



UNIVERSITY OF GOTHENBURG
DEPT OF SWEDISH

CLT

Språk
BANKEN



Swedish multiword expressions and sublanguage parsing

DIMITRIOS KOKKINAKIS

Dep. of Swedish, the Swedish Language Bank &
the Centre for Language Technology (CLT)

University of Gothenburg

Sweden

dimitrios.kokkinakis@svenska.gu.se





OVERVIEW

- Types of MWE
- Textual data
- Related work
- Parsing method
- MWE knowledge intergation
- Goal: event extraction
- Results and future work

Swedish multiword expressions and sublanguage parsing – PARSEME WG2

Dimitrios Kokkinakis
Centre for Language Technology & Språkbanken
University of Gothenburg
Box 200, SE 405 30 Göteborg

ABSTRACT

Motivation: multilayer, linguistic annotation of textual corpora (in specialized fields) is important for empirically-based, data-driven language processing and human language technologies. One such annotation is at the syntactic and functional level (parsing) which, apart from the challenge it poses, is often required as a supporting technology in e.g. information extraction/retrieval, terminology management, semantic search and knowledge acquisition.

Goal: in sublanguages, where specialized processing tools have to be adapted which is can be expensive for many applications and/or languages, means for reducing parsing complexity needs to be explored. Here, means of reducing parsing complexity in Swedish (scientific) medical texts.

MWE: used as a pre-processing step for reducing complexity and ambiguity at the various levels of the linguistic processing.

Sublanguage Parsing: using finite-state cascades, constituent based shallow parsing is applied on large samples of Swedish medical texts. The parser is made aware of both the shallow semantic annotations from various domain-specific medical resources, and other multiword expressions (named entity recognition ; compound function words). This results into the effect of slightly increasing the number of rules, but decreasing their complexity.

Evaluation: based on the extraction of the syntactic relations Subject and Object. We manually evaluated and compared the results with and without the MWE identification on a random sample of 200 lengthy sentences, using the syntactic relations as a benchmark in Subject-Verb-Object-triplets. Results show that pre-processing improves accuracy for these relations (18%), with a total PR for Subject 54.3%/66.9% and for Object 88.4%/92.7%.

Results: parsing performance can be improved by applying a number of pre-processing steps based on MWE. MWE have positive impact on syntactic analysis and should be identified prior to parsing. However, there are several types of MWE and several resources that can be safely used in order to achieve better parsing performance, independent of the application in mind, or textual characteristics.

Future: increasing use of MWE, also with the addition of "new" types of suitable resources (at least for Swedish), particularly a construction (a DB of Swedish constructions under development). These constructions are to quite a large extent of restricted, semi-productive and tend to be highly problematic for language technology, and also seem to be quite common, often neglected in both grammars and dictionaries, which focus on general rules and individual words, respectively. Hence, a comprehensive account of construction-specific patterns is so far lacking. We therefore started to develop such a resource, aiming to make it descriptively adequate, simple enough for large-scale coverage, and formalized to enable computational uses.

<http://spraakbanken.gu.se/english/wg2>

PLETHORA OF MWEs

Figure from Boldea G. & Even S.: "Multiword Expressions: A pain in the neck of lexical semantics"

PARSING AND MWE

```

[00] svensk-nyheter | 00-nyheter.nor |
[01] [00] svensk-nyheter | 00-nyheter.nor |
[02] [00] svensk-nyheter | 00-nyheter.nor |
[03] [00] svensk-nyheter | 00-nyheter.nor |
[04] [00] svensk-nyheter | 00-nyheter.nor |
[05] [00] svensk-nyheter | 00-nyheter.nor |
[06] [00] svensk-nyheter | 00-nyheter.nor |
[07] [00] svensk-nyheter | 00-nyheter.nor |
[08] [00] svensk-nyheter | 00-nyheter.nor |
[09] [00] svensk-nyheter | 00-nyheter.nor |
[10] [00] svensk-nyheter | 00-nyheter.nor |
[11] [00] svensk-nyheter | 00-nyheter.nor |
[12] [00] svensk-nyheter | 00-nyheter.nor |
[13] [00] svensk-nyheter | 00-nyheter.nor |
[14] [00] svensk-nyheter | 00-nyheter.nor |
[15] [00] svensk-nyheter | 00-nyheter.nor |
[16] [00] svensk-nyheter | 00-nyheter.nor |
[17] [00] svensk-nyheter | 00-nyheter.nor |
[18] [00] svensk-nyheter | 00-nyheter.nor |
[19] [00] svensk-nyheter | 00-nyheter.nor |
[20] [00] svensk-nyheter | 00-nyheter.nor |
[21] [00] svensk-nyheter | 00-nyheter.nor |
[22] [00] svensk-nyheter | 00-nyheter.nor |
[23] [00] svensk-nyheter | 00-nyheter.nor |
[24] [00] svensk-nyheter | 00-nyheter.nor |
[25] [00] svensk-nyheter | 00-nyheter.nor |
[26] [00] svensk-nyheter | 00-nyheter.nor |
[27] [00] svensk-nyheter | 00-nyheter.nor |
[28] [00] svensk-nyheter | 00-nyheter.nor |
[29] [00] svensk-nyheter | 00-nyheter.nor |
[30] [00] svensk-nyheter | 00-nyheter.nor |
[31] [00] svensk-nyheter | 00-nyheter.nor |
[32] [00] svensk-nyheter | 00-nyheter.nor |
[33] [00] svensk-nyheter | 00-nyheter.nor |
[34] [00] svensk-nyheter | 00-nyheter.nor |
[35] [00] svensk-nyheter | 00-nyheter.nor |
[36] [00] svensk-nyheter | 00-nyheter.nor |
[37] [00] svensk-nyheter | 00-nyheter.nor |
[38] [00] svensk-nyheter | 00-nyheter.nor |
[39] [00] svensk-nyheter | 00-nyheter.nor |
[40] [00] svensk-nyheter | 00-nyheter.nor |
[41] [00] svensk-nyheter | 00-nyheter.nor |
[42] [00] svensk-nyheter | 00-nyheter.nor |
[43] [00] svensk-nyheter | 00-nyheter.nor |
[44] [00] svensk-nyheter | 00-nyheter.nor |
[45] [00] svensk-nyheter | 00-nyheter.nor |
[46] [00] svensk-nyheter | 00-nyheter.nor |
[47] [00] svensk-nyheter | 00-nyheter.nor |
[48] [00] svensk-nyheter | 00-nyheter.nor |
[49] [00] svensk-nyheter | 00-nyheter.nor |
[50] [00] svensk-nyheter | 00-nyheter.nor |
[51] [00] svensk-nyheter | 00-nyheter.nor |
[52] [00] svensk-nyheter | 00-nyheter.nor |
[53] [00] svensk-nyheter | 00-nyheter.nor |
[54] [00] svensk-nyheter | 00-nyheter.nor |
[55] [00] svensk-nyheter | 00-nyheter.nor |
[56] [00] svensk-nyheter | 00-nyheter.nor |
[57] [00] svensk-nyheter | 00-nyheter.nor |
[58] [00] svensk-nyheter | 00-nyheter.nor |
[59] [00] svensk-nyheter | 00-nyheter.nor |
[60] [00] svensk-nyheter | 00-nyheter.nor |
[61] [00] svensk-nyheter | 00-nyheter.nor |
[62] [00] svensk-nyheter | 00-nyheter.nor |
[63] [00] svensk-nyheter | 00-nyheter.nor |
[64] [00] svensk-nyheter | 00-nyheter.nor |
[65] [00] svensk-nyheter | 00-nyheter.nor |
[66] [00] svensk-nyheter | 00-nyheter.nor |
[67] [00] svensk-nyheter | 00-nyheter.nor |
[68] [00] svensk-nyheter | 00-nyheter.nor |
[69] [00] svensk-nyheter | 00-nyheter.nor |
[70] [00] svensk-nyheter | 00-nyheter.nor |
[71] [00] svensk-nyheter | 00-nyheter.nor |
[72] [00] svensk-nyheter | 00-nyheter.nor |
[73] [00] svensk-nyheter | 00-nyheter.nor |
[74] [00] svensk-nyheter | 00-nyheter.nor |
[75] [00] svensk-nyheter | 00-nyheter.nor |
[76] [00] svensk-nyheter | 00-nyheter.nor |
[77] [00] svensk-nyheter | 00-nyheter.nor |
[78] [00] svensk-nyheter | 00-nyheter.nor |
[79] [00] svensk-nyheter | 00-nyheter.nor |
[80] [00] svensk-nyheter | 00-nyheter.nor |
[81] [00] svensk-nyheter | 00-nyheter.nor |
[82] [00] svensk-nyheter | 00-nyheter.nor |
[83] [00] svensk-nyheter | 00-nyheter.nor |
[84] [00] svensk-nyheter | 00-nyheter.nor |
[85] [00] svensk-nyheter | 00-nyheter.nor |
[86] [00] svensk-nyheter | 00-nyheter.nor |
[87] [00] svensk-nyheter | 00-nyheter.nor |
[88] [00] svensk-nyheter | 00-nyheter.nor |
[89] [00] svensk-nyheter | 00-nyheter.nor |
[90] [00] svensk-nyheter | 00-nyheter.nor |
[91] [00] svensk-nyheter | 00-nyheter.nor |
[92] [00] svensk-nyheter | 00-nyheter.nor |
[93] [00] svensk-nyheter | 00-nyheter.nor |
[94] [00] svensk-nyheter | 00-nyheter.nor |
[95] [00] svensk-nyheter | 00-nyheter.nor |
[96] [00] svensk-nyheter | 00-nyheter.nor |
[97] [00] svensk-nyheter | 00-nyheter.nor |
[98] [00] svensk-nyheter | 00-nyheter.nor |
[99] [00] svensk-nyheter | 00-nyheter.nor |
[100] [00] svensk-nyheter | 00-nyheter.nor |

```

```

[00] svensk-nyheter | 00-nyheter.nor |
[01] [00] svensk-nyheter | 00-nyheter.nor |
[02] [00] svensk-nyheter | 00-nyheter.nor |
[03] [00] svensk-nyheter | 00-nyheter.nor |
[04] [00] svensk-nyheter | 00-nyheter.nor |
[05] [00] svensk-nyheter | 00-nyheter.nor |
[06] [00] svensk-nyheter | 00-nyheter.nor |
[07] [00] svensk-nyheter | 00-nyheter.nor |
[08] [00] svensk-nyheter | 00-nyheter.nor |
[09] [00] svensk-nyheter | 00-nyheter.nor |
[10] [00] svensk-nyheter | 00-nyheter.nor |
[11] [00] svensk-nyheter | 00-nyheter.nor |
[12] [00] svensk-nyheter | 00-nyheter.nor |
[13] [00] svensk-nyheter | 00-nyheter.nor |
[14] [00] svensk-nyheter | 00-nyheter.nor |
[15] [00] svensk-nyheter | 00-nyheter.nor |
[16] [00] svensk-nyheter | 00-nyheter.nor |
[17] [00] svensk-nyheter | 00-nyheter.nor |
[18] [00] svensk-nyheter | 00-nyheter.nor |
[19] [00] svensk-nyheter | 00-nyheter.nor |
[20] [00] svensk-nyheter | 00-nyheter.nor |
[21] [00] svensk-nyheter | 00-nyheter.nor |
[22] [00] svensk-nyheter | 00-nyheter.nor |
[23] [00] svensk-nyheter | 00-nyheter.nor |
[24] [00] svensk-nyheter | 00-nyheter.nor |
[25] [00] svensk-nyheter | 00-nyheter.nor |
[26] [00] svensk-nyheter | 00-nyheter.nor |
[27] [00] svensk-nyheter | 00-nyheter.nor |
[28] [00] svensk-nyheter | 00-nyheter.nor |
[29] [00] svensk-nyheter | 00-nyheter.nor |
[30] [00] svensk-nyheter | 00-nyheter.nor |
[31] [00] svensk-nyheter | 00-nyheter.nor |
[32] [00] svensk-nyheter | 00-nyheter.nor |
[33] [00] svensk-nyheter | 00-nyheter.nor |
[34] [00] svensk-nyheter | 00-nyheter.nor |
[35] [00] svensk-nyheter | 00-nyheter.nor |
[36] [00] svensk-nyheter | 00-nyheter.nor |
[37] [00] svensk-nyheter | 00-nyheter.nor |
[38] [00] svensk-nyheter | 00-nyheter.nor |
[39] [00] svensk-nyheter | 00-nyheter.nor |
[40] [00] svensk-nyheter | 00-nyheter.nor |
[41] [00] svensk-nyheter | 00-nyheter.nor |
[42] [00] svensk-nyheter | 00-nyheter.nor |
[43] [00] svensk-nyheter | 00-nyheter.nor |
[44] [00] svensk-nyheter | 00-nyheter.nor |
[45] [00] svensk-nyheter | 00-nyheter.nor |
[46] [00] svensk-nyheter | 00-nyheter.nor |
[47] [00] svensk-nyheter | 00-nyheter.nor |
[48] [00] svensk-nyheter | 00-nyheter.nor |
[49] [00] svensk-nyheter | 00-nyheter.nor |
[50] [00] svensk-nyheter | 00-nyheter.nor |
[51] [00] svensk-nyheter | 00-nyheter.nor |
[52] [00] svensk-nyheter | 00-nyheter.nor |
[53] [00] svensk-nyheter | 00-nyheter.nor |
[54] [00] svensk-nyheter | 00-nyheter.nor |
[55] [00] svensk-nyheter | 00-nyheter.nor |
[56] [00] svensk-nyheter | 00-nyheter.nor |
[57] [00] svensk-nyheter | 00-nyheter.nor |
[58] [00] svensk-nyheter | 00-nyheter.nor |
[59] [00] svensk-nyheter | 00-nyheter.nor |
[60] [00] svensk-nyheter | 00-nyheter.nor |
[61] [00] svensk-nyheter | 00-nyheter.nor |
[62] [00] svensk-nyheter | 00-nyheter.nor |
[63] [00] svensk-nyheter | 00-nyheter.nor |
[64] [00] svensk-nyheter | 00-nyheter.nor |
[65] [00] svensk-nyheter | 00-nyheter.nor |
[66] [00] svensk-nyheter | 00-nyheter.nor |
[67] [00] svensk-nyheter | 00-nyheter.nor |
[68] [00] svensk-nyheter | 00-nyheter.nor |
[69] [00] svensk-nyheter | 00-nyheter.nor |
[70] [00] svensk-nyheter | 00-nyheter.nor |
[71] [00] svensk-nyheter | 00-nyheter.nor |
[72] [00] svensk-nyheter | 00-nyheter.nor |
[73] [00] svensk-nyheter | 00-nyheter.nor |
[74] [00] svensk-nyheter | 00-nyheter.nor |
[75] [00] svensk-nyheter | 00-nyheter.nor |
[76] [00] svensk-nyheter | 00-nyheter.nor |
[77] [00] svensk-nyheter | 00-nyheter.nor |
[78] [00] svensk-nyheter | 00-nyheter.nor |
[79] [00] svensk-nyheter | 00-nyheter.nor |
[80] [00] svensk-nyheter | 00-nyheter.nor |
[81] [00] svensk-nyheter | 00-nyheter.nor |
[82] [00] svensk-nyheter | 00-nyheter.nor |
[83] [00] svensk-nyheter | 00-nyheter.nor |
[84] [00] svensk-nyheter | 00-nyheter.nor |
[85] [00] svensk-nyheter | 00-nyheter.nor |
[86] [00] svensk-nyheter | 00-nyheter.nor |
[87] [00] svensk-nyheter | 00-nyheter.nor |
[88] [00] svensk-nyheter | 00-nyheter.nor |
[89] [00] svensk-nyheter | 00-nyheter.nor |
[90] [00] svensk-nyheter | 00-nyheter.nor |
[91] [00] svensk-nyheter | 00-nyheter.nor |
[92] [00] svensk-nyheter | 00-nyheter.nor |
[93] [00] svensk-nyheter | 00-nyheter.nor |
[94] [00] svensk-nyheter | 00-nyheter.nor |
[95] [00] svensk-nyheter | 00-nyheter.nor |
[96] [00] svensk-nyheter | 00-nyheter.nor |
[97] [00] svensk-nyheter | 00-nyheter.nor |
[98] [00] svensk-nyheter | 00-nyheter.nor |
[99] [00] svensk-nyheter | 00-nyheter.nor |
[100] [00] svensk-nyheter | 00-nyheter.nor |

```

GOAL: SRL and EVENT EXTRACTION

The 2nd General Meeting of PARSEME (WG-2), March 10-11, 2014, Athens