Case Grammar-tagging in the VISL Constraint Grammar
WORK IN PROGRESS 1.1

Søren Harder

October 9, 2001/ February 13, 2002

Abstract

This is both a quickly written guide to the very first code in my PhD-project (see: http://visl.sdu.dk/visl/about/soeren.html) and a very first draft of a potential chapter in the thesis. The system described is accessible on http://visl.sdu.dk/visl/da/parse.html (i.e. 'Flat structure'), if you set the parser to 'Parser with case roles'. *Please do not quote this text!*

1 VISL and constraint grammar

This paragraph is still not written:

Something about the purpose of CG, introducing 'running text' and 'tags'

Constraint Grammar is not, as most other approaches to language named *'XX Grammar'* are, a theory of grammar, but a formalism for rules enabling linguists to construct algorithms that correctly tag words in running text. The tags used in Constraint Grammar could be taken from some specific scientific theory of grammar, but usually comes from what may be called 'standard' or 'school grammar', the 'theory' of grammar used in primary and secondary schools and also usually seen as the basis that theoretical linguistics builds on. This includes part-of-speech (noun, verb, adjective etc.) and sentential function (subject, object, main verb, finite verb, adverbial etc.). These levels have already been treated within the VISL-project; the purpose of the project described below is, given a text already automatically tagged with POS and sentential function, to tag it with 'cases' as those introduced by Charles Fillmore's theory of Case Grammar (Fillmore 1970, Fillmore 1987), which I will give an introduction to in section 2.1. This procedure also demonstrates how the output of one constraint grammar may be the input of another, so that for example one constraint grammar tags a raw text with parts of speech-tags, the next grammar uses these tags to tag the text with grammatical functions, which again is the input to my case-tagger.[1] The purpose of this analysis is, among others, to be used as part of the source language analysis component of a machine translation system.

1.1 Strategies in Constraint Grammar

When a constraint grammar analyses a text, it does so in two steps: mapping and constraints. In the mapping-stage you define the 'tag space' that the constraints work in. For each word in the text you map one or (usually) more potential tags. The word 'fast', for example, could be an adjective or an adverb (in the meaning:

[1] PRELIMINARY NOTE: AND AS I DO (IF I DO IT) WITH ANAFOR.CG

```
REMOVE ADV  (-1 ART) (+1 N); #remove Adv-reading between ART and N
REMOVE V  (-1 ART) (+1 N); #remove V-reading between ART and N
SELECT ADV (+1 >>>) (NOT *-1 <vk>); #if the last word: Adv (except if
the sentence contains a copula, ``<vk>'')
SELECT VFIN (0 V) (*-1 <<< LINK NOT *1 VFIN); #if it is the only
possible finite, it is a finite.
```

'quick') or a noun or a verb (in the meaning: 'abstaining from food'). In the constraint-stage you remove the incorrect tags one at a time depending on the tags of the words in its immediate context. In the phrase 'the fast car' the word 'fast' occurs between an article and a noun, so it can be neither a verb nor an adverb[2], thus constraining the word's ambiguity from a quadruple ambiguity to a simple ambiguity between noun and adjective. Lexical mapping, as that of 'fast' above, happens in the lexicon, but syntactic mapping as the mapping of '@SUBJ' to potential subjects or '@>N' to potential premodifiers of nominals is done by mapping-rules in the grammar. In figure 1.1 I show how the rules that handle the ambiguity of 'fast' may look. The first word, REMOVE or SELECT, tells whether to remove the particular interpretation (the second word) from consideration, or 'select' it, i.e. remove all other readings. The rest of the rule tells in which context to the left ('-') and to the right ('+') the rule applies. The text after '#' is comments for the human reader only. These rules are not actual rules; they may not achieve the intended, and may have to be modified or, more likely, supplemented with other rules handling more specific contexts.

The mapping rules may refer to the context, though usually not in the degree that constraint rules do. This means that it is very much up to the grammar writer how to divide the workload between mapping and constraint rules. In principle you could, in one extreme, have a mapping component that maps all parts of speech to each and every word in the text, making the work for constraints a cumbersome one. In the other extreme, you could in theory make each mapping rule so complexly context-dependent that the mapping component only maps one — the correct — tag to each word, making the constraint component unnecessary. An example of this last strategy can be seen in my treatment of %TOP-DIST on page 4.1.[3] The grammar writer will have to find his own golden middle between these two extremes. In section 4.1 I will describe the choices made in this project.

2 Case Grammar

This paragraph is still not written:

Something about Fillmore's case grammar.

2.1 The Case tag-set used

This paragraph is still not written:

The tagset used in this project will be described. It is a somewhat larger set than Fillmores. Below the relevant comments from the code. This is not the final tag-set. The headers 'nominale' etc. are not to be seen as exclusive, but typical, instantiation, the %AG could be realised as an @ADVL and the @SUBJ may realise a %TEMP-PL.

[2]Exceptions to the rule, like 'the very man' and 'the then king of England' may be written into the rule

[3]PRELIMINARY NOTE: GET THE CODE UP TO DATE

```
# udgangspunkt: form - funktion - rolle, dvs. en rolle til hver
#    konstituent på sætningsplan, bortset fra konjunktioner

### nominale @SUBJ, @ACC, @DAT, @PIV

# %AG = agent (human or institution or the like, +physical act)
# %PAT = patient (object acted on, +physical entity)
# %BEN = beneficiary (+phys. ent.)
# %TH = theme
#       I've used this for 2 diff. types: 1) 'internal' objects,
#       i.e. objects, that are not entities acted on (%PAT) or
#       created (%RES), i.e. a) non-physical entities (often
#       clauses) b)e.g. complements of Adj, 2) subjects of copula
#       and other verbs that do not denote physical or mental acts:
#       e.g. stå, komme. (better examples needed)
# %EXP = experiencer, subj of cogn./perc. verbs
# %RES = factitive, resultative (She knitted a sweater) = %FACT


### attributive/prædikative @SC, @OC, (@N<PRED, @APP), @PRED

# %QUAL-RES = resulting quality, identity (han blev præsident,
#             de valgte ham til præsident)
# %QUAL-STATE = quality, identity (han er syg, han vendte hjem
#             som storryger)


### adverbielle @ADVL, @SA, @OA

# %TOP-PL = place/topological, route?!
# %TOP-GL = goal
# %TOP-SRC = source
# %TOP-DIST = distance (han løb 3 kilometer)

# %TEMP-PL = point or "area"/periode in time
#            (på et tidspunkt, før/efter, under krigen)
# %TEMP-GL = time/temporal goal
# %TEMP-SRC = time/temporal source
# %TEMP-DUR = durative (han arbejdede [i] 3 uger, på 3 timer)

# %INSTR = instrumental (vha., med)
# %MOD = mode (måde)
# %PURP = finality (hensigt, for at)
# %COND = conditional (betingelse, hvis, medmindre, undtagen)
# %CONC = concessive (indrømmende, selvom, på trods af)
# %CAUSE = cause (årsag, fordi)
# %EFF = effective (så, derfor)

# %COM = comitative, "combinatorial" (sammen med, med)

# %UNDEF = undefined (i stedet for)

### verbale @V
```

```
# %EV = event, det logiske (inderste) prædikat (som regel MV),
#       copulas are unmarked and all kinds of modality ignored
```

3 Phenomena covered

As of now the only syntactic functions accounted for in my case grammar are nominal subjects and objects. Other functions as well as prepositional phrases, clauses etc. will follow as the grammar is extended.

I have up to this point used three one-page excerpt of simple narrative as benchmark: 'Little Red Riding Hood' (see app. A), 'The Golden Compas' (Pullman 1996) (translated children's litterature: app. B) and 'Silas and The Black Mare' (Bødker 1967?) (original children's litterature: app. C). I have tried to write the rules in a general style though, and they should, even at this early stage, be able to handle other texts with some adequacy, and in the long run handle any given text.

In the next section I will be a bit more specific about the phenomena the grammar handles and give the full code of the grammar with explanatory comments.

4 Strategies in tagging

4.1 Mappings

Subjects are mapped to 'agent', 'patient' (for passive subjects), 'experiencer', 'instrument' and 'theme' and objects are mapped to 'patient', 'theme' and 'resultative'.[4]

The Danish 'sig'-reflexive is given a special treatment. Reflexive pronouns (including first and second person pronouns) are not mapped, because they, in my opinion, denote middle voice and not a seperate entity in the case frame. Only 'licensed' reflexives, i.e. reflexives that are governed by reflexive verbs, are exempt[5] case-marking; non-licensed reflexives as in 'He caught himself a big fish', *do* get case-marking (typically %BEN), as do emphatic[6] reflexives ('sig selv'). Two mapping rules make sure that first and second person pronouns do get a case when they do not function as reflexives, but as normal objects. This is done by checking, that they are not co-referent with the subject.

```
### Mapping-rules

MAP (%AG %PAT %EXP %TH) TARGET (@SUBJ>); #Subjects standing to the
                                         # left of the verb
MAP (%AG %PAT %EXP %TH) TARGET (@<SUBJ); #Subjects standing to the
                                         # right of the verb

MAP (%PAT %FACT %TOP-DIST) TARGET (@<ACC) IF (NOT 0 REFL);
                                         #refl. have no case-roles
MAP (%PAT) TARGET (@<ACC) IF (0 PERS-1-ACC LINK *-1 @SUBJ LINK
    NOT 0 PERS-1-NOM); #...neither in 1st person
MAP (%PAT) TARGET (@<ACC) IF (0 PERS-2-ACC LINK *-1 @SUBJ LINK
    NOT 0 PERS-2-NOM); #...nor in 2nd person
```

[4] Objects are also mapped to 'distance', which is the case-role I use for 'routes' (and not 'place', %TOP-PL, as suggested above). This is to handle the phrase "gå den lige vej" (litt: to go the straight path/way). As the grammar is extended there will be a lot of specialised mappings like this. The mapping rules assigning these will probably refer to very specific contexts.

[5] PRELIMINARY NOTE: RIGHT WORD??

[6] PRELIMINARY NOTE: TERM??

4.2 'Chipping off of the block'

The mappings create a 'brutto-analysis', in which you are certain that you have given the correct analysis – along with a lot of incorrect ones. To use a metaphor that is often used in Constraint Grammar, this is like a boulder before the stonecutter starts chiseling out the statue within it, by removing all the superfluous rock around it. Now we have to distinguish the correct from the incorrect analyses.

I do this in three steps, three sub-algorithms run one after the other:

1. idiosyncratic collocations
2. 'chipping off'
3. cleaning up

4.2.1 Idiosyncratic collocations

The first sub-algorithm handle lexical exceptions, where verbs, which normally assign one case to its subject or object, assign another, because they occur with particles, special complements or are metaphoric in some degree. The verb "se" ('look'), e.g., normally has an experiencer as subject, but to "se sig om" ('look around') is an action with an agent. At the current state four of these idiosyncratic collocations are written into the grammar, all pertaining to the case of the subject. As the grammar is tested on more text, more of these structures wil be found, and this part of the grammar will grow a lot.

These structures could also be handled in the lexicon.[7]

For all the different idiosyncratic subject structures, there are two rules: one for subjects to the left of the finite verb (non-inverted subjects, tagged with @SUBJ> by earlier levels of Constraint Grammar) and one for subjects to the right of the finite verb (inverted subjects: @<SUBJ). Even more rules will be needed for each construction, if we are to handle for example a relative clause "breaking" one of these constructions.[8]

```
#0#Idiomatisk leksikale constraints ###############
CONSTRAINTS

# se sig om/for  AG, ikke EXP
SELECT (%AG) (0 @SUBJ>)  (*1 ("se") BARRIER @MV LINK 1 REFL LINK
       *1 ("om") OR ("for") BARRIER NON-ADV);
SELECT (%AG) (0 @<SUBJ)  (-1 VFIN LINK -1 ALL LINK *1 ("se")
       BARRIER @MV - ("se") LINK *1 REFL BARRIER NON-SUBJ-MAT LINK
       *1 ("om") OR ("for") BARRIER NON-ADV);\nb{This rule should be:
SELECT (%AG) (0 @<SUBJ)  (-*1 VFIN BARRIER NON-PRE-MOD LINK -1 ALL LINK *1 ("se")
       BARRIER @MV - ("se") LINK *1 REFL BARRIER NON-SUBJ-MAT LINK
       *1 ("om") OR ("for") BARRIER NON-ADV);
change this in alle rules.}

# døren går op/i (men ikke 'gå op i')
SELECT (%PAT) (0 @SUBJ>) (*1 ("gå") BARRIER @MV LINK *1 ("op") OR
       ("i") BARRIER NON-ADV LINK NOT 1 ("i"));
SELECT (%PAT) (0 @<SUBJ) (-1 VFIN LINK -1 ALL LINK *1 ("gå") BARRIER
```

[7] Preliminary note: This is a very important discussion, but it is late, and I still have a lot to write.

[8] Preliminary note: Somewhere I need to write something about 'the limits of CG' esp. w.r.t. relative clauses. E.g. Anafor.cg.

```
       @MV - ("gå") LINK *1 ("op") OR ("i") BARRIER NON-ADV/SUBJ-MAT
       LINK NOT 1 ("i"));

#jeg holder af dig/at svømme
SELECT (%EXP) (0 @SUBJ>) (*1 ("holde") BARRIER @MV LINK *1 ("af")
       BARRIER NON-ADV) ;
SELECT (%EXP) (0 @<SUBJ) (-1 VFIN LINK -1 ALL LINK *1 ("holde")
        %BARRIER @MV - ("holde") LINK *1 ("af") BARRIER NON-ADV/SUBJ-MAT);

# blive underlig til mode
SELECT (%EXP) (0 @SUBJ>) (*1 ("blive") OR ("være") BARRIER @MV LINK
       *1 ("underlig") OR ("mærkelig") BARRIER NON-ADV LINK 1 ("til")
       LINK 1 ("mode")) ;
SELECT (%EXP) (0 @<SUBJ) (-1 ("blive") OR ("være") OR ("har") LINK *1
       ("underlig") OR ("mærkelig") LINK 1 ("til") LINK 1 ("mode")) ;
```

4.2.2 'Chipping off'

The second, main, algorithm does the 'normal' Constraint Grammar processing, removing and selecting readings as the context becomes more and more disambiguated. The first, and largest, part of this algorithm handles cases mapped to subjects. The rules are grouped after which case they , and whether they select or remove a case.[9] I will in the following explain each rule. The rules are usually doubled to handle both non-inverted and inverted subjects, that will be found respectively to the left and the right of their finite verb, in the same way as the rules described in last section. For each of these rule-pairs there is also an extra rule to handle subjects to the right of the finite verb, but to the left of the main verb, i.e. inversions in complex tenses like 'Derfor kan Rødhætte ikke genkende sin bedstemor' (litt.: Therefore can Little Red Riding Hood not recognise her grandmother'). Extra rules (usually one) are also added to handle up to one embedded clause (i.e. one relative clause with no embedded clause) modifying a non-inverted subject.[10] These two kind of rules (inversion in complex tense and relative clause) ought also be written for each of the idiosyncratic verb uses above, but the likelihood of these specific constructions occuring has been deemed to small.

Relative clauses Some further comments on the treatment of relative clauses may be necessary. A piece of standard code is inserted in the beginning of a copy of each rule that needs to handle relative clauses. This code basically says 'skip the relative clause and ...'. The code is: *1 CLB BARRIER VFIN LINK *1 VFIN LINK 0 @FS-N< LINK *1 VFIN BARRIER @SUBJ LINK -1 ALL. This code finds a clause boundary (CLB) to the right (with no intruding finite verbs) followed somewhere to the right by a finite verb with the tag @FS-N<, indicating that it is the head of a finite clause (in Danish: 'finit sætning' FS) modifying a noun to its left, i.e. a relative clause. Having identified the relative clause, we now want to find the main verb of the matrix clause. We cannot just find the next *main* verb to the right, as it may be part of either the relative clause (having composite tense) or the matrix sentence. But we are certain that the next *finite* verb belongs to the matrix sentence, but only if there are no further embedded sentences, either

[9] PRELIMINARY NOTE: EXPLAIN SELECT AND REMOVE SOMEWHERE TOGETHER WITH THE SYNTAX OF THE CG-RULES.

[10] Inverted subjects pose no problems, as the relative clause do not come between the head of the subject and the main verb in that construction. The same is the case for objects.

another relative clause or a clause modifying something in the first relative clause.[11] To avoid further embeddings we prohibit any subjects between the two finite verbs ('... BARRIER @SUBJ ... '). Having found the matrix finite verb, which may or may not be the main verb, we go left one word ('... LINK -1 ... '), regardless of what we find in that location ('... ALL ... '), and we are now ready to search for the matrix main verb to the right, which may be either the finite verb, that we just found, or an non-finite main verb further right. It is this main verb that assigns case to the subject[12] before the relative clause.

Theme The case 'theme' is removed for the subject, if there is no theme-assigning verb, typically a copula, but it may also be other verbs denoting situatedness as 'stand' and 'lie'. These rules are so general, that they do not remove anything even if the theme-assigning is in an embedded sentence. The functionality of this group of rules heavily overlaps with the equivalent group of rules in the 'clean up'- section. I will return to this below.[13]

```
#############
# REMOVE %TH

REMOVE (%TH) (0 @SUBJ>) (NOT *1 V-THEME LINK 0 @MV);
REMOVE (%TH) (0 @<SUBJ) (NOT *-1 V-THEME LINK 0 @MV);
REMOVE (%TH) (0 @SUBJ>) (*1 CLB BARRIER VFIN LINK *1 VFIN LINK
       0 @FS-N< LINK *1 VFIN BARRIER @SUBJ LINK -1 ALL LINK NOT
       *1 V-THEME BARRIER CLB/SB LINK 0 @MV);
```

Patient and Resultative The case 'patient' is selected for the subject, if the main verb is passive, and removed if the main verb is a transitive verb with an object.[14] A special rule handles OVS, not by looking for an object, but by removing 'patiens' if you are certain that the element is subject of a transitive verb; this rule is written to handle direct speech,'"Who is it?", he said' which is OVS in Danish, and is only triggered with speech-verbs. Another rule removes the patiens-reading from subjects even if the object is not expressed, but is a suppressed object pronouns in a relative clause, e.g. 'The woman, (that) the wolf ate, was old'. The last group of rules unconditionally removes the patiens reading from subjects of intransitive verbs; no conditions on this rule have been necessary yet, but will be added as the grammar grows, depending on the semantics of the verb.

The patient-selection rules has been copied for resultative, with an extra condition that the main verb has to belong to a special small class of verbs (V-RES: build, make etc.). No copying of remove-tags has been made. Instead removal is handled in clean-up, where a rule removes all resultative tags, that are not the only possible analysis. This may turn out to be too crude a solution.

```
#############
# SELECT %PAT

SELECT (%PAT) (0 @SUBJ>) (*1 @MV BARRIER CLB/SB LINK 0 PAS) ;
                              # ulven blev fanget, ulven fangedes
SELECT (%PAT) (0 @<SUBJ) (*-1 @MV BARRIER CLB/SB LINK 0 PAS) ;
```

[11] Preliminary note: There are exceptions to this, where you have conjoined finite verbs in the relative clause. Can they be dealt with?

[12] Preliminary note: Remember to write about the Constraint Grammar necessity of assigning functions to heads of phrases, not to phrases

[13] Preliminary note: This has been changed!!

[14] Preliminary note: or just a transitive verb

```
                              # derfor fangedes ulven
SELECT (%PAT) (0 @<SUBJ) (*1 @MV BARRIER CLB/SB LINK 0 PCP2 LINK 0
        %PAS) ;               # derfor blev ulven fanget
SELECT (%PAT) (0 @SUBJ>) (*1 CLB BARRIER @MV LINK *1 VFIN LINK
       0 @FS-N< LINK *1 VFIN BARRIER @SUBJ> LINK -1 ALL LINK *1 @MV
       LINK 0 PAS) ; # ulven, som ..., blev fanget; ulven, som ...,
       fangedes; kan ikke klare 'ulven, som ..., fordi ..., fangedes'

#############
# REMOVE %PAT

REMOVE (%PAT) (0 @SUBJ>) (*1 @MV BARRIER CLB/SB LINK 0 <vt> LINK
       *1C @<ACC& BARRIER @MV) ; # ulven spiste kaninen
REMOVE (%PAT) (0 @<SUBJ) (*-1 @MV BARRIER CLB/SB LINK 0 <vt> LINK
       *1C @<ACC& BARRIER @MV) ; # derfor spiste ulven kaninen
REMOVE (%PAT) (0 @<SUBJ) (*1 @MV BARRIER CLB/SB LINK 0 <vt> LINK
       0  @ICL-AUX< LINK *1C @<ACC& BARRIER @MV) ;
                                 # Derfor har ulven spist kaninen
REMOVE (%PAT) (0 @<SUBJ) (*-1 @MV BARRIER CLB/SB LINK 0 <vt> LINK
       NOT 0 <vi>) ;             # "Hvem er det?" spurgte hun. :

REMOVE (%PAT) (0 @SUBJ>) (*1 @SUBJ> BARRIER VFIN LINK *1 VFIN LINK
       *1 VFIN BARRIER @SUBJ> LINK NOT -1 KC LINK *2C @<ACC
       BARRIER VFIN); # Ulven, der blev fanget, havde spist kaninen
REMOVE (%PAT) (0 @<SUBJ) (*-1 @MV BARRIER CLB/SB LINK 0 (<vt>) LINK
       *1C @MV LINK *1C CLB LINK *1 @<ACC& BARRIER @MV) ;
       # Derfor spiste ulven, der kun havde fanget det ene dyr, kaninen

REMOVE (%PAT) (0 @SUBJ>) (*1 @MV BARRIER CLB/SB LINK 0 (<vt>) LINK
       1 ALL LINK *-1 VFIN BARRIER NON-V LINK 0 @FS-N<);
       # vt i bis. uden obj. "..., hun havde sparket, ...

REMOVE (%PAT) (0 @SUBJ>) (*1 @MV BARRIER CLB/SB LINK 0 <vi>); #Manden gik
REMOVE (%PAT) (0 @<SUBJ) (*-1 @MV BARRIER CLB/SB LINK 0 <vi>);
                                   #derfor gik manden
REMOVE (%PAT) (0 @<SUBJ) (*1 @MV BARRIER CLB/SB LINK 0 <vi> LINK
       0 @ICL-AUX<);               #derfor er manden gået
REMOVE (%PAT) (0 @SUBJ>) (*1 CLB BARRIER VFIN LINK *1 VFIN LINK 0
       @FS-N< LINK *1 VFIN BARRIER @SUBJ LINK -1 ALL LINK *1 @MV
       LINK 0 <vi>);               #Manden, som ..., gik
```

Experiencer Experiencer is selected for subjects of mental verbs. Mental verbs are marked as such in the lexicon, which marks all verbs according to Halliday (1994)'s process types. Some of the verbs marked as mental in the lexicon are ignored, such as "gå" "have" and "trykke", as these are verbs with very general meanings, that only have their mental meaning in specific contexts. These contexts have to be treated as idiosyncratic. Likewise experiencer is selected as the case for subjects with predicative mental adjectives, such as "glad", "lykkelig", "tilfreds", "utilfreds", "ked" og "sur" (engl.: 'happy', 'contented', 'sad', 'angry').

```
#############
# SELECT %EXP
```

```
SELECT (%EXP) (0 @SUBJ>) (*1 @MV BARRIER CLB/SB LINK 0 V-MENT );
                                       #Han tænkte
SELECT (%EXP) (0 @<SUBJ) (*-1 @MV BARRIER CLB/SB LINK 0 V-MENT );
                                       # -"- inversion
SELECT (%EXP) (0 @<SUBJ) (*1 @MV BARRIER CLB/SB LINK 0 V-MENT LINK
       0 @ICL-AUX<);                   # -"- inv. + sammensat tid
SELECT (%EXP) (0 @SUBJ>) (*1 CLB BARRIER VFIN LINK *1 VFIN LINK
       0 @FS-N< LINK *1 VFIN BARRIER @SUBJ LINK -1 ALL LINK *1 @MV
       BARRIER CLB/SB LINK 0 V-MENT ); #Manden, som ..., tænkte


SELECT (%EXP) (0 @SUBJ>) (*1  @MV BARRIER CLB/SB LINK 0 <vk> LINK
       *1 A-MENT BARRIER CLB/SB);      # Han var glad
SELECT (%EXP) (0 @<SUBJ) (*-1 @MV BARRIER CLB/SB LINK 0 <vk> LINK
       *1 A-MENT BARRIER CLB/SB);      # -"- inversion
SELECT (%EXP) (0 @<SUBJ) (*1 @MV BARRIER CLB/SB LINK 0 <vk> LINK
       0  @ICL-AUX< LINK *1 A-MENT BARRIER CLB/SB);
                                       # -"- inv. + sammensat tid
SELECT (%EXP) (0 @SUBJ>) (*1 CLB BARRIER VFIN LINK *1 VFIN LINK 0
       @FS-N< LINK *1 VFIN BARRIER @SUBJ LINK -1 ALL LINK *1 @MV
       BARRIER CLB/SB LINK 0 <vk> LINK *1 A-MENT BARRIER CLB/SB);
                                       # Manden, som ..., var glad
```

Before I turn to functions of objects, I would like to turn the reader's attention to the fact that there are no rules selecting or removing the typical subject function of agent, neither in this or in the 'clean up'-subalgorithm. This is the function left when all others have been removed. I do not even need to select it in 'clean up', as we will see.

Objects The object section is marked by the lack of work put into it, as there are only one general, and two very specific, rules. But there are grammatical reason for its small size as well. Objects typically occur to the right of finite/main verbs, where subjects may occur to the right, between them or to the left. This also means that relative clauses do not intervene. The general rule says that objects of a restricted class of verbs ("sige", "lave" and "bygge" so far) are resultatives (or "factitives").[15] The two specific rules are the 'if' and 'only if' parts of the rule: an object is a distance/route if and only if the main verb is 'to go': "Han gik en tur", "Han gik tre kilometer".[16]

```
##############
# SELECT %FACT

SELECT (%FACT) (0 @<ACC) (*-1 @MV BARRIER CLB/SB LINK 0 V-FACT);
                                          # Jack built a house.

##################
# IFF %TOP-DIST

SELECT (%TOP-DIST) (0 @<ACC) (*-1 @MV LINK 0 ("gå"));
       #"gå en tur", er der andre der tager top-dist som direkte obj?
```

[15] PRELIMINARY NOTE: THIS IS CHANGED
[16] PRELIMINARY NOTE: THIS HAS BEEN MOVED TO MAPPING

```
REMOVE (%TOP-DIST) (0 @<ACC) (*-1 @MV LINK NOT 0 ("gå")); # iff
```

4.2.3 Cleaning up

In the central algorithm described above I have typically either used select- or remove-rules, not both, for a given case (with patiens as an exception). The cleaning up algorithm works by the principle: if you have tried to prove something, and cannot, it must be false, and if you have tried to falsify something, but cannot, it must be true.

The rules of the previous subalgorithm pertaining to theme have been mirrored in this subalgorithm. They probably only belong one place, but I haven't yet decided which. On the other hand it may be necessary to keep them both places.[17]

```
##########################
# SUBJECTS
##########################

#############
# REMOVE %EXP

REMOVE (%EXP); #hvis ikke du har selected exp. ovenfor, er den ikke

#############
# SELECT %TH

SELECT (%TH) (0 @SUBJ>) (*1 V-THEME  LINK 0 @MV);
       #i normale tilfælde: subj for kopula = TH, bemærk betingelsen
       #  i denne regel er 'dobbelt' i 1 REM-TH; skal slettes her
SELECT (%TH) (0 @<SUBJ) (*-1 V-THEME LINK 0 @MV);
       #d.o.; hele denne regel slettes

##########################
# OBJECTS
##########################

#############
# SELECT %PAT

SELECT (%PAT) (0 @<ACC); # hvis ikke andet gælder er obj. PAT

END
```

A Text 1: Rødhætte

Rødhætte så sig om. Solstrålerne skinnede mellem bladene ned på de mange, kønne blomster, og hun tænkte: "Bedstemor vil sikkert blive glad for en buket blomster. Det er så tidligt, så jeg kan godt nå at plukke nogle." Hun gik nu ind mellem træerne og begyndte at plukke nogle, men når hun havde plukket en, syntes hun, at der stod nogle meget kønnere længere borte, og på den måde kom hun dybere og dybere ind i skoven. Ulven gik imidlertid den lige Vej til bedstemoderens hus og bankede på:

[17] Preliminary note: this has been changed, hasn't it, otherwise it deserves a more thorough discussion

"Hvem er det?" spurgte hun. "Det er mig, lille Rødhætte, med kager og vin til dig," svarede ulven. "Tryk på klinken," råbte bedstemoderen, "jeg er så dårlig, at jeg ikke kan stå op." Ulven trykkede på klinken, så døren gik op, og uden at sige et ord gik den lige hen til sengen og slugte den gamle bedstemor. Så tog den hendes klæder på, lagde sig i sengen og trak forhænget for. Rødhætte havde imidlertid plukket lige så mange blomster, hun kunne bære, og kom langt om længe hen til huset. Hun kunne ikke forstå, at døren var åben, og da hun kom ind i stuen, blev hun så underlig til mode, og det kunne hun slet ikke begribe, for hun holdt ellers så meget af at være hos sin bedstemor. "Goddag," sagde hun, men der var ingen som svarede. Hun gik hen til sengen og trak forhænget til Side, og der lå bedstemoderen med natkappen trukket dybt ned i panden og så så underlig ud.

B Text 2: Det Gyldne Kompas

Lyra og hendes daimon bevægede sig varsomt gennem den mørknende spisesal og passede særligt på ikke at blive set fra køkkenet. De tre langborde, der nåede fra den ene ende til den anden, stod allerede dækkede til middag. Sølvtøjet og krystalglassene fangede dagens sidste lys i de pudsede og slebne overflader. Bænkene var trukket ud for de kommende gæster. Olieportrætter af tidligere rektorer hang højt på de dunkle vægge. Lyra nåede uset frem til højbordet og kastede et blik tilbage mod køkkendøren. Der var ingen at se; hun trådte op på gulvets forhøjning og betragtede bordet og de fornemme sæder. Her sad kun de højestrangerende blandt kollegiets lærde herrer. Deres bord var ikke dækket med sølvtøj, men bestik af guld. De fjorten siddepladser var ikke egetræsbænke, men mahognistole med fløjlspuder. Lyra standsede ved Rektors stol og slog på det største krystalglas med en fingernegl. Den sprøde lyd klingrede tydeligt gennem det store rum.

C Silas og den sorte hoppe

C.1 Chapter 1

Han kom sejlende nedad floden i en sær lille brednæset båd, og han sejlede ikke siddende oprejst som andre, brugte ikke årerne, men lod strømmen føre sig som den havde lyst. Selv lå han nede i bunden, og man måtte tro det var en der havde god tid, det gik kun fremad med strømmens fart, og på afstand lignede det fuldstændig en tom båd.

Oppe på loftet over den lange staldbygning på bredden gik hestehandler Bartolin og forkede hø ned gennem et hul i gulvet. Det støvede vældigt, og med mellemrum stak han hodet ud gennem lugen for at få renset lungerne. Udenfor var det en smuk dag med lidt diset sol, og lige nedenfor stalden løb floden forbi i et bredt sving, han kunne følge den med blikket et godt stykke vej til begge sider.

Men pludseligt stivnede han og kneb øjnene sammen for lyset. Hvad var nu det der kom drivende? - En båd?

Og det så ovenikøbet ud som om der ikke var nogen i den. Kunne man tænke sig den var herreløs?

Bartolin trak sig tankefuldt i sit buskede hageskæg, han måtte nok hellere undersøge nærmere hvad det var, hans gamle pram havde nemlig længe trængt til en afløser.

C.2 Chapter 2

Da Silas sent om eftermiddagen igen kom til syne i døren til hestehandlerens private gemak, virkede han yderst vågen og veloplagt. Han så sig om i stalden med klare øjne, og Bartolin kunne ikke undgå at mærke hans interesse for dyrene.

Men nu var der også ganske mange heste for tiden, tænkte han, stalden var næsten helt fuld, og det var et flot syn når man kastede et blik ned langs rækken der nåede fra den ene ende af det lange hus til den anden.

Bartolin kunne med rette kro sig. Men det var jo godt nok et større arbejde at passe dem,

Nå, nu ser du lidt mere kvik ud, sagde han da Silas kom hen til ham.

Nu wer der heller ikke mere tilbage, svarede drengen tilfreds og gnubbede sig på maven.

Hvad er det der ikke er mere tilbage af?

Hestehandlerens stemme var fuld af bange anelser.

Af maden, sagde Silas ligeud. Nu er jeg bare så tørstig, har du ikke noget jeg kan drikke.

Har du ædt det hele?

Bartolin ville ikke tro sine egne ører. Har du ikke levnet noget til mig?

Jamen jeg var da sulten, fastholdt Silas.

References

Bødker, C.: 1967?, *Silas og den sorte hoppe*, Branner & Korch.

Fillmore, C.: 1970, The case for case, *in* E. Bach and R. T. Harms (eds), *Universals in Linguistic Theory*, Holt, Rinehart and Winston, London, pp. 0–88. Reprinted in Fillmore (1987).

Fillmore, C.: 1987, *Fillmore's Case Grammar: A Reader*, Vol. 16 of *Studies in Descriptive Linguistics*, Julius Groos Verlag, Heidelberg.

Halliday, M.: 1994, *An Introduction to Functional Grammar*, second edn, Edward Arnold, London/Sydney.

Pullman, P.: 1996, *Det Gyldne Kompas*, Gyldendal, København.