A phase-based approach to Scandinavian definiteness marking

Abstract

We propose a syntactic approach to apparent blocking effects in the realization of definiteness marking in the Scandinavian languages. The claim is that the differences in definiteness marking can be attributed to a requirement that a definiteness feature ([def], a property of N) must be located at the left edge of the DP phase in order to be PIC-accessible for probes outside of the DP. As a result, [def] can be spelled out on N if N is the only element within DP and [def] is therefore part of DP's edge domain (giving rise to suffixal marking). In contrast, the presence of an (overt) adjectival modifier (at the left edge of DP) requires feature movement of [def] to D, which is then realized as a prenominal article (with additional spell-out of the lower copy of [def] in Swedish). The paper also addresses the (slightly different) behavior of definiteness marking in the context of relative clauses and certain issues pertaining to the interpretation of the different strategies.

Keywords: Scandinavian, definiteness, phase (impenetrability condition), copy theory, feature movement

1. Background

The present paper proposes a syntactic account of definiteness marking in Danish and Swedish based on the independently motivated notion of a phase in the sense of Chomsky (2001). The claim is that prenominal definiteness marking in Scandinavian is the result of movement of a definiteness-feature [def] to the edge of a DP-phase in order to evacuate [def] from the impenetrable domain of the phase. The discussion is restricted to standard Swedish and Danish, following Embick & Noyer (2001) and Hankamer & Mikkelsen (2005). For relevant discussion of other languages or dialects, see Delsing (1988, 1993), Julien (2005), Anderssen (2006), Alexiadou (2007), among others (cf. also, in particular, Kramer 2007, who independently proposes an analysis of definiteness marking in Amharic that is very close to the present one in many respects). Crucially, the analysis that we propose in section 3 does without the concepts of competition and blocking, both of which are involved in the two alternative analyses sketched in section 2.

The core facts of Swedish and Danish definiteness marking are the following. To begin with, definiteness on unmodified nouns in Swedish and Danish is marked by a suffix, not by a prenominal definiteness-marker (see [1a,b] and [1c, d]).

(1a) Swedish; simple noun
    mus-en
    mouse-DEF

(1b) Swedish; adjectival modifier
    mus-fiol
    mouse-yellow

(1c) Danish; simple noun
    hest-en
    horse-DEF

* We would like to thank Hans-Martin Gärtner, Line Mikkelsen, Sten Vikner, and an anonymous reviewer for helpful comments.
However, in the context of a prenominal adjective the prenominal definiteness-marker becomes obligatory (see [2b] and [3c]); Swedish (in contrast to Danish) also retains the suffix (see [2b] vs. [2c] and [3b] vs. [3c]).

In what follows, we shortly outline two recent analyses of definiteness marking in Swedish and Danish, Embick & Noyer (2001) on the one hand and Hankamer & Mikkelsen (2005) on the other. We then present an alternative account of definiteness marking in terms of \[\text{[def]}\] movement, arguing that it is able to account straightforwardly for all the observations made by Hankamer & Mikkelsen. The paper concludes by illustrating that the present account is also able to account for interpretational issues of definiteness marking (that came up only recently in the literature), which the theories of Embick & Noyer (2001) and Hankamer & Mikkelsen (2005) cannot embrace for principled reasons.

2. Two previous analyses


According to Embick & Noyer (2001) the feature \[\text{[def]}\] is a lexical property of D, sometimes spelled out as \text{-en}. They assume that N-to-D movement in Scandinavian (as in [5a]; see Delsing 1993) is the default that applies in order to satisfy the constraints in (4a, b). Furthermore, only if an A blocks N-to-D in the syntax as in (5b) due to the Head Movement Constraint (HMC, Travis 1984), the following post-syntactic morphological repair operations become active: First, \text{-en} on D is supported by an expletive \text{d}-host (see ① in [5c]); second, \[\text{[def]}\] on D is marked on N by insertion of a “dissociated” morpheme \text{-en} (see ② in [5c]). In contrast to Swedish, Danish lacks the second strategy (i.e., ②) in this context.

\begin{align*}
(4a) & \quad \text{N-def: An N is marked as definite when D is \[\text{[def]}\].} \\
(4b) & \quad \text{D-def: A D that is \[\text{[def]}\] must have a host.}
\end{align*}
Under this analysis, d-support and insertion of a dissociated -en are last resort operations that do not apply unless N-to-D is blocked due to the presence of an intervening adjective. In other words, if nothing forces these last resort operations, N-to-D movement applies as the default while application of the other strategies is blocked.

The notions of last resort, blocking, and default indicate a hidden competition in Embick & Noyer’s (2001) analysis. Moreover, condition N-def appears to be violated in (5c) in Danish. The first point becomes even more apparent if Embick & Noyer’s (2001) theory is expressed in terms of optimality theoretic constraint interaction (Prince & Smolensky 2004), which also makes understandable the violation of N-def in (5c). An explicit reconstruction in these terms could look as follows. In addition to the constraints in (4a, b), there are constraints against d-support (see [6a], from Grimshaw 1997, proposed to control do-support in English), and against the insertion of dissociated morphemes (see [6b]).

(6a) **FULL-INTERPRETATION:**
No d-support.

(6b) **DISSOCIATION:**
No dissociated morphemes.

The rankings for Swedish and Danish that derive the facts illustrated in (1)–(3) are given in (7).

(7a) Swedish ranking
N-def, D-def, HMC ≫ *dissociation, full-int

(7b) Danish ranking
D-def, HMC ≫ *dissociation, Full-int ≫ N-def

### 2.2. A lexical account: Hankamer & Mikkelsen (2005)

The theory of Hankamer & Mikkelsen (2005) postulates a lexicon rule D that converts an N into a definite D, which is spelled out as N-en (see [8]). Alternatively, they assume that a lexical definite D and an N can combine to form a DP in the syntax. Since As only merge with Ns, but not Ds, it follows that an N converted by D cannot co-occur with an A. Thus the impossibility of A + N-en in Danish is derived.

(8) \[
\begin{array}{c}
\text{PF } \alpha \\
\text{CAT } N
\end{array}
\Rightarrow
\begin{array}{c}
\text{PF } \alpha + \text{en} \\
\text{CAT D} \\
\text{DEF +}
\end{array}
\]

Moreover, Hankamer & Mikkelsen (2005) assume that syntactically formed DPs are blocked by semantically equivalent DPs that are formed in the lexicon (see Di Sciullo & Williams 1987 and Ackema & Neeleman 2004 on blocking). This prevents the prenominal marker to appear in contexts without A. Ns that are modified by As cannot be blocked because there are no semantically equivalent nominals in the lexicon. To account for double marking in Swedish, Hankamer & Mikkelsen (2005) stipulate that the category change N → D by D is optional (making co-occurrence of A + N-en possible after all).\(^1\)

---

\(^1\) This raises the following question: If the category change does not take place and A merges with a so formed definite NP, then why does the prenominal marker enter the stage? Hankamer & Mikkelsen (2005) assume that this is the case because such definite NPs select a prenominal marker.
To summarize, just like the account of Embick & Noyer (2001), the theory of Hankamer & Mikkelsen (2005) inherently involves the concepts of competition and blocking: The lexical strategy is preferred to (i.e., blocks) the syntactic strategy if both are available in principle.

3. A phase-based account

3.1. Assumptions

The present phase-based approach, which does without competition and blocking, is based on the idea that features (in particular the definiteness-feature) can undergo movement (see Chomsky 1995). Suppose (again following Chomsky 1995) that movement in general, and feature movement in particular, involves copying, and that spell-out usually realizes the highest of a series of copies in a chain. Next, assume that alongside vP and CP, DP is a phase in the sense of Chomsky (2001) (see e.g., Svenonius 2004, Heck & Zimmermann 2004, Den Dikken 2007). Phases, and thus DPs, are subject to the Phase Impenetrability Condition (PIC, cf. Chomsky 2001), a version of which is given in (9).

(9) PHASE IMPENETRABILITY CONDITION:
Material within a phase XP is not accessible to operations at ZP (the next phase) unless it is within the edge domain of XP.

(10) Edge domain:\2:
The edge domain of a phase XP comprises the left area up to and including the left-most overt element within XP.

A general consequence of the PIC is that any goal-feature that is supposed to value a probe-feature in a higher phase must first move to the edge domain of the current phase in order to be accessible for either Agree or further movement. This leads to the following well-known problem: If all movement is feature-driven due to last resort (LR, Chomsky 1995, Collins 1997; see [11]), then what triggers successive cyclic movement to the phase edge?

(11) LAST RESORT:
Movement must result in immediate feature valuation.

A standard answer given in the literature is that edge movement is driven by particular edge features, see Collins (1997), Fanselow & Mahajan (2000), Chomsky (2001), McCloskey (2002). This view is not unproblematic, though; for instance, it might be argued that the insertion of such edge features violates Chomsky’s (2001) Inclusiveness Condition. In the present paper, we thus adopt an alternative approach to the above mentioned problem. Namely, we follow Heck & Müller (2000, 2007) in assuming that movement to the phase edge is not feature-driven after all and therefore violates LR. However, such a violation is

---

2 This notion of edge domain is non-standard in that it makes reference to the phonological overtness of elements rather than to particular positions within the phase. It is only overt elements that define the edge of a phase, thereby closing off its accessible domain. As a crucial consequence, even an N-head can be accessible within the DP provided the specifiers (if present) and the head of the DP are phonologically empty.
tolerated by the grammar if it helps to avoid a violation of the higher ranked constraint PHASE BALANCE (PB, see [12]).

(12) PHASE BALANCE:
For every probe in the numeration, a matching goal must be accessible in the current phase.

(13) Accessibility:
A matching goal is accessible if it is in the edge domain of the current phase.

The definitions used here somewhat trivialize the complications taken into account in Heck & Müller (2000, 2007), but they suffice for the present purpose.\(^3\) The core idea relevant for the present analysis is as follows. Assume that the definiteness-feature is a property of N in Scandinavian (cf. Delsing 1993; see also Borer 1999 on Hebrew). Next, suppose that [DEF] on an object must be accessible to a probe on v (the same holding for [DEF] on a subject and a probe on T), which is overtly signaled by verbal [DEF]-agreement in some languages (see Trommer 1995, Kiss 2002 on Hungarian). Now, if DP is a phase, then it follows that [DEF] must be in the edge domain of DP in order to be accessible for [DEF]-agreement. As we will see, this sometimes involves creating a higher copy of [DEF] by feature-movement, which is then spelled out as a prenominal [DEF]-marker.\(^4\)

3.2. Analysis

3.2.1. The core facts

In the absence of a prenominal A, N-[DEF] is the leftmost overt element within DP, and thus [DEF] is accessible. As a consequence, PB does not force [DEF]-movement, which is therefore blocked by LR. Since there is only one copy of [DEF] (namely on N), it trivially constitutes the highest one and is thus spelled out (namely as -en), see ① in (14a). This holds for both Danish and Swedish.

\(^3\) For instance, if a probe in the numeration is matched by two goals, one in the numeration and one in the current phrase marker, then the probe must be able to satisfy PB without triggering movement of the goal in the current phrase marker (because the probe counts on being satisfied by the goal in the numeration at some later point).

\(^4\) A reviewer objects that if [DEF]-movement is triggered by DP-external probes, then one should expect DPs in isolation to behave differently from DPs in clauses. Suppose the derivation started with a numeration that exclusively contained material that ended up DP-internally. Then at no point of the derivation would there be a probe in the numeration such that (12) could trigger DEF-movement. Consequently, many of the DPs in this paper (most of which are cited in isolation) should lack DEF-movement, contrary to fact. Note that this conclusion is only valid under the reviewer’s explicit assumption that “it is possible to determine the well-formedness of the relevant DPs in isolation”. We do not share this assumption, though. Rather we take it that a grammar only determines the well-formedness of root clauses. (In times when grammars were still characterized as sets of phrase structure rules this idea was expressed by the assumption that every derivation starts with the symbol S.) In fact, sometimes the well-formedness of a phrase cannot be determined in isolation, compare, for example, the instances of a book about in (ia) and (ib):

(ia) Who\(_2\) did you read a book about t\(_2\)?
(ib) *a book about

Cases where well-formedness appears to be determined in isolation are thus an illusion. (A consequence of this assumption is that all grammatical utterances that are superficially not root clauses de facto involve root clauses and ellipsis.) In this sense, the presentation in this paper is rather sloppy in that we give grammaticality judgements for DPs where we should have given judgements for root clauses. We thank the reviewer for urging us to clarify this point.
In the presence of a prenominal A, PB forces movement of [def] from N across A into the edge domain of DP (in violation of LR) in order to remain accessible, see ① in (14b). This creates two copies of [def], the higher copy of which is spelled out as an allomorph den/det/de, see ② in (14b). Again, this is the case for both Danish and Swedish. However, for Swedish we need the additional stipulation that the lower copy of [def] is spelled out, too.

(14a) (14b)

Note in passing that the present analysis does not depend on whether AP is the complement of D (see Abney 1987, Embick & Noyer 2001) or whether it is in SpecN (see Svenonius 1992, Hankamer & Mikkelsen 2005).

3.2.2. -ende-Nouns

Hankamer & Mikkelsen (2005: 97 f.) observe that deverbal “common gender” nouns that are formed with the suffix -ende require the prenominal definiteness-marker and are incompatible with the suffix -en in both Danish (see [15a, b]) and Swedish. In contrast, neuter Ns formed with -ende are well-behaved in that they lack the prenominal marker in the absence of an A, see (15c, d). Hankamer & Mikkelsen (2005) take this as an argument in favor of their lexical treatment, the idea being that a lexical rule can be sensitive to morphological properties while (post)-syntactic movement cannot.

(15a) Common gender -ende-noun
*studerrende-en
student-def
(15b) den studerrende
def student
(15c) Neuter -ende-noun
udseend-et
appearance-def
(15d) *det udseende
def appearance

Note that studerrende formally looks like a present participle. Provided that it is treated as a participle in the syntax, i.e., that it modifies an empty N as in (16a), it follows that its derivation involves movement of [def] from the empty N across the participle, resulting in a prenominal definiteness-marker. Neuter -ende-forms on the other hand could be assumed to be real Ns; thus they do not contain any (overt) modifier, see (16b). Movement of [def] is blocked by LR and no prenominal definiteness-marker is spelled out.

(16a) Deverbal common gender noun
[DP den [NP [AP studerrende ] [N Ø ]]]
(16b) Deverbal neuter noun
[DP D [NP anliggend-et ]]
A welcome consequence of this analysis is that it automatically explains why common gender -ende-Ns in Swedish never bear [def] as a suffix, just as in Danish (see HANKAMER & MIKKELSEN 2005: 102), in contrast to what is usually the case in Swedish. The idea is that spell-out of the lower copy is not possible in these cases, because the empty N is not an appropriate host to support the -en Suffix.

We hasten to add that HANKAMER & MIKKELSEN (2002) give arguments against the analysis in (16a), claiming that forms like studerende should be treated as Ns in the syntax. We suspect that the arguments can be challenged. But should they turn out to prevail, then the present approach must resort to another explanation for the behavior of deverbal -ende nouns.

3.2.3. Relative clauses

The following observations of HANKAMER & MIKKELSEN (2005: 107f.) concern relative clauses. Restrictive relatives occur with prenominal or post-nominal definiteness-marking in Danish and Swedish (see [17] and [18]). In contrast, appositive relatives always require post-nominal definiteness-marking (see [17a] and [18a]). Finally note that the double marking in Swedish becomes optional in the context of restrictive relative clauses (see [17b] and [17c]).

(17a) Swedish; restrictive and appositive
mus-en som vi såg
mouse-def that we saw

(17b) Swedish; restrictive only
den mus som vi såg
def mouse that we saw

(17c) Swedish; restrictive only
den mus-en som vi såg
def mouse-def that we saw

5 For instance, HANKAMER & MIKKELSEN (2002) observe that studerende cannot be modified by an adverbial as ivrigt (‘eagerly’), while bona fide prenominal participles can (see [i–a,b]), suggesting that it cannot be a modifier.

(ii) Jeg så en studerende på gangen
I saw a student in hallway
‘I saw a student in the hallway.’

To be precise, (ii) is ambiguous between the two relevant readings, while the presence of the adverb disambiguates (ii). The ambiguity could be traced back to two different structures, one with studerende as an N, the other with studerende as a prenominal participle. An alternative interpretation would be to say that the participle studerende can occupy two different positions of modification within the DP, a higher or a lower one, resulting in different interpretations (see CINQUE 1994). The adverb in (i) would then be compatible with only one of these.
To derive these facts in the present phase-based theory we make the following assumptions. First, suppose that restrictive relatives are merged pre-nominally as the sister of N (see Smith 1964, Stockwell et al. 1973, where this is proposed, albeit for another reason; cf. also Fanselow 1986). Restrictive relative clauses then obligatorily extrapose to the right (but are interpreted in-situ), an operation that we assume to be triggered by a feature on D (thus respecting LR). Further assume that appositive relatives are merged to the right of DP (or that they are not a proper part of their syntactic environment in narrow syntax to begin with, see, e.g., Edmonds 1979, Cinque 1982).

Finally, suppose that the phase head of the nominal domain is actually not D but n, a functional head between N and D (in the sense of Adger 2003). As a consequence, PB will be relevant on the nP-level already, while the probe that triggers extraposition does not enter the structure before D is merged.

Under these assumptions, two options arise at the nP-level of a derivation involving a restrictive relative clause. Either PB-driven [def]-movement across the relative clause to the edge of nP applies, in order to render [def] accessible within nP. As before, such movement creates a higher [def]-copy, spelled out as a prenominal marker. On the DP-level, feature-driven extraposition applies. Alternatively, PB can also be satisfied on the nP-level by early extraposition of the relative clause to the right of nP. Note that once extraposition has applied on the nP-level, [def]-movement to n is blocked by LR. Thus no higher [def]-copy is created. On the DP-level, string vacuous feature-driven extraposition applies. It follows that in the context of a restrictive relative either a prenominal or a post-nominal definiteness-marker shows up.

If appositive relatives are merged to the right of DP, they can never force [def]-raising to the edge of nP. Thus appositives only co-occur with a post-nominal definiteness-marker. We must leave open here why optional deletion of the lower [def] is possible in Swedish restrictive relatives (see [17b] and [17c]).

---

6 It follows that [def]-movement must then target the n-head instead of D. Everything else remains the same.

7 Interestingly, the argument presupposes that optimization applies locally, phrase-by-phrase. If it did not apply until the DP-level, then extraposition to DP would satisfy the probe on D as well as PB, crucially without violating LR. But then Danish and Swedish restrictive relatives should be incompatible with a prenominal definiteness-marker, contrary to fact. See Heck & Müller (2000, 2007) for more arguments that optimization in the syntax applies locally.
3.2.4. Post-nominal PPs

Finally, Hankamer & Mikkelsen (2005: 111 f.) observe that post-nominal PPs never license the prenominal definiteness-marker, but require the [def]-suffix. This holds for adjuncts as well as for complement PPs (see [20a, b] and [20c, d], respectively, for Danish).

(20a) Post-nominal adjunct PP
gris-en med blå pletter
pig-def with blue spots

(20b) *den gris med blå pletter
def pig with blue spots

(20c) Post-nominal complement PP
forfatter-en til bog-en
author-def to book-def

(20d) *den forfatter til bog-en
DEF author to book-def

These facts follow without further ado from the present theory if both types of PPs are merged post-nominally. In this position, they will never render [def] inaccessible, and thus LR blocks [def]-movement. Hence, there is only one low copy, spelled out as -en.

3.2.5. Interpretational issues

It has been claimed more recently (see Julien 2005, Anderssen 2006, Alexiadou 2007) that the prenominal and the post-nominal definiteness-marker are interpreted differently: Prenominal [def] denotes uniqueness, post-nominal [def] denotes specificity. Provided that this is correct, then a theory of definiteness marking should at least in principle be able to account for it. However, as we see it, the theories of both Embick & Noyer (2001) and Hankamer & Mikkelsen (2005) lack this capability for principled reasons. Here is why.

In the theory of Embick & Noyer (2001) the dissociated definiteness marker is copied post-syntactically, i.e., at a point of the derivation that is located behind the branching off to LF; therefore, dissociated definiteness-markers cannot feed the LF-side of the derivation and, consequently, they cannot contribute any meaning. For Hankamer & Mikkelsen (2005) it is crucial that blocking of a DP with prenominal definiteness-marker by a lexically formed DP with post-nominal [def] is only possible if their interpretation is the same. This, of course, is incompatible with the idea that prenominal and post-nominal definiteness-markers are interpreted differently.

For the phase-based approach proposed here no such principled problem arises: First, [def] is copied in the syntax and can thus feed LF; second, the present analysis does not presuppose semantic equivalence of the different variants of definiteness marking. Rather, we assume that the syntax simply provides structures, which are then accordingly interpreted. Depending then on its position in a structure, [def] is interpreted differently (see Diesing 1992, Chomsky 2001).
4. Conclusion

To conclude, a syntactic (as opposed to a post-syntactic or a lexical) analysis of [def]-spread in Scandinavian is not only feasible, but attractive. Most of Hankamer & Mikkelsen’s (2005) observations fall out from the independently motivated movement to the edge of a phase, once the following assumptions are accepted: DPs (or nPs for that matter) are phases; [def] must value a probe on v (and T) and must therefore be part of DP’s edge domain; movement creates several copies, the higher of which (usually) undergoes spell-out.

References


Heck, Fabian & Müller, Gereon (2000): Successive cyclicity, long-distance superiority, and local opti-


