



DELPHIN

Swedish Grammars in LKB

Lars Ahrenberg and Sara Stymne
NLPLab
Department of Computer and Information
Science, Linköpings universitet

1

Overview

- Background
 - The Scandinavian Matrix Network
- SweCore – A Swedish Core Grammar
- More background – related work
- English-Swedish verb frame divergencies
- Expectations / Contributions (interspersed in oral presentation)

2

Scandinavian Matrix Network

- Goals

- ☐ Create a Matrix for Scandinavian languages
- ☐ Create moderately large grammars for Danish, Norwegian and Swedish
- ☐ ...

3

The SweCore Grammar

- Based on NorSource

- ☐ NorSource type and rule files with only very small modifications
- ☐ Swedish inflectional morphology and (small) lexicon

4

Norwegian-Swedish differences treated

Phenomena	Norwegian	Swedish
Gender	Fem, Masc, Neut	Common, Neut
Definiteness and abstract NP:s	<i>den oppvoksende slekt</i>	<i>det uppväxande släktet</i>
Definiteness and demonstratives	<i>denne hunden</i>	<i>denna hund</i>
Possessive attributes	<i>katten hans, hans katt</i>	<i>hans katt, *katten hans</i>
s-passives	n.a. in all tenses: <i>*katten jagets, *har jagets</i>	available for all tenses: <i>katten jagades, har jagats</i>
Negative imperatives	<i>Ikke kom for sent, Kom ikke for sent</i>	<i>*Inte kom för sent, Kom inte för sent</i>
Particle placement	<i>legge fabriken ned, legge ned fabriken</i>	<i>*lägga fabriken ner, lägga ner fabriken</i>

5

Some observations and conclusions

- NorSource rules and types generally worked very well for Swedish
 - Norwegian more varied than Swedish
 - delete and restrict rather than add new types and features
 - NorSource treatment of word order for verb complements found suboptimal

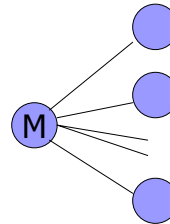
6

Some observations and conclusions

Matrix methodology

To model n languages,

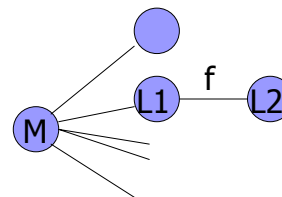
- model what is common to them and then
- model what is specific for each of them



Complementary methodology

To model a language,

- find a model of a similar language
- define a mapping from that language



7

SweCore developments

- Include a core lexicon
 - all function words
 - ~2000 content words
- Use/develop methods for lexical acquisition from (application) corpora

8

Related work - projects

- Recent projects on translation technologies
 - PLUG (joint with Uppsala, Göteborg)
 - knowledge-lite word alignment algorithms
 - Swedish, English, German, French, (Italian)
 - Transmap
 - a parallel English-Swedish treebank (database)
 - parsed (Connexor's dependency parsers)
 - aligned at word level, being checked for correctness
 - a specification of close English-Swedish translation
 - Koma (joint with Uppsala)
 - Corpus-Based Machine Translation

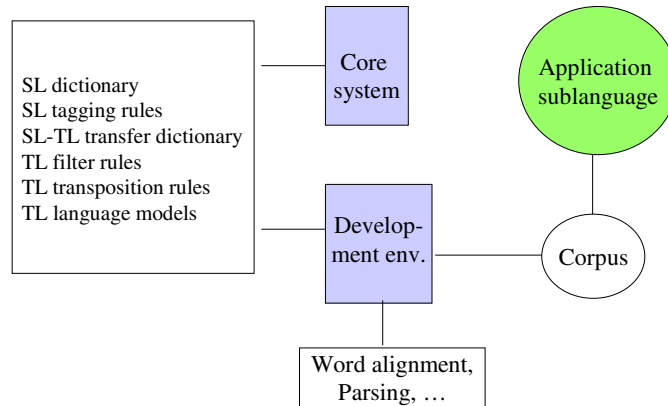
9

Related work – cont.

- Data generation from parallel corpora
 - bilingual word alignment
 - lexicon acquisition
 - term extraction
 - construction data
- Machine translation (T⁴F)
 - tokenize, tag, transfer, transpose, filter (and rank)

10

The T4F framework



11

T4F underspecified TL representation

```

1  subj(2)  he    {han}
2  v-ch(4)  is    {är, NULL}
3  adv(2)   not   {inte}
4  main     watching {tittar+på, titta+på, tittade+på,
                                     att+titta+på, ser+på, se+på, ... }
5  obj(4)   you   {du, dig, ni, er }
  
```

Results (after transpositions, filterings):

Han {tittar, ser} inte på {dig, er}

12

Why deep, why shallow?

- Why shallow representations?
 - don't compute what cannot be computed correctly (e.g. quantifier scope)
 - don't compute what is not needed for the task at hand (e.g. quantifier scope restrictions)

13

Why deep, why shallow

- Why deep representations?
 - Supports (safe) paraphrasing
 - I don't have any money
 - *Jag har inte några pengar*
 - But also: *Jag har inga pengar*
 - Provides more information for resolving lexical ambiguities
 - This man we know we should never have picked.

14

Predicate relations vs. words

■ PRO:

- supports **modularity**, and **bidirectionality** i.e. the classical argument for interlinguas
- supports inference (at least potentially) and thus more general paraphrasing

■ AGAINST:

- the translation relation is different from semantic equivalence

15

English-Swedish verb frame divergencies

■ Goals (Sara Stymne's MA thesis)

- Descriptive
 - Collect and classify which verb frame divergences exist between Swedish and English
- (Machine) Translation theoretical
 - Find out what cases can be handled by a common semantic representation (MRS)
 - Find the border where this approach does not work any more and some other approach, for instance, semantic transfer, would be needed
- Practical
 - Implement a Matrix based grammar for Swedish and a parallel one for English covering the described divergences
 - Find out what parts of the grammar are the same for Swedish and English, and what the differences are

16

Examples

- Prepositional object vs simple object
 - trust someone / lita på någon (trust on someone)
 - look at something / betrakta något (look something)
- Idiomatic particle verb vs plain verb
 - remember / komma ihåg
- Reflexive verb vs non-reflexive verb
 - revenge oneself on someone / hämnas på någon (revenge on someone)
- Combinations of the above
 - endure something / stå ut med något (stand out with something)
 - succeed / komma sig upp (come oneself up)
- Ditransitive verbs with different valence patterns
 - tell a story to someone berätta en historia för någon
 - tell someone a story *berätta någon en historia

17

More examples

- Absence or presence of infinitive marker
 - want to go / vill gå (want go)
 - will go / kommer att gå (come to go)
- Support verbs
 - decide / fatta beslut (make decision)
 - want something / vill ha något (want have something)
- Head inversions
 - finish packing / packa färdigt (pack ready)
- Verb vs. copula + adjective
 - owe / vara skyldig
 - be able to / kunna

18

Empty vs. non-empty items

	Empty	Non-empty
Reflexive pronoun	perjure <i>oneself</i>	kill <i>oneself</i>
Particle	throw <i>up</i> (vomit)	get <i>up</i>
Pronoun	depend <i>on</i> something	sit <i>on</i> something

19

Tentative solutions

- Particles and prepositions
 - have a feature PFORM, used by other constituents to choose a correct prep./part.
 - PFORM-values are organised in an hierarchy making it possible to choose either a specific prep./part. or a class of prepositions such as locational prepositions.
- Verbs
 - Particle verbs have a feature GOVPA
 - Verbs with prepositional complements have a feature GOVPR
 - These are used to choose the correct pform(s) of the prep./part.

20

Tentative solutions

- Pure reflexive verbs
 - take an obligatory reflexive pronoun as a complement
perjure oneself / **someone else*
 - the reflexive complement is empty, since it is seen as part of the verb sense
 - the reflexive pronoun is constrained to have the same PNG-value as the subject.
- Regular transitive objects
 - can take a reflexive pronoun as complement
Shave oneself / *someone else*
 - problematic since in Swedish the accusative and reflexive forms are the same in the first and second person (mig-me, mig-myself). This means that Swedish “jag rakar mig” is ambiguous between “I shave myself” and “I shave me” unless “mig” is resolved as either reflexive or accusative.

21

Tentative solutions

- Empty complements are on the comps list
- This means that many new base types facilitating empty comps have to be introduced, e.g.
 - Basic-four-arg, basic-five-arg
 - Trans-empty2arg-lex-item, intrans-empty2-3ndarg-lex-item
- Alternative approach
 - Introduce a new list in the valence for empty complements
 - Make sure all existing rules handle information on this list correctly

22

Implemented grammar

- Small parallel grammars for English and Swedish using Matrix v 0.8

- ☐ verb frame differences as above
- ☐ yes-no questions
- ☐ main and subordinate clause word order
- ☐ lexicon includes case representatives only (~ 70 entries)
- ☐ LANG-feature to separate the languages

swedish-only-rule := headed-phrase &

[SYNSEM.LANG #sw & sw,
HEAD-DTR.SYNSEM.LANG #sw].

swedish-only-lex := basic-lex &

[SYNSEM.LANG sw].