Case Marking and Subject Extraction in Danish[1]

Bjarne Ørsnes
Copenhagen Business School

Proceedings of the LFG02 Conference
National Technical University of Athens, Athens
Miriam Butt and Tracy Holloway King (Editors)
2002
CSLI publications
HTTP://CSLI-PUBLICATIONS.STANFORD.EDU/

[1]For helpful discussion, I am grateful to Mary Dalrymple, Line Mikkelsen, Sten Vikner and the audience at LFG02.

1 Introduction

The empirical domain of case theory involves four types of phenomena ([Lee, 2002b]): The marking of core grammatical functions, semantically induced case marking, dependency effects and domain effects. Common to the first three of these phenomena is that case is determined on the basis of local head-dependent relationships, but also "case Stacking" which is given as an illustration of a domain effect in [Lee, 2002b], can be argued to involve a local head-dependent relationship between a nominal and its nominal adjuncts. In this paper I discuss another kind of domain effect where case assignment, however, can not be determined on the basis of a head-dependent relation. Instead, case assignment is determined by the specific syntactic construction. The case in point is the accusative marking of a (non-locally) extracted pronominal subject in Danish.

In Danish pronominal subjects are assigned nominative case, but non-locally extracted pronominal subjects (subjects extracted across a clause boundary) are assigned accusative case. The use of nominative for an extracted subject inevitably brings out a reading of the pronominal as the matrix subject. It is argued that this use of the accusative for an extracted pronoun has a clearly disambiguating function. Accusative signals that the fronted constituent is not the subject of the matrix clause (but possibly the subject of an embedded clause). The paper shows how this generalization can be represented in LFG and, concomitantly, how this kind of constructional case assignment can be accommodated in a lexicalist framework such as LFG.

A further challenge is to account for the distribution of the personal pronouns in a language which otherwise does not employ morphological function specification. Only the personal pronouns exhibit a morphological distinction between nominative and accusative. From a monolingual point-of-view, it is thus empirically inadequate to postulate the existence of nominative and accusative case for other kinds of nominals. A crucial point in the present analysis is to restrict the account to the relevant class of lexical items. I demonstrate how the *Constructive Case*-approach of [Nordlinger, 1998] is capable of accomplishing exactly this.

The accusative marking of non-locally extracted subjects is observed in an informal register of Danish while standard Danish does not seem to allow extraction of pronominal subjects at all. The use of an accusative pronoun to mark an extracted subject is associated with a certain stylistic effect (a colloquial register), and this use is not accepted by all speakers. I will show how this pattern of variation can be accounted for by a small difference in the lexical entries of the personal pronouns. In addition, it is shown how the observed variation in extraction of pronominal subjects can be accommodated within the *Constructive Case*-approach.

Central to the discussion is the question whether an extracted subject is identified in terms of c-structure properties or in terms of f-structure properties. This discussion sheds important light on the syntax of extraction in Danish and I conclude that the extracted subject can only be identified in f-structure terms. It emerges from the analysis that an extracted object is associated with an empty category while an extracted subject is not.

The present analysis has been implemented in XLE (Xerox Linguistic Environment) as part of a broad-coverage LFG-grammar for Danish.[2]

[2]This work is supported by a grant from Nordisk Ministerråd as part of the language technology programme.

2 The Distribution of the Personal Pronouns

Apart from genitive which is only used to mark nominal attributes, case no longer serves to identify grammatical functions in Danish. Grammatical functions are identified on a configurational basis: objects occur in VP or in I ("object-shift", [Sells, 2001]) while the subject canonically occurs in the specifier position of IP. However, Danish is a V2 language and allows for the topicalization of almost any kind of constituent. In these cases, grammatical functions can frequently only be determined on the basis of selectional restrictions. Morphological function specification is only observed in conjunction with the personal pronouns whose distribution is sensitive to their grammatical function.[3] A distinction between nominative and accusative is observed in the 1st and 2nd person pronouns. Among the 3rd person singular pronouns, the natural-gender specific pronouns, *han*/'he' and *she*/'she', have an accusative form and only take human antecedents. The grammatical-gender specific pronouns *det*/'it' and *den*/'it' have no accusative form and only take non-human antecedents. The 3rd person plural pronoun *de*/'they' has an accusative form and take both human and non-human antecedents.

The basic generalization about the distribution of the personal pronouns in Danish is that the nominative forms are used for subjects (SUBJ) while the accusative forms are used for all other grammatical functions OBJ, OBJ2, XCOMP and in "case-less" positions, i.e. positions which are not associated with an argument function (e.g. "adjoined-to"-positions).

(1) *han* vinder (SUBJ)
he.NOM wins

(2) det er *ham* der vinder (XCOMP)
it is *him*.ACC who wins
'he is the one who is winning'

(3) hun giver *ham* præmien (OBJ2)
she gives *him*.ACC the award

(4) hun giver præmien til *ham* (OBJ)
she gives the award to *him*.ACC

(5) *dig*, du kan gå din vej ("case-less" position)
You.ACC, you.NOM can go your way
'As for you, you can go your own way'

Contrary to the proposed markedness-hierarchies of case forms in [Woolford, 2001] and [Lee, 2002a], the accusative form seems to be the most unmarked form in Danish. The accusative form is used unless there is good reason not to. In this respect the Danish personal pronouns pattern with the English personal pronouns. [Hudson, 1990] claims that case no longer exists in English. Instead, we find a (closed) class of lexical items whose distribution is determined by their grammatical function. The generalization in Hudson is that the nominative forms only occur with finite verbs. Since the nominative forms impose a constraint on their governing head, i.e. that it be finite, Hudson analyses the distribution of the nominative

[3] The distribution of the reflexive pronouns could also be determined by grammatical function. The present analysis can straight-forwardly be extended to cover these pronouns as well. Since, however, their distribution also can be claimed to follow from syntactic constraints on coreferentiality they are excluded from the present discussion.

pronouns as head-marking. In Danish the use of the nominative forms is even more restricted: nominative forms are only used when the pronominal is a local subject. A non-local subject is realized with its accusative form giving rise to a kind of movement paradox:

(6) Peter tror *han* vinder
Peter thinks *he* wins
'Peter thinks he is going to win'

(7) *ham* tror Peter *e* vinder
him thinks Peter wins
'he is the one of whom Peter believes that he is going to win'

(8) * *han* tror Peter *e* vinder
he thinks Peter wins
'he is the one of whom Peter believes that he is going to win'

In (7), the pronoun *ham*/'him' is the subject of the embedded verb *vinder*/'wins'. Example (8) is only possible on a reading with *han*/'he' as the matrix subject.[4]

In non-subject extraction configurations, the pronominal retains its case:

(9) *ham* tror jeg ikke jeg kender *e*
him.ACC think I not I know *e*
'As for him I don't think I know him'

The extraction of a pronominal subject is not accepted by all speakers. Speakers who reject (7) (i.e. the accusative marking of an extracted subject), also reject (8). For those speakers extraction of a pronominal subject is not possible at all. The accusative marking in (7) belongs to an informal register, but it is not restricted to spoken language. On the contrary, it is common on internet-pages from where the following examples are extracted.

(10) *Dem* håber jeg vil komme og besøge mig i det nordsjællandske.
them hope I will come and see me in North Zealand
'I hope they will come and see me in North Zealand'
(www.hjem.get2net.dk/vmf/side_7.htm)

(11) *dem* ved jeg er gode!
them know I are good
'I know they are good'
(home.worldonline.dk/~ejstrups/kokkenhaven_juni.htm)

(12) *ham* ved jeg ikke lige hvem er
him know I not exactly who is
'I don't quite know who he is'
(strikeforce.boomtown.net/phpBB/viewtopic.php?topic=414&forum=2)

(13) *dem* ved jeg ikke hvordan smager
them know I not how taste
'I don't know how they taste'
(www.fyldepennen.dk/tekster/520)

[4] The *e* indicates the intended reading of *han*/'he' as an extracted subject. It does not necessarily correspond to an empty category. Cf. the discussion in section 6.2.1.

3 Previous Approaches

3.1 Stylistic Variation: Hansen 1972

The use of the nominative and the accusative forms of the personal pronouns is subject to considerable variation. This is only to be expected given that case no longer serves to identify grammatical functions (a point made in [Nordlinger, 1998]). The most common contexts of variation are:

- Coordinated structures

  (14) Peter og *jeg/mig* har været i vandet i dag
  Peter and I/me have been swimming today
  (example from [Hansen, 1972])

- Object of preposition

  (15) mange af de/dem udefra ved i virkeligheden bedre besked
  many of *they/them* from outside are better informed actually
  (example from [Hansen, 1972])

- Pronouns with restrictive modification

  (16) *de/dem* her ser da meget bedre ud
  they/them here look a lot better, don't they?
  (example from [Hansen, 1972])

To account for this variation, [Hansen, 1972] distinguishes two different registers associated with social connotations, register 1 being "standard" and register 2 being "informal/colloquial". These two registers employ different rules for the use of the pronominal forms. The main difference between the two registers is that pronouns in obligatorily stressed positions are in the accusative case in discourses in register 2, regardless of their grammatical function (i.e. also subjects as in (14) and (16) above).

Topicalization of constituents other than the matrix subject is always associated with stress, so the extraction configuration in (7) is in accordance with Hansen's generalization about discourses in register 2. However, there is a crucial differences between the extraction context in (7) (which is not discussed in Hansen) and the variation contexts in (14) through (16) above. The variation contexts permit a choice between the nominative and the accusative form, and the choice points to a certain register. In (7) there is no choice as to the form of the extracted pronominal. An extracted pronominal subject must be marked with the accusative form, or it can not be extracted at all.

Furthermore, there are exceptions to the generalization that the accusative form is used for obligatorily stressed pronouns. Pronominal subjects occurring with focus adverbials are obligatorily stressed, and yet only the nominative form is possible:

(17) kun/netop/også han/*ham kan klare det
only/exactly/also he/*him can do it

Hansen claims that this construction does not constitute a counter-example to the generalization about the use of the pronoun forms in register 2 on the grounds that pronouns with focus adverbials do not occur in register 2 discourses at all. For this reason no variation is observed in this construction.

The same kind of argument could be made about the extraction configuration: subject extraction does not occur in discourses in register 1, therefore no variation is observed. The picture is, however, more complicated. Subject extraction seems to occur in both registers, depending on the nature of the extracted element. Subject extraction involving lexical items without case marking, as in (18) and (19) below, is accepted by all speakers.

(18) *det* tror jeg ikke *e* passer
that think I not *e* is true
'I really don't think that is true'

(19) det er ham *som*/e jeg tror vinder
it is him *who*.REL I think wins
'he is the one who I think is going to win

In (18) the neuter subject pronoun *det*/'that' is extracted, and selectional restrictions prevent it from being interpreted as the matrix subject. (19) illustrates subject relativization with the conjunction *som*/'as' in C or an empty C position. In this case, the matrix subject is identified by means of its position.

Subject extraction as such does not seem to be restricted to different registers or discourses, even though there may be a difference in frequency. Rather the variation in acceptability between e.g. (18) and (7) stems from the case-marking of the extracted item. Some speakers do not allow the use of the accusative form to mark an extracted subject, and using the nominative form leads to a different interpretation than the intended one. Summing up,the use of the accusative form to mark an extracted subject can not be accounted for by appealing to different registers and an associated difference in the use of the pronominal forms in obligatorily stressed positions.

3.2 "Default Case": Schütze 2001

Schütze ([Schütze, 2001]) argues that the morphological case of the pronouns in English calls for a notion of default case. DPs are licensed by abstract Case (*structural licensing*) while morphological default case is assigned to DPs which fail to be assigned case otherwise. Languages vary as to which case counts as the default case. In English and Danish, accusative is the default case, as hinted at above. The most important environments where default case applies, are environments without a case assigner for the DP (as in (5) above), or where the pronoun does not occupy the head position D of the DP. On Schütze's analysis, default case is licensed on modified pronouns since modified pronouns do not occupy the head position D (example (16) above). Case spreading in coordinated structures is subject to parametric variation, and on the assumption that case does not spread in English and Danish coordinated structures, default case is licensed in these structures accounting for (14) above. Schütze's analysis essentially accounts for the variation contexts above, but his analysis does not extend to the accusative marking of a non-locally extracted subject. The extraction configuration in (7) is not a default-case environment on Schütze's account. Example (7) contains a case assigner for the accusative subject, i.e. the finite verb of the embedded clause, and the pronoun, being a simple bare pronoun, occupies the head position D. What gives this use of the accusative a flavour of being "default", is the fact that the accusative marks any grammatical function except for the subject, and the subjecthood of the preposed constituent in (7) is somewhat obscured by the fact that the pronoun receives its grammatical relation from an embedded predicate. This use of "default" is, however, not syntactic since there is no syntactic motivation for using a default case in (7). Rather the intuition behind the use of the accusative to mark a non-locally extracted subject, is that the constituent is not the subject of the closest finite verb. This is the intuition which is given a precise formulation in section 6.1.

3.3 Case Neutralization: Taraldsen 1981

[Taraldsen, 1981] develops an account of pronominal extraction in Norwegian within a transformational framework. In Norwegian, the extracted pronominal retains its case as shown in (20) below from [Taraldsen, 1981].

(20) Han hadde de trodd *e* ville komme forsent
He had they thought *e* would be late

Taraldsen, however, notes that only 3rd person pronouns can be extracted. 1st and 2nd person pronouns are barred from extraction: [5]

(21) + Jeg hadde de trodd *e* ville komme forsent
I had they thought *e* would be late

On Taraldsen's account, the extracted subject leaves a trace in its original position (within S, i.e. in the specifier of IP) and in its intermediate landing site (COMP in S', i.e. specifier of CP). Taraldsen gives the following structural representation:

(22) han_i *hadde de trodd* $[_{S'}\ t_i\ [_{S}\ t_i$ *ville komme forsent*$]]$

Taraldsen's analysis crucially relies on case theory. The trace of the topicalized subject is assigned nominative case by INFL in its canonical position, and the trace is assigned accusative case by the matrix verb *trodd*/'believed' in its landing site in S'. Conflicting case requirements are thus imposed on the preposed constituent in the matrix clause. The constituent has to spell out the feature matrix [+NOM,+ACC] which is only possible through a neutralized form. According to Taraldsen, the 3rd person pronouns are neutralized between nominative and accusative in modern Norwegian, while 1st and 2nd person pronouns are not. In this way, only the 3rd person pronouns can resolve the conflicting case requirements resulting from movement of the embedded subject, and the examples with 1st and 2nd person pronouns are ruled out.

Taraldsen's analysis is interesting in that it assumes the possibility of exceptional case marking from the matrix verb, thus accounting for the accusative form of the preposed constituent. But an account based on case theory faces several difficulties in Danish.

As will be shown in section 4, there is no evidence that the matrix verb in e.g. (7) assigns case to the subject of the embedded predicate.

[5] Taraldsen uses the symbol '+' to indicate ungrammaticality.

More importantly, the extracted subject can not be associated with an empty category in the specifier position of the embedded CP since this position may be occupied by an overt *wh*-phrase. Cf. the examples in (12) and (13) repeated below for convenience.[6]

(23) *ham* ved jeg ikke lige *hvem* er
him know I not exactly *who* is
'I don't quite know who he is'
(strikeforce.boomtown.net/phpBB/ viewtopic.php?topic=414&forum=2)

(24) *dem* ved jeg ikke *hvordan* smager
them know I not *how* taste
'I don't know how they taste'
(www.fyldepennen.dk/tekster/520)

Finally, no neutralization is observed in the 3rd person pronouns in Danish (outside the specific variation contexts mentioned in section 3.1).

(25) han/*ham vinder
he/*him wins

Interestingly, however, 1st and 2nd person pronominal subjects do not extract so easily in Danish either, though it does not seem to be impossible:

(26) mig forventer de laver det hele
me expect they do everything
'they expect me to do everything

This behaviour of the 1st and 2nd person pronouns can not be explained by appealing to neutralization, not even if the variation contexts in section 3.1 are considered reflections of neutralization. These patterns of variation pertain to 1st, 2nd and 3rd person pronouns alike and so do not point to an asymmetry in the behaviour of the pronouns under extraction.

I have no explanation for the apparent decrease in acceptability pertaining to extraction of 1st and 2nd person pronouns. It is, however, striking that this restriction pertains to the deictic pronouns as opposed to the anaphoric pronouns. Possibly this behaviour is related to discourse semantic factors, but this issue awaits further study.

To sum up the discussion of this section: Hansen's account of the variation in the use of nominative and accusative pronouns does not explain all the intricacies of the accusative marking of extracted subjects since this construction does not allow a choice between a nominative and an accusative form. Taraldsen case- and trace-based approach can not account for the Danish data as there can be no trace in the specifier of the embedded CP. The origin of the accusative form remains unexplained in the Danish data.

In the following section, I return to this question. Is the preposed constituent really an extracted constituent or does it receive its grammatical relation from the matrix verb as in raising constructions?

[6]To account for crossover phenomena in German [Berman, 2000] also assumes that non-locally extracted constituents are associated with an empty category in the specifier of the embedded in CP. Again, this can not be the case in Danish, since the specifier may be occupied by an overt constituent.

4 Subject Extraction or Object Raising

The phenomenon that a matrix constituent is coreferential with a subject of an embedded clause, is known from raising and equi contexts. In examples such as (7), the displaced constituent is not a semantic argument of the matrix verb, but we still need to consider the possibility that it is assigned its grammatical relation by the matrix verb as in raising-constructions. This would explain the presence of the accusative form, as hinted at in the analysis in [Taraldsen, 1981].

There are several differences between the extraction construction in (7) and raising constructions. First of all, embedded verbal predicates in raising constructions are generally infinite while the verbal predicate in (7) is finite.

Secondly, the extracted constituent can not occur in object position, neither within VP not in I, which is the position of unstressed pronominal objects ("object-shift") ([Sells, 2001]). In raising contexts, the raised constituent can occur in object position as shown below. Note that in (27b) the pronoun must be stressed.

(27) a. * Peter forventer [I' jo [VP ham [IP vinder]]]
Peter expects as you know him wins

b. Peter ser [I' jo [VP ham [IP flygte]]]
Peter sees as you know him escape

(28) a. * Peter forventer [I' [I ham] jo [VP [IP vinder]]]
Peter expects him as you know wins

b. Peter ser [I' [I ham] jo [VP [IP flygte]]]
Peter sees him as you know escape

The extracted constituent can only occur in the specifier position of the matrix CP:

(29) [CP ham [C forventer] Peter jo vinder]
him expects Peter as you know wins

Finally, the extracted constituent can not raise to subject under passivization while an accusative controller can:

(30) a. * han forventes vinder
he is_expected wins

b. han ses flygte
he is_seen escape

There is no evidence that the dislocated constituent receives its grammatical relation from the matrix verb or that it is is assigned case by Exceptional Case Marking. The use of the accusative form of the pronoun is constructionally determined and can not be the result of dependent-marking.

5 Two Hypotheses about the Use of the Nominative and Accusative Forms

Before stating the generalization about the use of the nominative and accusative forms of the personal pronouns, I am going to explore two alternative hypotheses. The first one is that the accusative forms of the personal pronouns serve to mark discourse functions (TOPic or FOCus). The second hypothesis is that case marking is tied to phrase structure positions, i.e. that case marking is a mere c-structure annotation. In the latter case, position in itself serves to identify grammatical functions and the use of morphological case thus constitutes a more or less redundant piece of information.

I will show that both of these hypotheses face serious theoretical and empirical difficulties, and that the use of the nominative and accusative forms serves a disambiguating function.

5.1 Accusative as a Marker of Discourse Functions

One of Hansen's ([Hansen, 1972]) insights about case marking in the colloquial register in Danish was that pronouns in obligatorily stressed positions are in the accusative case. Obligatory stress is an indication of discourse prominence. The use of the accusative form could thus in itself signal discourse prominence so that discourse functions are associated with accusative case (wherever this applies).

Further support for this hypothesis comes from the fact that pronouns left-adjoined to the matrix CP are generally in the accusative case, cf. (31) and (32) below.

(31) *Dig*, du kan gå din vej
You.ACC, you.NOM can go your own way

(32) *ham der*, han brokker sig altid
him there, he is always complaining
'that man, he is always complaining'

Example (32) furthermore illustrates the tendency to use the accusative form of a restrictively modified pronoun as in (33). This use is, however, subject to variation (cf. (16) above).

(33) *dem der kommer for sent*, skal sidde på første række
those.ACC *who are late*, have to sit in the first row

In LFG, this use could be accounted for by assuming an implicational statement to the effect that an f-structure is associated with accusative case if it is the value of a discourse function.

$(\text{DF}\downarrow) \Rightarrow (\downarrow\text{CASE})=\text{ACC}$

This analysis faces a number of difficulties. First of all it seems counter-intuitive to use the most unmarked form to indicate discourse prominence. More importantly, the nominative form may also be associated with discourse prominence (signalled by (contrastive) stress) as in (34) below. The accusative form is excluded in this case.

(34) *HAN/*HAM* kommer i hvert fald ikke
*HE/*HIM* comes certainly not
'He certainly won't be here'

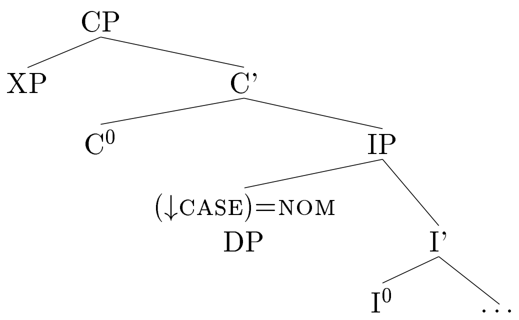
Finally, the nominative form may be associated with a focus adverbial as in (35).

(35) *kun hun* kan gøre det
Only she can do it
'she alone can do it'

Accusative as a marker of discourse functions only applies to constituents which are not simultaneously matrix subjects. This in itself suggests that this is not the right generalization.

5.2 Case-marking as a C-structure Annotation

As noted, case no longer serves to identify grammatical functions in Danish. The canonical position of the subject is the specifier position of IP while extracted constituents are in the specifier of CP. If the specifier position of IP is associated with nominative case, subjects are only associated with nominative in their canonical position. The relevant piece of c-structure would look as depicted below.



One disadvantage of this approach is that all DPs in the specifier position of IP are associated with nominative case in the f-structure. Since common nouns are unspecified for case in the lexicon, the association of nominative case with the specifier of IP does not prevent common nouns from occurring in this position. From a monolingual point of view, however, there is no empirical motivation for postulating nominative or accusative case for common nouns in Danish. What we need to ensure, is that common nouns may appear in the specifier of IP while only nominative personal pronouns may appear in this position. This can be accomplished by associating the following disjunction with the specifier of IP:[7]

$$\neg(\downarrow\text{CASE}) \vee (\downarrow\text{CASE}) =_c \text{NOM}$$

This disjunction licenses a DP in the specifier of IP if the DP bears no case specification, or if it is associated with nominative case. Lexical entries for common nouns contain no case information while the lexical entry for the pronoun *han*/'he' specifies that it is nominative:

han D ($\uparrow$CASE)=NOM
$\vdots$

In this way, all common nouns but only nominative personal pronouns are licensed in subject position. Accordingly, other nominal c-structure positions must be constrained to require accusative case or no case specification.

This proposal crucially hinges on the split IP/CP-hypothesis, i.e. on the hypothesis that the subject always occurs in the specifier of IP, also in non-inverted declarative main clauses.

[7]This approach has been suggested to me by Mary Dalrymple (p.c.).

This is the stance adopted in [Sells, 2001], while others argue that all V2-clauses are CPs (e.g.[Vikner and Schwartz, 1996]). On either account, a unique position for the subject can not be identified. On Sells' analysis constituents bearing a grammaticalized discourse function appear in the specifier of CP. Subjects in non-inverted main clauses may thus appear in the specifier of CP if they are associated with a discourse function (indicated with prosodic prominence, e.g. contrastive stress) (p. 35). In these cases, however, the subject retains its nominative case, as we have already seen:

(36) *HAN/*HAM* kommer i hvert fald ikke
*HE/*HIM* comes certainly not
HE certainly won't be here'

In light of examples such as (36), it is not possible to maintain that nominative pronouns are always in the specifier of IP, and that constituents associated with prosodic emphasis are always in the specifier of CP. The pronoun in (36) would have to be in two different positions at the same time. On Sells' account, the matrix subject may be in the specifier of either IP or CP (according to its status as a discourse function). On the account in [Vikner and Schwartz, 1996], the subject is always in the specifier of CP alongside all other kinds of topicalized constituents. On either account there is no unique position for the subject, and the hypothesis that case marking is an annotation associated with a specific c-structure position can not be maintained.

6 Accounting for the Distribution of the Nominative and Accusative Forms

6.1 Generalization about the Use of the Nominative and Accusative Forms

The discussion so far has established that the use of the nominative and accusative forms is not associated with discourse prominence. Nor is it a mere c-structure annotation.

The generalization that emerges from the data can be stated as follows (ignoring the variation contexts mentioned in section 3.1):

Nominative The DP is the subject of the immediately containing f-structure

Accusative The DP is *not* the subject of the immediately containing f-structure (but possibly the subject of an embedded f-structure)

The immediately containing f-structure, in turn, is defined as below.

Immediately containing f-structure The immediately containing f-structure is the f-structure associated with the closest functional projection (either CP or IP) dominating the DP

From a processing perspective, an initial accusative pronouns signals that the pronominal DP is not the subject of the matrix clause, and that it must be related to its canonical position later in the clause, either as an argument of the matrix predicate or as an argument of an embedded predicate. The nominative, in turn, signals the the DP is the subject of the matrix clause. In this way the use of the nominative and the accusative pronouns serves a disambiguating function. Note that case marking here is the only means to disambiguate the sentence. In main clauses there is no unique position for the subject, as detailed above. Furthermore, selectional restrictions do not suffice to disambiguate in most of these cases. "Bridge"-verbs are generally associated with human subjects, but an extracted personal pronoun evidently is also human apart from the 3rd person plural pronoun which may also (anaphorically) refer to non-humans. In far the most cases, the involved constituents consequently share the same relevant semantic features (e.g. [+HUMAN].). Nothing indicates that the default-association of TOP(ic) and matrix subject is overridden, except for the case marking. The accusative form prevents the topicalized constituent from being interpreted as the matrix subject.[8]

It follows from the generalization above that the case form is not determined by a local head-dependent relation alone, but also by the construction itself. The challenge is to account for constructionally determined case assignment, while restricting case distinctions to a specified subset of the nominals as detailed above. In the next section I address the question of how to identify a non-locally extracted subject. Two approaches present themselves: case assignment mediated through an empty category and *Constructive Case* ([Nordlinger, 1998]).

I conclude that *Constructive Case* is to be preferred on empirical as well as on theoretical grounds.

6.2 Case Assignment and Empty Categories

The discussion so far has revealed that the choice between the nominative or the accusative form of the pronoun is sensitive to the status of the subject as a local or a non-local subject. It has also been established that a non-locally extracted subject can not be identified in terms of phrase structure position since the specifier of CP is also the position of topicalized local subjects. Another way to identify extracted elements in terms of c-structure properties, is through their association with an empty category.

According to [Bresnan, 2001], non-local extraction always involves an empty category as a last resort to identify grammatical functions (p. 202). Languages vary, however, as to whether local extraction involves empty categories. In configurational languages such as English and Danish, local extraction of non-subjects involves empty categories while local extraction of subjects does not since the SUBJ(ect) is associated with the TOP(ic) by default. On this account, an extracted pronoun is associated with an empty category in exactly the cases where it appears in accusative case (i.e. extracted non-subjects and non-locally extracted subjects). Consequently we can impose the requirement on an empty category that it be identified with a discourse function bearing accusative case as shown below.

[8]A corollary of this analysis is that extraction involving two proper nouns can not be so disambiguated. Selectional restrictions are of no use since both nominals are human, and no case marking is available:

(1)	Peter	tror	Louise	kommer
	Peter	believes	Louise	is coming

(1) is (potentially) ambiguous between a reading where *Peter* is the matrix subject and the subject of the embedded predicate *kommer*/'is coming'. Out of context, only a reading where *Peter* is the matrix subject and the topic of the clause seems to be available. It does, however, remain to be investigated whether intonation and/or contextual information suffice to bring out the other reading in examples such as (1). Subject extraction involving nominals with no case marking and common semantic features, is strikingly rare. This seems to suggest that such constructions are generally avoided.

$$\begin{array}{c} e \\ ((x\uparrow)\ \text{DF}) = \uparrow \\ ((x\uparrow)\ \text{DF CASE}) = \text{acc} \end{array}$$

This account, however, crucially hinges on the assumption that non-locally extracted subjects are associated with an empty category. To what extent is this assumption empirically justified?

6.2.1 Subject/Object-Asymmetry in crossover

In [Bresnan, 2001], binding in crossover is used as a diagnostic for detecting empty categories. Binding is associated with linear precedence in many languages, and linear precedence pertains to c-structure as do empty categories.

As the examples in (37a) through (38b) illustrate, subjects and objects behave differently in their binding potential in crossover in Danish.

(37) a. $Peter_i$ tror $han_{i/j}$ vinder
Peter thinks *he* wins

b. $ham_{*i/j}$ tror $Peter_i$ vinder
him thinks *Peter* wins

(38) a. $Peter_i$ tror ikke de får fat på $ham_{i/j}$
Peter thinks not they get hold of *him*
'Peter doesn't think they will get hold of him

b. $ham_{\%i/j}$ tror $Peter_i$ ikke de får fat på
him thinks Peter not they get hold of
'Peter doesn't think they will get hold of him'

In (37a), the matrix subject *Peter* may be coreferential with the subject pronoun *han*/'he' in the embedded clause. If the pronoun is extracted, as in (37b), coreference is excluded. Note further that the very same pattern is observed with two pronouns as in (39a) and (39b) below. The ungrammaticality of the co-referential reading of (37b), thus, does not reduce to a principle C violation:

(39) a. han_i tror $han_{i/j}$ vinder
he thinks *he* wins

b. $ham_{*i/j}$ tror han_i vinder
him thinks *he* wins

In (38a), the matrix subject *Peter* can be coreferential with the object pronoun in the embedded clause. If the object is extracted, as in (38b), coreference is still possible for most speakers.[9]

[9] As indicated, not all speakers accept the coreferential reading in (38b). It has been suggested to me by Sten Vikner (p.c.) that this variation may be associated with whether one allows preposing of reflexive anaphors as in (1).

(1) % sig selv elsker Peter mest af alle
himself loves Peter most of all

It seems to be the case that speakers who accept (1) also accept (38b). This issue awaits further study.

These data suggest that subjects and objects behave differently under non-local extraction, viz-à-viz binding. It is, however, a matter of discussion whether the binding patterns above are subject to syntactic constraints. [Bresnan, 2001] argues that coreference relationships of the kinds illustrated above, are not governed by syntactic principles, but rather by semantic or discourse semantic factors. In detecting empty categories, Bresnan instead relies on semantic binding, i.e. binding of pronouns by semantic operators such as quantifiers. She argues that semantic binding is governed by syntactic principles, and that binding relationships involving quantifiers can be used to detect empty categories.

In operator binding the very same pattern as above is observed. An extracted object may be bound by a matrix subject operator, but an extracted subject may not.

(40) a. * *dem selv$_i$* tror *ingen$_i$* bliver opdaget af skattevæsenet
themselves thinks *noone* will_be discovered by the tax office
'noone thinks they will be discovered by the tax office'

b. % *dem selv$_i$* tror *ingen$_i$* skattevæsenet får fat på
themselves thinks *noone* the tax office gets hold of
'noone thinks the tax office will get hold of them'

In Danish as in English, an operator must linearly precede and syntactically outrank the bindee. [Bresnan, 2001] gives the following definitions:

- The domain of a binder excludes any pronominal that f-precedes it
- The domain of a binder excludes any pronominal contained in a constituent that outranks the binder

F-precedence in turn is defined as below:

(41) F-precedence

f f-precedes g if the rightmost node in $\phi^{-1}(f)$ precedes the rightmost node in $\phi^{-1}(g)$.

The condition on operator binding are given in (42).

(42) **Condition on operator binding**

An operator must outrank and f-precede the bindee.

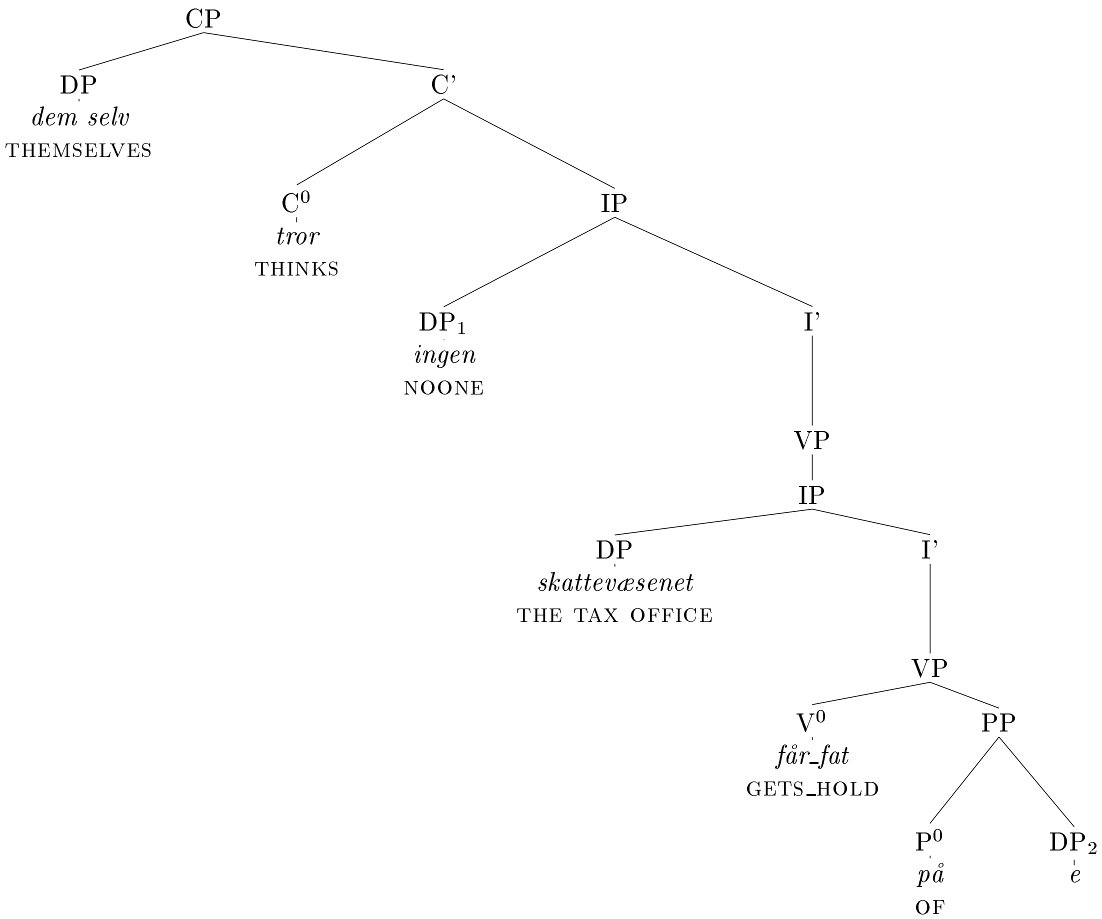
Lets us consider the binding conditions for each of the examples in (40a) and (40b) in turn, beginning with the case of object extraction.

The f-structure for (40b) is given below:

$$\left[\begin{array}{ll} \text{TOP} & \left[\text{“THEMSELVES}_i\text{”}\right] \\ \text{SUBJ} & \left[\text{“NOONE}_i\text{”}\right] \\ \text{PRED} & \textbf{‘think}\left\langle(\uparrow\text{SUBJ}),(\uparrow\text{COMP})\right\rangle\text{’} \\ \text{COMP} & \left[\begin{array}{ll} \text{SUBJ} & \left[\text{PRED ‘TAX\_OFFICE’}\right] \\ \text{PRED} & \textbf{‘get\_hold}\left\langle(\uparrow\text{SUBJ})(\uparrow\text{OBL}_{th})\text{’}\right\rangle \\ \text{OBL}_{th} & \left[\;\;\right] \end{array}\right] \end{array}\right]$$

Since the SUBJ of the matrix clause, *ingen*/'noone', outranks the COMP containing the OBL_{th}, *dem selv*/'themselves', the requirement on syntactic rank is fulfilled.

Consider next the c-structure for (40b):



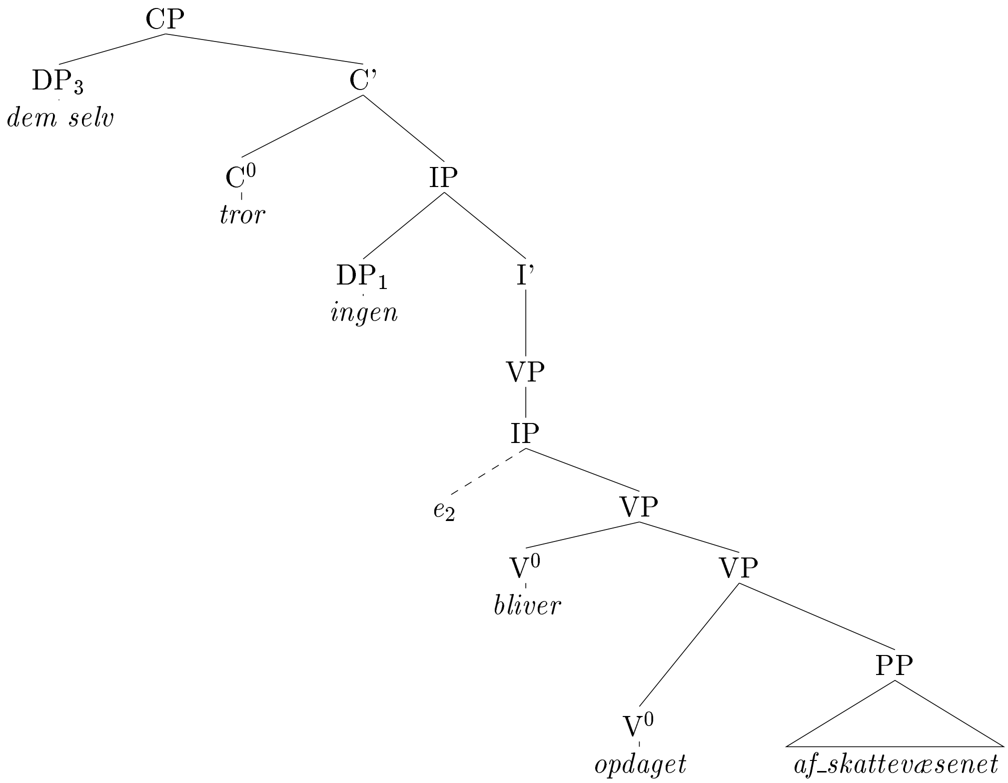
The right edge of the operator (the binder) is DP_1 in the figure above. The right edge of the bindee is DP_2, i.e. the DP dominating the empty category. The right edge of the binder linearly precedes the right edge of the bindee which means that the operator f-precedes the bindee. Since both requirements in (42) are met, the binding pattern is predicted to be acceptable.

Consider next the f-structure for (40a):

$$\left[\begin{array}{ll}
\text{TOP} & \left[\text{“THEMSELVES}_i\text{”}\right] \\
\text{SUBJ} & \left[\text{“NOONE}_i\text{”}\right] \\
\text{PRED} & \textbf{‘think}\left\langle(\uparrow\text{SUBJ}),(\uparrow\text{COMP})\right\rangle\text{’} \\
\text{COMP} & \left[\begin{array}{ll}
\text{SUBJ} & [\ \ ] \\
\text{PRED} & \textbf{‘discover}\left\langle(\uparrow\text{SUBJ})\ (\uparrow\text{OBL}_{ag})\text{’}\right\rangle \\
\text{OBL}_{ag} & \left[\text{“TAX OFFICE”}\right]
\end{array}\right]
\end{array}\right]$$

The SUBJ of the matrix clause outranks the COMP containing the SUBJ of the extracted constituent. The outranking condition is fulfilled.

The c-structure is given below.



If the extracted subject is associated with an empty category as suggested in [Bresnan, 2001], the right edge of the extracted constituent is the node (e_2), and DP_1 of the operator f-precedes the bindee. Since both requirements in (42) are met, binding is erroneously predicted to be acceptable. If, however, the extracted subject is not associated with an empty category, the right-edge of the pronoun is DP_3 and so the operator no longer f-precedes the bindee. This means that the second condition on operator binding is not met, and the example is correctly ruled out.

What these data suggest is that subject extraction is not associated with an empty category (neither locally nor non-locally), while object-extraction is associated with an empty category (locally and non-locally). These data thus lends support to the analysis of extraction in [Falk, 2001]. [Falk, 2001] claims that subject extraction is not associated with an empty category and suggests that "bridge"-verbs are associated with an optional functional uncertainty equation identifying the discourse function of the clause with the subject function of a COMP.

6.2.2 A Comparison with a Traceless Account

[Dalrymple et al., 2001] develops an account to binding in weak cross-over which does not rely on the presence of empty categories.

On their account, f-precedence is defined as follows:

F-precedence f_1 f-precedes f_2 if and only if all c-structure nodes corresponding to f_1 precede all nodes corresponding to f_2.

In determining syntactic prominence and linear precedence, Dalrymple et al. invoke the notion of co-argumenthood: i.e. arguments of the same predicate. Syntactic prominence must hold between co-arguments, i.e. the co-arguments containing the operator and the pronominal, respectively. Linear precedence must hold between the co-argument of the operator and the pronominal itself. The relevant definitions are given below.

Let CoargOp and CoargPro be coargument f-structures such that CoargOp contains O and CoargPro contains P. Then:

Syntactic Prominence An operator O is more prominent than a pronoun P if and only if CoargOp is at least as high as CoargPro on the functional hierarchy

Linear Prominence An operator O is more prominent than a pronoun P if and only if CoargOp f-precedes P

In both example (40a) and (40b), the *CoargOp* is the matrix subject, *ingen*/'noone', and the *CoargPro* is the COMP of the matrix verb *tror*/'believes'. Since the SUBJ(ect) is more prominent than the COMP on the relational hierarchy, the syntactic prominence condition is fulfilled. The condition on linear prominence must hold between the the c-structure nodes corresponding to the *CoargOp* and the c-structure nodes corresponding to the pronoun P. In both (40a) and (40b), the pronoun linearly precedes the *CoargOp*. Since the condition on linear precedence is not met, the traceless account of Dalrymple et al. rules out both (40a) and (40b), contrary to fact.

6.2.3 Conclusion on Case Marking and Empty Categories

The lesson to learn from the lengthy discussion on empty categories is that an extracted subject can not be identified through the presence of an empty category since a non-locally extracted subject is not associated with an empty category. As a consequence, accusative marking of extracted constituents can not be enforced by means of an empty category. Since an extracted subject has been shown not to be associated with a unique c-structure position, it seems that an extracted subject can not be identified in c-structural terms at all.

6.3 *Constructive Case*

[Nordlinger, 1998] develops an account to morphological function specification where the case bearing constituent projects its f-structure environment directly. The formal means is that of an inside-out function equation, i.e. the case bearing element is associated with an equation stating that the f-structure of the constituent must be the value of a certain GF. On this approach the German singular, accusative, masculine determiner *den*/'the' is associated with the following (simplified) lexical entry:

den DET (OBJ$\uparrow$)

This equation states that the f-structure of the determiner (and of its co-head nominal) must be the value of the OBJ attribute.

Two features of the *Constructive Case*-approach makes it particularly advantageous for the present analysis.

Constructive Case allows the morphological function specification to be lexically associated with the relevant lexical items. In the grammar of Danish, only the personal pronouns are associated with case-related constraints on their syntactic environment. No such constraints are associated with common nouns whose distribution is independent of their grammatical function.

Secondly, inside-out functional equations allow the use of functional uncertainty. It is possible to specify paths of variable length through the f-structure. Functional uncertainty thus allows for the implementation of constructionally dependent morphological function specification while maintaining a lexical association.

The most straight-forward lexical representation of the personal pronouns would be to associate the equation below with the accusative form of the personal pronouns stating the the pronoun is not a subject.

$\neg(\textsc{subj}\uparrow)$

This approach, however, will not yield the desired result. An inside-out functional uncertainty may denote several f-structures, even with a fixed path ([Dalrymple, 2001]). For the example in (7), the equation above denotes the f-structure of the entire clause and the f-structure of the COMP. But the equation above states that there can be no f-structure in which the f-structure of an accusative pronoun is the value of the SUBJ attribute. Since the accusative pronouns is the SUBJ of the f-structure associated with the COMP, the equation above erroneously rules out example (7).

Instead we need a disjunctive statement saying that either the accusative pronoun is not a subject, or it is an extracted subject. In f-structural terms, a subject is (non-locally) extracted if the f-structure associated with the pronoun is structure-shared with the discourse function of an f-structure denoted by an inside-out functional equation starting in the f-structure of the SUBJ attribute and leading through at least one occurrence of the COMP attribute.[10] The pronoun *ham*/'him' is associated with the following lexical entry:

$$
\begin{array}{lllcl}
\textit{ham}: & \mathrm{N} & (\neg(\textsc{subj}\ \uparrow)\ \vee & & \\
 & & ((\textsc{comp}^{+}\ \textsc{subj}\ \uparrow)\ \textsc{df}) & = & \uparrow\ ) \\
 & & (\uparrow\ \textsc{pred}) & = & \text{`PRO'} \\
 & & (\uparrow\ \textsc{pers}) & = & 3\textsc{rd} \\
 & & (\uparrow\ \textsc{sex}) & = & \textsc{masc} \\
 & & (\uparrow\ \textsc{nb}) & = & \textsc{sing}
\end{array}
$$

The equations state that either the f-structure of the pronoun is not the value of a SUBJ attribute, or it is structure-shared with the discourse function of an f-structure in which the pronoun is the subject of a COMP attribute (which in turn may be contained in a COMP itself).

[10] Actually this account is somewhat simplified since the subject may also be extracted from the complement clause of a preposition in Danish. To enhance readability I have left out this part of the equation in the lexical entry above. The fully expanded equation takes the following format:

(1) $((\{\textsc{obj}|\textsc{comp}\}^{+}\ \textsc{subj}\ \uparrow\ )\ \textsc{df}){=}\uparrow$

I follow [Dalrymple and Lødrup, 2000] in assuming that the complement clause of a preposition bears the function OBJ(ect).

As far as the nominative pronoun *han*/'he' is concerned, we need an equation to the effect that the pronoun must be a subject and that it can not be an extracted subject. Thus we need two conjunctively specified equations (conjunction is implicit in lexical entries). The lexical entry is given below:

$$
\begin{array}{llll}
\textit{han}: & \mathrm{N} & (\textsc{subj}\uparrow) & \\
 & & \neg(((\textsc{comp}^{+}\ \textsc{subj}\uparrow)\ \textsc{df}) & = \ \uparrow) \\
 & & (\uparrow\ \textsc{pred}) & = \ \text{'PRO'} \\
 & & (\uparrow\ \textsc{pers}) & = \ 3\textsc{rd} \\
 & & (\uparrow\ \textsc{sex}) & = \ \textsc{masc} \\
 & & (\uparrow\ \textsc{nb}) & = \ \textsc{sing}
\end{array}
$$

The equations state that the f-structure of the nominative pronoun is the value of a SUBJ attribute and that it can not be an extracted subject, i.e. it can not structure-shared with the discourse function of an f-structure in which the pronoun is the subject of a COMP attribute. These lexical entries capture the generalization in section 6.1 and account for the data. Only the accusative pronoun can occur as a non-locally extracted subject, and the nominative pronoun can not be non-locally extracted at all.

Let us return briefly to the pattern of variation mentioned in the beginning. As noted, not all speakers accept extraction of a pronominal subject. This fact can easily be accommodated in the present analysis. These speakers simply lack the second half of the disjunctive functional uncertainty equation in the lexical entry for *ham*/'him': $((\textsc{comp}^{+}\ \textsc{subj}\uparrow)\ \textsc{df}) = \uparrow)$. In the absence of this second half of the equation, the entry states that an accusative pronoun can never be a subject. Since a nominative pronoun can not be an extracted subject (as shown in the lexical entry above), non-local extraction of a personal pronoun is blocked altogether. A grammatical-gender specific pronominal subject such as *det*/'that' may still be extracted since the pronoun is associated with no constraints on its syntactic function. A possible stylistic restriction on the extraction of pronominal subjects (as hinted at in the discussion in (3.1)) is thus tied to the shape of the lexical entry for the accusative pronouns.

7 Conclusion

The present analysis has shown that the use of the nominative and accusative forms of the personal pronouns in Danish is constructionally determined (in addition to being determined by the grammatical function of the pronouns). The use of accusative to mark a non-locally extracted subject has a clearly disambiguating function since neither phrase structure position nor selectional restrictions serve to identify the matrix subject in these cases. The construction-specific constraints on the distribution of the personal pronouns can not be identified in terms of c-structure (neither through position in the phrase structure nor through association with an empty category). The constraints, however, can be stated in terms of f-structure properties, represented in the lexical entries of the pronouns by means of inside-out functional uncertainty equations. This approach has the advantage that it associates morphological function specification directly with the relevant lexical items without postulating case marking in other parts of the grammar where it is not empirically justified. Furthermore it allows for a straight-forward representation of (stylistic) variation in this particular construction and it accounts for the observed subject/object-asymmetry of operator binding in crossover.

References

[Berman, 2000] Berman, J. (2000). *Topics in the Clausal Syntax of German.* PhD thesis, Universität Stuttgart.

[Bresnan, 2001] Bresnan, J. (2001). *Lexical-Functional Syntax.* Basil Blackwell.

[Dalrymple, 2001] Dalrymple, M. (2001). *Lexical-Functional Grammar*, volume 34 of *Syntax & Semantics.* Academic Press.

[Dalrymple et al., 2001] Dalrymple, M., Kaplan, R. M., and King, T. H. (2001). Weak Crossover and the Absence of Traces. In Butt, M. and King, T. H., editors, *Proceedings of the LFG01 Conference.* CSLI Publications.

[Dalrymple and Lødrup, 2000] Dalrymple, M. and Lødrup, H. (2000). The Grammatical Function of Complement Clauses. In Butt, M. and King, T. H., editors, *Proceedings of the LFG2000 Conference.* CSLI Publications.

[Falk, 2001] Falk, Y. (2001). *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax.* Number 126 in CSLI Lecture Notes. CSLI Publications.

[Hansen, 1972] Hansen, E. (1972). Dr. Jekyll og Mr. Hyde i Dansk Grammatik. University of Copenhagen.

[Hudson, 1990] Hudson, R. (1990). *English Word Grammar.* Basil Blackwell.

[Lee, 2002a] Lee, H. (2002a). Crosslinguistic Variation in Argument Expression and Intralinguistic Freezing Effects. In Ionin, T., Ko, H., and Nevins, A., editors, *MIT Working Papers in Linguistics*, volume 43, pages 103–122. Cambridge, MIT Linguistics department.

[Lee, 2002b] Lee, H. (2002b). A Lexicalist OT Approach to Case. University of North Carolina at Chapel Hill.

[Nordlinger, 1998] Nordlinger, R. (1998). *Constructive Case: Evidence from Australian Languages.* Stanford, CA: CSLI Publications.

[Schütze, 2001] Schütze, C. T. (2001). On the Nature of Default Case. *Syntax*, 4(3):205–238.

[Sells, 2001] Sells, P. (2001). *Structure, Alignment and Optimality in Swedish.* Stanford Monographs in Linguistics. CSLI Publications.

[Taraldsen, 1981] Taraldsen, K. T. (1981). Case-conflict in Norwegian topicalization. In Burke, V. and Pustejovsky, J., editors, *NELS 11: Proceedings of the Eleventh Annual Meeting of the North Eastern Linguistics Society*, pages 377–398. Graduate Students Linguistics Association of the University of Massachussetts.

[Vikner and Schwartz, 1996] Vikner, S. and Schwartz, B. (1996). The Verb Always Leaves IP in V2 clauses. In Belletti, A. and Rizzi, L., editors, *Parameters and Functional Heads. Essays in Comparative Syntax*, pages 11–61. Oxford: Oxford University Press.

[Woolford, 2001] Woolford, E. (2001). Case Patterns. In Legendre, G., Grimshaw, J., and Vikner, S., editors, *Optimality-Theoretic Syntax.* Cambridge, Mass.: MIT Press.