Project Number LRE 61029: Large Scale Grammars for EC Languages
Deliverable E-D6-DK

Design of Syntax Implementation for Danish Phrase Structure and Predicate Argument Structure

Costanza Navarretta

February 5th, 1996

Contents

Foreword

This report contains an overview of the MLAP treatment of Danish syntax and a discussion of the implementability of the suggested formalization. A description of the chosen implementation will be given as part of the documentation of WP9.

The author would like to thank the members of the Danish MLAP group and Brad Music for discussing and correcting the report.

Chapter 1

Introduction

In Povlsen, Jørgensen & Music (1995) priority lists for both core coverage of linguistic phenomena and extensions for the LSGRAM project were given[1]. The priority lists for the linguistic phenomena to be implemented were the following :

- Core coverage:
  1. Active and passive main clauses
  2. NP and PP constructions, also as complements
  3. Complement finite and non-finite subclauses
  4. Coordination
  5. Simple negation
- Extensions:
  1. *der*-constructions ('there')
  2. Modal verbs
  3. Verbal complements
  4. Topicalization
  5. Relative clauses, including participial clauses

In addition to the implementation of the above linguistic phenomena, the core coverage included the implementation of a TLM component and the extensions included treatment of messy details. Both have been implemented in workpackage 5 (Music 1995).

The coverage of the linguistic phenomena described by the Danish MLAP group (LINDA) has in the meanwhile been changed. The new list[2] of linguistic phenomena whose descriptions we have access to is the following:

- Active main clauses
- NP and PP constructions, also as complements
- AP and AdvP constructions
- Complement finite and infinite subclauses

[1] The definition of core coverage was based on the analysis of an area-specific corpus.
[2] The elements in the list are not given any priority here.

- Coordination
- Verbal complements
- Relative clauses (not including participial clauses)

Active main clauses, NP, PP, AP and AdvP constructions, finite and infinitive complement subclauses, agreement and relative clauses are described in Povlsen, Paggio & Underwood (1995). Predicate-argument structure (PAS) has been investigated in Pedersen Jørgensen & Ørsnes (1995).

Determination is treated in Neville & Povlsen (1995) and coordination is investigated in Underwood (1995). These two reports will be included as part of the section on Phrase Structure in the final LINDA manual.

In the following we will first look at the lean approach which the Danish LINDA project has used to formalize linguistic structures (chapter 2).

We will then discuss the description and the formalization of Danish linguistic phenomena in Povlsen, Paggio & Underwood (1995) (chapter 3), Neville & Povlsen (1995) (section 3.4), Underwood (1995) (chapter 4) and Pedersen & Jørgensen (1995) (chapter 5).

Finally we will make a new priority list for the linguistic phenomena to be treated in the next LSGRAM implementation phase (chapter 6).

Chapter 2

The Lean Approach

The Danish LINDA group has followed the so called "lean approach", described in Bennett & Schmidt (1995), as a basis for their formalizations. The concept "lean approach" is used in the LINDA reports to cover both solutions necessitated by implementational matters and formal modifications of the "original" HPSG formalism (e.g. the discussion of the features *substantive* and *functional* and the treatment of relative clauses).

In the lean approach the original HPSG *phrasal sign* has been modified. A new type, *structure*, has been introduced to provide structural information, e.g. projection levels and right and left headedness.

To represent the different projections, X^0, $\overline{X}$,$\overline{\overline{X}}$,X^{max}, the two binary features MAX and MIN are used. They permit switching between projection levels and controlling recursion. The two *structure* features HEADED and HEADING, taking the values *left* and *right*, are introduced in order to control the direction of recursion in the binary rules. Within the same rule they permit switching to another type of binary configuration, a condition for the recursive application of a rule being that the mother node and the head daughter are 'heading' in the same direction. The above two features are used in the whole implementation.

The attribute CONSTR, also of type *structure*, is used to characterize the syntactic function of the daughters in a given phrasal sign, such as *head_subj*, *head_compl*, corresponding to the head-schemata in HPSG[1].

For handling the interleaving of adjuncts and complements in Danish, the lean approach's principle of splitting the COMPS-list into 'head' and 'tail' is followed.

The lean approach gives a different interpretation of the CATEGORY feature than the one provided by Pollard and Sag (1994). In the lean approach the type *head* represents word classes, while functional and substantive are coordinated subtypes at *category* level. This is done because, given the HPSG definition of functionals as words lacking arguments and modifiers, different parts of speech can have both a substantive and a functional use. Substantives have the two subcategorisation features SUBJ and COMPS (see Borsley (1987) and Pollard & Sag (1994) cap. 9) while functionals have the feature SPEC.

In Povlsen, Paggio & Underwood (1995) the following Danish parts of speech which can be used as both substantives and functionals are given:

[1] The *funct_att* schema, which is not a head-schema in HPSG is added to the list.

Part of Speech	Substantive	functional
v	verb	auxiliary
n	noun	expletive *det*, *der*
p	preposition	complementiser *for*
det	determiner	articles

In addition to these, the complementiser "at" might be included, though it is not clear whether the infinitive marker *at* should be treated as a defective auxiliary or a complementiser (see also Pollard & Sag (1994) on the infinitive marker *to* pp. 125-128).

We will consider it a complementiser to avoid the use of an empty marker in infinitive complement clauses.

One of the main issues in the formalization of Danish relative clauses and of PAS has been, in the spirit of the lean approach, to reduce the number of lexical rules.

2.1 Conclusion

Obviously we must in the main lines follow the lean approach, given the Alep formalism is itself lean.

We will implement binary rules, because this is more effective (see the German implementation).

The two features MAX og MIN will not be used for phrasal signs. The saturation principle will be used instead. HEADING and HEADED will only been used when necessary (e.g. for nominal signs for attaching postmodifying structures first). We will implement the attribute CONSTR The distinction between functional vs. substantive parts of speech will only be followed so long it is most efficient implementationally

Chapter 3

Phrase Structure

In this chapter we review the specifications for basic phrase structure (sections 3.1 and 3.2), agreement (section 3.3) and relative clauses (section 3.5) given in Povlsen, Paggio & Underwood (1995). In section 3.4 we discuss the formalization of determination in Neville & Povlsen (1995).

3.1 Clausal Constructions

For describing relatively fixed word order in Danish, the Field Grammar language model developed by Diderichsen (1946) has been followed. The LINDA group suggests to express topological knowledge with binary rules (Povlsen, Paggio & Underwood (1994), p. 24):

> As the main idea is to combine topological knowledge in the structure rules, in order to identify for instance the syntactic subject of a sentence, the serialization constraints will be explicitly stated as head features in the phrasal sign.

A HEAD feature, NEX, is introduced to indicate the word-class order in the Actualisation Field. This feature is used in the report in the following two cases:

- main clauses (NEX has value *vna*)
- subordinate clauses (NEX has value *nav*).

We will discuss the use of the NEX feature in section 3.1.2 on main clauses and in section 3.1.3 on subordinate clauses.

3.1.1 Auxiliaries

It is suggested to attach auxiliaries to main verbs after the attachment of complements and adjuncts to the main verb. Following Bennett and Schmidt (1995) auxiliaries are regarded as functionals.

A *head* feature AUX is introduced, to distinguish between auxiliaries and other verbs. The VMARK feature having the same value as the verbal form should be used as a marking principle after Van Eynde & Schoenmakers (1995). All possible VMARK values must be worked out.

Separate lexical heads must be coded for auxiliary and main verb readings of verbs.

3.1.2 Main clauses

The linear order for Danish main clauses usually corresponds to the following schema (Povlsen, Paggio & Underwood (1995), p. 5):

Base Field	Actualisation Field	Content Field	Extra position
VAR	v(fin) NP(subj) *Adv	*v(nonfin) *Compl *Adv	Clause(inf)

The VAR in the Base Field indicates that different elements can occupy it (there are some restrictions, e.g. the finite verb can never fill the Base Field).

The following three special cases are considered :

- finite main verb in first position
- finite main verb in the Actualisation Field
- interleaving of adjuncts and complements

Finite main verb in first position

In the group of main clauses with a finite verb in first position are included interrogative and imperative clauses (the report refers also to the special case of topicalized finite subclauses in which the subordinated conjunction *hvis* (if) is omitted). In this kind of clauses the base field is considered empty.

To this group can also be added interrogative sentences which are introduced by a *wh-* (*hv-*) interrogative word[1].

The sample formalization of interrogative sentences consists of a phrasal rule in which the subject comes after the main verb. The same formalization is suggested for imperative clauses, in which case the SUBJ-list for the verbal head will be saturated and the value of MOOD in the verbal head will not be *ind*, but *imp*.

Finite main verb in the Actualisation Field

When the main verb fills the first place in the Actualisation Field, the Base Field can be occupied by the subject or by other elements.

In the implementation we can choose to consider the "normal" position of the subject to be in the Base Field (SVO) or after the main verb in the Actualisation Field. In Danish sentences such as *Hunden spiser katten* are ambiguous[2]. Topicalization is discussed briefly, but formal specifications for unbounded dependencies are not available, thus we have to decide how to treat these sentences. Main clauses containing a non-subject element in the Base Field (topicalization) are formalized with a separate rule in which the SUBJ-element comes after the main verb.

The head of main verbs contains the NEX feature *vna* (see also section 3.1).

To the above group of main clauses interrogative clauses beginning with an *wh*-interrogative word could be included.

Interleaving of adjuncts and complements

In some sentences there is a violation of the general linear precedence rules. The most common cases are the following:

[1] Such as *hvornår*, *hvem*, *hvad*.

[2] The two possible meanings of the sentence are *The dog eats the cat.* and *The cat eats the dog.*

- nominal phrases which, when negated, are moved from the Complement Field to the Actualisation Field
- Object-complements realised as non-stressed personal pronouns, such as *mig*, *hende* (*me*, *her*), are moved to the Actualisation field, even though there is a valency bound particle in the sentence

The interleaving of adjuncts and complements is handled with the use of binary rules (see also in chapter 2).

Approaches to the order in which adjuncts appear in the Content Field (e.g. heavy and light elements, place and temporal adjuncts etc.) and the order of adverbs in the Actualization field must be determined during implementation.

3.1.3 Subordinate Clauses

Subordinate clauses are divided into finite and infinitive subclauses.

The schema for subordinate finite clauses is the following (Povlsen, Paggio & Underwood (1995), p. 5):

Conj. field	Actualisation field	Content field	Extra Position
subconj	NP(subj) *Adv v(fin)	*v(nonfin) *Compl *Adv	Clause(inf)

The corresponding schema for infinitive subclauses is as follows (Povlsen, Paggio & Underwood (1995), p. 8):

Conj. field	Actualisation field	Content field
subconj	*Adv v(infin)	*v(nonfin) *Compl *Adv

The NEX feature with value *nav* indicates the identification of a subclause(see also in section 3.1).

It must be decided whether the infinitive marker *at* must be considered a subordinate conjunction as in *at ville tage hjem imorgen* (Povlsen, Paggio & Underwood (1995) p. 8) or not as in *for at kunne spare penge på længere sigt* (p. 9).

In formalizing subclauses the *funct_att* schema is used.

3.1.4 Conclusion

In treating main clauses we propose to distinguish the following two cases:

- The Base Field is empty (the verb is in first position in the clause): the (possibly empty) subject follows the main verb in the Actualisation Field. This covers simple interrogative clauses, imperative clauses (subjectless), comment main clauses[3].
- The Base Field is not empty (the main verb is after the element in the Base Field). This covers all the other cases. The subject can stay in the Base Field, or after the main verb

[3] In old texts it was also common in main clauses which were closely linked to the previous main clause. (see Diderichsen (1957) p. 193)

in the Actualization Field if another element (non-subject) fills the Base Field as effect of topicalization. One can discuss whether the subject is topicalized when it fills the Base Field, or whether the "normal" position of the subject is to the left of the main verb. This would of course influence implementation.

Whether the normal position of the subject should be in the Base Field or in the Actualisation Field could be decided according to the frequency of the two different phenomena in written language, though from an implementational point of view we need not resolve this issue.

The use of the NEX feature is limited to distinguish main clauses (value *vna*) from subordinate clauses (value *nav*). We will extend the use of the NEX feature to distinguish between main clauses with the subject on the first position in the sentence (NEX-value *nva*) and main clauses with the subject after the finite verb (NEX-value *vna*).

Unbounded dependencies and passive clauses are very common in Danish and also within the text corpus chosen, Povlsen, Jørgensen & Music (1995). Thus approaches to these will have to be developed.

Time permitting, we will investigate the possibility of implementing simple cases of topicalization/unbounded dependencies according to the description of relative clauses in Povlsen, Paggio & Underwood (1995) and "simple" passive constructions such as *Æblet spises.*[4] and *Æblet spises af nogen*[5].

Auxiliaries will be treated as substantives, because it is more implementationally efficient in our case. The VMARK feature will thus not been implemented.

Subclauses as complementizers will be handled as in HPSG with a HEAD_MARKER schema and a MARKING feature (see also in section 3.4. Both the complementizer *at* and the infinitive marker *at* are treated as functionals.

3.2 NP, AP, AdvP and PP Constructions

3.2.1 NP Constructions

Post-nominal constituents can function both as complements and adjuncts, thus different rules should be generated to cover post-nominals. In Danish post-nominal complements can be PPs or sentential complements.

Adjuncts can be PPs, ADJPs, participles and attributive relatives. A sample formalization of PPs is given.

Pre-nominal constituents are ordinals, cardinals, determiners, genitive nominal phrases and adjectival phrases. An example of a rule for premodifying adjectives is given.

3.2.2 AP Constructions

Adjectives can occur in attributive or predicative use and they can be modified by preposed adverbs of degree. The non-recursive rule for premodified adjective phrases is given (see also in section 5.3).

[4] The apple is eaten.
[5] The apple is eaten by somebody.

3.2.3 AdvP Constructions

The only adverbs considered are those which modify adjuncts and various subtypes of adverbs (manner and quantifier adverbs). Other adverbial phrases, both those which occur in the Actualisation field and in the Content field are frequent and will have to be formalized, time permitting.

3.2.4 PP Constructions

Prepositions can precede verbs in infinitives, nominals and prepositional phrases. In Danish they can also govern finite verbs and adverbial phrases (see also in Pedersen and Jørgensen & Ørsnes (1995) and section 5.4).

3.2.5 Conclusion

For the formalization of PP,AdvP and AdjP as adjuncts, the Linda specifications are a good starting point. In formalizing PP we will both follow the descriptions in Povlsen, Paggio & Underwood (1995) and in Pedersen and Jørgensen & Ørsnes (1995).

3.3 Agreement

Agreement in Danish is present in the following cases:

- within NPs: agreement of determiners, adjectives and nouns
- within copula constructions: agreement between subject and adjective
- pronoun antecedent agreement

In HPSG agreement is considered both a syntactic and a semantic phenomenon and this interpretation is followed by the LINDA group. The discussion about "grammatical" vs. "gender" languages and about Danish as a language in-between has been taken up. The problem regarding pragmatic restrictions on the sex of the referent referred to, is left open (eg. *jeg mødte barnet, da 'han/hun/det' kom op ad trappen.*).

Agreement for attributive and predicative adjectives are handled as in HPSG.

The Danish MLAP group has decided to make SPEC a non-head feature.

3.3.1 Conclusion

We will formalize SPEC as an head-feature and implement agreement as part of the SPEC value.

3.4 Determination

In Neville & Povlsen (1995) determiners include articles, demonstratives, possessives, quantifiers, cardinals and ordinals. According to their respective position in the NP, determiners are classified into predeterminers, central determiners and postdeterminers.

The overview of Danish determiners is the following (Neville & Povlsen (1995) p. 2):

	predeterminer	central determiner	postdeterminer
article		den, en	
demonstrative		denne (her), den her, den (der)	
possessive		min, din, hans ...	
quantifier	al, begge	nogen, hver, enhver, ingen	lidt, meget, få, mange, adskillige
lx rep cardinal			en, to, tre, ...
ordinal			første, anden, tredje, ...

Singular count nouns are said to subcategorise obligatorily for a central determiner. This is generally true, but there are many exceptions. Singular count nouns do not take a central determiner in the following cases[6]:

- if used as a concept, e.g.: *Har I bil? De fik kylling til frokost. Bruger du deodorant? Jeg har fået mand, hus og hund. Hus med have er hans højeste ønske. Hun skulle gå med hat. Marie løb ud uden frakke. Værket foreligger i manuskript.*
- instruments and past-times: *De hører radio. Hun spiller guitar. Han læser avis.*
- nouns for nationality, profession etc. in predicative position: *Han er student. Marie er dansker.* (but *Han er en dygtig studerende.*)
- after *som*: *som professor, som ung mand, som uerfaren politiker*
- other: *med opslået krave, med opknappet frakke, udleveres kun mod/på recept*

In our opinion many of the above cases can be reduced to the first group. Following the LINDA group we will not formalize the above listed exceptions.

There are two different ways to solve the above cases:

- Singular count nouns do not subcategorize for a central determiner.
- Singular count nouns do subcategorize for a central determiner, unless they are used as concepts in which case they became a mass noun (this could be expressed by a lexical rule or one could express it in the lexicon in the most relevant cases, e.g. in connection with verbs like "have, få, bruge, spise", prepositions like "med, uden", after "som"). All the other cases could be handled in the lexicon as special cases and/or fixed phrases.

In the LINDA formalization pre- and postdeterminers are treated as adjuncts while central determiners are treated as specifiers.

A *category*-feature, DETMARK, is introduced to constrain the combinatorics of Danish determiners (both specifiers and adjuncts). The use of DETMARK is explained as follows (Neville & Povlsen (1995),p. 16):

> In any structure the determiner enters into, it will select a head sister N^p which is marked for a specific boolean expression over various determiner types, and at the same time it will project onto the mother node its own lexically specified value for DETMARK. N^p represents a generalisation over NP and $\overline{N}$. The selection constraints and project values for the various types of determiner are given in the following:[7].

[6] We do not claim the list to be complete.

[7] Attributive, qualifying adjectives are assumed to select *unmarked* $\overline{N}$s and project the value *unmarked* onto their mother $\overline{N}$.

determiner	selects	projects
prequantifier	*~(prequant V cquant V postquant)*	*prequant*
central quantifier	*(ord V unmarked)*	*cquant*
postquantifier	*(ord V unmarked)*	*postquant*
cardinal	*(ord V unmarked)*	*card*
definite	*~(prequant V cquant V definite)*	*definite*
ordinal	*(card V unmarked)*	*ord*

How *hver* should handled when combined with possessives (e.g. *vi fik hver vores hat*) is not discussed, thus we will not cover these clauses.

Genitive nominal phrases, which have a function similar to specifiers, are not mentioned in the report.

In the report on determination, the authors do not address the issue of which determiners are functionals and which are substantives, but in practice central determiners are treated as functionals because they have the SPEC feature, although some of them, e.g. enhver, can be modified. Pre- and postdeterminers are treated as adjectives, which are substantives, although it can be discussed whether a quantifier as *alle* (all) is an adjective. The two head-schemata for handling determiners are the head-specifier schema (central determiners) and the head-adjunct schema (pre- and postdeterminers). They are used as in Pollard & Sag (1995), thus SPEC is a *head* feature.

Whether the interrogative pronoun *hvilken* should be considered a determiner is an open question (e.g. *Hvilken bog har du læst?* (*Which book have you read?*).

3.4.1 Conclusion

In the main lines we will follow the classification of determiners given in Neville & Povlsen (1995), but we will treat *enhver* as a quantifier, and not an article. In the same way we will treat *min, sin* etc. as pronouns and not articles.

We will also follow Neville & Povlsen (1995) in treating SPEC as an *head* feature. Time permitting we will implement genitive phrases as definite determiners following Pollard and Sag (1994) pp. 53 and 54.

The DETMARK feature is implemented as MARKING feature also used to implement complementizers (see also section 3.1.4).

3.5 Relative Clauses

In the lean approach relative clauses are treated quite differently than in HPSG. The main difference between the two theories is that relativers in the lean approach are not treated as a special class, but just as a wh-movement phenomenon. It is reasonable to do this and we can probably use part of this formalization when implementing topicalization.

A new attribute of the *head* type, TYPE (we propose the name RELTYPE), is introduced, with the values *full* or *empty* for handling relative clauses with no relative pronoun introducing them.

An empty relative and an empty trace are formalized. These could give problems in implementation.

Semantic aspects such as *animate* and *inanimate* are introduced, and the CASE *pobj* (prepositional object) is used. The feature CASE with values NOM, DAT, ACC, GEN is used.

Four rules are given for building the top of relative clauses and three for building the middle and the bottom of the relative dependency.

3.5.1 Conclusion

The description of relative clauses appears to be completely formalized, and will be used as is with the exception of type TYPE (to be renamed) and reconsideration of empty traces in implementational terms.

Chapter 4

Coordination

4.1 Coordination of like categories

In this chapter we discuss the specifications for basic coordination of like categories (i.e. elements of the same part of speech) in Danish given in Underwood (1995). By basic coordination is meant "coordination which does not involve gaps due to ellipsis or extractions" (Underwood (1995), p. 3).

The coordinated elements are called "conjuncts" while the coordinating ones are called "conjunctions". Coordinating conjunctions are divided into "conjunctions proper" (*og*, *samt*, *eller*, *men*[1]) which link the conjuncts and can function alone and "preconjunctions" (*både*, *enten*, *hverken*) which introduce coordinate structures and cannot function alone. Coordinate structures are called "binary" when they only contain two conjuncts, "iterative" if they contain more than two.

The coordinating conjunctions are grouped into the following three types:

- conjunctive:*og*, *samt*, *både*
- disjunctive: *eller*, *enten*, *hverken*
- adversative conjunctions: *men*

To the above conjunction are added commas which can replace the non final conjunctions *og* and *eller* in iterative constructions.

The linear precedence of coordinate conjunctions is given in the following two tables (Underwood (1995) p. 7):

Binary coordination

first conjunct	second conjunct	example	English translation
enten	eller	enten A eller B	either A or B
hverken	eller	hverken A eller B	neither A nor B
-	eller	A eller B	A or B
både	og	både A og B	both A and B
-	og	A og B	A and B
-	men	A men B	A but B

[1] It is though added that *men* is sometimes used as iterative conjunction in spoken clauses.

Iterative coordination

first conjunct	intermediate conjuncts	final conjunct	example	English translation
enten	,/eller	eller	enten A, B eller C	either A, B or C
hverken	,/eller	eller	hverken A, B eller C	neither A B nor C
-	,/eller	eller	A, B eller C	A, B or C
både	,/ og	og/samt	både A, B og C	both A, B and C
-	,/ og	og/samt	A, B og C	A, B and C

For formalizing conjunction the category feature CONJ is introduced, having the following structured value (Underwood (1995) p. 13):

$$\begin{bmatrix} \text{FIRSTCONJ} & \textit{boolean} \\ \text{LASTCONJ} & \textit{boolean} \\ \text{CONNECT} & \end{bmatrix}$$

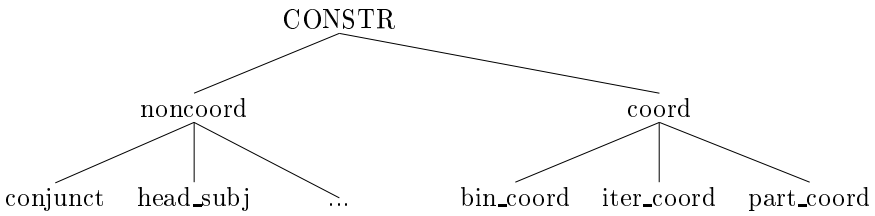
Conjuncts are treated as the union of a conjunction and the element conjoined. A leftmost conjunct which is not introduced by a preconjunction is considered a "nil conjunct". All lexical heads must be considered "nil conjuncts" having the following CONJ value:

$$\begin{bmatrix} \text{FIRSTCONJ} & + \\ \text{LASTCONJ} & - \\ \text{CONNECT} & \end{bmatrix}$$

CONJ being a *category* and not a *head* feature, the following modification of the Marking Principle is introduced (Underwood (1995) p. 13):

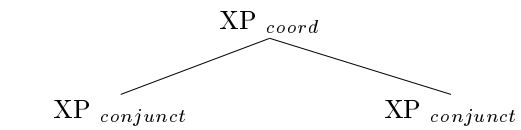
> In a headed phrase, the MARKING value and the CONJ value are shared with the MARKER-DTR if any, and with the HEAD-DTR otherwise.

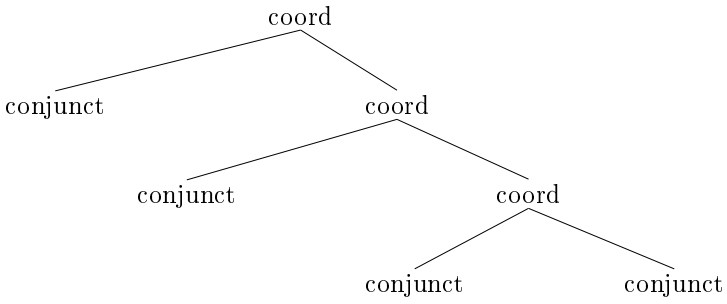
The following new type hierarchy for the attribute CONSTR is proposed (Underwood (1995) p. 19):



The scope of CONSTR is extended to apply to both lexical and phrasal signs. All lexical heads must have the default value [CONSTR noncoord].

To avoid different interpretations of the scope of coordinate structures, without claiming that the given solution is the correct one, binary and iterative coordinate structures have respectively the following structures (p. 19):





In both cases values for HEAD, SUBJ and COMPS must be structure-shared between the mother node and the two immediate daughter conjuncts.

The above solution can be problematic when the element conjoined are not maximal phrases.

The Danish LINDA group treats commas as proper conjunctions which cannot occur as a final conjunct (this is expressed in the rules). It is proposed to distinguish between real "nil" conjuncts which can only occur as the leftmost conjunct in coordinate structures and those built with commas, relying on the availability of text handling procedures which can convert such commas into lexical items.

To handle coordinate nominals having conflicting types for case (genitive), such as "Peter og Maries hus", rules are given in which the only legal construction is one in which the righthand daughter is genitive and the left one is not.

4.2 Conclusion

Identification of 'conjunctive' commas will be difficult to do as part of text handling. This is a minor problem, however, and on the whole the specification will be implemented as given.

Chapter 5

Predicate Argument Structure

In this chapter we discuss the formalization of PAS in Danish by the LINDA group (sections 5.1, 5.2 and 5.3). The LINDA report on PAS, Pedersen, Jørgensen & Ørsnes (1995), also contains a treatment of Danish prepositions in general and of non-valency bound prepositions. In section 5.4 we discuss this treatment.

5.1 PAS for Verbs

The description of PAS for Danish verbs in Pedersen, Jørgensen & Ørsnes (1995) is, in the main lines, taken over from Eurotra. The event-argument is added.

The main groups for verbs are the following:

- zero-valent
- mono-valent: unergatives and unaccusatives
- divalent: strict transitives, divalent prepositional, divalent with weakly bound prepositions, divalent with a finite clause as second argument, divalent with an infinite clause as second argument: control verbs (equi and raising verbs),ditransitive verbs with an additional predicative complement
- trivalent: strict ditransitive, trivalent with a strongly bound preposition, trivalent with one weakly bound preposition, trivalent with two prepositions, trivalent with a finite clause as second argument, trivalent with an infinite clause: control verbs (equi and raising verbs)
- tetravalent

The authors suggest implementing optionality alternations, such as object deleting verbs, by means of different lexical entries, while active-passive alternations and dative shift alternations will be treated by means of lexical rules.

5.2 PAS for Nouns

Also the description and the proposed formalization of PAS for Danish nouns is taken over from Eurotra. We will not implement what in Eurotra was called "specifier-nouns", e.g. *en kop kaffe*, because they will not be described by the Danish MLAP group.

Nouns differ from other classes taking arguments, because nouns in their most prototypical form do not take arguments and because arguments to nouns are in most cases optional.

Nouns are divided into *simple nouns* and *derived nouns*. The latter comprise *deadjectival nouns* and *deverbal nouns* which inherit the PAS from the corresponding adjectives and verbs. But derivation has not been implemented in the Danish morphology.

Deverbal nouns can be divided into *function* and *predicative* nominals, the former describing nouns that have lost their dynamic content and thus do not take arguments, e.g. *bygning*, the latter describing nouns which denote dynamic processes or events, e.g. a *nomen actionis* like *anvendelse*, and not individuals. In-between are the so called *nomen agentis* which refer to entities, but which can take arguments, such as *bruger*.

The nominal sign being basically referential introduces referential indices.

Function nominalisations bear an INDEX feature which is referential. Predicative nominalisations denote events and contain an event-argument in the PAS structure. The INDEX for predicative nominalisations is an event. The RELN value for event-denoting nouns is the noun.

The optionality of complements is assumed to be a syntactic phenomenon, so that nominals will have the same PAS assignment whether the arguments are syntactically realized or not.

5.3 PAS for Adjectives

Also the description and formalization of PAS for adjectives in LINDA is taken over from Eurotra.

Adjectives are classified according to their distribution in the clause, i.e. whether they occur attributively or predicatively and whether they occur as raising or non-raising adjectives. The following classification is given:

- non-raising adjectives: monovalent, divalent, trivalent
- raising adjectives: monovalent, divalent

Following Eurotra, raising adjectives are given a PAS of type *arg2*. The syntactic alternation between extraposition and raising structures is formalized with lexical rules.

Most adjectives can occur in both predicative and attributive position. The LINDA group proposes to encode all adjectives as predicatives, while a lexical rule should provide the attributive use. As for verbs and nouns optionality of arguments can be treated by assigning adjectives different subcategorization types. Raising adjectives can both occur with the anticipatory explicit subject *det* and without it. A lexical rule should take care of cases without the anticipatory *det*.

5.4 Prepositions

In the Eurotra framework subordinate conjunctions were treated as prepositions and multi-word prepositions were considered as one unit. The former approach is not followed by the LINDA group because it is incompatible with the way HPSG treats conjunctions. Multi-word prepositions are considered as one unit also in the present framework.

The following four syntactic environments for prepositional phrases are distinguished:

- Prepositions as heads of weakly bound complements of predicators
- Prepositions as heads of PP adjuncts

- Prepositions as case-markers of objects
- Prepositions as heads of predicative phrases

The MLAP group proposes to have a separate lexical entry for the preposition in each of the above cases.

5.5 Conclusion

The description of PAS and will on the whole be implemented as the authors suggest.

We will not define different lexical entries for each preposition, according to the different syntactic functions of prepositional phrases for reasons of implementational efficiency. Instead,ALEPs refinement phase will be exploited for realizing these distinctions. No specifier nouns will be implemented.

Chapter 6

Revised priority list

Given the available specifications and the conclusions/considerations regarding implementation of them within this document, the following is a priority list of the linguistic phenomena to be implemented:

1. Determination
2. Pre-modified and post-modified NP constructions
3. Active main clauses
4. PP constructions, also as complements
5. Complement finite and non-finite subclauses
6. AP and AdvP constructions
7. Verbal complements
8. PAS
9. Relative clauses
10. Coordination

The following additional phenomena will be implemented time permitting:

1. Passivisation
2. Other PP,AdjP, AdvP constructions
3. Other post-modified NP constructions
4. Topicalization/long-distance dependencies
5. Simple negation
6. Der/det constructions

Bibliography

[1] P. Bennett and P. Schmidt. WP5: Phrase Structure. MLAP-93-15: Deliverable D5a, IAI, 1995.

[2] R. Borsley. Subjects and Complements in HPSG. Technical Report 107-87, CSLI, Stanford. center for the Study of Language and information, 1987.

[3] P. Diderichsen. *Elemænter Dansk Grammatik*. Gyldendal, København, 1946.

[4] F. Van Eynde and A. Schoenmakers. Mood, tense and aspect. Mlap93/15 - reference manual, MLAP, 1995.

[5] B. Music. Danish Morphology in ALEP. LSGRAM - LRE 61029: Deliverable E-D5-DK, Center for Sprogteknologi, Copenhagen, 1995.

[6] A. Neville and C. Povlsen. Specifications for Determination in Danish. LINDA - MLAP93-09 Deliverable 8, CST, Copenhagen, 1995.

[7] B.S. Pedersen and L.D. Jørgensen and B. Ørsnes. Specifications for Predicate-Argument Structure in Danish. LINDA - MLAP93-09: Deliverable 5.1 Final Version, Center for Sprogteknologi, Copenhagen, 1995.

[8] C. Pollard and I.A. Sag. *Head-Driven Phrase Structure Grammar*. CSLI, The University of Chicago Press, Chicago, 1994.

[9] C. Povlsen, L.D. Jørgensen, and B. Music. Danish corpus analysis and priority list. LSGRAM - LRE 61029: Deliverable E-D2-DK, Center for Sprogteknologi, Copenhagen, 1995.

[10] C. Povlsen, P. Paggio, and N.L. Underwood. Specifications for Phrase Structure in Danish. LINDA - MLAP93-09: Deliverable 4.2, Center for Sprogteknologi, 1995.

[11] N.K. Simpkins. ET-6/1 Linguistic Formalism. Technical report, Cray Systems, 1994.

[12] N. Underwood. Specifications for Basic Coordination in Danish. LINDA - MLAP93-09 Deliverable 7, CST, Copenhagen, 1995.