

The ‘Specifier’ in an HPSG grammar implementation of Norwegian

Lars Hellan and Dorothee Beermann

NTNU, Trondheim, Norway

{lars.hellan,dorothee.beermann}@hf.ntnu.no

Abstract

We present a principled background for the adoption of a category ‘Specifier’ in the analysis of noun phrases, and show how, under certain constraining assumptions, it can be successfully employed in the implementation of an HPSG grammar of noun phrases for Norwegian. How widely the assumptions can be applied on empirical grounds cross-linguistically, is still a matter for further investigation.

1 Introduction

Although the structure of NPs varies considerably across languages, the types of items encountered inside NPs seem on the whole to be the same. This could make the ‘noun phrase’ (among other construction types) a candidate for serving as a useful module in multilingual grammar development: as such, its specifications could potentially be reused in grammars from language to language, and semantic outputs could be harmonized, so as to simplify the construction of semantically-based applications cross-linguistically. In this paper we address an NP-module as suggested in computational HPSG-grammars using the LKB platform cf. (Copestake, 2002) and the ‘HPSG Grammar Matrix’ cf. (Bender et al., 2002). Here we focus on the notion ‘Specifier’ as proposed in the latter and in the HPSG tradition at large. Analyzing a central set of phenomena in Norwegian related to the notion ‘Specifier’, we conclude, that this construct is a useful one for a Norwegian grammar, but with some essential provisos.

Among the growing family of LKB based grammars related to the ‘HPSG Grammar Ma-

trix’, is a grammar for Norwegian, NorSource,¹ which constitutes a background and test bed for proposals under discussion.

2 Formalism and assumptions

Central among the components delivered through the Matrix is the module Minimal Recursion Semantics (‘MRS’) for semantic composition cf. (Copestake et al., 2001) and (Copestake et al., submitted). MRS representations are ‘flat’ representations of the elementary predications that compositionally represent the meaning connected to individual constructions. The formalism seems a promising candidate for becoming a widely accepted exchange format within computational semantics. An example is given in figure 1 below, showing the MRS (as produced in a standard English grammar) for the sentence *The boy throws the ball*. The RELS list is a ‘bag’ of those elementary predications (EPs) that are expressed by the sentence; in this case there are six EPs, of which one reflects *throw*. The subject and object arguments of the verb are reflected by the coindexation of the ARG1/ARG2 of the verb with the ARG0 of the determiner and the noun, corresponding to the ‘bound variable’. The remaining EP represents the ‘message type’. Scope properties are expressed in the HCONS list, ‘x QEY y’ meaning essentially that x scopes over y. HCONS thus records the scopal tree of the constituent

¹See (<http://www.ling.hf.ntnu.no/forskning/nor-source>). This grammar was initiated in 2002, and had its first period of development inside the EU project DeepThought (<http://www.project-deepthought.net>). It takes part in the Delph-in initiative (<http://www.delph-in.net>), which is a cooperative effort within deep language processing, using LKB and many other types of software communicating with LKB, thereamongst the ‘HPSG Grammar Matrix’.

in question, as outlined in (Copestake et al., submitted).

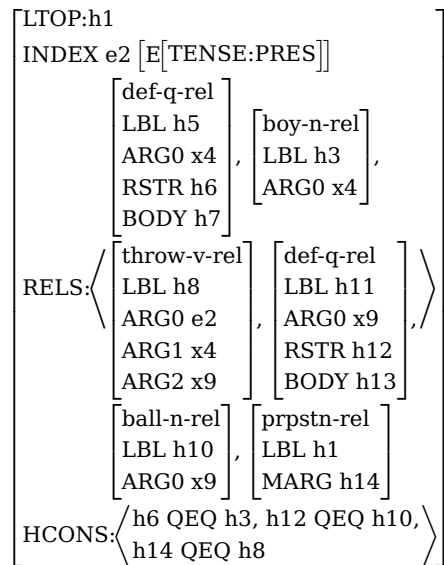


Figure 1: MRS-structure for the sentence *The boy throws the ball*

A standard MRS representation of an NP will include a quantifier and a restrictor predicate, as exemplified with *def-q-rel* and *boy-n-rel* in figure 1. EP-predicates used in MRSs can be either lexically induced, like *throw-v-rel* and the other predicates in figure 1, or constructionally induced, examples being provided below.

Background assumptions for the analysis of NPs in this framework come from many sources, including mainstream logical traditions arising with Frege and Russell, and generative syntax, here represented by HPSG, as, e.g. in (Pollard and Sag, 1994). Among important assumptions that are currently baked into the Matrix grammar design, we can mention:

From predicate logic:

- There can be only one variable per simple NP.
- There can be only one quantifier binding all instances of that variable in the representation of an NP.

From syntactic tradition

- In addition to complements and adjuncts, NPs also contain specifiers.

Specifiers enter the feature structure specifications in the following ways:

- A noun (N) signals the need for a specifier through a feature

SPR <[...] >

where properties of the specifier are stated

inside '[...]'.
 - An item whose typical function is to serve as a specifier, signals its need for something to specify through a feature

SPEC <[...]>

where properties of what is 'specified' are stated inside '[...]'.
 A specifier item can be a word like *all*, *both*, *every*; *some*, *no*, *this*, *the*, *a* or a phrase like a prenominal genitive, as in *my best friend's money*, or a measure phrase, as in *tre gule krukker genever* ('three yellow jars (of) genever').

When a specifier item is encoded in a lexical entry, it is typically represented as introducing a quantifier relation in MRS, with a scope corresponding to the constituent 'specified'. When the specifier item is a phrase (e.g., a possessor or a measure NP), the quantifier relation in MRS has to be introduced as part of the operation combining that phrase with the constituent 'specified'. Either way, necessarily one, but only one, occasion is provided in the phrasal combinatorics for the introduction of a quantifier, and only one item is assigned the status as a specifier per NP.

Let us refer to these assumptions as *Spec-unique*. Theoretically, it represents a way of reconciling insights behind the 'DP-analysis' of NPs - cf. (Hellan, 1986), (Abney, 1997) - with the traditional notion of 'noun' being the head of NPs. Like a head relative to its complement, a specifier can determine certain properties of the noun (or attributes of the noun); still, the noun can be argued to be the overall head of the NP. Being thus counted as a type of 'attribute' of the noun, as are complements and adjuncts, the Specifier has a position intermediate between complements and adjuncts: complements are valence-bound and restricted in number, adjuncts are not valence-bound and unrestricted in number, whereas specifiers are not valence bound, but restricted in number. That this number is set to 'one' is partly conditioned by items like those mentioned above being mutually exclusive in English; but more importantly, it reflects the standard assumption of every NP being represented by just one (as 'type') variable, and hence just one quantifier binding that variable. This being hence a primarily theoretically motivated design, let us first briefly comment on its correctness as an

empirical generalization across languages.

Spec-unique as ‘There must be a Specifier’: The first part of *Spec-unique* says that there *must* be a Specifier. English is in effect among the very few languages in the world with a requirement that there must be a word/phrase-type Spec item in an NP: most languages allow, e.g., a bare nominal form to constitute an NP, also when the referent is countable. That is, the ‘only one’ part of Spec-Unique is typologically dubious.

Spec-unique as ‘There can be at most one Specifier’:

The second part of *Spec-unique* says that there can be at most one Specifier. An example from Ga (spoken in Ghana) is suggestive of the possibility of there being many potential ‘specifier’ items in an NP (from (Dakubu, 2005):

- (1) nakai gbekɛ gbɛɛi nɛ lɛ fɛɛ tɛɛ shi
 that child dog that DEF all go down
 Id N Poss N Dem Def Quant V N
 ‘all those FAM [dogs of the child]’

Possible Specifiers here include items categorized as ‘Identifier’, ‘Demonstrative’, ‘Definite (article)’, ‘Quantifier’ and ‘Possessor’, that is, altogether five items. To fully assess the impact of the construction in the present setting would of course require an investigation by itself, and we mention the example here only as a pointer to a possibly significant counter case to the ‘at most one Spec’ hypothesis.

In Norwegian, cases that *could* be construed as multiple-Specifiers include:

- (2) alle mine tre gutter
 ‘all my three boys’
- (3) de tre energiske guttene
 ‘the three energetic boys’

If we want to maintain the ‘at most one’-part of *Spec-unique* for Norwegian, then presumably numerals should not count as Specifiers; then the definite article *de* will be unique in having Spec status in (3), but in (2)

there is still a choice between the universal quantifier and the possessor NP having Spec status. In the following we consider a set of phenomena in the Norwegian NP with a view to the possible overall fruitfulness of the ‘at most-one’ part of *Spec-unique* for its grammar, and in conjunction with this, we return to the question concerning (2). The phenomena to be considered are:

- Definiteness
- Genericity
- the Referential/Attributive distinction
- Boundedness

3 Definiteness in Norwegian

‘Definiteness’ is standardly recognized as having two morpho-syntactic instantiations in Norwegian:

- the definite article suffix, as in *gutten*
- the weak adjective form, as in
den snille gutten
min snille gutt

Reference to definiteness in the grammar is also made in constraining which NPs can occur as (direct) objects in presentational constructions (the so-called ‘indefiniteness requirement’). NP types *not* allowed in such constructions include the forms above, but also universally quantified NPs counting as indefinite relative to the above criteria, like *hver katt* (‘each cat’), as exemplified in **det sitter hver katt i haven* (‘there sits each cat in the garden’).

Semantically, the two first instantiations of definiteness seem to reside in an assumption of shared familiarity with the item referred to - that is, it signals that the referent is familiar to the speaker, and it signals an assumption on the part of the speaker that the referent is familiar to the hearer as well. For the time being, we include the ‘familiarity’ factor in the standard MRS representation. One way in which this inclusion could take place is as a boolean attribute, for instance inside the INDEX attribute; another is as a predicate constituting its own EP, and we have chosen the latter, with the relation type *familiar-rel*.

The third instantiation of definiteness (as in *hver gutt* (‘each boy’)) does not carry the pre-

supposition of familiarity. Typical of these NPs is that they express universal quantification, and they align with the Russellian notion of uniqueness in that their referent in a sense is unique, namely being the total set (in question) of instances of the restrictor predicate. Included in the ‘Matrix package’ of assumptions and specifications is a subtype of quantifier relations called *def-q-rel*, supposed to appear in the MRS of any NP counted as ‘definite’; an instance was seen in Figure 1, and we adopt this convention. ‘Indefinite’ universally quantified NP like ‘each cat’ are characterized by the relation *univ-q-rel*, and their ‘definiteness’ has to be marked by features which we will not enter further into here.

Accordingly, there is no MRS relation common to all the types of definite NPs. As for those NPs represented through *familiar-rel*, the latter is not a quantifier relation, and given the formal requirement mentioned above to the effect that there be a quantifier in the representation of every NP, we implement this constraint by using also an EP with the relation type *def-q-rel* for the NPs instantiating the first two shapes of definiteness.

3.1 Constructing a grammar of definiteness in Norwegian

Since in the most standard case, *gutten*, definiteness is expressed through suffixation on the noun, it is reasonable to let the suffixed definite article consistently introduce the definite quantifier, represented in MRS as *def-q-rel*. The definite article suffix, as in *gutt-en*, is identified through the feature ‘DEF-MORPH +’.

The second case, as in *min snille gutt*, will have a noun marked ‘DEF-MORPH -’, but the N projection as such still needs to be marked as definite, and this ‘guise’ of definiteness will be marked ‘DEFINITE +’. ‘DEF-MORPH +’ entails ‘DEFINITE +’, but not vice versa. Thus, the noun form *gutten* will be marked as

$$\left[\text{synsem} \mid \text{local} \mid \text{cat} \mid \text{head} \begin{bmatrix} \text{def-morph} & + \\ \text{definite} & + \end{bmatrix} \right]$$

The NP *min snille gutt* will be marked as

$$\left[\text{synsem} \mid \text{local} \mid \text{cat} \mid \text{head} \begin{bmatrix} \text{def-morph} & - \\ \text{definite} & + \end{bmatrix} \right]$$

and the lexical specification of *gutt* by itself will be ‘DEF-MORPH -’, leaving the value of

‘DEFINITE’ open for decision in the structure it enters.

In a lexically oriented framework, it is reasonable to let the assignment of the EP with *def-q-rel* to the form *gutten* be done at lexeme level, rather than as part of the phrasal composition of the NP. In the case of *min gutt*, however, which also receives *def-q-rel*, this will have to happen in the combinatorial rule combining the possessive *min* with *gutt*.

The weak adjective suffix *-e*, as in *den snille gutten* and *min snille gutt*, induces the following specification on the adjective it gets suffixed to:

$$\left[\dots \mid \text{head} \mid \text{mod} \left\langle \left[\text{local} \mid \text{cat} \mid \text{head} \mid \text{definite} & + \right] \right\rangle \right]$$

Here, what comes after ‘MOD’ constrains the noun that the adjective can modify. This specification thus allows both *snille gutten* and *snille gutt*, but the latter only for a case where a definite article word or a possessive precedes the adjective, as in *den snille gutten* and *min snille gutt*, respectively. When this adjective combines with the noun, thus, the resulting phrase N’ requires combination with one of these items at the next combinatorial stage. We encode this requirement through the specification

$$\left[\text{synsem} \mid \text{local} \mid \text{cat} \mid \text{val} \mid \text{spr} \left\langle \left[\right] \right\rangle \right]$$

on the N’, i.e., as a requirement that it combines with a specifier. Notably, this is a specification not inherent in the noun, but arising through a combinatorial rule. For nouns as lexical entities, in Norwegian, there is no reason to provide them with the requirement of a non-empty Specifier-list.

In cases like *den snille gutten*, the Specifier induced by the weak adjective is what we may call the *left-edge-def-word den*. It co-occurs with the suffixed article, and since the latter carries the definite quantifier and an NP can only have one quantifier, the left-edge word can only carry a non-quantificational relation, viz. *familiar-rel*, which in the MRS of *den snille gutten* thus occurs twice. (Thus, in the only case where Norwegian does have an obligatory Specifier, the lexical item filling this position lacks the quantifier-inducing capacity that is otherwise presupposed in the assumption-cluster of *Spec-unique*.) The ‘left-edge’ *den* thus has the specification

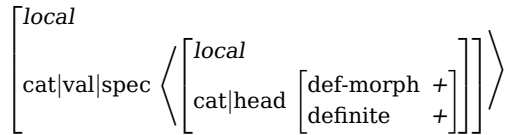
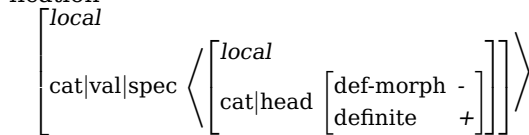


Figure 2: Partial feature-structure for left-edge *den*

whereby the occurrence of the def-suffixed noun is required, with its *def-q-rel*.

3.2 Generic and attributive readings

Another def-article word *den* should now be noted, which has a restricted use in Norwegian, but has a syntax corresponding to the normal *den* in Danish. It may be given the initial specification



by which it induces *def-q-rel*. However, there is more to its semantics, pertaining to factors such as Genericity and Referentiality; consider the following examples:

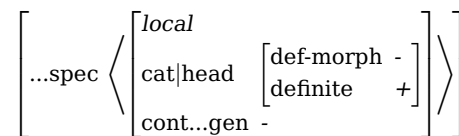
- (4) den bengalske tiger er utryddet
'the Bengal tiger is extinguished'
- (5) ? den bengalske tigeren er utdÅydd
- (6) den bengalske tigeren sitter bak deg
'the Bengal tiger sits behind you'
- (7) ??* den bengalske tiger sitter bak deg
- (8) ?* den sultne tiger sitter bak deg
'the hungry tiger sits behind you'
- (9) den siste tiger du ser kan du springe fra
'the last tiger you see you may run away-from'
- (10) den siste tigeren du ser kan du springe fra

In (4), the *den* under consideration induces a generic reading, a reading required by the adjective *utryddet*, and hence (4) is well-formed whereas in (5), the left-edge *den* is less felicitous in combination with such a generic verb. With the 'instantiated' status induced by *sitte bak* in (6), the left edge *den* goes fine, whereas the other *den* does not (cf. (7)). An 'instantiated' adjective like *sulten* enhances

this effect of incompatibility with the second *den* further, as seen in (8). Thus, we may call this second *den* a 'situationally detached' *den*, or, for short, we label it *den-aloof*. As (9) shows, this *den* also functions well with an attributive reading (in the sense of (Donnellan, 1966)); but unlike the genericity factor, an attributive reading can also obtain with the left-edge *den*, as seen in (10).

Our account so far yields identical MRSes for, and also does nothing to distinguish the grammaticality status between, (4) and (5). Suppose that we let *utryddet* require of its ARG1 that it have a feature 'GEN(ERIC) +', a feature entered under the path SYNSEM|LOCAL|CONT|HOOK|INDEX|SORT. A definite noun is unspecified for this feature, but the left-edge *den* would require of its SPEC item that it have the value '-' for this feature, i.e., it would have the following specification, rather than the one in figure 2:

Left-edge *den*:



The *den-aloof*, in contrast, would require 'GEN +', i.e:

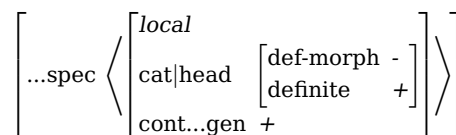
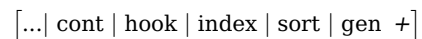
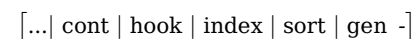


Figure 3: Partial feature-structure for *den-aloof*

(5) is then ruled out by clashing INDEX values, since *utryddet* requires the path



in the specification of its subject, while the NP in (5) actually has



induced by the left-edge *den*. Notice that the following is accepted:

- (11) tigeren er utryddet
'the tiger is extinguished'

since the noun form *tigeren* as such is neutral with regard to the GEN feature - only the two *dens* have particular concerns about this feature.

Conversely, in (7) and (8), the verb (helped by the adjective in (8)) requires of its subject that it be

[...| cont | hook | index | sort | gen -]

but the *den-aloof* here induces

[...| cont | hook | index | sort | gen +]

accounting for the ungrammaticality of (7) and (8).

As for the 'Attributive' reading possible in (9) and (10), what can be concluded from this is that even with non-generic verbs, *den-aloof* is acceptable if the reading can be construed as attributive. If we mark such a reading of an NP as 'ATTR +', and accept all generic readings as also attributive, then one full representation of *den-aloof* becomes:

Den-aloof-1:

$$\left[\dots \text{spec} \left\langle \left[\text{loc} \left[\begin{array}{l} \text{cat|head} \\ \text{cont} \dots \text{sort} \end{array} \right] \left[\begin{array}{l} \text{def-morph -} \\ \text{definite +} \\ \text{gen +} \\ \text{attr +} \end{array} \right] \right] \right\rangle \right]$$

and another becomes

Den-aloof-2:

$$\left[\dots \text{spec} \left\langle \left[\text{loc} \left[\begin{array}{l} \text{cat|head} \\ \text{cont} \dots \text{sort} \end{array} \right] \left[\begin{array}{l} \text{def-morph -} \\ \text{definite +} \\ \text{gen -} \\ \text{attr +} \end{array} \right] \right] \right\rangle \right]$$

to allow the non-generic status of (9). Conceivably, then, the 'general' *den-aloof* can be stated as follows, rather than as in figure 3:

Den-aloof-General:

$$\left[\dots \text{spec} \left\langle \left[\text{loc} \left[\begin{array}{l} \text{cat|head} \\ \text{cont} \dots \text{sort} \end{array} \right] \left[\begin{array}{l} \text{def-morph -} \\ \text{definite +} \\ \text{attr +} \end{array} \right] \right] \right\rangle \right]$$

This will allow a generic reading when such a reading is induced by the verb, a possibility blocked for left-edge *den*. Such assignments will be possible, though, only if the grammar

is able to make a decision for any NP whether it is attributive or referential; this may well be even harder to obtain than a computation of genericity, however.

4 Boundedness

As is well established in the literature (cf., e.g., (Smith, 1991)) an indefinite plural NP induces atelicity in the verbal construction; examples include, with NPs in object position:

- (12) Jon leste artikler i flere dager
'Jon read articles for many days'
- (13) ??Jon leste Goethes diktsamlinger i flere dager
Jon read Goethe's poem collections for many days
- (14) ?*Jon leste tre artikler i flere dag
Jon read three articles for many days'

According to an analysis implemented in this same grammar, as described in (Beermann and Hellan, 2000), in the verbal construction, unboundedness can be marked by a feature specification 'DELIMITED -' in the event-index of the verbal projection. An NP inducing this value can be marked 'BOUNDED -' in the ref-index of the nominal projection. Both annotations are exposed in MRS. Relative to the system we have so far outlined, the following algorithm can be used:

Adhering to the same lexicalist approach as above of representing semantically contentful inflection at lexeme level, non-boundedness is associated with a plural indefinite noun form as such. - The representation includes the feature 'BOUNDED -' already mentioned, and a quantificational relation *plur indef-q-rel*. - Doing this bottom-up, we need to ensure that the specifications mentioned are located in a constructional build-up leading to constructions of the pattern in (12), and not to patterns like (13) or (14). - Essential to ensure this is constraining the SPR list of a non-bounded noun form to be empty: combination with adjectives is possible, e.g., but not with quantifiers; only for nouns co-occurring with the latter will the combinatorics of a noun be checked via SPR-list cancellation. Thus, indefinite plurals will have the specification

$$\left[\dots \text{spec} \left\langle \left[\text{loc} \left[\begin{array}{l} \text{cat|val|spr} \\ \text{cont|index|bounded -} \end{array} \right] \right] \right\rangle \right]$$

This analysis illustrates one more respect in which, by partial involvement of the construct ‘Specifier’, we are in a position to control the assignment of fairly subtle semantic values.

However, as will be recalled from the discussion in section 2 concerning the following examples,

- (15) alle mine tre gutter
‘all my three boys’
- (16) de tre energiske guttene
‘the three energetic boys’

to adhere to the *Spec-unique* assumption, it would be desirable to count numerals as not being Specifiers. On that assumption, (14) is still not accounted for. As it seems, there is no non-ad hoc way of both preserving *Spec-unique* and regulating the boundedness phenomenon *exclusively* through the use of the Specifier mechanism: what we need (and as is implemented) is an additional feature which will distinguish those indefinite plurals that do combine with a cardinality expression from those which don’t. Hence, the Spec mechanism is not sufficient to encode all the boundedness distinctions needed.

We may also notice that for the case in (15), if we want to maintain the ‘at most one’-part of *Spec-unique*, and we maintain the treatment of definiteness inducers as Specifier items, then we cannot also treat the quantifier *alle* as a Specifier. In the grammar in question, *alle* is consequently treated as a head and the following nominal projection as its complement; semantically, this configuration is treated analogously to a partitive construction, such that both the definite Specifier and the quantifier can introduce a variable, one each. This is then an analysis complying with the overall desiderata concerning NPs and quantifiers mentioned initially, and, what is crucial in the present connection, in such a way as not to be in conflict with the role designed for the Specifier category.

5 Conclusion

Although there are some clear limitations to how pervasively one can apply the Specifier category in the analysis of Norwegian, our discussion warrants a conclusion to the effect that the

Specifier, as conceived under the ‘at most one’ part of the *Spec-unique*, is a useful construct, motivated not only by the *à priori* concerns outlined in the beginning, but also in the analysis of rather subtle facts pertaining both to syntax, morphology and semantics of NPs. Having demonstrated this for the case of Norwegian, it of course remains to provide similar demonstrations for other languages. Motivation for trying this would reside in part in the theoretical background of the notion mentioned earlier, and in part in reusability advantages that such a potential cross-grammatically valid module might provide for multilingual grammar engineering. However, the linguistic facts must come first, and the ‘Specifier’ will have to earn its status through accommodation of the facts language by language.

References

- S. Abney. 1997. *The English Noun Phrase in its Sentential Aspect*. PhD dissertation, MIT.
- D Beermann and L. Hellan. 2000. A treatment of directionals in two implemented hpsg grammars. In St Mueller, editor, *Proceedings of the HPSG04 Conference, Katholieke Universiteit Leuven*, pages –. CSLI Publications.
- E. Bender, D. Flickinger, and S. Oepen. 2002. The grammar matrix: An open-source starterkit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In NN, editor, *Proceedings of the Workshop on Grammar Engineering and Evaluation, Coling 2002, Taipei*, pages 8–14. NN.
- A. Copestake, A. Lascarides, and D. Flickinger. 2001. An algebra for semantic construction in constraint-based grammars. In *Proceedings of the 39th Annual Meeting of the ACL*, pages –.
- A. Copestake, D. Flickinger, I. Sag, and C. Pollard. submitted. Minimal recursion semantics: an introduction. Submitted December 2003.
- A. Copestake. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford.
- M.E. Kropp Dakubu. 2005. Noun phrase structure in ga. Talk given at NTNU Spring 2005.
- K. Donnellan. 1966. Reference and definite descriptions. *The Philosophical Review*.

- L Hellan. 1986. The headedness of noun phrases in norwegian. In P. Muysken and H. van Riemsdijk, editors, *Features and Projections*, pages -. Foris Publications.
- C. Pollard and I. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago University Press.
- C. Smith. 1991. *The Parameter of Aspect*. Kluwer Publishers.