), 'Sentiment analysis' (checked), 'Named entity recognition' (unchecked), and 'Readability metrics' (checked). There are also buttons for 'Show advanced settings', 'Reset default settings', and a green 'Upload and run!' button.

Figure 5: File upload mode