

A PRELIMINARY COMBINATORY CATEGORIAL GRAMMAR ANALYSIS OF NORWEGIAN LANGUAGE

Onur Çobanoğlu

Middle East Technical University

onur.cobanoglu@ceng.metu.edu.tr

Abstract. Combinatory categorial grammar (CCG) has been proved to be a powerful tool in explaining a wide-variety of linguistic phenomena in human languages. In this work, we present a CCG model of a simple subset of Norwegian language, which correctly predicts certain constructions and gives sound explanations of certain behaviors of the language, based on the lexical properties of the language. Among the studied constructions are basic word order, semantic asymmetries, argument extractions out of complement and relative clauses, control verbs, coordination, node raising and gapping. Besides these, as a guidance for future research, some unique and interesting constructions and behaviors of the Norwegian language are presented.

1. INTRODUCTION

Combinatory Categorial Grammar (CCG) ((Ades & Steedman, 1982), (Dowty, 1979), (Bach, 1979), (Bach, 1988)) is a linguistic framework, which models the language as syntactic and semantic categories of its lexical entries and predicts the phrase structure by combination of categories according to some determined combinatory rules. Although its being a highly lexicalized theory, recent extensions on CCG like multi modal CCG (Baldrige, 2002) and type hierarchical CCG (McConville, 2005) increased the control of lexicon over derivational control much beyond the capabilities of the original framework.

Up to now, CCG is applied to the analysis of various languages like English, Turkish, Dutch, Tagalog, Toba Batak, Japanese etc. In this work, a preliminary analysis of the Norwegian language using CCG is presented. Scope of the analysis covers the most frequently inspected topics in CCG analyses: In Section 2, an analysis of the basic word orders of Norwegian in declarative, interrogative and subordinate clauses are given. Section 3 presents some logical form constraints in order for model to predict certain semantic asymmetries observed in the language. Section 4 includes categories of lexical items taking place in control constructions in Norwegian. Finally, in Section 5, coordination behaviors of the language are tried to be explained according to the constructed model.

Before beginning the analysis, there is one point to be clarified: There is not a homogenous single Norwegian. Besides two wide dialects Nynorsk (meaning ‘new Norwegian’) and Bøkmål (meaning “bookish language”) dominating the Norway, there are many minor dialects throughout the country. In this work, lexicon may be originated from different dialects, but to the best of our knowledge, in the scope of this work all dialects show similar behavior.

In this work, we will use both featural properties of type hierarchical CCG (T-CCG) and modalities introduced in multi-modal CCG (MM-CCG). The rule set that is used in this work, presented below, is completely taken from (Baldrige, 2002):

- *Forward Application (>)*

$$X/*Y:f \quad Y:x \quad \rightarrow \quad X:fx$$

- *Backward Application (<)*

$$Y:x \quad X/*Y:f \quad \rightarrow \quad X:fx$$

- *Forward Type Raising (>T)*

$$X:x \quad \rightarrow \quad Y/iY\backslash_i X: \lambda f.fx$$

- *Backward Type Raising (<T)*

$$X:x \quad \rightarrow \quad Y\backslash_i Y/i X: \lambda f.fx$$

- *Forward Harmonic Composition (>B)*

$$X/\diamond Y:f \quad Y/\diamond Z:g \quad \rightarrow \quad X/\diamond Z: \lambda x.f(gx)$$

- *Backward Harmonic Composition (<B)*

$$Y\backslash\diamond Z:g \quad X\backslash\diamond Y:f \quad \rightarrow \quad Z\backslash\diamond X: \lambda x.f(gx)$$

- *Forward Crossed Composition (>B_x)*

$$\begin{array}{llll} X/\times Y:f & Y\backslash Z:g & \rightarrow & X\backslash Z: \lambda x.f(gx) \\ X/\times Y:f & Y\backslash_{\times} Z_{+ANT}:g & \rightarrow & X\backslash_{\times} Z_{+ANT}: \lambda x.f(gx) \quad (>B_x^{ant}) \\ X/\leftarrow\times Y:f & Y\backslash Z_{+ANT}:g & \rightarrow & X\backslash Z_{+ANT}: \lambda x.f(gx) \quad (>B_x^{ant}) \\ X/\leftarrow\times Y:f & Y\backslash_{\times} Z_{+ANT}:g & \rightarrow & X\backslash_{\times} Z_{+ANT}: \lambda x.f(gx) \quad (>B_x^{ant}) \end{array}$$

- *Backward Crossed Composition (<B_x)*

$$\begin{array}{llll} Y/\times Z:g & X\backslash Y:f & \rightarrow & X/\times Z: \lambda x.f(gx) \\ Y/\times Z_{+ANT}:f & X\backslash_{\times} Y:g & \rightarrow & X/\times Z_{+ANT}: \lambda x.f(gx) \quad (>B_x^{ant}) \\ Y/\leftarrow\times Z_{+ANT}:f & X\backslash Y:g & \rightarrow & X/\leftarrow\times Z_{+ANT}: \lambda x.f(gx) \quad (>B_x^{ant}) \\ Y/\leftarrow\times Z_{+ANT}:f & X\backslash_{\times} Y:g & \rightarrow & X/\leftarrow\times Z_{+ANT}: \lambda x.f(gx) \quad (>B_x^{ant}) \end{array}$$

Here +ANT feature is used to indicate the antecedent government (Steedman, 2000), (Baldrige, 2002) of the argument. Following the general convention, we will abbreviate \backslash_{\times} as \times and \backslash_{\leftarrow} as \leftarrow .

2. BASIC WORD ORDER

Norwegian language belongs to the Germanic languages family, and like some other members of the family (e.g. Dutch and German) it exhibits V2 word order in main clauses, but unlike Dutch and German, subordinate clause word order is strictly SVO (as in English) (Ellingsen, 2004). V2 constraint imposes that the finite verb must be the second constituent in a declarative clause, and in such a clause first constituent always topicalizes:

- (1) Jeg liker en bok
I like.PRES a book
'I like a book'
- (2) en bok liker jeg
a book like.PRES I
'I like a book'

Obviously, (1) and (2) brings an ambiguity question of how to determine subject and object if their relative order is undetermined. In her work, depending on the statistical results obtained from Oslo Corpus, (Øvrelid, 2004) states that in general subject is higher than or equal to object in animacy and definiteness. Since in (1) and (2), “a book” is clearly less animate than “I”, there is no ambiguity about subject and object.

V2 constraint does not imply that relative order of verb’s arguments are free. In fact, relative order of arguments is strict and what breaks that order is topicalization of one of the arguments. This can be well understood from the following ditransitive examples:

- (3) Gyrd ga en bok til Inge (SVdOiO)
 Gyrd give-PAST a book to Inge
 ‘Gyrd gave a book to Inge’
- (4) En bok ga Gyrd til Inge (dOVSiO)
 ‘Gyrd gave a book to Inge’ (‘a book’ emph.)
- (5) Til Inge ga Gyrd en bok (iOVSDo)
 ‘Gyrd gave a book to Inge’ (‘to Inge’ emph.)

As careful readers may see, argument order pattern is SdOiO, and to get a main clause one of these arguments topicalizes and comes to the first position. Arguments except the topicalized one must maintain their relative orders. For example, the following sentence is invalid:

- (6) * Gyrd ga til Inge en bok

Argument order can be best seen in question forms, where finite verb is in the first position:

- (7) Ga Gyrd en bok til Inge
 ‘Did Gyrd give a book to Inge?’

Note that there are no head words (like “did” in English) indicating the question form. Question form is obtained by only moving the finite verb to first question.

So, in transitives argument order is SO and in ditransitives argument order is SdOiO. Note that indirect objects of ditransitives are prepositional phrases (like shifted datives of English ditransitives) with head “til”. However, in subordinate clauses, there is a strict order rather than relative order of untopicalized verbal arguments. Order is SV in intransitives, SVO in transitives and SvdOiO in ditransitives:

- (8) ...at Gyrd ga en bok til Inge (SVdOiO)
 that Gyrd give-PAST a book to Inge
 ‘... that Gyrd gave a book to Inge’
- (9) *... at en bok ga Gyrd til Inge (dOVSiO)
 ‘... that Gyrd gave a book to Inge’ (‘a book’ emph.)
- (10) *... at til Inge ga Gyrd en bok (iOVSDo)
 ‘... that Gyrd gave a book to Inge’ (‘to Inge’ emph.)

In the next sections, we will present the categories and rules to obtain these verb orders.

2.1. DECLARATIVE/INTERROGATIVE CLAUSES

In our model, verbal categories are determined according to interrogative clause order, and declarative clauses are derived from interrogatives by topicalization of a single element (actually this is what goes on Norwegian):

$$\text{sitter: } S_{\text{inter}}/NP_{\text{nom}} := \lambda x. \text{sit}'x$$

koerer: $S_{\text{inter}}/\text{NP}_{\text{acc}}/\text{NP}_{\text{nom}} := \lambda x \lambda y. \text{drive}'yx$

ga: $S_{\text{inter}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}/\text{NP}_{\text{nom}} := \lambda x \lambda y \lambda z. \text{give}'zyx$

A constituent topicalizes by one of the following rules:

(11) *Norwegian Forward Noun Phrase Extraposition Rule (>fxp-np)*

$\text{NP}:a' \rightarrow_{\text{fxp-np}} S_{\text{decl}}/\$i/\phi(S_{\text{inter}}/\$i/\text{NP}): \lambda f. fa'$

(12) *Norwegian Forward Prepositional Phrase Extraposition Rule (>fxp-pptil)*

$\text{PP}_{\text{til}}:a' \rightarrow_{\text{fxp-pptil}} S_{\text{decl}}/\$i/\phi(S_{\text{inter}}/\$i/\text{PP}_{\text{til}}): \lambda f. fa'$

Effects of these rules on the derivations can be seen in the following sentences:

(13) $\frac{\text{Gyrd}}{\text{NP}: \text{Gyrd}'}$ $\frac{\text{ga}}{S_{\text{inter}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}/\text{NP}_{\text{nom}}: \lambda x \lambda y \lambda z. \text{give}'zyx}$ $\frac{\text{boka}}{\text{NP}: \text{book}'}$ $\frac{\text{til Inge}}{\text{PP}_{\text{til}}: \text{Inge}'}$

$\xrightarrow{\text{fxp-np}} S_{\text{decl}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}/\phi(S_{\text{inter}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}/\text{NP}_{\text{nom}}): \lambda f. f\text{Gyrd}'$

$\xrightarrow{S_{\text{decl}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}: \lambda y \lambda z. \text{give}'zy\text{Gyrd}'}$

$\xrightarrow{S_{\text{decl}}/\text{PP}_{\text{til}}: \lambda z. \text{give}'z\text{book}'\text{Gyrd}'}$

$\xrightarrow{S_{\text{decl}}: \text{give}'\text{Inge}'\text{book}'\text{Gyrd}'}$

(14) $\frac{\text{Boka}}{\text{NP}: \text{book}'}$ $\frac{\text{ga}}{S_{\text{inter}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}/\text{NP}_{\text{nom}}: \lambda x \lambda y \lambda z. \text{give}'zyx}$ $\frac{\text{Gyrd}}{\text{NP}: \text{Gyrd}'}$ $\frac{\text{til Inge}}{\text{PP}_{\text{til}}: \text{Inge}'}$

$\xrightarrow{\text{fxp-np}} S_{\text{decl}}/\text{PP}_{\text{til}}/\phi(S_{\text{inter}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}): \lambda f. f\text{book}'$

$\xrightarrow{S_{\text{decl}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}: \lambda y \lambda z. \text{give}'zy\text{Gyrd}'}$

$\xrightarrow{S_{\text{decl}}/\text{PP}_{\text{til}}: \lambda z. \text{give}'z\text{book}'\text{Gyrd}'}$

$\xrightarrow{S_{\text{decl}}: \text{give}'\text{Inge}'\text{book}'\text{Gyrd}'}$

(15) $\frac{\text{til Inge}}{\text{PP}_{\text{til}}: \text{Inge}'}$ $\frac{\text{ga}}{S_{\text{inter}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}/\text{NP}_{\text{nom}}: \lambda x \lambda y \lambda z. \text{give}'zyx}$ $\frac{\text{Gyrd}}{\text{NP}: \text{Gyrd}'}$ $\frac{\text{boka}}{\text{NP}: \text{book}'}$

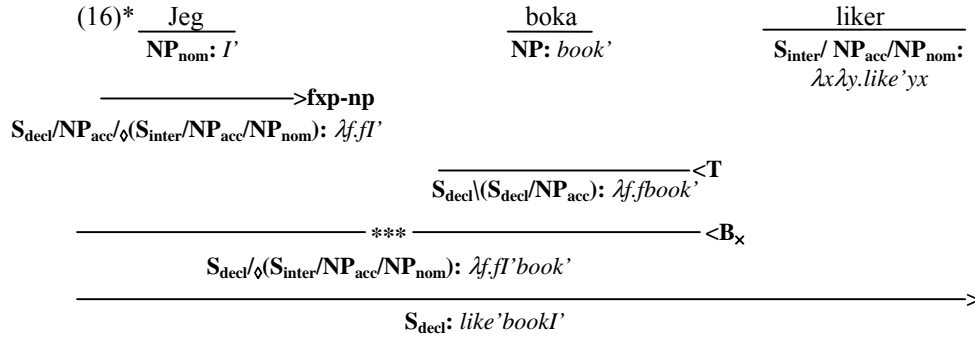
$\xrightarrow{\text{fxp-pptil}} S_{\text{decl}}/\phi(S_{\text{inter}}/\text{PP}_{\text{til}}): \lambda f. f\text{Inge}'$

$\xrightarrow{S_{\text{decl}}/\text{PP}_{\text{til}}/\text{NP}_{\text{acc}}: \lambda y \lambda z. \text{give}'zy\text{Gyrd}'}$

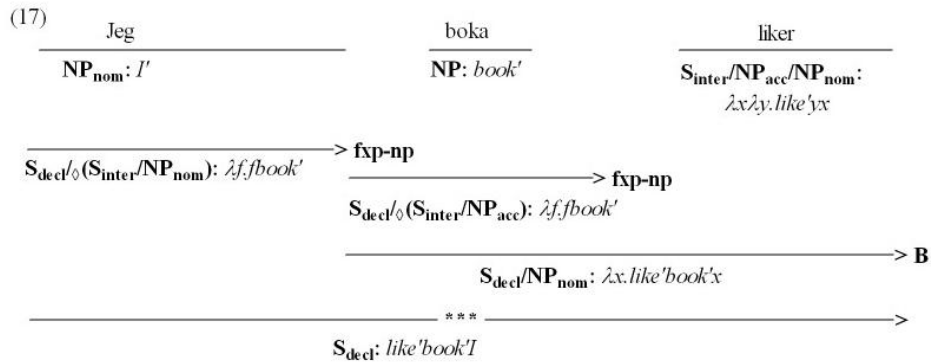
$\xrightarrow{S_{\text{decl}}/\text{PP}_{\text{til}}: \lambda z. \text{give}'z\text{book}'\text{Gyrd}'}$

$\xrightarrow{S_{\text{decl}}: \text{give}'\text{Inge}'\text{book}'\text{Gyrd}'}$

Diamond modality on the forward slash ‘/’ in the lexically raised category (occurring after forward extraposition) is necessary to block some unwanted combinations like the following:



Since these topicalization rules require an interrogative form to convert to declarative form, these rules cannot be applied twice subsequently in a simple clause. In this way V2 constraint is preserved by guaranteeing that not more than one element is topicalized, and one element must be topicalized in order to obtain declarative form. Below it can be seen how the parse of a doubly topicalized clause is blocked:



Note that current model is unable to decide the subject and object in main clauses. Since this is a semantic problem originating from ontology, we will show how to handle this in section 3.2. For now, assume that invalid parses of a sentence like “Gyrd liker Pizza” finding “Gyrd” as object and “Pizza” as subject are not produced by the model, and model correctly finds subject and object.

Besides its natural appearance to us, there is one more important reason why we choose to derive declarative forms from interrogative forms: What comes to the first position and topicalizes does not have to be one of the verb’s arguments. Following two examples illustrate this fact:

(18) i dag koerer hun bilen
 Today drive.PRES she car.THE
 ‘Today she drives the car’

(19) Gyrd tror jeg at ga boka til Inge
 Gyrd think.PRES I that give.PAST book.THE to Inge
 ‘I think that Gyrd gave the book to Inge’

When topicalized are not among its arguments, it's hard for a verb to guess which and how many elements come to the left of it, in order to preserve V2 constraint. In addition, in such a situation a mechanism that makes the verb take all of its arguments to the right must be developed. We believe that such options are not as linguistically sound as topicalization rules. (Steedman, 2000) and (Baldrige, 2002) use this approach as well, in their analysis of Dutch, which exhibits almost the same characteristics with Norwegian in main clauses.

2.2. SUBORDINATE CLAUSES

Since subordinate clause order is strictly SVO, there is no need for topicalization mechanisms like in declarative clauses. In model, a lexical rule for verbal categories is defined in order to obtain subordinate clauses:

(20) *Norwegian Subordinate Derivation Rule (inter-to-sub)*

$$S_{\text{inter}}/\$i/\text{NP}_{\text{nom}}:f \rightarrow_{\text{inter-to-sub}} S_{\text{sub}}/\$i\backslash\text{NP}_{\text{nom}}:f$$

In fact, this is no different from defining an additional ‘subordinate’ category for all verbs (as (Steedman, 2000) and (Baldrige, 2002) did in their Dutch analysis).

In our language model, we define the categories of the verb “tro” (“think” in English) and complementizer “at” as follows:

(21) tro: $S_{\text{inter}}/\blacktriangleleft S_{\text{sub}}/\text{NP}_{\text{nom}} := \lambda x \lambda s. \text{think}'sx$

(22) at: $S'_{\text{sub}}/\blacktriangleleft S_{\text{sub}} := \lambda f.f$

Like all verbs, when “tro” is used in a main clause it should obey the V2 constraint. However, current topicalization rules are insufficient to bring a constituent before “tro” in a main clause, since modality on the slash ‘/j’ in forward extraposed category $S_{\text{decl}}/\$i/\diamond(S_{\text{inter}}/j\$i/\text{NP})$ is incompatible with \blacktriangleleft modality in the “tro” category. One possibility is (like (Steedman, 2000) and (Baldrige, 2002) did) modifying the topicalization rules as follow:

(23) $\text{NP}:a' \rightarrow_{\text{fxp-np}} S_{\text{decl}}/\$i/\diamond(S_{\text{inter}}/\$i/\text{NP}): \lambda f.fa'$

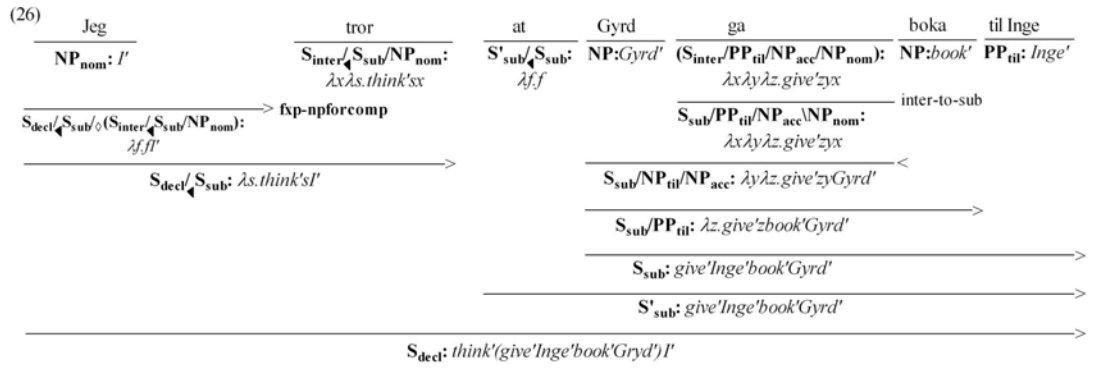
(24) $\text{PP}_{\text{til}}:a' \rightarrow_{\text{fxp-pptil}} S_{\text{decl}}/\$i/\diamond(S_{\text{inter}}/\$i/\text{PP}_{\text{til}}): \lambda f.fa'$

For this reason, rather than modifying the existing rules, we adopted to put a new forward extraposition rule for noun phrases, which applies only to the category of ‘tro’ (so to the similar words):

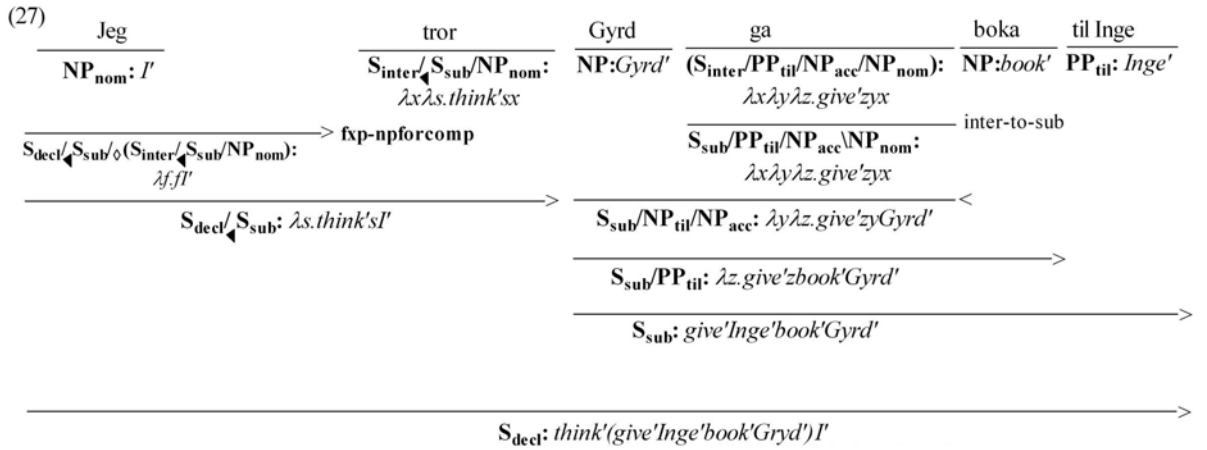
(25) *Norwegian Forward Noun Phrase Extraposition Rule for Subordinating Verbs (>fxp-npforsub)*

$$\text{NP}:a' \rightarrow_{\text{fxp-npforsub}} S_{\text{decl}}/\blacktriangleleft S_{\text{sub}}/\diamond(S_{\text{inter}}/\blacktriangleleft S_{\text{sub}}/\text{NP}_{\text{nom}}): \lambda f.fa'$$

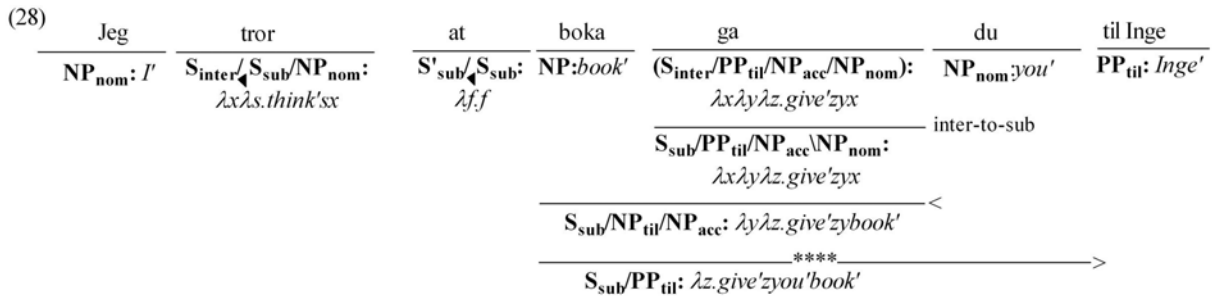
Now we are ready to show some sample derivations of simple sentences having a subordinate clause:



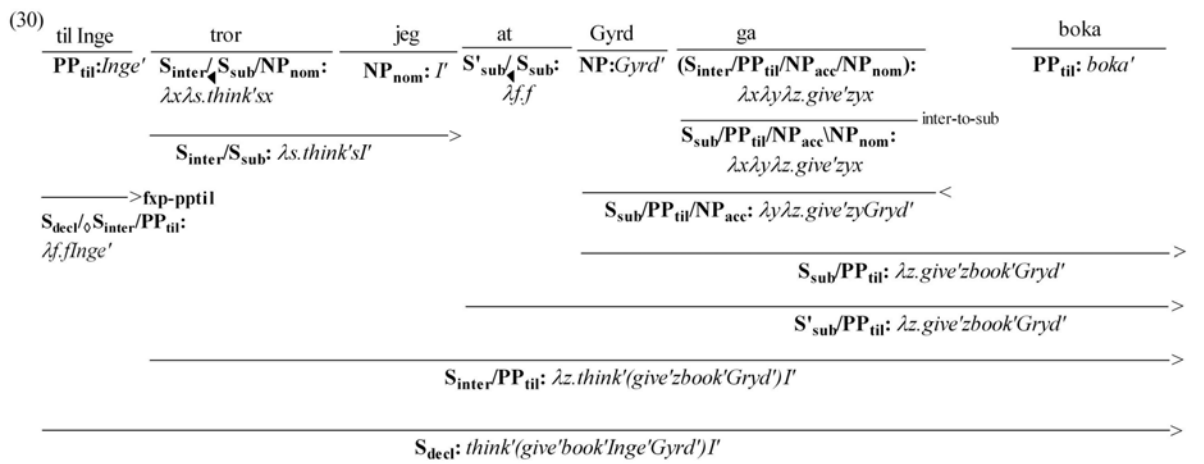
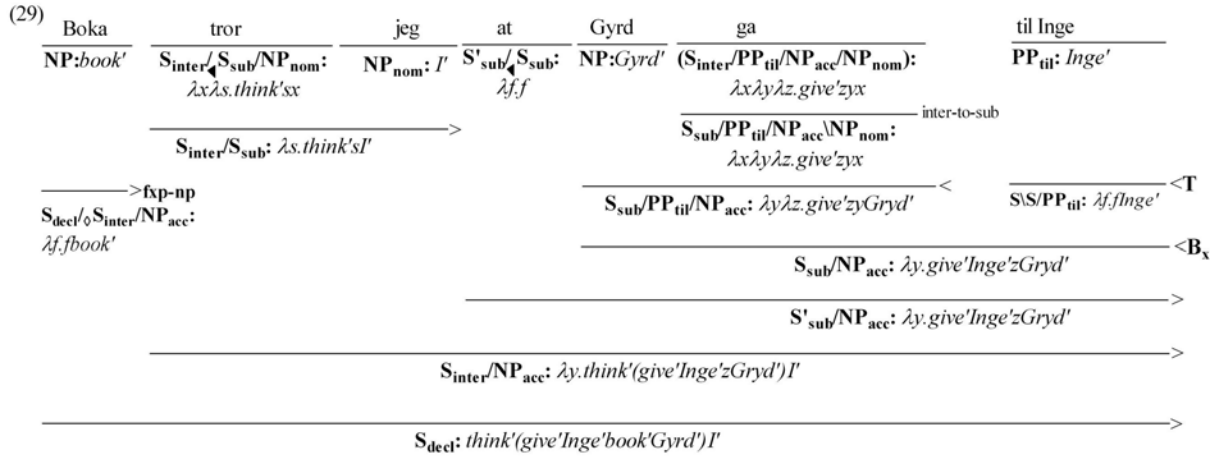
S'_{sub} means ‘complemented subordinate sentence’. In this respect, ‘at’ takes an uncomplemented subordinate clause and produces the complemented version of it. Note that S_{sub} argument of ‘tro’ does not specify whether the argument is complemented or not, so we can drop the complementizers as can be done in Norwegian:



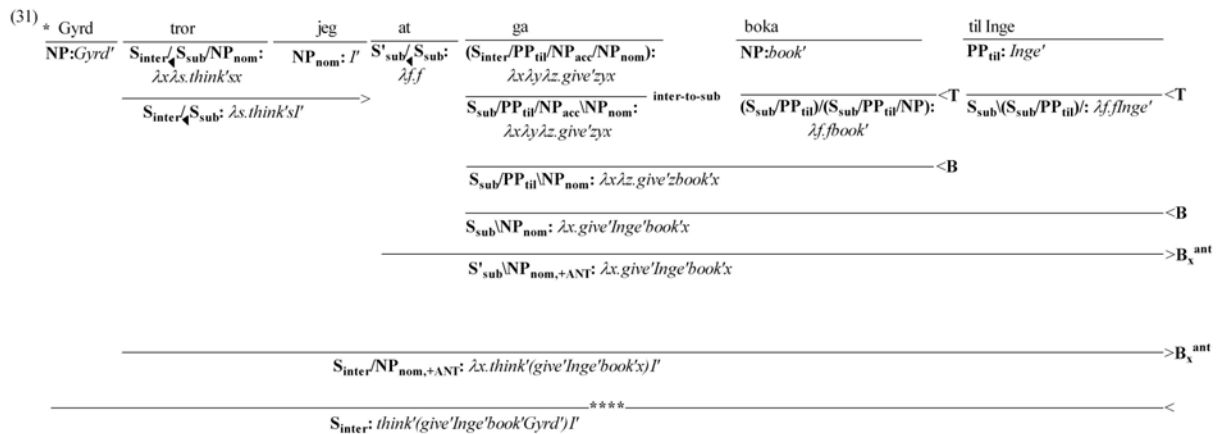
Word order of subordinate clause verbs correctly rules out the non-SVO orders, like the following:



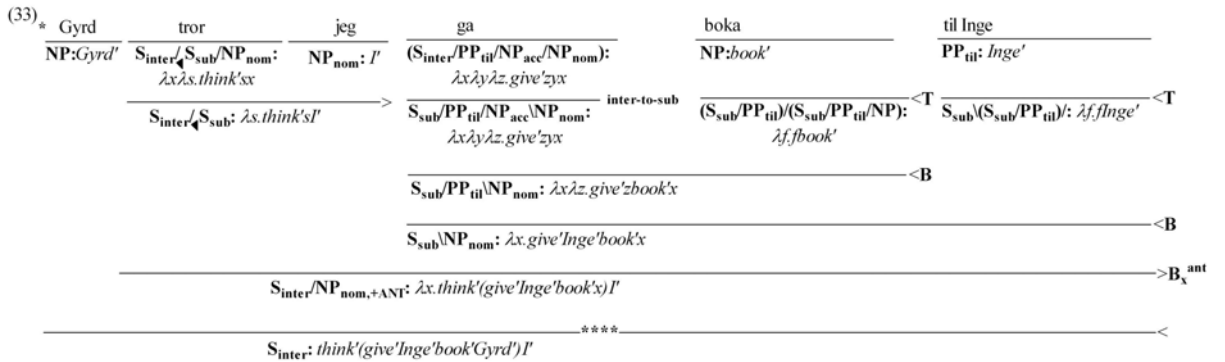
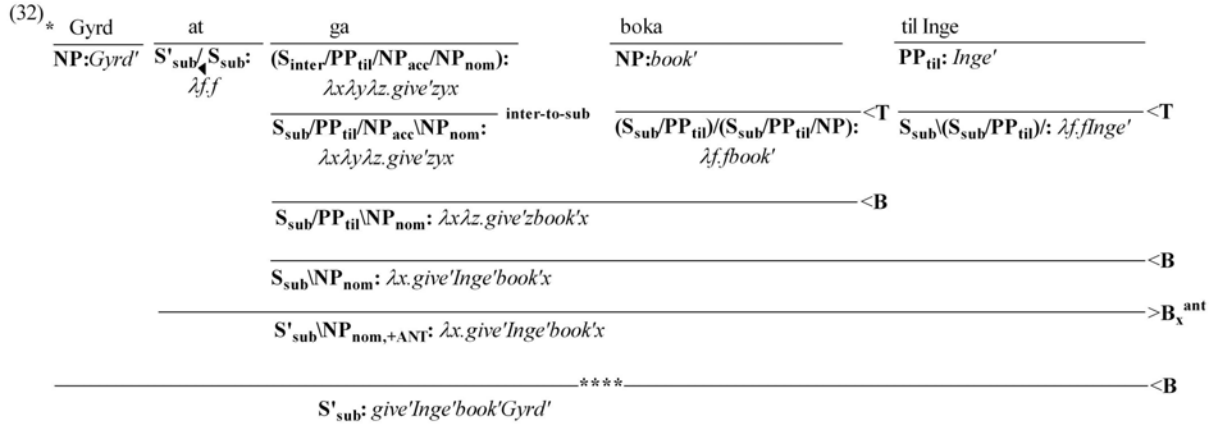
As can be seen in (19), it is also possible that one of the arguments of the verb in the subordinate clause is extracted to the first position in the main clause and topicalize, like in (29) and (30):



In order to extract and topicalize the subject of the subordinate clause, we will need some more rules since subordinate clause with missing subject will look for an antecedent to the left of it and will not go under simple functional composition with any NP (following the model that NP categories are marked as -ANT [non-antecedent] default, and remain -ANT when type-raised, as given in (Steedman, 2000) and (Baldrige, 2002)), as illustrated:



Antecedent government is the reason why we use ◀ modality in the categories of “tro” and “at”. In this way unwanted combinations with non-antecedent noun phrases and some other unwanted parses can be prevented, as illustrated below:



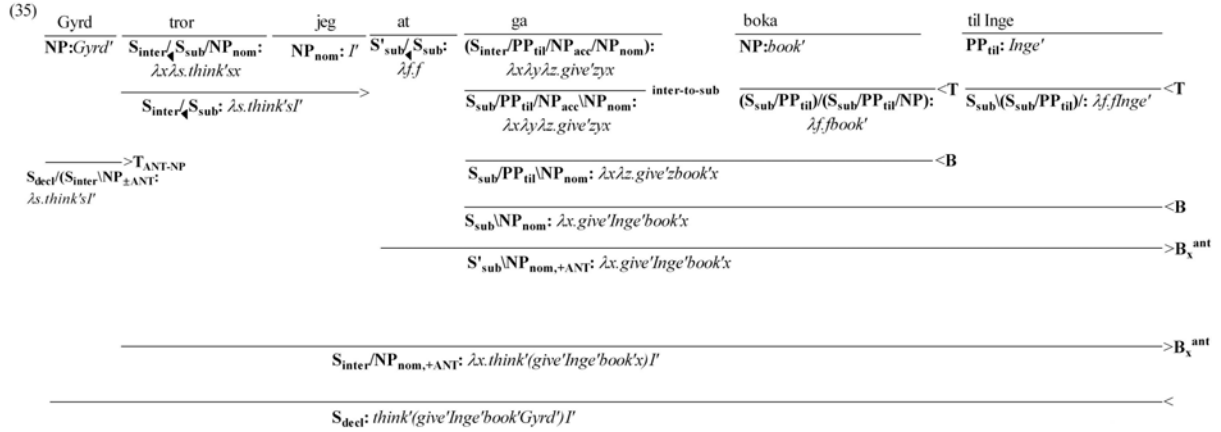
As can be guessed from (33), ◀ modality is used in the category of “tro” for cases with missing complementizers. Even in such situations antecedent government must be handled.

In order for antecedents to combine with categories looking for them, we add the following forward raising rule to the model:

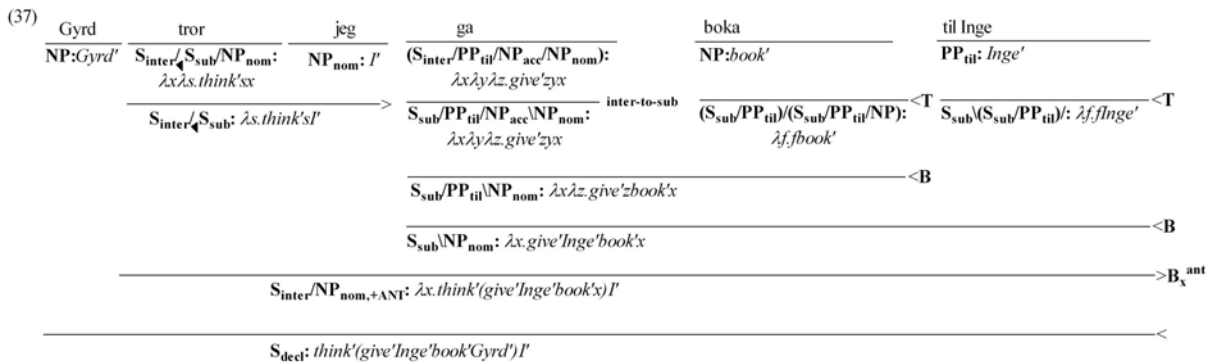
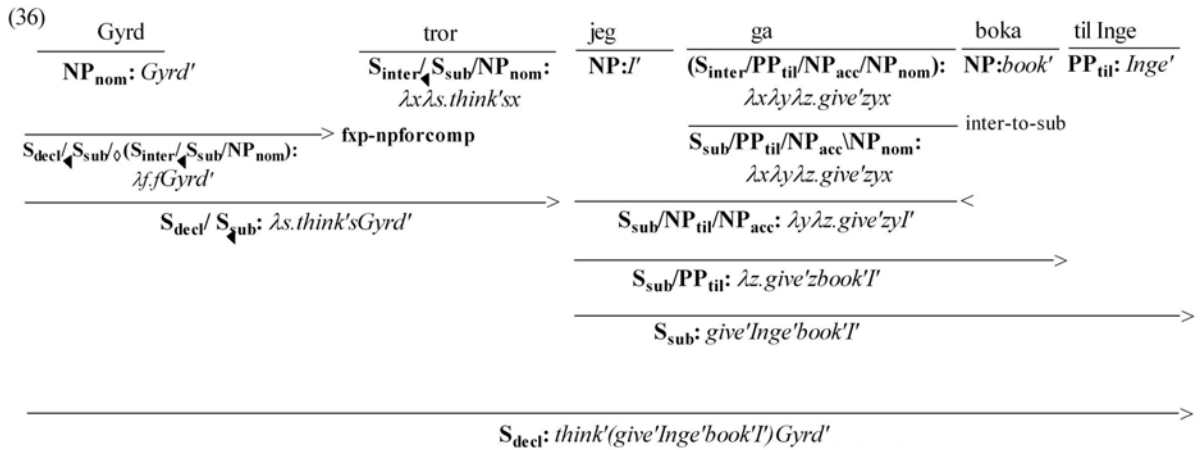
(34) Norwegian Forward Type Raising Rule for Antecedent NPs (>T_{ANT-NP})

$$\text{NP:a'} \rightarrow_{>T_{ANT-NP}} \text{S_{decl}/(S_{inter}/NP_{\pm ANT}): } \lambda f.f a'$$

Note that by taking an interrogative form and producing a declarative form, this rule also handles the topicalization of the antecedent NP, hence the V2 constraint. Now we can parse the following example:



Extractability of the subject creates an ambiguity when complementizer misses from the above sentence: “Gyrd tror jeg ga boka til Inge” can be interpreted as both “Gyrd tror at jeg ga boka til Inge” or “Gyrd tror jeg at ga boka til Inge”. This ambiguity is approved by native speakers (ambiguity seems to be resolved prosodically, in the written Norwegian “Gyrd, tror jeg, ga boka til Inge” unambiguously gives the second meaning). Fortunately, our model can guess this ambiguity as can be seen in (36) and (37):



As the final word, we should say that the whole subordinate clause can topicalize and come before the verb “tro”:

(38) At Gyrd ga boka til Inge tror jeg
That Gyrd give.PAST book-THE to Inge think.PRES I
'I think that Gyrd gave the book to Inge'

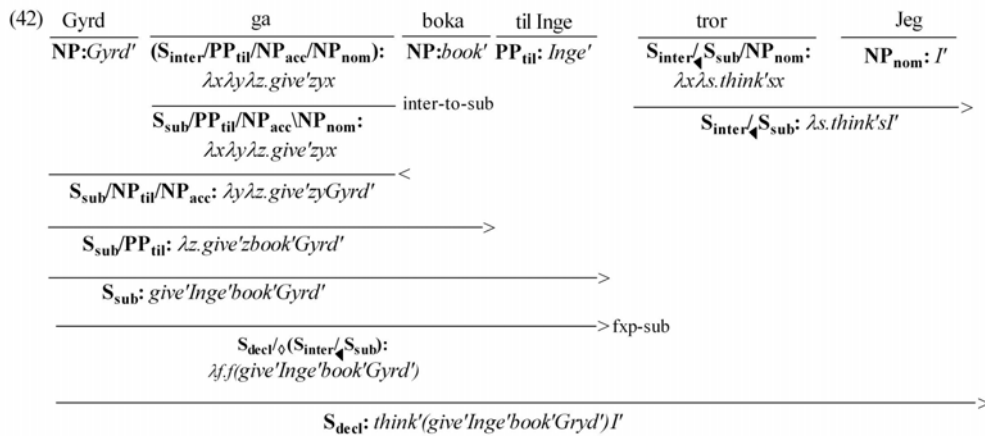
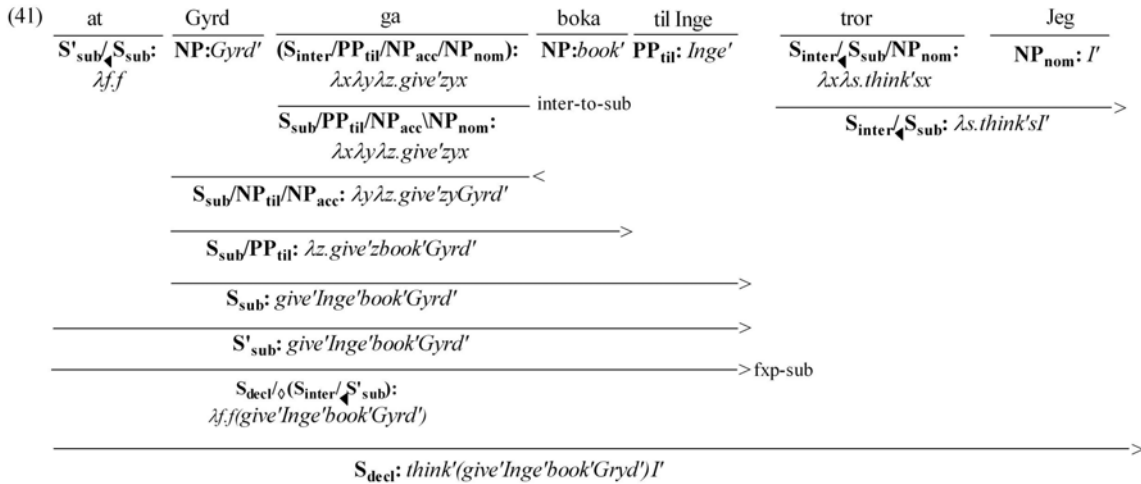
(39) Gyrd ga boka til Inge tror jeg
Gyrd give.PAST book-THE to Inge think.PRES I
'I think that Gyrd gave the book to Inge'

In order to handle such situations, we add a topicalization rule for subordinate clauses:

(40) *Norwegian Forward Subordinate Clause Extraposition Rule (>fxp-sub)*

$$S_{sub}:a' \rightarrow_{fxp-sub} S_{decl}/\diamond(S_{inter}/\blacktriangleleft S_{sub}): \lambda f.fa'$$

Now our model can give parses of (41) and (42) as follows:



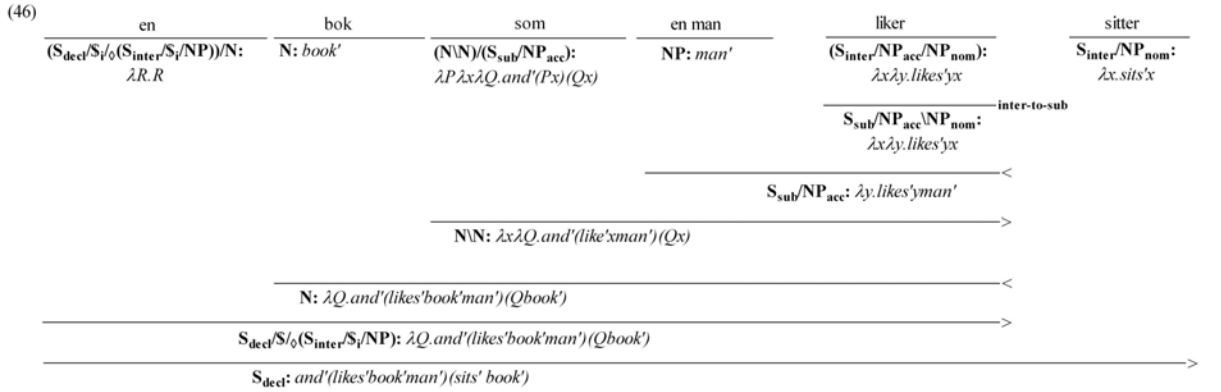
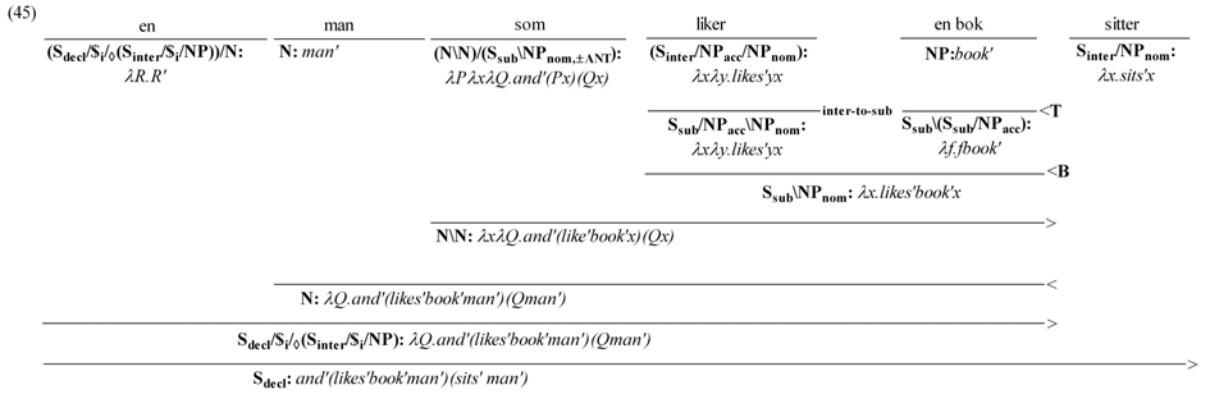
2.3. RELATIVE CLAUSES

Following the convention which counts relative clauses as noun modifiers (Steedman, 1996), we give the categories of the relative pronoun “som” as follow:

$$(43) \quad \text{som: } (N \setminus N) / (S_{\text{sub}} \setminus \text{NP}_{\text{nom}, \pm \text{ANT}}) := \lambda P \lambda x \lambda Q. \text{and}'(Px)(Qx)$$

$$(44) \quad \text{som: } (N \setminus N) / (S_{\text{sub}} / \text{NP}_{\text{acc}}) := \lambda P \lambda x \lambda Q. \text{and}'(Px)(Qx)$$

Since relative clauses are subordinate clauses, we reflect this expectation to the category of “som” by marking the ‘S’ argument as ‘S_{sub}’. Two simple examples including relative clauses are parsed below:



Note that we used the additional category (47) for indefinite particle ‘en’ (meaning ‘a’ in English). This category is inspired from its English counterpart introduced in (Steedman, 1996). However, since whatever it produces must topicalize when it becomes the first constituent in a main clause (due to V2 constraint), like in the examples above, category is arranged accordingly:

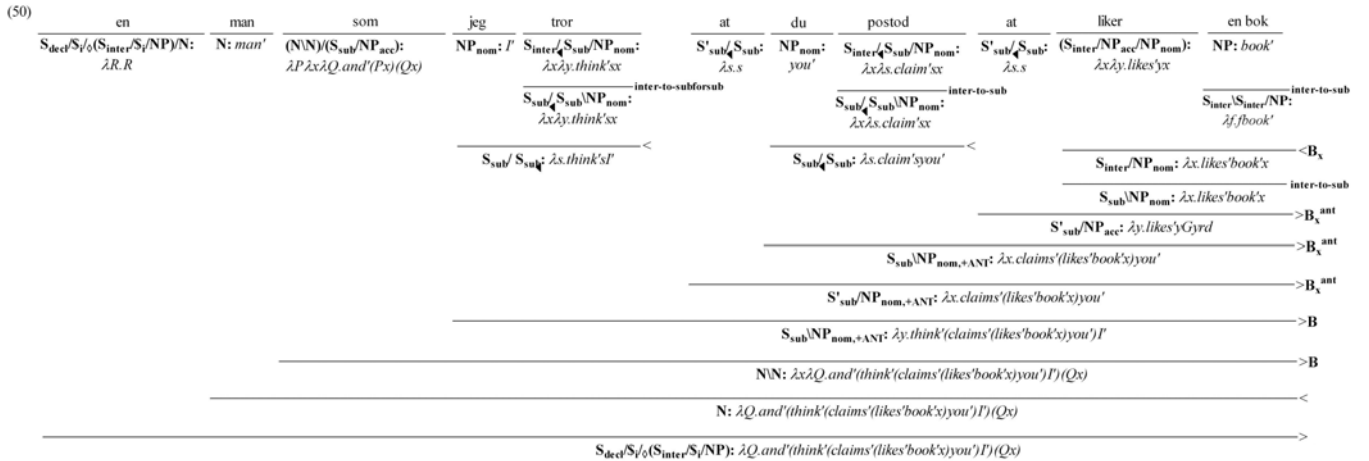
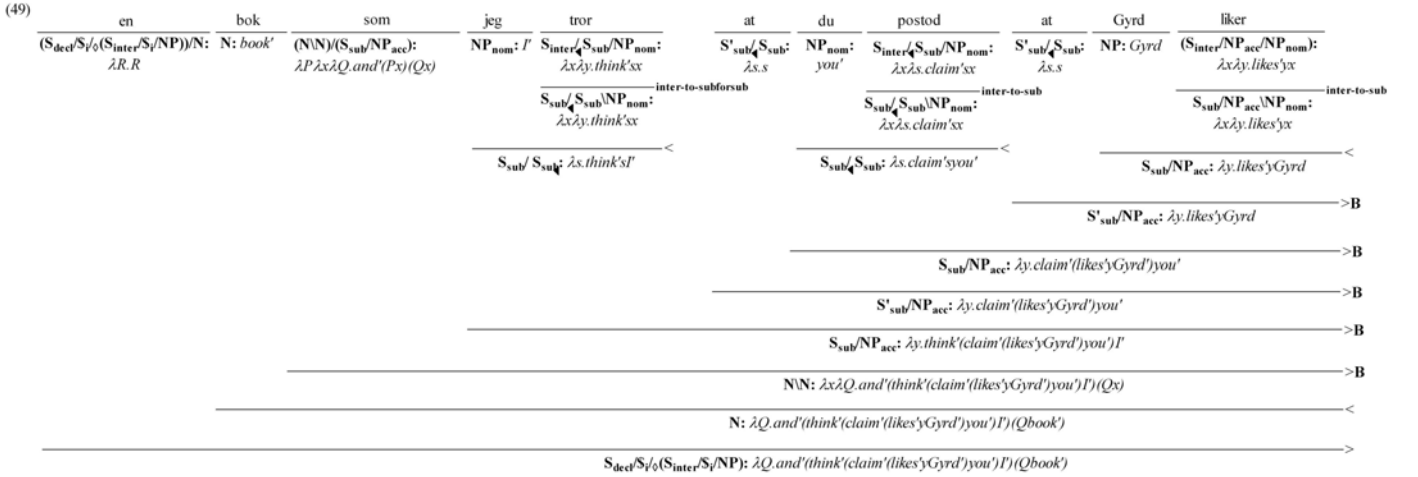
$$(47) \quad \text{a: } (S_{\text{decl}}/S_i/\phi(S_{\text{inter}}/S_i/\text{NP}))/N := \lambda R.R$$

Unlike English, both subjects and objects can be long distance extracted out of relative clauses. Current state of model implies this naturally, since we can extract both subject and object out of complement clauses. However, before predicting this syntactic symmetry, we should add another rule to derive subordinate clauses using subordinating verbs like ‘tro’:

(48) Norwegian Subordinate Derivation Rule for Subordinating Verbs (inter-to-subforsub)

$$S_{\text{inter}} / \leftarrow S_{\text{sub}} \backslash \text{NP}_{\text{nom}} : f \quad \rightarrow_{\text{inter-to-subforsub}} \quad S_{\text{sub}} / \leftarrow S_{\text{sub}} \backslash \text{NP}_{\text{nom}} : f$$

Now we can show both kinds of long distance extractions:



3. SEMANTIC ASYMMETRIES

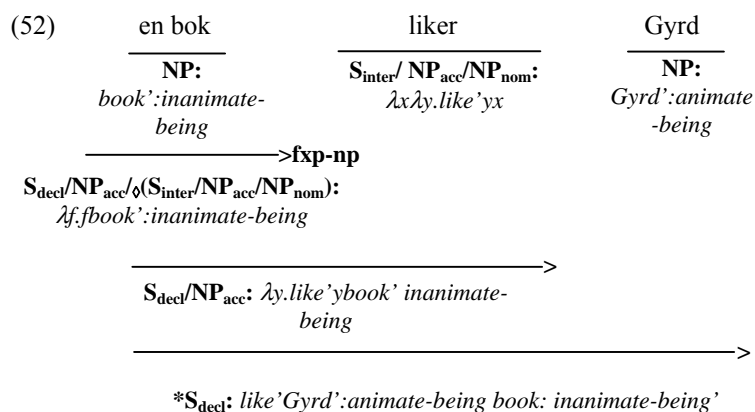
3.1. BINDING

3.2. SUBJECT-OBJECT AMBIGUITY

As stated before, VSO+topicalization creates two options (SVO & OVS) for interpretation, syntax provides no solution to solve the ambiguity in sentences. (Øvrelid, 2004), addressing the same problem, published some statistics from Oslo Corpus comparing animacy and definiteness properties of subjects and objects in sentences. Having seen that in 97.6% of the sentences object is not higher than subject in animacy, she proposed the following constraint:

- (51) *Norwegian Animacy Constraint*: In a sentence, object cannot be higher than subject in animacy.

Subject and object show a statistical asymmetry in definiteness as well, but since definiteness results are more uncertain, we will not give place to definiteness in our model's semantic constraints. Since we can use final logical form to identify the subject and object, we can apply the constraint to obtained sentences. In order to specify animacy for constituents, we can use T-CCG (McConville, 2005) to define an ontology of the world, which is reflected to arguments as features. A very simple ontology discriminating only animate and inanimate objects would work in parsing the following sentence.



Here, the reading *like'Gyrd': animate-being book: inanimate-being'* is blocked due to the fact that subject is lower than object in animacy.

4. CONTROL CONSTRUCTIONS

Norwegian control verbs shows similarity to English in structure. There are two kinds of control verbs, one for subject control and one for object control:

- (53) Jeg overbeviste Gyrd om aa bade Inge
 I persuade.PAST Gyrd (about) to bathe Inge
 'I persuaded Gyrd to bathe Inge'
- (54) Jeg lovte Gyrd aa bade Inge
 I promise.PAST Gyrd to bathe Inge
 'I promised Gyrd to bathe Inge'

Different from English, infinitival complements cannot undergo heavy noun phrase shift:

- (55) *Jeg overbeviste om aa bade Inge Gyrd
 I persuade.PAST (about) to bathe Inge Gyrd
 'I persuaded Gyrd to bathe Inge'
- (56) *Jeg lovte aa bade Inge Gyrd
 I promise.PAST to bathe Inge Gyrd
 'I promised Gyrd to bathe Inge'

As reader may notice, the only syntactic difference between subject and object control verbs is head of their infinitival complements: While object control verbs use infinitival complements with head “om”, subject control verbs use the head “aa” instead.

Control verbs are no exception to the V2 constraint. Like all other verbs, their arguments have a strict relative order (this order is subject, object and infinitival complement, respectively) and any of the arguments may topicalize:

- (57) Overbeviste jeg Gyrd om aa bade Inge
persuade.PAST I Gyrd (about) to bathe Inge
'Did I persuade Gyrd to bathe Inge?'
- (58) Jeg overbeviste Gyrd om aa bade Inge
I persuade.PAST Gyrd (about) to bathe Inge
'I persuaded Gyrd to bathe Inge'
- (59) Gyrd overbeviste jeg om aa bade Inge
Gyrd persuade.PAST I (about) to bathe Inge
'I persuaded Gyrd to bathe Inge' (Gyrd emphasized)
- (60) om aa bade Inge overbeviste jeg Gyrd
(about) to bathe Inge persuade.PAST I Gyrd
'I persuaded Gyrd to bathe Inge' ("to bathe Inge" emphasized)
- (61) Lovte jeg Gyrd aa bade Inge
promise.PAST I Gyrd to bathe Inge
'Did I promise Gyrd to bathe Inge?'
- (62) Jeg lovte Gyrd aa bade Inge
I promise.PAST Gyrd to bathe Inge
'I promised Gyrd to bathe Inge'
- (63) Gyrd lovte jeg aa bade Inge
Gyrd promise.PAST I to bathe Inge
'I persuaded Gyrd to bathe Inge' (Gyrd emphasized)
- (64) aa bade Inge overbeviste jeg Gyrd
to bathe Inge persuade.PAST I Gyrd
'I persuaded Gyrd to bathe Inge' ("to bathe Inge" emphasized)

An important point is that infinitival complements are islands in a sense, i.e., no constituent inside an infinitival complement can be extracted and topicalized:

- (65) * om aa bade overbeviste jeg Gyrd Inge
(about) to bathe persuade.PAST I Gyrd Inge
- (66) *Inge overbeviste jeg Gyrd om aa bade
Inge persuade.PAST I Gyrd (about) to bathe
- (67) * aa bade Inge overbeviste jeg Gyrd om
to bathe Inge persuade.PAST I Gyrd (about)
- (68) * aa bade lovte jeg Gyrd Inge
to bathe persuade.PAST I Gyrd Inge
- (69) *Inge lovte jeg Gyrd aa bade
Inge promise.PAST I Gyrd to bathe

Now we will present the new categories capturing the control behavior of the language:

4.1. OBJECT CONTROL

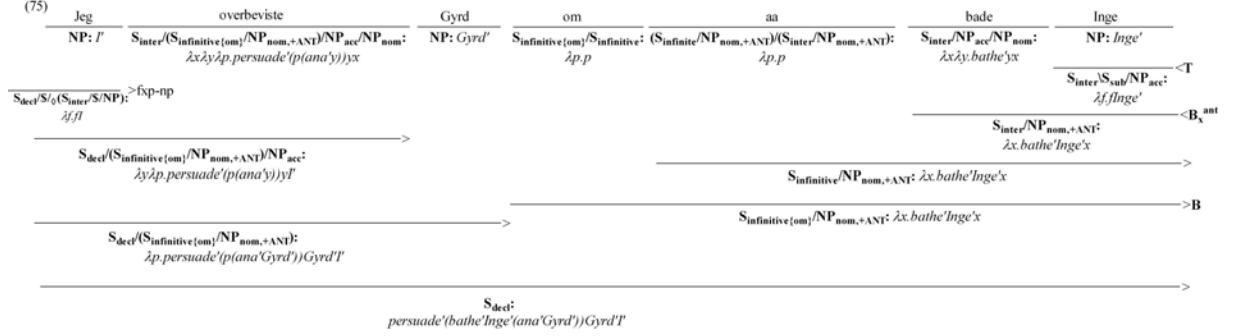
Our presentation for the object control verb category is somewhat similar to the one proposed in (Steedman & Baldrige, 2005) with two important differences: 1) Due to the structure of the language, directionalities of the category are different, 2) In (Steedman & Baldrige, 2005), argument category of

The head “om” does not make any change in the structure and functionality of the infinitival clause, so we designed its category to reflect this behavior:

$$(74) \quad \text{om: } S_{\text{infinitive}\{\text{om}\}}/S_{\text{infinitive}} := \lambda f.f$$

Here we must note a hidden assumption done throughout the text that head of any infinitive clause is by default ‘aa’, unless specified otherwise. This assumption is important since there are times when head of the infinitival clause is vital to block unwanted combinations (as we will see).

Now we are ready to parse a basic sentence with object control:



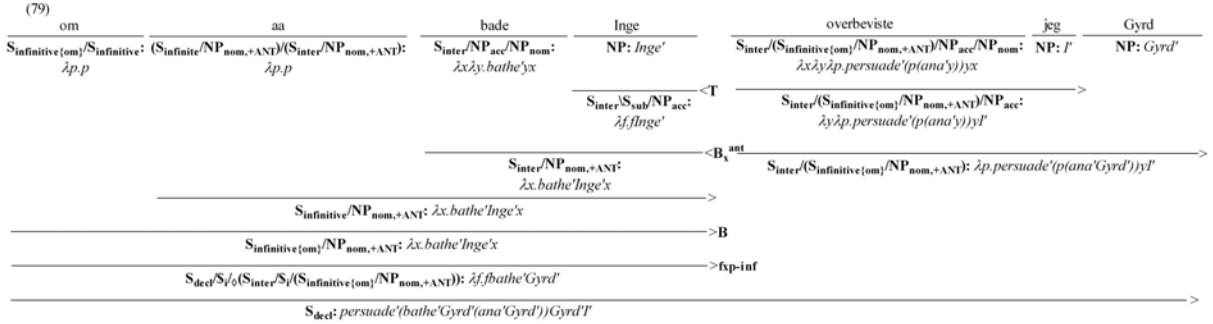
To topicalize the whole infinitival complement, we must add a new topicalization rule as follows:

(78) *Norwegian Forward Infinitive Clause Extraposition Rule (>fxp-inf)*

$$S_{\text{infinitive}}/NP_{\text{nom,+ANT}}: \lambda x.fx \rightarrow \text{fxp-inf}$$

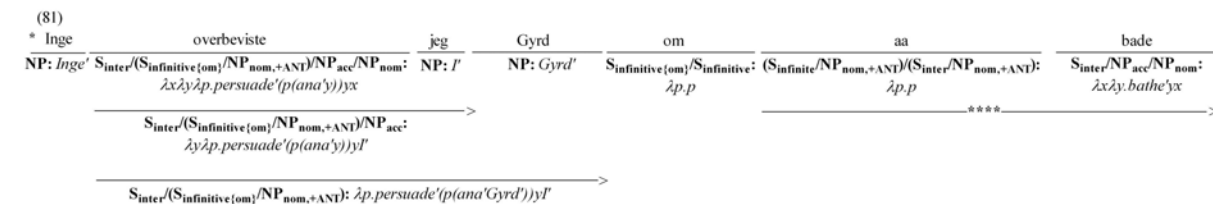
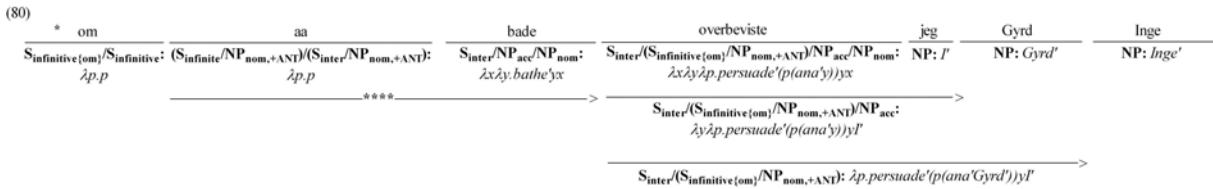
$$S_{\text{decl}}/S_i/\phi(S_{\text{inter}}/S_i/(S_{\text{infinitive}}/NP_{\text{nom,+ANT}})): \lambda g.gf$$

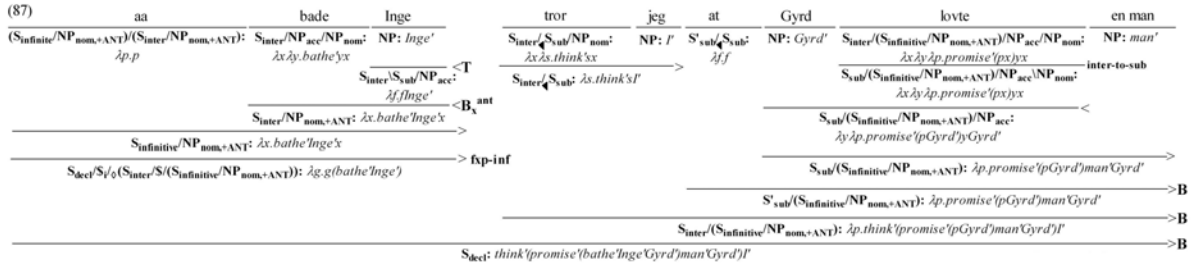
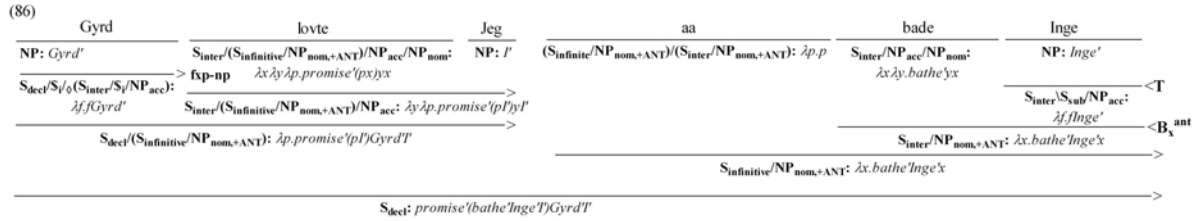
Semantic part of the rule may come weird, but the example below may show the necessity:



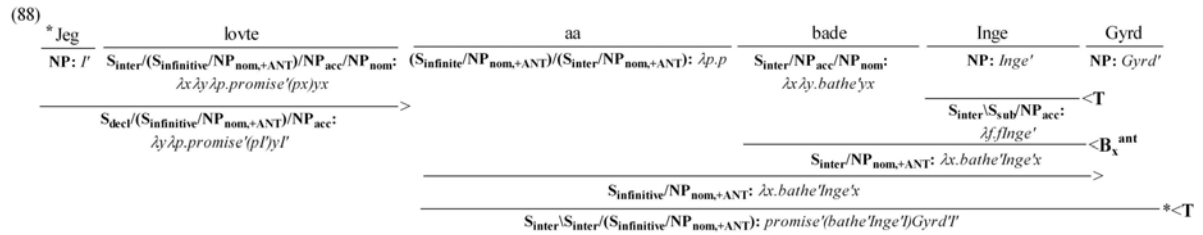
Since missing subject NP of the infinitival clause will be never found, λx argument should be handled somehow. Following the fact that the semantic category of the control verb already provides a subject for the infinitival clause not necessarily looking for a subject, what we need is to get rid of x argument in the semantic.

Following examples illustrates how the model blocks the parses with extracted elements of the infinitival complement. In the first two examples are blocked due to that infinitive marker takes a verb phrase with only a subject argument, so incomplete verb phrases cannot combine with it. In the third example, forward extraposition of infinitival complement does not work due to the head disagreement: Head of the topicalized infinitival clause is “aa”, whereas the control verb looks for one with head “om”:

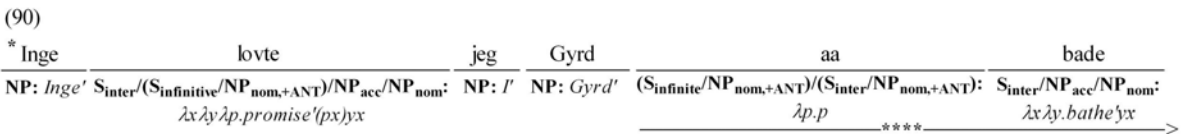
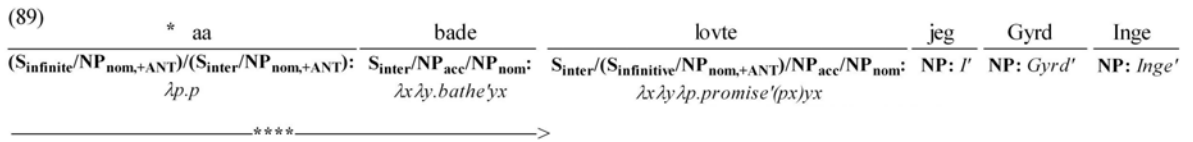




As in subject control, constraint on the heavy noun phrase shift of arguments can be predicted via non-availability of type-raising for infinitive clause:



Islandhood of the infinitival complement is provided like in (80) and (81)



5. COORDINATION

5.1. COORDINATE STRUCTURE CONSTRAINT

5.2. NODE RAISING

5.3. GAPPING

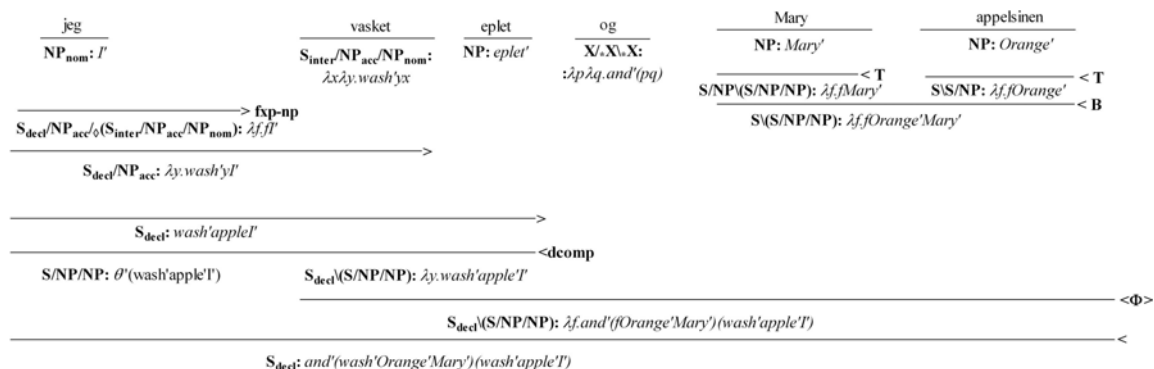
Verb gapping behavior of the Norwegian language clearly shows two characteristics:

1. There is no backward verb gapping
2. If there is forward verb gapping, in order for sentence to be grammatical ungapped conjunct must obey V2 constraint

Since our model generates only VSO+topicalization and SVO verb orders, this result is not surprising. Following (Steedman, 2000), we have used decomposition rules to analyze gapping in main clauses verb-medial appearance of main clauses make a node raising analysis impossible. A sample parse can be seen as follows:

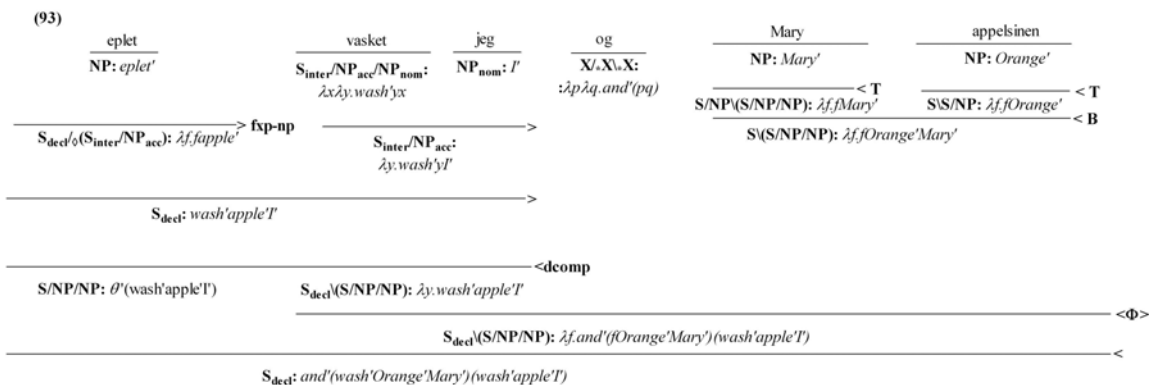
- (91) Jeg vasket eplet og Mary appelsinen
 I washed apple.THE and Mary orange.THE
 ‘I washed the apple and Mary (washed) the orange’

□



- (92) Eplet vasket jeg og Mary appelsinen
 apple.THE washed I and Mary orange.THE
 ‘I washed the apple and Mary (washed) the orange’

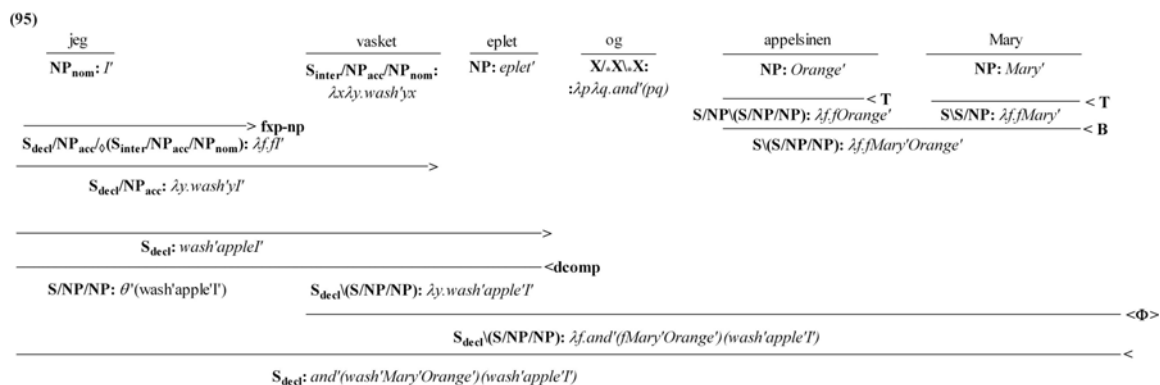
We can see that this is natural when we look at the parse of this sentence:



Since what defines the subjectness and objectness of the arguments in the gapped conjunct is their type raisings and combination, it does not change with the changes in the gapped conjunct. In fact, looking at derivations we can see that the only way that “A B” produce an argument cluster is A’s being subject and B’s being object. Following sentence approve this result:

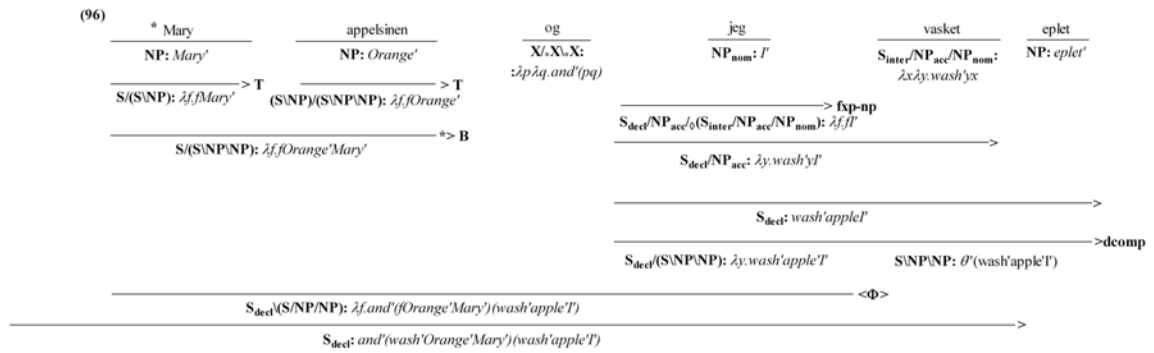
(94) Jeg vasket eplet og appelsinen Mary
 I washed apple.THE and orange.THE Mary
 ‘I washed the apple and orange (washed) Mary’

Native speakers unambiguously interpretes the sentence this way, although they find it quite strange (possibly because of the violation of animacy constraint). However, CCG model correctly predicts this:



As stated, if ungapped conjunct is not V2, sentence is invalid. This is also an expected result, since the ungapped conjunct cannot combine into an S, to later decompose

That backward verb gapping is ungrammatical in Norwegian is as expected either, since only verb orders in Norwegian are VSO and SVO. Since neither is verb final, using forward decomposition rule, together with clustering of gapped left conjunct into a category looking for a verb taking all its arguments to the left (which is illustrated in (96)) will fail “by the restrictions on type-raising to raising to parametrically licensed categories permitted by Type Transparency” (Steedman, 2000).



Note that Norwegian behaves differently from Dutch and German (which are very close to Norwegian), since SOV verb order of subordinate clauses of Dutch and German enables the parametrical licensing of S/(S\NP\NP)

6. REFERENCES

- (Ades & Steedman, 1982) Ades, Anthony, and Mark Steedman. 1982. "On the Order of Words." *Linguistics and Philosophy*, 4, 517–558.
- (Bach, 1979) Bach, Emmon. 1979. "Control in Montague Grammar." *Linguistic Inquiry*, 10, 513–531.
- (Baldrige, 2002) Baldrige, Jason. 2002. *Lexically Specified Derivational Control in Combinatory Categorical Grammar*. Ph.D. thesis, University of Edinburgh.
- (Dowty, 1979) Dowty, David. 1979. "Dative Movement and Thomason's Extensions of Montague Grammar." In Steven Davis and Marianne Mithun, eds., *Linguistics, Philosophy, and Montague Grammar*, 153–222. Austin, TX: University of Texas Press.
- (Dowty, 1988) Dowty, David. 1988. "Type-Raising, Functional Composition, and Nonconstituent Coordination." In Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler, eds., *Categorical Grammars and Natural Language Structures*, 153–198. Dordrecht: Reidel.
- (Ellingsen, 2004) Ellingsen, Liv. 2004. "Norwegian word order in HPSG". In *Proceedings of the 11th International Conference on HPSG*, Leuven, Belgium, 2004
- (Lodrup, 2001) Lodrup, Helge: *Pseudocoordinations in Norwegian and Control Theory*, Proceedings of the 6th International Lexical-Functional Grammar Conference
- (McConville, 2005) McConville, Mark. 2005. "Type-hierarchical CCG." Ms., University of Edinburgh.
- (Øvrelid, 2004) Øvrelid, Lilja: *Disambiguation of syntactic functions in Norwegian*, Proceedings of 20th Scandinavian Conference of Linguistics, 2004
- (Steedman, 1987) Steedman, Mark. 1987. "Combinatory Grammars and Parasitic Gaps", *NLLT*, 5, 403–439.
- (Steedman, 1988) Steedman, Mark. 1988. "Combinators and Grammars", in Richard Oehrle et al. (eds.), *Categorical Grammars and Natural Language Structures*, D. Reidel, Dordrecht, pp. 417–42.

(Steedman, 1990) Steedman, Mark. 1990. "Gapping as Constituent Coordination", *Linguistics and Philosophy*, 13, 207–263.

(Steedman, 1996) Steedman, Mark. 1996. *Surface Structure and Interpretation*, MIT Press, Cambridge, Massachusetts.

(Steedman, 2000) Steedman, Mark. 2000. *The Syntactic Process*. Cambridge, MA:MIT Press.

(Steedman & Baldrige, 2005) Steedman, Mark, and Jason Baldrige. 2005. "Combinatory Categorical Grammar." In Kirsti Börjars, ed., *Non-transformational syntax: a guide to current models*. Blackwell. To appear.