

# Automatic Variation of Swedish Text by Syntactic Fronting

Kenneth Wilhelmsson  
Department of Linguistics  
University of Gothenburg  
kw@ling.gu.se

## Abstract

Ongoing work with a prototype implementation for automatic fronting of primary (main clause) constituents in Swedish input text is described. Linguistic constraints and some technical aspects are also discussed.

## 1 Introduction

Automatic variation of Swedish text is a relatively unexplored area. Variation by lexical means was tested in an experiment by Rosell (2005) using *Folkets synonymlexikon* (Kann and Rosell, 2005). The program used the fact that synonymy was expressed as a matter of degree (expressed numerically), to vary a threshold value for admitting lexical substitution. The lack of cases of true lexical synonymy, however, seemed to be an important factor, as shown in the evaluation. Producing truth-preserving (*salva veritate*) paraphrases by syntactic means from textual input is a task that has been undertaken in two experimental projects. Pascoe & Ullner (2006) described the process of automatic shift of voice in sentences analyzed by *CassSwe* (Kokkinakis & Johansson Kokkinakis, 1998), producing active sentences from their passive counterparts – a transformation motivated by readability. Lindberg & Svensson (1992) earlier made use of Diderichsen’s topological clause description of Nordic languages (Diderichsen, 1946), see table 1. The work dealt with syntactic fronting using a Prolog implementation for achieving truth-preserving variants of hand-picked sentences analyzed by the *MorP Parser* (Källgren, 1992). This paper describes ongoing work with a similar

approach to that of the latter, but for free text, using the syntactic analysis described in Wilhelmsson (2008).

Fundamental field	Nexus field			Content field		
	Finite v.	Subject	Adv.	Non-finite v.	Obj/pred.	Adv.
<i>Atomstorleken</i>	<i>skulle</i>	[ - ]	<i>ju</i>	<i>peka</i>		<i>på motsatsen.</i>
* <i>Ju</i>	<i>skulle</i>	<i>atomstorleken</i>	[ - ]	<i>peka</i>		<i>på motsatsen.</i>
<i>På motsatsen</i>	<i>skulle</i>	<i>atomstorleken</i>	<i>ju</i>	<i>peka</i>		[ - ]
<i>Motsatsen</i>	<i>skulle</i>	<i>atomstorleken</i>	<i>ju</i>	<i>peka</i>		<i>på</i> [ - ]

**Table 1:** An adaptation of Diderichsen’s main clause schema showing basic Swedish declarative word order together with fronting of different positional content, including fronting of the prepositional complement *motsatsen* of the adverbial *på motsatsen*.<sup>1</sup>

## 2 Generation of Paraphrases by Fronting in Input Text

The basic procedure for fronting of any constituent in simple declarative sentences is to place a currently fronted constituent at its *canonical* (or, at least, at an *acceptable*) position according to the sentence schema, whereafter any constituent that it is possible to topicalize may be fronted. The implementation is focused on the task of immediate paraphrase generation in the act of writing to facilitate correct reformulations. It lets a user point at an unbounded full syntactic constituent in the main clause (i.e. subject, object, predicative or adverbial, thus not the fourth example in Table 1), which appear fronted. Thus,

<sup>1</sup> “Prepositional objects” are seen as a type of adverbials, in accordance with e.g. Teleman *et al* (1999).

the parsing is done in parallel with user input. The prototype implementation is made in (uncompiled) JavaScript. The inner representation is an XML-like code, like below.<sup>2</sup>

```
<subjekt>Atomstorleken</subjekt>
<pfv>skulle</pfv>
<adverbial>ju</adverbial>
<adverbial>därmed</adverbial>
<piv>peka</piv>
<adverbial>på motsatsen</adverbial>
<tom>.</tom>
```

A number of restrictions in this straightforward procedure can be noted, of which some are discussed in Lindberg & Svensson (1992).

- Particles, reflexive pronouns and some other primary constituents including a group of adverbials, like *ju* in Table 1 and back-referring expressions (“*vilket var bra*”) cannot be fronted.
- Very long constituents can be fronted, but may make sentences seem clumsy or even unnatural.
- A number of verbs will, if not forming an auxiliary verb construction, as in Table 1, result in a potential violation of the truth-preserving, through subject/object ambiguity. *Bilden föreställer tavlan* will easily introduce a different meaning of a text if transformed into *Tavlan föreställer bilden*.<sup>3</sup>

This type of transformation relies heavily on high accuracy of the syntax analysis, where *exact matching* of primary constituents (including all attributes) is necessary for grammatical output – neither more than one constituent or parts of constituents can be fronted (with a few exceptions such as prepositional complements in Table 1).

A key idea behind the parsing method used is to rely less on matching of unbounded (recursive) primary constituents (subject, object/predicative and adverbials), while identifying bounded ones (e.g., verbs, see Wilhelmsson 2008), thereby delimiting fields in the schema. This particular parsing method, and output format, seems to be

<sup>2</sup> *Pfv* and *piv* here stand for ‘primary finite verb’ and ‘primary non-finite verb’, respectively.

<sup>3</sup> Note also, that fronting of a nominal constituent, thereby producing a correct paraphrase, without having made the correct subject/object identification from the start, often is possible.

appropriate, or even necessary, for the task described. Currently, a POS tagger with an estimated accuracy of 95.3 % is used.

## References

- Paul Diderichsen, 1946. *Elementær Dansk Grammatik*. Gyldendal
- Viggo Kann and Magnus Rosell, 2005. Free Construction of a Free Swedish Dictionary of Synonyms. In Proc. 15th Nordic Conference on Computational Linguistics – (NODALIDA 05)
- Dimitrios Kokkinakis and Sofie Johansson Kokkinakis, 1998, A Cascaded Finite-State Parser for Syntactic Analysis of Swedish, Research Reports from the Department of Swedish, GU-ISS-99-2, University of Gothenburg
- Gunnel Källgren, 1992. Making maximal use of surface criteria in large-scale parsing: the MorP parser, Papers from the Institute of Linguistics, University of Stockholm, 60. Stockholm: Univ. Institute of Linguistics.
- Janne Lindberg and Carin Svensson, 1992. Topikaliserings som skrivstöd. En implementering med satschema: NADA, KTH.
- Maria Pascoe and David Ullner, 2006. VOICEover, Att automatisk aktivera en passiv sats i svenskan, (Datalingvistikprogrammet, University of Gothenburg)
- Magnus Rosell, 2005, Automatisk synonymvariering av text (Course project, Språkgranskningsverktyg, KTH)
- Ulf Telemann, Erik Andersson and Staffan Hellberg. 1999. *Svenska Akademiens grammatik*
- Kenneth Wilhelmsson, 2008. Heuristic Schema Parsing of Swedish Text, The Second Swedish Language Technology Conference (SLTC 2008), KTH, Stockholm