The noun phrase in Mainland Scandinavian: an efficient implementation of a topological analysis

Anders Søgaard
Center for Language Technology

Petter Haugereid
Norwegian University of Science and Technology

Contents

1 Introduction

The goal of this report is twofold. One is to develop a theory about the noun phrase of Mainland Scandinavian in which the implementation of topological fields in typed feature structure is important, and another is to document the development of a Scandinavian Grammar Matrix, which relies on this topological implementation. Briefly, a Grammar Matrix is meant to be a set of generalized type declarations that are cross-linguistically useful, so that a language particular grammar gains

Table 1: Word Order in the Mainland Scandinavian Noun Phrase

P1	P2	P3	P4	P5	P6	P7
alle	disse	mine	sidste fire	mindste	gule	tæpper
kun	dette	universitets	ene		næsten vellykkede	projekt

from inheriting from the Grammar Matrix. In the case of Mainland Scandinavian, this idea is very appealing, and as is evident from this work, a Grammar Matrix for closely related languages can be very useful, since it ensures compatibility between the language particular grammars and can be employed for multilingual applications. The topological approach, i.e. direct declarations of word order, was a very straight-forward way to identify the commonalities and differences between the languages included in the Scandinavian Grammar Matrix; namely, Danish, Norwegian and Swedish (Mainland Scandinavian).

In recent years, mathematical linguistics has paid some interest to linear notions of word order, that is, non-hierarchical ones. One immediate problem encountered by such approaches to word order concerns theoretical efficiency (in the Chomskyan sense), i.e. it is really necessary to employ a specialized formal machinery to constrain word order appropriately? The answer to this question, at least wrt. Mainland Scandinavian, is negative. A field- or position-based theory of word order *can* be implemented in a conventional and mathematically sound typed feature structure grammar without further enrichment of the formal machinery or other complications of the theory. The article documents an implementation of sentence schemata theory in typed feature structure logic and explores its consequences. In addition, it introduces a novel theory of embedded phrases in the noun phrase, radically re-interpreting the inventory of syntactic categories for Mainland Scandinavian. The gains are two-fold. The grammar simplifies earlier accounts of noun phrase syntax considerably and thus provides important generalizations about this small set of languages (i.e. Danish, Norwegian and Swedish). The authors of this article strongly encourage its readers to read it as a contribution to both grammar engineering and general linguistics.

2 A Type-Hierarchical Implementation of Word Order Positions

The basic word order of the Mainland Scandinavian noun phrase is given in Table 1.[1] The schema is based on work by Lohmann (1956) and Underwood et al. (2000). In addition, superlatives precede positive adjectives (Neville, 2000). Only considering simple noun phrases, the schema ignores the positions of embedded phrases, i.e. conjuncts, measure noun phrases, prepositional phrases. Also the postnominal comparative is ignored in this section.

How are these facts captured in a phrase structure grammar? In the simple framework of context-free phrase structure grammars, there is no efficient way of doing so. Order constraints are inherent in rewriting rules, but these rules suffer from a considerable loss of generality. The standard extension of the context-free formalism to explain word order phenomena is linear precedence constraints. Unfortunately, there is no independent motivation for this autonomous syntactic

[1] One important exception is the distribution of pronominal possessives in Norwegian Bokmål. Pronominal possessives can also be postnominal in this language:

(1) gutten min seiler

In the implementation, two solutions to this problem were tested. A complex notion of contextual determination of category values was compared to the addition of a separate phrase for postnominal attachment of pronominal possessives (Søgaard and Haugereid, 2004). From an engineer's point of view, the separate phrase costs little in terms of efficiency. On theoretical grounds, the more complex treatment may be preferred. Both of these treatments avoid lexical duplication, which is the major problem with more traditional accounts such as Hellan and Haugereid (2004).

Table 2: Definitions and Translations of Basic Positioned Categories.

Position	Definition	Usual name
P1	$\begin{bmatrix} \text{D} = d1 \\ \text{V} = + \end{bmatrix}$	Quantifier
P2	$\begin{bmatrix} \text{D} = d2 \\ \text{V} = + \end{bmatrix}$	Determiner
P3	$\begin{bmatrix} \text{D} = d3 \\ \text{V} = + \end{bmatrix}$	Possessive
P4	$\begin{bmatrix} \text{D} = d4 \\ \text{V} = + \end{bmatrix}$	Numeral
P5	$\begin{bmatrix} \text{D} = d5 \\ \text{V} = + \end{bmatrix}$	Superlative
P6	$\begin{bmatrix} \text{D} = d6 \\ \text{V} = + \end{bmatrix}$	Adjective
P7	$\begin{bmatrix} \text{D} = d7 \\ \text{V} = - \end{bmatrix}$	Noun

module that bears little similarity to the formal machinery of context-free phrase structure grammar. Consequently, every account of word order based on independently motivated theoretical constructs will outrank the theory of linear precedence on grounds of theoretical efficiency, if the account offers a minimal level of generalization. In this article, it is demonstrated how a field- or position-based linear notion of word order can be implemented in a typed feature structure grammar, a mathematically sound formalism with independent motivation, i.e. this formalism allows for linguistic generalizations of considerable depth (e.g. Flickinger, 2000).

Briefly, a typed feature structure grammar is a set of typed feature structures that are defined on a finite set of features Φ and a type hierarchy $\langle \Sigma_\tau, \sqsubseteq \rangle$ where Σ_τ is a set of types. In other words, it is a tuple $\langle Q, \top, \delta, \theta \rangle$ where:

- Q is a finite set of nodes
- $\top \in Q$
- $\delta : Q \times \Phi \rightarrow Q$ is a partial feature-value function
- $\theta : Q \rightarrow \Sigma_\tau$ is a partial typing function
- if $\nu_i \in Q$ and $\nu_i \neq \top$ then ν_i is a δ-descendant of $\top$

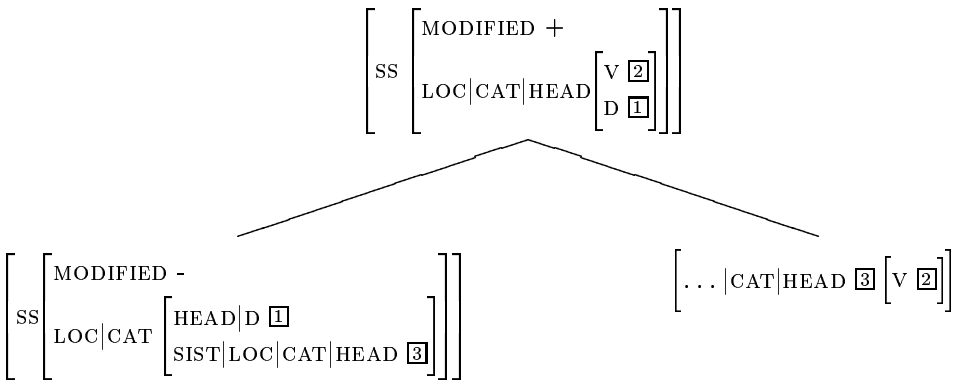
Unification of α and β is defined as $\alpha \bowtie \beta$, i.e. the greatest lower bound or infimum of α and β. If there is no infimum of α and β, $\alpha \bowtie \beta = \bot$. Efficiency depends on the size of Φ and Σ_τ. In addition, the number of phrasal types $\Pi \in \Sigma_\tau$ that the parser may instantiate as linguistic rules is important in this respect. Our intuition has been that it was more important to limit the number of features than to limit the number of types, and even more important to limit the number of rules.

Our theory of word order exploits the formal nature of type hierarchies. Briefly, a hierarchy is defined such that each position is incompatible with any other word order position, i.e. $p_i \bowtie p_j = \bot$, but at the same time, it introduces types that are compatible with different subsets of the set of positions. This enables the statement of a number of important precedence constraints. In some more detail:

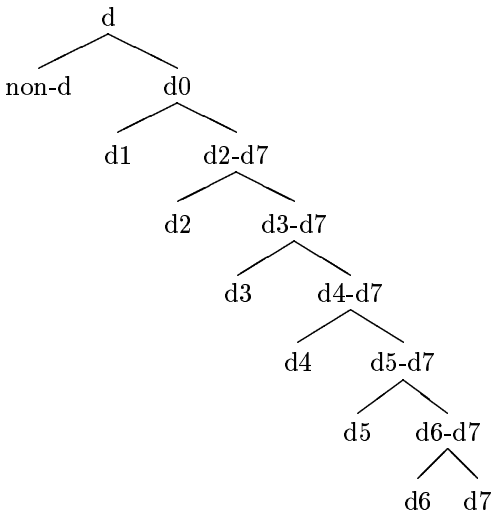
Syntactic categories are factorized, the two factors being the boolean feature V and the attribute D, which enables distinctions between four different determiner classes, adjectives and nouns.[2] The determiner category values are labelled d1, d2, d3, etc.. The four determiner classes are numbered

consecutively, adjectives are given the D values *d5* (for superlatives) and *d6*, while nouns are given the value *d7*. Only nouns have negative V values. This is resumed in Table 2.

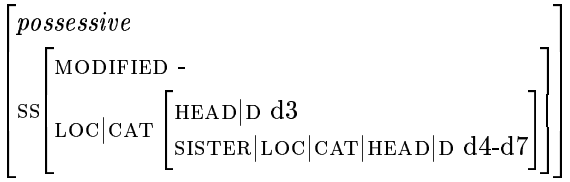
Assume this partial declaration of the *noun-phrase* rule (omitting agreement, semantic composition, etc.):



It is the *d*-hierarchy that enables the statement of word order constraints. Consider the hierarchy for a minute:



From Table 1 and 2 and *noun-phrase*, it is obvious how precedence constraints can be embedded in type declarations, e.g. the precedence of Danish (pronominal and genitival) possessives to numerals can be stated as a type *possessive*:



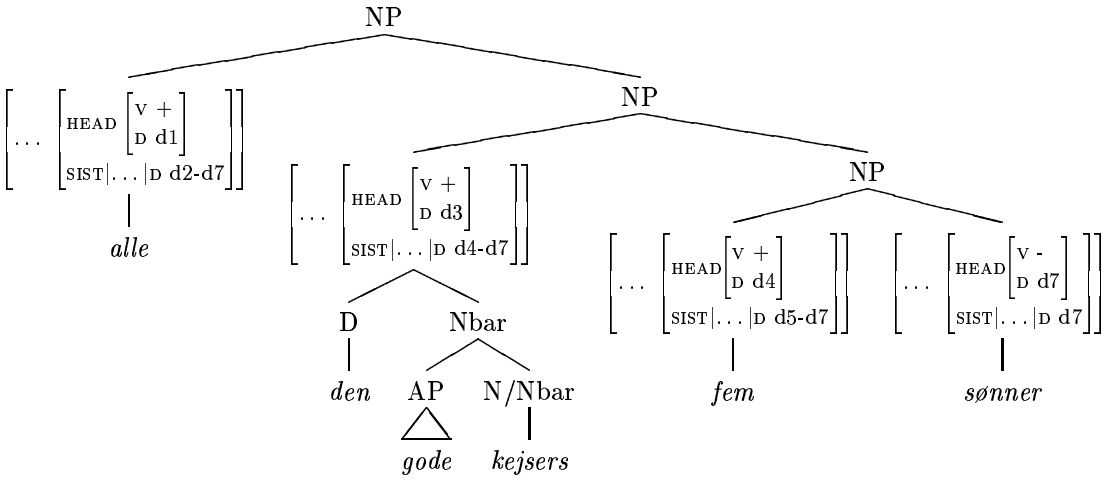
In *noun phrase*, *possessive* can "specify" (let's call this relation *D-specification*) numerals, superlatives, adjectives and nouns, since their categories are compatible with the sister category of the possessive type. On the other hand, the possessive type can only be D-specified by quantifiers and demonstratives. Similarly, demonstratives can only be D-specified by quantifiers, but D-specifies possessives, numerals, superlatives, adjectives and nouns.

This kind of implementation accounts for the grammaticality of (2):

[2] While Underwood et al. (2000:153) argue that Danish has three determiner classes, Lohmann (1956) argues that in Swedish there are six.

(2) alle den gode kejsers fem sønner

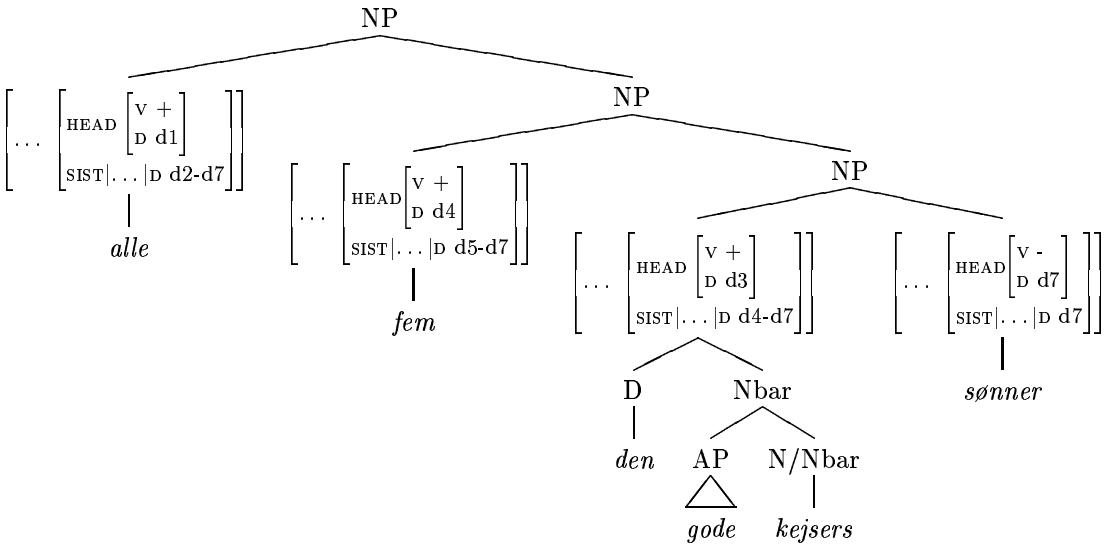
This is clear from the tree representation:



and the ungrammaticality of (3):

(3) *alle fem den gode kejsers sønner

Here's the tree:



Informally, *fem* cannot specify its sister because of the unifications in *noun-phrase*. The infimum of *d3* and *d5-d7* is $\perp$. The right-branching structure is the result of the following constraint on *noun-phrase*: The left daughter must always be non-modified, while the mother is modified. The right daughter is unspecified in this regard.

Imagine a one-rule grammar where *noun-phrase* is the only phrasal type, i.e. the grammar only parses by this rule, and call it G_1. Of course G_1 is very flexible. G_1 allows every combination of determiners, adjectives and nouns that obeys the implemented precedence constraints to function as noun phrases. One important *extension* to G_1 is a rule that licenses embedded noun phrases, a rule which will concern us later. One immediate restriction is the introduction of agreement constraints. Our account of agreement is largely conventional, i.e. we introduce agreement features on semantic indeces (see e.g. Underwood et al., 2000:182), but we adopt the reinterpretation of

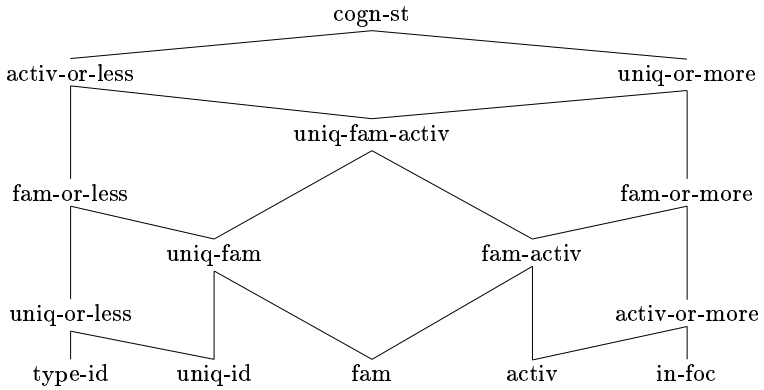
declension phenomena introduced by Borthen and Haugereid (2004). In addition, we introduce the possibility of quantifier agreement constraints.

While person and number agreement seem to be quite consistent across the Scandinavian languages, the gender systems differ. Norwegian Bokmål has three genders, while Danish and Swedish only have two, i.e. common gender and neuter. In addition, Bokmål has a feminine form. The use of feminine forms in Bokmål is often a matter of style. In Nynorsk, there is a three-way distinction between feminine, masculine and neuter. In Swedish, there may be morphological evidence of a mass-count distinction too (Delsing, 1993).

2.1 Cognitive status

The three conventional agreement features for Mainland Scandinavian are PERS(ON), NUM(BER) and GEN(DER). A fourth feature was adopted, namely COGN(ITIVE)-ST(ATUS), from Borthen and Haugereid (2004), and a fifth one, QUANT(IFIER)-TYPE, was used to compute quantificational structures. The constructional computation of cognitive status or givenness and quantificational structure is very important for the efficiency of the grammar and is also empirically motivated (see Søgaard, in prep.). The standard design where these features are specified for semantic indeces, may be abandoned, if coindexation of variables, e.g. for grammmatical anaphora or coordinative structures, is necessary. The features can be moved to SYNSEM|LOCAL|AGR, or the more complicated design of Borthen and Haugereid (2004) can be adopted.

The COGN-ST attribute and type hierarchy below was designed to account for various grammaticalizations of *givenness* (in the sense of Gundel et al., 1993), e.g. in light pronouns and bare singulars (Borthen and Haugereid, 2004).



In our grammar design, this hierarchy covers or makes sense of a variety of phenomena, including:

- declension[3]
- "nounless" indefinites
- vocatives
- specification of givenness in logical form

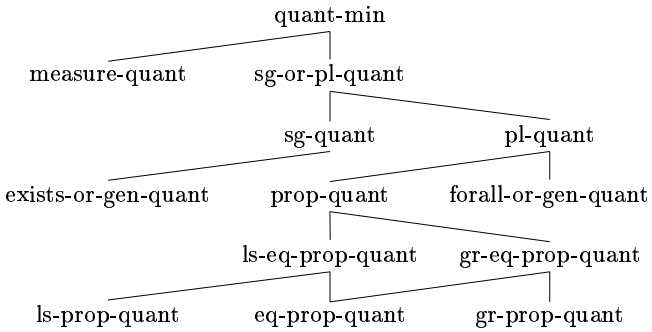
2.2 Quantification

Certain assumptions about quantificational structure are inherent in the MRS algebra. Most are rather conventional, i.e. that all quantifiers have both a restriction and nuclear scope. In addition

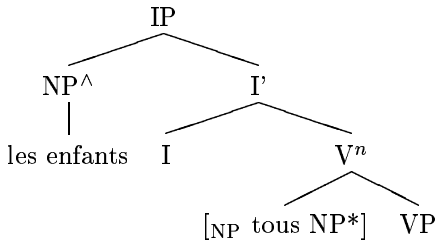
[3]Netter's (1994) boolean DECL feature has been suggested as an adequate place to restrict adjectival agreement in Mainland Scandinavian (e.g. Neville, 2000; though see Mikkelsen (1998) for some discussion). The boolean distinction has a direct translation into the *cogn-st* hierarchy, namely the distinction between *uniq-or-more* and *type-id*.

to these assumptions, we propose to introduce quantifiers at the constructional level, rather than at the lexical level. The motivations for this are both empirical of nature and relates to efficiency.

QUANT-TYPE is our primary mechanism for computing the quantifier of a noun phrase. In a sense, while the quantifier is chased up the syntactic tree, it is forced down some type hierarchy, always looking for a greatest lower bound:



If you introduce quantifier phrases, or if you introduce a flat right branching analysis, traditional implementations of Head-Driven Phrase Structure Grammar (e.g. the English Resource Grammar, Flickinger, 2000, and the Babel system, Müller, 2004) are forced to introduce lexical duplication or lexical rules to reflect the quantificational and non-quantificational "uses" of specifiers in the context of multiple specification. This can be avoided if every element in the noun phrase is allowed to constrain its quantificational force, but the quantifier is only inserted at the constructional level. The insertion is lexically dependent, so only one noun phrase construction is needed. Empirical motivation for a constructional account of quantification comes from so-called "floated quantifiers". In Head-Driven Phrase Structure Grammar, the external argument is available for verbs and adverbs. Sportiche (1988) argued quite convincingly for a stranding analysis for French, i.e. that quantifiers strand when the subject moves from the original subject position:



There are various reasons why this analysis does not apply well to Mainland Scandinavian: (i) If the floated quantifiers are related to the prenominal quantifier position, movement effects grammaticality, cf. (4a,b), (ii) on this analysis, there is no explanation why only a limited set of quantifiers float, (iii) the fact that the floated quantifier does not strand with the original subject position in passives; cf. (4c)

(4) (a) Drengene blev alle ti arresteret.
(b) *Alle ti drengene blev arresteret.
(c) *Drengene blev arresteret alle.

(i) can be remedied by relating floated quantifiers to partitive constructions; this is also Sportiche's proposal. Sportiche still needs to address the question why only a limited set of quantifiers are left stranded, since all quantifiers occur in partitives. In fact, the quantifiers which most frequently strand, *alle* (all) and *begge* (both), are somewhat odd in partitive contexts:

(5) ?Alle/begge af drengene blev arresteret.

If these are in fact ungrammatical, this is an even stronger argument against Sportiche's stranding analysis. On the analysis which is implemented in the extension of our topological noun phrase grammar, floated quantifiers are just quantifiers which function somewhat like subject-oriented adverbs (modifying the external arguments). Their distribution and behavior is semantically motivated, and the analysis depends on the constructional account of quantification. See Søgaard (in prep.).

2.3 Double articulation

In simple contexts, the double definite in Norwegian Bokmål has the following distribution:

(6) (a) en gutt seiler
(b) *en gutten seiler
(c) den gutten seiler
(d) *den gutt seiler
(e) alle gutter seiler
(f) alle guttene seiler
(g) *alle de gutter seiler
(h) alle de guttene seiler

Basically, the restriction seems to be that whenever you have a *d2-word* with [COGN-ST = *uniq-or-more*] specifying the noun, it must be definite. In principal, the distribution in (6) could be explained by agreement in cognitive status, if we assumed the non-definite form to be [COGN-ST = *type-id*] . But we can't. Consider:

(7) mine gutter seiler

The boys referred to here are at least as contextually salient as the boys in (6h). Consequently, the uninflected plural form must be underspecified with respect to cognitive status. Instead we introduce a separate feature NFORM to restrict the inflection of the noun; cf. the use of NFORM in NorSource (Hellan and Haugereid, 2003). In Danish, the non-definite form is also underspecified wrt. to cognitive status. The definite form, the distribution of which is the result of a lexical change in *d*-value (to *d2*), is specified to be *uniq-id*. For a similar analysis, see Hankamer and Mikkelsen (2002).

2.4 Some Preliminary Semantics

What is the translation of (8) into logical form?

(8) alle guttene seiler

The output Scoped MRS has the following form:

(9)
```
proposition_m_rel(forall-or-gen-quant(x4, sail_rel(e2) ∧
arg1_rel(e2,x4), quantifier_rel(x4) ∧ boy_rel(x4) ∧
uniq-or-more(x4) ∧ count(x4)))
```

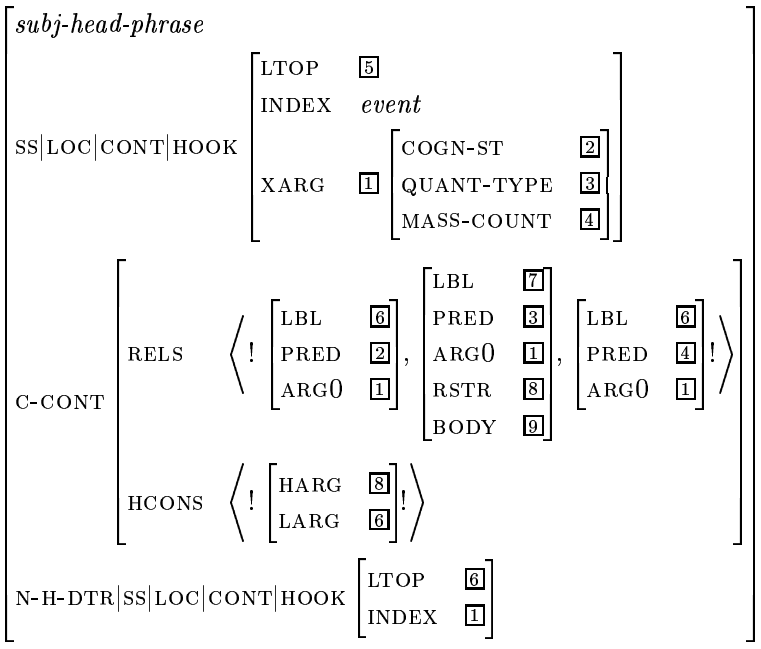
Since Scoped MRSs present quantifiers as 3-place relations over variables, nuclear scopes and restrictions, respectively, (9) corresponds to (10), omitting the propositional relation, i.e. the classification into sorts of speech acts:

(10) $Q^{\forall}_{\Gamma} x_1.\mathbf{quantifier}(x_1) \wedge \mathbf{boy}(x_1) \wedge \mathbf{uniq\_or\_more}(x_1) \wedge \mathbf{count}(x_1) \rightarrow \mathbf{sail}(e_1) \wedge \mathbf{arg1}(e_1, x_1)$
where $Q^{\forall}_{\Gamma}$ is the supertype of the universal and the generic quantifier.

$\lambda x_1.\mathbf{quantifier}(x_1) = \top$ for all $\{x_1|g(x_1)\}$, where g is the identifier of the assignment function. $\{x_1|g(x_1)\} \in \lambda x_1.\mathbf{count}(x_1) \vee \mathbf{mass}(x_1)$, i.e. the two characteristic functions partition the set of individuals in U.

Note: There is no quantification over events in the MRS formalism. Ontologically, this is by no means a safe move; for some discussion, see Partee (2000). While we wish to stay agnostic as to whether or not quantification over events is necessary, e.g. in examples such as *in every burning, oxygen is consumed*, we adopted the original design so as not to complicate matters. A MRS-based LKB grammar with quantification over events is described in Søgaard (2004).

Let's consider the composition of (8), omitting the propositional relation:



The MRS formalism allows for underspecification, but for clarity of exposition we've only considered fully specified forms. The quantifier gets it scope via the handle constraint on the HCONS list in the constructional contents. While the verbal semantics end up in the nuclear scope of the quantifier, the elementary predicates of the non-head daughter are put in the restriction; cf. the unification indicated by the 6 tag. Since the values of both COGN-ST, QUANT-TYPE and MASS-NOUN are subtypes of *predsort*, they can be unified with the PRED values to compute, respectively, the predicate specifying givenness, the quantifier and whether the referent is an object or some matter; cf. Link (1986).

In simple noun phrases, most constituents translate into 1-place relations. A couple of familiar exceptions include:

- numerals
- the quantifier *begge/båda*
- pronominal possessives

It has been attempted to provide first order analyses of these phenomena. This is motivated by efficiency and compatibility with postprocessing tools such as theorem provers and model builders. The first order analysis of numerals is of this form, e.g.:

(11) fem pojkar seglar

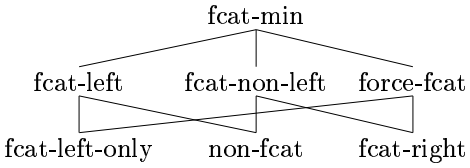
(12) $\exists x_2.(\forall x_1.\mathbf{boy}(x_1) \rightarrow \mathbf{five}(x_2) \wedge \mathbf{member\_of}(x_1, x_2)) \rightarrow \mathbf{sail}(e_1) \wedge \mathbf{arg1}(e_1, x_2)$

The implementation of numerals complicates the semantics of the noun phrase rule, since the index is changed by the specifier. The representation in the English Resource Grammar (Flickinger, 2000) is much simpler, but second order. For practical purposes, e.g. generation, numerals can just be translated into 1-place predicates over individuals. If the more adequate first order proposal

is to be implemented the percolation of the index must not be licensed by the noun phrase rule, but most include lexical coindexation of values transported independently. The numerals can then break the chain and introduce a new semantic index. Of course it must then also introduce a quantificational structure for the index of its sister.

3 The Syntax and Semantics of Embedded Noun Phrases

To account for the syntax of embedded noun phrases, we introduce a new attribute FCAT on *cat-extended*. This feature has some independent motivation.



Intuitively, FCAT indicates the ability of constituents to appear in functional positions, such as that of conjunctions and genitival clitics. Most (if not all) closed class items have FCAT values that are subtypes of *force-fcat*, while most open class items are *non-fcat*. Certain kinds of adjectives and nouns also appear in functional positions in restricted contexts, e.g. measure noun phrases. Some phenomena of relevance are considered below.

Embedded noun phrases are all instances of the type *add-noun-phrase*. The essentials of this type are:

$$\left[\begin{array}{l} \text{SS|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \left[\text{D } \boxed{1}\right] \\ \text{FCAT non-fcat} \\ \text{SISTER} \left[\text{LOC|CAT|HEAD} \left[\text{D } \boxed{2}\right]\right] \end{array}\right] \\ \text{HEAD-DTR} \left[\text{SS|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \left[\text{D } \boxed{1}\right] \\ \text{FCAT force-fcat} \\ \text{SIST} \left[\text{LOC|CAT|HEAD} \left[\text{D } \boxed{2}\right]\right] \end{array}\right]\right] \\ \text{NON-HEAD-DTR} \left[\text{SS|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \left[\text{D d}\right] \\ \text{FCAT non-fcat} \end{array}\right]\right] \end{array}\right]$$

$$\left[\text{SS|LOC|CAT|FCAT fcat-left}\right] \quad \left[\text{SS|LOC|CAT|FCAT fcat-non-left}\right]$$

Embedded noun phrases are leftheaded or rightheaded. If they're rightheaded, the FCAT value of the head daughter should be unifiable with the infimum of *force-fcat* and *fcat-left*, and if they're leftheaded, the infimum of *force-fcat* and *fcat-non-left*. Similarly, for the non-head daughters.

For the embedded noun phrases to interact correctly with the basic noun phrase, the MODIFIED value of the mother must be specified. In our set-up, this value is dependent on the head daughter, i.e. the functional constituent.

[The subsection on coordination has been removed. A separate article will be published.]

3.1 Genitival Possessives

Danish and Swedish are the least interesting languages with respect to possessives. Consequently, these are the ones *we* are mostly interested in, since the possessive constructions of these two

languages is exactly the intersection of possessive constructions in Mainland Scandinavian (incl. dialects such as Northern Swedish and Western Jutlandic). In these languages, there are only two possessive constructions, apart from idiosyncratic prepositional possessives,[4] i.e. pronominal possessives and genitival possessives. The former are treated as prenominals with D value *d3*. This section is thus devoted to the syntax and semantics of the genitival ones.

(13) hundeejeres katte sover

(13) translates $\forall x_1.\mathbf{cat}(x_1) \wedge \exists x_2.\mathbf{dog\_owner}(x_2) \rightarrow \mathbf{poss\_by}(x_1, x_2) \rightarrow \mathbf{snore}(e_1) \wedge \mathbf{arg1}(e_1, x_1)$. Genitival clitics and syntactically genitival pronouns (in Norwegian) are of the type *gen-functional-lex-item*, which is a suptype of *functional-lex-item* and embodies the following additional restrictions:

$$\left[\begin{array}{l} \textit{gen-functional-lex-item} \\ \left[\begin{array}{l} \text{MODIFIED notmod} \\ \text{SS} \left[\begin{array}{l} \text{LOC}|\text{CAT} \left[\begin{array}{l} \text{FCAT fcat-right} \\ \text{HEAD} \left[\begin{array}{l} \text{V} + \\ \text{D d3} \end{array}\right] \\ \text{SIST} \left[\text{HEAD} \left[\text{D d4-d5}\right]\right] \end{array}\right] \end{array}\right] \end{array}\right] \end{array}\right]$$

To cope with the language-specific genitival possessives of Norwegian, we added a *norwegian-post-poss-phrase* to *norwegian.tdl*. In addition, we added the standard possessive preposition *til* to the lexicon. It translates into $\lambda x_2.\lambda x_1.\mathbf{poss\_by}(x_1, x_2)$. This enables the following generations:

(14) (a) gutten til speiderlederen seiler →
speiderlederens gutt seiler →
speiderlederen sin gutt seiler

(b) guttene mine seiler →
mine gutter seiler

The changes in definiteness are due to specific restrictions on the prenominal and postnominal possessives in Norwegian. Our general treatment of nominal definiteness is described above. Since transfer rules are costly in machine translation systems, the fact that synonomous constructions receive the same semantic representation enhances the efficiency of such a system, minimizes the number of transfer rules, i.e. language-internal transfer rules duplicate the number of language-external transfer rules, and thus eases the construction of a broad-coverage translation system. Our ability to generate the forms in (14) effortlessly is a clear advantage of our approach.

3.2 Other Kinds of Embedded Noun Phrases

In this section, we briefly discuss related phenomena, i.e. prepositional phrases, measure noun phrases, partitives and compounds. On our account, these are too all instances of *add-noun-phrase*. While some of these phrases are headed by closed class constituents which can be clearly marked as functional categories ([FCAT = +]), others, incl. the latter, are headed by open class constituents which are not (at least not unambigously) functional elements in the employed sense of the word. This is illustrated in the type hierarchy at page 10.

Prepositional phrases In our grammar, there is no difference between conjunctions and prepositions, except of the PRED value of the 2-place relation. Of course, there is no resolution algorithm for prepositions apart from ellipsis.

[4] Opposed to what Delsing (1993:150) calls *standard* prepositional possessives in Norwegian.

Measure noun phrases Measure noun phrases (or *quantifying nouns*; cf. Delsing, 1993) are somewhat more complicated. This is the only ∅-derivation lexical rule we admitted ourselves this far. Quantifying nouns behave very differently from standard common nouns: While you can turn *en flaske øl* into *en ølflaske* and *en bukett blommor* into *en blombukett*, the following compounds are ungrammatical:[5]

(15) (a) *ett turistantal
(b) *en vinliter

Similarly, while *en god kop dårlig kaffe* sounds like a contradiction, *en lang række korte udtalelser* does not. The latter certainly involves a new quantified structure, i.e. an *add-noun-phrase*, the pseudo-partitive construction may not. Currently, the pseudo-partitive is not allowed in our grammar, but a simple modification makes room for an intersective analysis: Simply let the sister of a noun have the same *d*-value as the noun itself.

The measure noun (as a functional category) has the following specifications:

$$\left[\textit{measure-noun-lex-item}\ \left[\text{SS}\ \left[\text{MODIFIED notmod}\ \text{LOC}|\text{CAT}\ \left[\text{FCAT fcat-right}\ \text{HEAD}\ \left[\begin{array}{l}\text{V -}\\ \text{D d1}\end{array}\right]\ \text{SIST}|\text{LOC}|\text{CAT}\ \left[\text{HEAD}\ \left[\begin{array}{l}\text{V +}\\ \text{D d6-n}\end{array}\right]\ \text{FCAT fcat-non-left}\right]\right]\right]\right]\right]$$

Modification of measure adjectives When it comes to modification of measure adjectives, as in (16a-b), the picture looks slightly different. Apparently, it's not grammatical when the adjective is modified by a complete noun phrase. In schematic terms, only noun phrases consisting of **P4-6** constituents can modify measure adjectives.

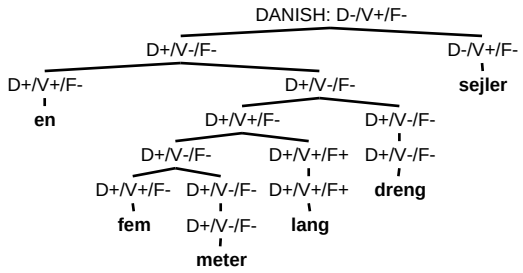
(16) (a) en fem meter lang dreng sejler
(b) *en alle fem meter lang dreng sejler

This restriction was implemented in the following way: Since numerals are underspecified with regard to their quantifier type, but determiners are not, the measure adjective constrains the possible quantifier type of the preceding noun phrase to a type incompatible with all determiners. This quantifier type is called *measure-quant*; cf. the hierarchy at page 7.

This move also reduces ambiguity in a very natural way. Alternatively, we would have to say that (17) was somehow ambiguous between the standard reading, which we argue *is* grammatical (see above), and a reading where the measure adjective licenses a new quantifier. Of course, only the standard reading is interesting.

(17) et langt snorker

The syntactic tree for (16) is given here:



[5] On the notion of ungrammaticality wrt. compounds, see Søgaard (2004).

Embedded Phrases Headed by *af/av* The partitive construction, i.e. an embedded phrase headed by *af/av*, is pretty simple. The functional constituent *af/av* has the following specifications:

$$\left[\begin{array}{l} \textit{part-functional-lex-item} \\ \text{SS}\left[\begin{array}{l} \text{MODIFIED hasmod} \\ \text{LOC}|\text{CAT}\left[\begin{array}{l} \text{FCAT fcat-non-left} \\ \text{HEAD}\left[\begin{array}{l} \text{V -} \\ \text{D conj} \end{array}\right] \end{array}\right] \end{array}\right] \end{array}\right]$$

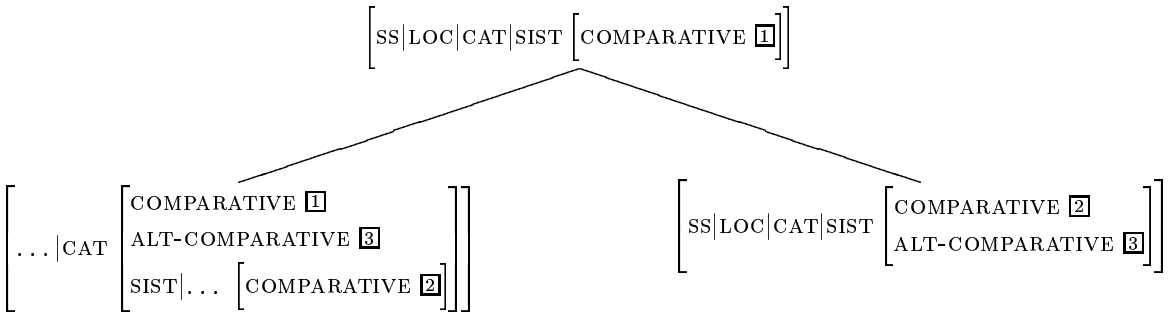
Consequently, the embedded phrase branches off to the right, and there are no restrictions on the categories appearing to the left off and within the added phrase:

(18) (a) de fleste af de fem drenge sejler

Comparative Phrases An embedded phrase with a complicated distribution is the comparative phrase. Consider the following data:

(19) (a) *en dreng hurtigere sejler
(b) *en dreng end en pige sejler
(c) en dreng hurtigere end en pige sejler
(d) en hurtigere dreng end en pige sejler
(e) *en hurtig dreng end en pige sejler
(f) *en hurtigst dreng end en pige sejler

Because our grammar does not allow subcategorization within the noun phrase, there is really no way of stating the dependency between the comparative adjective and the embedded noun phrase. Instead a pair of boolean features is added, COMPARATIVE and ALT-COMPARATIVE. First, the following percolation of values was added to *noun-phrase*:



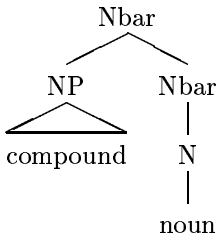
The ungrammaticality of (19) is caused by the general restriction on nouns that their right sisters are non-verbal ([V = -]). Since the comparative adjective can occur alone in Mainland Scandinavian, when it is premodifying, the ALT-COMPARATIVE feature allows the comparative adjective to restrict its right sister to be comparative, blocking:

(20) *en dreng hurtigere og en pige sejler

The restriction on the embedded phrases headed by *end/enn/än* that they must be preceded by comparative adjectives, is the result of the lexical specification of comparative adjectives: While every other lexical item unifies the value of SISTER... COMPARATIVE and COMPARATIVE, the comparative adjectives do not, thus breaking the chain of percolation. If the comparative phrase is then specified to be [COMPARATIVE = +] , we can simply specify the subject to have a negative COMPARATIVE value.

Of course we're not too happy about introducing *two* features for comparatives, and it should be possible to somehow conflate them and still get the right predictions for (19).

Some Remarks on Compound Nouns If we underspecify the FCAT value of nouns, it should be possible to implement compound nouns as instances of *add-noun-phrase*, conforming to the standard analysis; cf. Radford (1980):



If we want to say that a noun, rather than an *NP*, expands the $\bar{N}$, this amounts to saying that the left sister of the noun in the *add-noun-phrase* should be $[\text{LEX} = +]$.

3.3 Modification of Modifiers

In our specification, we have added a few small phrases for modification of modifiers, two of which are common to all the Scandinavian languages; namely, the *adj-mod-phrase* and the *num-mod-phrase*. Many features of these two phrases are expressed in their common supertype, *non-quant-mod-phrase* (for non-quantifier-introducing phrase). The third special phrase is the *norwegian-post-poss-phrase*, found in the *norwegian.tdl* type file. The latter is not really necessary, in that the postnominal pronominal possessive could be accounted for in the *noun-phrase*. However, since we were not writing a Norwegian grammar, but a specification of the common aspects of Mainland Scandinavian, we did not want to complicate the general syntax of the noun phrase on the expense of the two other languages.

Consider the supertype *non-quant-mod-phrase*:

$$\begin{bmatrix} \textit{non-quant-mod-phrase} \\ \text{MODIFIED notmod} \\ \text{FCAT non-fcat} \\ \text{SISTER|HEAD} \left[\text{D d5-d6}\right] \\ \text{H-D|MODIFIED notmod} \\ \text{N-H-D|MODIFIED notmod} \end{bmatrix}$$

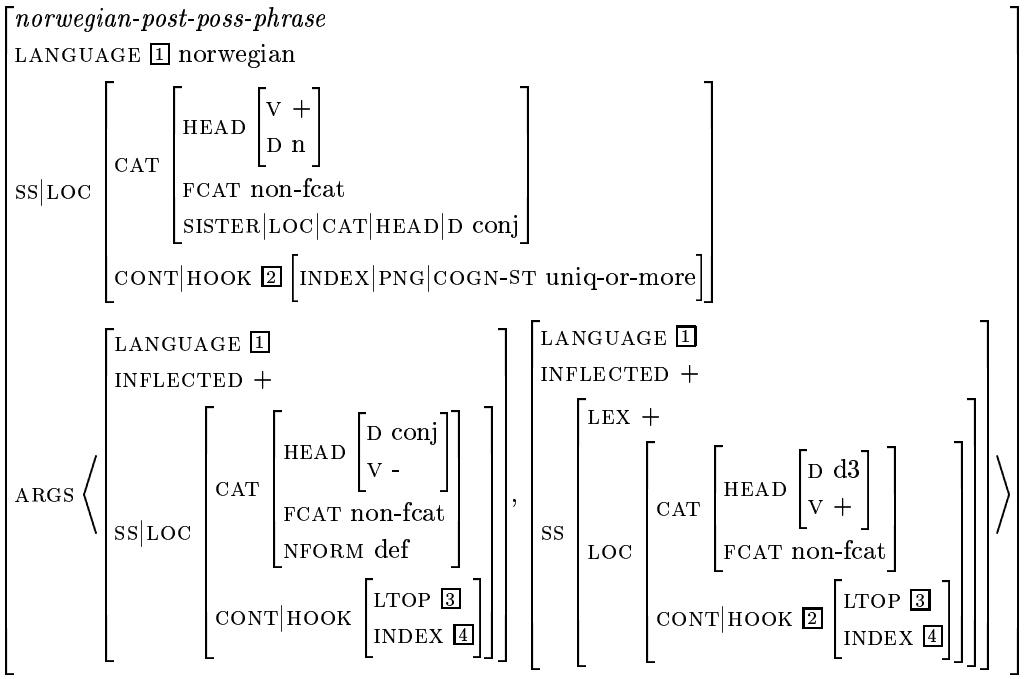
The subtypes mostly concern the semantics. In addition, *num-mod-phrase* further specificies that:

$$\begin{bmatrix} \textit{num-mod-phrase} \\ \text{HEAD} \left[\text{D d4}\right] \\ \text{N-H-D pl-adj-word} \end{bmatrix}$$

The further specification of *adj-mod-phrase* is:

$$\begin{bmatrix} \textit{adj-mod-phrase} \\ \text{MODIFIED notmod} \\ \text{FCAT non-fcat} \\ \text{SISTER|HEAD} \left[\text{D d6-n}\right] \\ \text{H-D|HEAD} \left[\text{D d6}\right] \\ \text{N-H-D} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{D non-d} \\ \text{V} + \end{bmatrix} \\ \text{FCAT fcat-left-only} \end{bmatrix} \end{bmatrix}$$

Consider finally the *norwegian-post-poss-phrase*.



This type accounts for the possessive construction in (21):

(21) gutten min seiler

The various phrases licensing modification of modifiers can of course be collapsed into one where something like the MOD feature is used to constrain the daughters pair-wise.

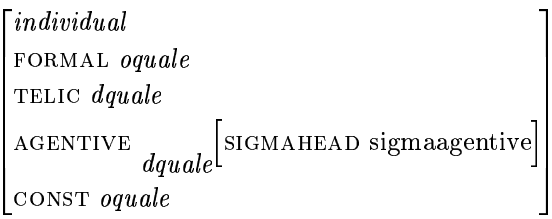
4 Complex Issues in the Semantics of Noun Phrases

The current grammar adopts a simplistic lexical semantics. Some phenomena exploit richer lexical structures, e.g. a qualia structure:

- compound nouns (Johnston and Busa, 1999; Bouillon and Bassac, 2001)
- presuppositional spatial prepositions, e.g. *at the blackboard* vs. *by the blackboard*
- some kinds of adjectives (cf. Pustejovsky, 1993)

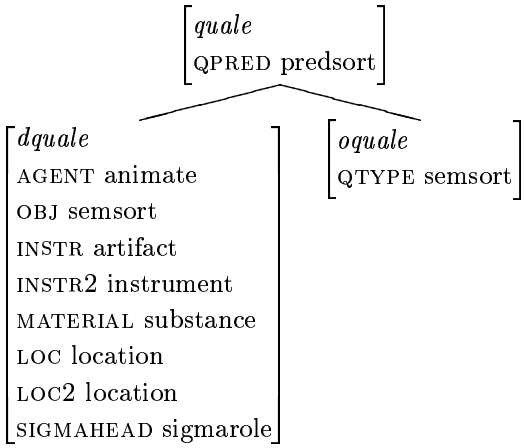
The implementation of a generative lexicon in the LKB system is described in Søgaard (2004). For Mainland Scandinavian, ressources for automatic derivation of lexicons are provided by the SIMPLE dictionaries. (Converters are available upon request.)

Since Søgaard (2004), some improvements have been made to the qualia implementation. Most importantly, it is now possible to generate from qualia-derived MRSs. The basic feature geometry is the same. First SORT on *individual* is substituted with the basic attributes of qualia structure, i.e.:[6]



[6]In Søgaard (2004), a fifth quale is introduced for constraining *contour* values of semantic types; the quale is omitted here for clarity of exposition.

The values of the different quales are defined in the following type hierarchy:[7]



5 Lexicon and Inflection

5.1 Lexical Entries

Lexemes are specified with regard to four features: STEM, SYNSEM|LKEYS|KEYREL|PRED, LANGUAGE and INFLECTION. The value of STEM is the orthography of the lexical item. The value of SYNSEM|LKEYS|
KEYREL|PRED is the PRED value of the main relation of the lexeme. In our grammar all PRED values are given in English. The LANGUAGE and INFLECTION features are there in order to keep track of which language they belong to and which inflectional pattern they have. The lexical entry for *dreng* has *danish* as value of LANGUAGE and *c3* as value of INFLECTION. The entry for *gutt* has *norwegian* as value of LANGUAGE and *c1* as value of INFLECTION, and the entry for *pojke* has *swedish* as value of LANGUAGE and *c5* as value of INFLECTION:

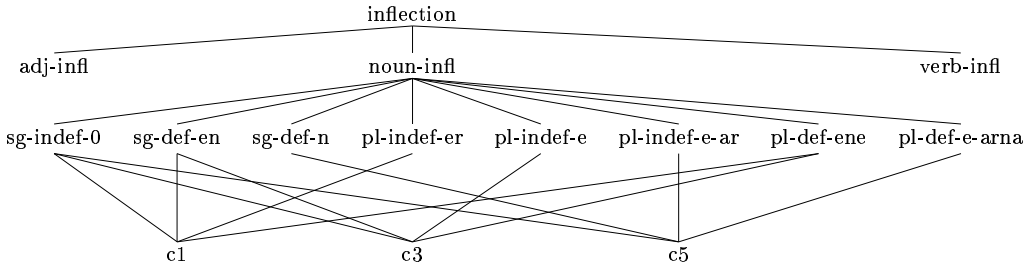
```
dreng := n-comm-le &
  [ STEM < ‘‘dreng’’ >,
    SYNSEM.LKEYS.KEYREL.PRED ‘‘_boy_n_rel’’,
    LANGUAGE danish,
    INFLECTION c3 ].

gutt := n-comm-le &
  [ STEM < ‘‘gutt’’ >,
    SYNSEM.LKEYS.KEYREL.PRED ‘‘_boy_n_rel’’,
    LANGUAGE norwegian,
    INFLECTION c1 ].

pojke := n-comm-le &
  [ STEM < ‘‘pojke’’ >,
    SYNSEM.LKEYS.KEYREL.PRED ‘‘_boy_n_rel’’,
    LANGUAGE swedish,
    INFLECTION c5 ].
```

5.2 Inflectional Rules

c1, *c3* and *c5* are codes which constrain the number of inflectional rules that apply to a lexeme. In the hierarchy below, the intermediate types are constraints that the inflectional rules put on their daughters. If a lexeme has an inflectional code type that is compatible with the constraint the inflectional rule puts on its daughter, the inflectional rule may apply. In the hierarchy below, the code *c3* for *dreng* is compatible with *sg-indef-0*, *sg-def-en*, *pl-indef-e* and *pl-def-ene*. We get the following inflections for *dreng* (*c3*), *gutt* (*c1*) and *pojke* (*c5*):

[7] Bouillon and Ballac (2001) suggest to represent modality in the qualia structures. It is not clear to us if this is necessary or even justified. Some discussion of this will follow in forthcoming papers.

Sing.Indef.	Sing.Def	Pl.Indef	Pl.Def
dreng	drengen	drenge	drengene
gutt	gutten	gutter	guttene
pojke	pojken	pojkar	pojkane



The code *sg-indef-0* means that the singular indefinite form doesn't add any inflection. The code *sg-def-en* means that the suffix *-en* is added by the inflectional rule. The code *pl-indef-e-ar* means that an *e* is removed from the end of the word and the suffix *-ar* is added. In this way, the inflectional mechanism can produce *gutten* from *gutt* and *pojkar* from *pojke*. The inflectional code constraint is placed on the daughter of the inflectional rule.

```
sg-indef-noun_infl_rule :=
sg-indef-noun-word &
  [ STEM #1,
    LANGUAGE non-danish,
    DTR [ STEM #1,
          INFLECTION sg-indef-0 ] > ].

d1-sg-def-noun-en_infl_rule :=
%suffix (* en)
infl-sg-def-noun-ltow-rule &
  [ DTR.INFLECTION sg-def-en ].

pl-indef-noun-e-ar_infl_rule :=
%suffix (e ar)
infl-pl-indef-noun-word &
  [ LANGUAGE swedish,
    DTR.INFLECTION pl-indef-e-ar ].
```

In addition to constraining the daughter with regard to inflectional code, the inflectional rules may also constrain which language they apply to. The *pl-indef-noun-e-ar_infl_rule* only applies to Swedish. And the *sg-indef-noun_infl_rule* doesn't apply to Danish. Instead of *sg-indef-noun_infl_rule*, *sg-noun_infl_rule* accounts for the non-inflected singular form in Danish.

The derivation of *pojkar* looks like this:

$$\left[\begin{array}{l} \textit{pl-indef-noun-word} \\ \text{STEM} \left\langle \text{“pojkar”} \right\rangle \\ \text{SYNSEM } \boxed{2} \\ \text{LANGUAGE } \boxed{1}\ \textit{danish} \end{array}\right]$$

$$\Big|$$

$$\left[\begin{array}{l} \textit{n-comm-le} \\ \text{STEM} \left\langle \text{“pojke”} \right\rangle \\ \text{SYNSEM } \boxed{2} \left[\text{LKEYS} \mid \text{KEYREL} \mid \text{PRED “\_boy\_n\_rel”}\right] \\ \text{LANGUAGE } \boxed{1} \\ \text{INFLECTION } \textit{c5} \end{array}\right]$$

5.3 Irregular Inflection

Irregular inflections like:

Pos.	Comp.	Sup.
lang	lengre	lengst
lang	længre	længst
lång	längre	längst

are stated in the *irregs.tab* file.

6 Differences Between the Three Languages

Phenomena	Danish	Norwegian	Swedish
Double articulation	-	+	+
Denna-determiners	-	-	+
Postnominal poss.	-	+	-
Stand. prepositional poss.	-	+	-
Quantifiers after genitival poss.	-	-	+
Sin-poss.	-	+	-

Most differences within the noun phrase relate to possessive constructions. Of course, there are other more subtle differences, e.g. quantification over measure nouns seem better in Danish than in Norwegian:

(22) (a) alle halvfems kilo kage blev transporteret væk

(b) ?alle nitti kilo(?ene) kake ble borttransportert

A similar difference (which is not yet implemented in the grammar) is the following difference in grammaticality:

(23) (a) alle ilts kemiske egenskaber blev da beskrevet

(b) ?alle oksygens kjemiske egenskaper ble da beskrevet

The pattern here seems to be that Norwegian does not allow a mass noun to constitute a genitival possessive. More generally, mass objects seem to be dispreferred in this construction. (23b) can easily be blocked by specifying [ARG1|PNG|MASS-NOUN = *count*] on *til*.

7 The Merits of Our Approach

Our major concern here was to combine efficiency and adequacy in a way that seemed preferable for a broad coverage grammar system. The merits of our grammar in this respect include that it contains:

- only one lexical entry per word, i.e. there's no duplication
- only two backbone rules, i.e. the search path of the parser is very restricted
- the constructional selection mechanisms, which is both empirically and theoretically motivated

In addition, the grammar has a certain intuitive appeal because of its topological nature. The grammar is also bidirectional and thus enables generation (see below).

Note on Coverage While our coverage (in terms of constructions) is not exactly *broad*, we still cover a lot of phenomena not covered by existing "broad coverage grammars" for Mainland Scandinavian; e.g. NorSource does not cover any of the possessive constructions covered by our grammar; further it does not cover measure noun/adjective phrases or comparatives.

8 Practical Applications

8.1 Grammar Engineering and Machine Translation

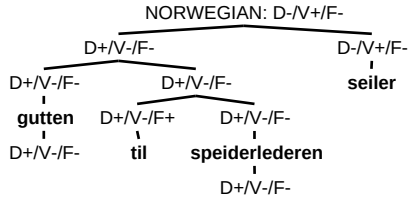
The specification provides the initial tool-kit for constructing broad-coverage grammars for Danish, Norwegian and Swedish. The language-specific grammar should consist of additional types in the language-specific type files. In addition to this furnishing of the process of writing a grammar for one of these languages, the specification also secures compatibility of computational grammars adhering to its general format. For one thing, this will ease the design of cross-Scandinavian machine translation systems.

In the present grammar, we added a LANGUAGE attribute taking the hierarchy of Scandinavian languages as possible values. This feature allows multilingual parsing and generation in a straightforward fashion. Since we use the same predicate names for equivalent relations, and since we generally rely on similar constructions in the different subgrammars, the grammar in a sense includes a small machine translation system: If you parse a sentence, you can get the equivalent sentences in all three languages from generation.

Let's walk through an example, namely the generation in (14a). The first step is parsing the sentence with the standard prepositional possessive, repeated here for convenience:

(24) gutten til speiderlederen seiler

The sentence (as most sentences covered by our grammar) gets only one parse. The syntactic tree is given here:



Generation use the MRSs as the intermediate language, not the Scoped MRSs. An important consequence is that not only the predicate names, but also the features on *individual* constrain the output of the generation process. The MRSs are given in abbreviated form as so-called Indexed MRSs:

```
<h1,e2:SEMSORT:TENSE:ASPECT:MOOD,
{h3:boy_rel(x4:UNIQ-OR-MORE:MASS-OR-COUNT:QUANT-MIN:PRS:SG:COMM:SEMSORT),
h5:poss_by_rel(x4, x6:SEMSORT:QUANT-MIN:MASS-OR-COUNT:UNIQ-OR-MORE:PRS:SG:COMM),
h7:scout-leader_rel(x6),
h3:quant-min(x6, h5, h8),
h7:uniq-or-more(x6),
h7:mass-or-count(x6),
h9:sail_rel(e2),
h9:arg1_rel(e2, x4),
h1:proposition_m_rel(h10),
h11:quant-min(x4, h9, h12),
h3:uniq-or-more(x4),
h3:mass-or-count(x4)},
{}>
```

The grammar generates (25) from (24), since they have exactly the same MRS representation:

(25) speiderlederens gutt seiler

How come such different constructions can produce exactly the same (detailed) semantic representations? This is made possible by the special architecture of our grammar. Since prepositional phrases and genitival possessives are instances of the same phrasal type, *add-noun-phrase*, equivalent and, more importantly, identical semantic forms are produced, if the functional constituents in these phrases are assigned the same predicate names.

8.2 Other Applications

Of course as any other computational grammar, the specification presented here may function as a component in all sorts of implementations, incl. text generation, information retrieval and dialogue systems.

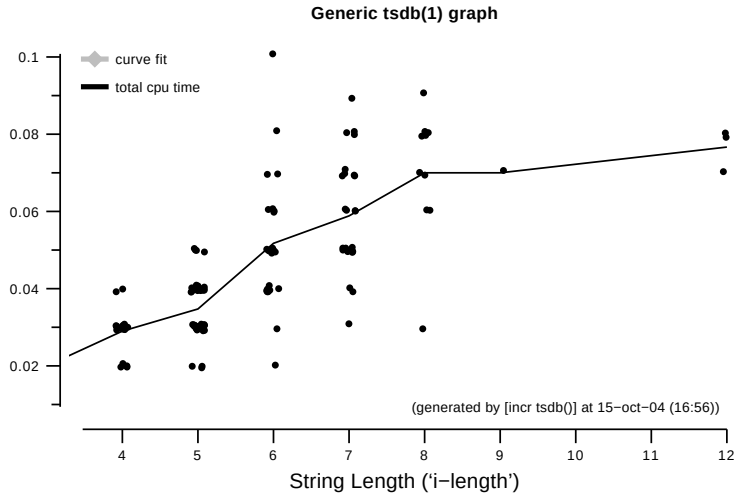
9 Appendices

Our test-suites are given below. The test-suites are covered by the current version of the grammar. This is also evident from the second column of the table that presents the test-suites.

Aggregate	total items #	positive items #	word string Ø	lexical items Ø	parser analyses Ø	total results #	overall coverage %
i–length in [10 .. 15]	3	3	12.00	16.67	1.00	3	100.0
i–length in [5 .. 10]	98	66	6.21	14.02	1.05	66	100.0
i–length in [0 .. 5]	37	14	3.57	6.86	1.21	14	100.0
Total	**138**	**83**	**5.98**	**12.90**	**1.07**	**83**	**100.0**

(generated by [incr tsdb()] at 15–oct–04 (16:55))

Unfortunately, the grammar is still too small for us to do meaningfull evaluation of its efficiency. However, while the performance tests of `tsdb` does not exactly *proove* the efficiency of our grammar, it does not bring bad news either. Here's its performance on our small test-suite:



The test-suites are given in Table 3.

Table 3: Our Test Items.

No.	Item	Readings	Edges
1	en dreng sejler	1	7
2	*en dreng dreng sejler	0	8
3	*fem dreng sejler	0	5
4	fem drenge sejler	1	7
5	*fire fem drenge sejler	0	8
6	*fem høj drenge sejler	0	8
7	fem høje drenge sejler	1	14
8	alle disse mine fem høje drenge sejler	1	37
9	*disse alle mine fem høje drenge sejler	0	30
10	*alle mine disse fem høje drenge sejler	0	25
11	alle disse disse kedelige spejderledere-s drenge sejler	1	39
12	alle fem høje seje drenge sejler	1	27
13	alle fem hurtigste høje drenge sejler	1	29
14	*alle fem høje kedeligste drenge sejler	0	19
15	alle to meter høje drenge sejler	2	52
16	alle fire to meter høje drenge sejler	1	49
17	*alle alle to meter høje drenge sejler	0	54
18	alle halvfems kilo dreng sejler	1	26
19	alle høje og seje drenge sejler	1	33
20	alle piger og drenge sejler	1	15
21	*og drenge sejler	0	6
22	*piger og og drenge sejler	0	9
23	*piger og sejler	0	4
24	alle spejderledere i denne division sejler	1	22
25	*alle i denne division spejderledere sejler	0	17
26	alle denne division-s spejderledere sejler	1	22
27	alle denne division og disse spejderledere s drenge sejler	1	62
28	alle seje og høje drenge og piger sejler	1	44
29	alle seje høje drenge og piger sejler	1	30
30	disse sidste fire drenge sejler	1	19
31	*den sidste fire drenge sejler	0	16
32	den utroligt hurtige dreng sejler	1	22
33	*den hurtige utroligt dreng sejler	0	12
34	de fleste af disse drenge sejler	1	15
35	*de fleste af og disse drenge sejler	0	14
36	begge af disse drenge sejler	1	15
37	fire af fem to meter høje drenge sejler	1	51
38	halvfems kilo to meter høj dreng sejler	1	55
39	alle drenge hurtigere end disse fem drenge hurtigere end disse fire sejler	1	63
40	*alle spejderledere-s drengene sejler	0	14
41	drengen i den division sejler	1	16
42	*drenge end piger sejler	0	9
43	hurtigere drenge end piger sejler	1	14
44	*hurtige drenge end piger sejler	0	15
45	drenge hurtigere end piger sejler	1	14

No.	Item	Readings	Edges
46	*drenge hurtigere sejler	0	6
47	*drenge hurtigere og piger sejler	0	10
48	den hurtigste af de tre drenge sejler	1	29
49	en gutt seiler	1	7
50	*en gutt gutt seiler	0	8
51	*fem gutt seiler	0	5
52	fem gutter seiler	1	7
53	*fire fire gutter seiler	0	8
54	*fem høy gutter seiler	0	8
55	fem høye gutter seiler	1	14
56	alle mine fem høye gutter seiler	1	29
57	*alle mine fem guttene seiler	0	16
58	alle de fem guttene mine seiler	1	27
59	alle disse fem guttene mine seiler	1	26
60	*disse fem gutter seiler	0	10
61	*gutter mine seiler	0	5
62	gutten til jenta seiler	1	10
63	*en gutt til jenta seiler	0	10
64	*disse alle mine fem høye gutter seiler	0	30
65	*alle mine disse fem høye gutter seiler	0	23
66	*alle disse mine høye fem gutter seiler	0	28
67	alle de fem høye tøffe guttene seiler	1	36
68	alle de fem raskeste høye guttene seiler	1	39
69	*alle de fem høye raskeste guttene seiler	0	26
70	alle to meter høye gutter seiler	2	52
71	alle de fire to meter høye guttene seiler	1	55
72	*alle alle de to meter høye guttene seiler	0	57
73	alle de nitti kiloene gutt seiler	1	24
74	alle høye og tøffe gutter seiler	1	33
75	alle jenter og gutter seiler	1	15
76	*og gutter seiler	0	6
77	*jenter og og gutter seiler	0	9
78	*jenter og seiler	0	4
79	alle speiderledere i denne foreningen seiler	1	22
80	*alle i denne foreningen speiderledere seiler	0	17
81	alle denne foreningen-s speiderledere seiler	1	21
82	alle denne foreningen og disse speiderlederne-s gutter seiler	1	50
83	alle tøffe og høye gutter og jenter seiler	1	44
84	alle tøffe høye gutter og jenter seiler	1	30
85	disse siste fire guttene seiler	1	19
86	*den siste fire guttene seiler	0	16
87	den utrolig raske gutten seiler	1	22
88	*den raske utrolig gutten seiler	0	12
89	de fleste av disse guttene seiler	1	15
90	*de fleste av og disse guttene seiler	0	14

No.	Item	Readings	Edges
91	begge av disse guttene seiler	1	15
92	*begge av disse gutter seiler	0	10
93	fire av fem to meter høye gutter seiler	1	51
94	nitti kilo to meter høy gutt seiler	1	56
95	alle gutter høyere enn disse fem seiler	1	26
96	alle gutter høyere enn disse fem guttene raskere enn disse fire seiler	1	63
97	alle speiderledere-s gutter seiler	2	20
98	gutten i den foreningen seiler	1	16
99	en pojke seglar	1	7
100	*en pojke pojke seglar	0	8
101	*fem pojke seglar	0	5
102	fem pojkar seglar	1	7
103	*fyra fyra pojkar seglar	0	8
104	*fem lång pojkar seglar	0	8
105	fem långa pojkar seglar	1	14
106	*dessa alla mina fem långa pojkarna seglar	0	22
107	alla långa pojkar seglar	1	14
108	alla de fem långa pojkarna seglar	1	27
109	alla de fem snabbaste långa pojkarna seglar	1	36
110	*alla de snabbaste pojkar seglar	0	13
111	alla tvåmeter långa pojkar seglar	2	46
112	alla fyra tvåmeter långa pojkar seglar	1	45
113	*alla alla tvåmeter långa pojkar seglar	0	47
114	nitti kilo pojke seglar	1	21
115	alla långa och tuffa pojkar seglar	1	32
116	alla flickor och pojkar seglar	1	14
117	*och pojkar seglar	0	6
118	*flickor och och pojkar seglar	0	9
119	*flickor och seglar	0	4
120	alla scoutledare i denna kår seglar	1	28
121	*alla i denna kår scoutledare seglar	0	18
122	alla denna kår-s scoutledare seglar	1	47
123	alla tuffa och långa pojkar och flickor seglar	1	43
124	alla tuffa långa pojkar och flickor seglar	1	29
125	dessa sista fyra pojkarna seglar	1	19
126	*den sista fyra pojkarna seglar	0	16
127	*den otroligt snabba pojke seglar	0	20
128	den otroligt snabba pojken seglar	1	22
129	*den snabba otroligt pojke seglar	0	12
130	de flesta av dessa pojkarna seglar	1	15
131	de flesta av dessa och dessa pojkarna seglar	1	24
132	båda av dessa pojkarna seglar	1	14
133	fyra av fem tvåmeter långa pojkar seglar	1	49
134	nitti kilo tvåmeter lång pojke seglar	1	54
135	alla pojkar snabbare än dessa fem pojkarna snabbare än dessa fyra seglar	1	60
136	alla scoutledare-s pojkar seglar	3	26
137	pojken i den kår seglar	1	20
138	pojken i denna kår seglar	1	19

10 References

Bouillon, Pierre; Ballac, Christian. 2001. The telic relationship in compounds in French and Turkish. Proceedings of the First International Workshop on Generative Approaches to the Lexicon. Geneva, Switzerland.

Borthen, Kaja; Haugereid, Petter. 2004. A grammar component for referential properties of nominals. Unpublished manuscript.

Delsing, Lars-Olof. 1993. The internal structure of noun phrases in the Scandinavian languages. Doctoral dissertation, University of Lund, Sweden.

Flickinger, Dan. 2000. On building a more efficient grammar by exploiting types. Natural Language Engineering 6: 15-28.

Gundel, Jeanette; Hedberg, N.; Zacharski, R. Cognitive status and the form of referring expressions in discourse. Language 69 (2).

Hankamer, Jorge; Mikkelsen, Line. 2002. A morphological analysis of definite nouns in Danish. Journal of Germanic Linguistics 14: 137-75.

Hellan, Lars; Haugereid, Petter. 2003. NorSource - an exercise in the Matrix grammar building design. Proceedings of the ESSLLI workshop on Ideas and Strategies for Multilingual Grammar Development. Vienna, Austria.

Johnston, Michael; Busa, Federica. 1999. Qualia structure and the compositional interpretation of compounds. In Evelyne Viegas (ed.), Breadth and depth of semantic lexicons. Dordrecht: Kluwer.

Link, Godehard. 1983. The logical analysis of plural and mass terms. In Rainer Bäuerle, Christoph Schwarze and Arnim von Stechow (eds.), Meaning, use, and interpretation of language. Berlin: de Gruyter.

Mikkelsen, Line. 1998. An HPSG analysis of agreement in Danish noun phrases. EUCCS-RP-1998-5, Edinburgh.

Partee, Barbara. 2000. Some remarks on linguistic uses of the notion of "event". In Carol Tenny and James Pustejovsky (eds.), Events as grammatical objects. Stanford: CSLI.

Pustejovsky, James. 1993. Type coercion and lexical selection. In James Pustejovsky (ed.), Semantics and the lexicon. Dordrecht: Kluwer.

Radford, Andrew. 1980. Transformational grammar. Cambridge: Cambridge University Press.

Søgaard, Anders. 2004. The Compound Matrix. Proceedings of HPSG'04. Leuven, Belgium.

Søgaard, Anders. in prep. Do quantifiers really float? Unpublished manuscript.

Underwood, Nancy; Povlsen, Claus; Paggio, Patrizia; Neville, Anne; Pedersen, Bolette; Jørgensen, Lise; Ørsnes, Bjarne; Braasch, Anna. 2000. The LINDA manual, typed feature-based specifications for a core grammar of Danish. Copenhagen: CST Working Papers.