Linguistic Applications of Multimodal and Polymorphic Categorial Grammar

Hans Leiß
(editor)

DYANA-2

Dynamic Interpretation of Natural Language
ESPRIT Basic Research Project 6852

Deliverable R1.3.C
September 1995

Linguistic Applications of Multimodal and Polymorphic Categorial Grammar

Hans Leiß
Universität München
(editor)

authors:
Martin Emms
Esther Kraak
Koen Versmissen

DYANA-2

Dynamic Interpretation of Natural Language
ESPRIT Basic Research Project 6852

Deliverable R1.3.C
September 1995

ESPRIT Basic Research Project 6852 DYANA-2

Dynamic Interpretation of Natural Language

Universiteit van Amsterdam
Institute for Logic, Language and Computation (ILLC)

University of Edinburgh
Centre for Cognitive Science (ECCS)

Universität München
Centrum für Informations- und Sprachforschung (CIS)

Universitetet i Oslo
Department of Linguistics and Philosophy (ILF)

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung (IMS)

Universität Tübingen
Seminar für Sprachwissenschaft (SfS)

Universiteit Utrecht
Research Institute for Language and Speech (OTS)

For copies of reports, updates on project activities and other DYANA-related information, contact:

The DYANA-2 Project Administrator
ILLC/Department of Philosophy
University of Amsterdam
Nieuwe Doelenstraat 15
NL-1012 CP Amsterdam

http://www.fwi.uva.nl/research/illc/dyana/Home.html

Contents

Introduction

This deliverable to Task 1.3, Grammar Specification, contains contributions to the subtasks 'Reasoning about structured resources: empirical case studies in parametric variation' and 'Polymorphic treatments' (of syntactic phenomena). The papers contain applications of the technical extensions of Lambek's categorial grammar developed and investigated earlier in the project, namely the extension of Lambek's propositional grammar logic by modal operators and second order quantifiers. They are part of a recent ongoing effort to describe linguistic phenomena of a range and at a level of detail that goes beyond the capabilities of ordinary Lambek calculus; another part of this effort is the book by Glyn Morrill[2], for example. Theoretical extensions of categorial grammar developed over the last years by researchers inside and related to the Dyana project provided the tools for detailed linguistic description by a new generation of categorial grammars.

In summary, the contributions of this deliverable demonstrate:

1. Clitic adjuncts, though being strongly attached to the head that serves as the host, can be given the semantically adequate phrasal scope by means of interaction postulates relating head adjuction and phrasal composition. The identification of a head by the clitic adjunct can be expressed by a pair of 'dual' modal operators in the types of clitic and head.
2. Linear order restrictions on subexpressions —as far as they are determined by their categories and features— can be stated using modal operators and principles of distribution of (unary) modal operators over (binary) composition operations.
3. Bounded discontinuity —in the sense of merging the lexical items of a functor's argument expressions into a linear order which is incompatible with an ordering of these arguments— can be handled by explicitly allowing an argument to be discontinuous using modalized argument types in the functor's type.
4. An argument in the HPSG-literature about subcategorization of auxiliaries can be shown to be irrelevant in the context of second order Lambek calculus.
5. Scope ambiguities involving quantifiers, auxiliaries and negation can be handled in a second order Lambek calculus in various ways.

The contributions on clitics and word order both examplify the use of multimodal categorial grammar logic, whose theoretical properties have been worked out mainly in the previous years of Dyana-2 (see deliverable R1.1.B). The applications presented now are framed in multimodal systems with different *binary* composition operations —ordinary phrasal composition, head adjunction, infixation— and *unary* operations on categories, the modal operators. It is demonstrated how the specific properties of each operation as well as the relations between these operations interact with the properties of lexical items in a logical inference of quite subtle grammatical judgements.

The contribution on auxiliaries and negation uses second order quantifiers, i.e. quantifiers ranging over categories, to express polymorphism in the subcategorization of auxiliaries and in the types of quantifiers and negation.

A particular instance of word order constraint, the auxiliary ordering constraint, is shown to be a *consequence* of the otherwise motivated lexical type assignments. For me, this raises a question: in which cases is an explicit addition of ordering restrictions via modal operators necessary, and in which would it be just a convenient way of representing an ordering, and possibly stand in contradiction to the implicitly given one?

Concerning the description of linguistic phenomena, the first and third paper give derivations of grammatical examples, but also present reasons why some ungrammatical examples are underivable using the lexical type assignments chosen. In my opinion, the further development of categorial grammar needs some metatheorems that provide help in showing underivability of certain sequents. Particularly useful were results saying that sequents of a given form remain underivable under any extension of the lexical type assumptions, so that grammars designed to entail or not entail certain judgements could be combined or extended without loosing the desired properties.

Below follows a description of the content of the contributions.

French object clitics

The handling of head adjunction and the treatment of linear order restrictions are combined in Esther Kraak's contribution on French object clitics.

Basically, the attachment of an object clitic to a verb requires a different, 'stronger' mode of adjunction than that of an ordinary object. Clitics and verbs are combined by head-adjunction, and the allowed cooccurrence pattern of different clitics gives a partial order in terms of NP-features.

Adjuncts A generally are combined with a compound phrase by infixation $\bullet_f$ which recursively combines the adjunct with the head component H of the phrase, not the remaining constituents C. Hence, infixation $\bullet_f$ and ordinary phrasal composition $\bullet_p$ are related by interaction principles of 'mixed' associativity and commutativity

$$(M)\;\; A \bullet_f (H \bullet_p C) \longleftrightarrow (A \bullet_f H) \bullet_p C, \qquad A \bullet_f (C \bullet_p H) \longleftrightarrow C \bullet_p (A \bullet_f H),$$

These allow an adjunct to 'move' into its logical scope and attach itself to a lexical head.

The attachment to a lexical head needs more fine-tuning of the system at the lexical type assignments. For the clitics this is done roughly as follows. Clitics are distinguished from ordinary objects NP's in that the verb is considered the argument of the clitic, not vice versa, so that the lexical type of a clitic is a refinement of VP/(VP/NP). Actually, two different function type constructors $/_f$ and $/_p$ have to be used, related to the two composition operations $\bullet_f$ and $\bullet_p$ by residuation. Clitics may be attached to verbs in particular forms only — finite or infinite, but not past participle— and differ corresponding to features of the NP-object they stand for —person, case, gender. Hence the type of an object clitic is of the form

$$\Box'(\Box\mathrm{VP}/_f(\Box\mathrm{VP}/_p\Box'\mathrm{NP})),$$

with a modal operator $\Box$ representing the features of the verb and $\Box'$ for those of its missing noun phrase.

The combination of a clitic with a verb is neither infixation nor ordinary phrasal combination, but head-adjuction $\bullet_h$, both of whose arguments are heads. Types of heads generally have the form $\Box H$ for some type H, and the grammaticality of a combination $A \bullet_h B$ needs that A and B are of this form, i.e. $(A \bullet_h B) = (\Box' A' \bullet_h \Box'' B')$.

Sentencehood of a structured combination of lexical items —a well-bracketed product of items using head-adjunction and phrasal combination as product operations— is checked by deriving the goal category $\Box S$ from the structured combination C of the items's lexical types. Intuitively, a goal $\Box S$ could be paraphrased as 'derive a datum of type S and check the well-formedness of the head-configuration of that S'. The basic law of residuation tells us that we can derive $\Box S$ from C if we can derive S from $\Diamond C$. The modal operator $\Diamond$ is percolated down to the types of the lexical items in C, using the distributivity principles

$$\Diamond(A \bullet_p B) \to \Diamond A \bullet_p B, \qquad \Diamond(A \bullet_p B) \to A \bullet_p \Diamond B$$

to 'look for' a head-adjunction combination in either factor of a phrasal combination. On a head-adjunction combination, $\Diamond$ splits in two, according to

$$(K) \qquad \Diamond(A \bullet_h B) \to \Diamond' A \bullet_f \Diamond'' B, \qquad \text{if } \Box \prec \Box' \preceq \Box'',$$

provided a predefined partial order on the modal operators is respected. Principle (K) is used for two purposes.

First, the modal operators inherited to the components check the properness of the head-adjunct combination: only if this combination $(A \bullet_h B) = (\Box' A' \bullet_h \Box'' B')$ comes with the proper modal operators, one can use the reduction

$$\Diamond \Box X \quad \to \quad X$$

—a consequence of the residuation principle for the unary operators— to get access to the internal structures A' and B' of the head-types,

$$\Diamond(\Box' A' \bullet_h \Box'' B') \quad \to \quad \Diamond' \Box' A' \bullet_f \Diamond'' \Box'' B' \quad \to \quad A' \bullet_f B',$$

and continue reasoning with the functor-argument structure of $A' \bullet_f B'$.

Second, (K) is used to express the linear ordering restriction one finds in the cooccurrence of multiple french object clitics: multiple clitics can only occur if they respect a particular partial order $\preceq$ determined (largely) by their number and case features, so this can be taken as the partial order of the corresponding modals in (K). To handle the clitic+verb combinations, verbs get a lexical type with a $\preceq$-maximal dummy modal operator which forces the verb to be right-peripheral.

The paper lists the constructions in which clitics can occur or cannot, and shows for each how to set up the lexical types of clitics, verbs and auxiliaries so that the grammaticality of the constructions are provable resp. unprovable. Besides aspects of order, this includes islands to cliticization, clitic extraction from medial positions, attachment of clitics to auxiliaries or infinitives, and clitic climbing verbs. The apparatus of multimodal categorial grammar allows a fairly elegant treatment of all these.

Bounded discontinuity and word order domains

The contribution by Koen Versmissen on 'Word order domains in categorial grammar' picks up Mike Reape's theory of word order domains developed in Dyana-1 and shows how Reape's insights can be transported from the framework of Head Phrase Structure Grammar to multimodal Categorial Grammar.

Reape's theory derives word order not from the sequence of leaves of ordered phrase structure trees, but from ordered sequences of constituents, the *word order domains* associated to nodes of the functor-argument structure of phrases. In this approach, a non-functor child can be embedded into its parent's word order domain either as an element or by the operation of *domain union*, i.e. by an order-preserving merge of the elements of its word order domain with those of the parent's. This allows discontinous occurrence of the child's constituents within the bounds of the parent constituent, as well as partially free word order via domains that are not totally ordered.

Versmissen's coding of Reape's theory in categorial grammar terms is by means of modal operators on two levels: one to account for the structure of domains, and another to account for the linear ordering of a domain. Ordering constraints are described by systematic linking of a noncommutative and an associative and commutative combination.

Structure If the combination $\bullet$ is associative and commutative, an expression of type $A \bullet B$ can be obtained by merging the A and B material, i.e. domain union is the default operation. To enfoce the B expression to be an element of the word order domain of the $A \bullet B$, one has to bracket the B material. A modal version $\Diamond B$ can be used for the type of B-material that cannot be domain unioned, and $C/\Diamond B$ for a functor that takes bracketed B-material into something of type C. Type $\Box B$ corresponds to material 'missing' some structure to become of type B. Intuitively, heads are of type $\Box A$, and project domains $\Diamond\Box A$.

Modalities are labelled (by labels related to, but not identical to the types A) to distinguish between domains of different kinds. The labels occur in the linear precedence (LP) constraints used to specify the domain orderings, and so the constraints give a partial ordering $\preceq$ of the labels L resp. the modalities $\Diamond_L, \Box_L$.

Linear ordering The type-logical formulation of LP constraints uses a form of principle (K) above, but at a second level of modalities. Namely, for each *set* S of labels L, modal operators $\Diamond^S, \Box^S$ are introduced to test the word order constraints. To prove sentencehood of an ordered combination C, one (essentially) proves $\Diamond^{\mathcal{L}} C \rightarrow \Box_S S$, where $\mathcal{L}$ is the set of all labels. The checking of LP constraints is done by pushing modals $\Diamond^S$ down the structure of C using the distribution principle

$$(K') \qquad \Diamond^S(A \bullet B) \quad \rightarrow \quad \Diamond^T A \otimes \Diamond^U B, \qquad \text{if } T \lhd U, \quad T, U \subseteq S,$$

where $T \lhd U$ means that putting $\{t \prec u \mid t \in T, u \in U\}$ is consistent with the LP constraints on $\prec$, until one meets labelled modals $\Diamond_L$. Note that (K') says that an *ordered* combination $A \bullet B$ with labels from S is an admissible ordering of an *unordered* combination $\Diamond^T A \otimes \Diamond^U B$ in which A takes its labels from $T \subseteq S$ and B takes its labels from $U \subseteq S$, *if* the LP constraints are compatible with putting A before B.

On meeting a labelled modal operator $\Diamond_L$, it is simply checked whether (a domain with) this label is allowed (as an element of the current domain), using

$$\Diamond^S \Diamond_L A \to \Diamond_L A, \qquad \text{if } L \in S.$$

In particular, if this happens on meeting two word order domains (of the proper kinds $L \in T, M \in U$), the functor-argument structure of the embedded types can be accessed to continue the derivation:

$$\Diamond^T(\Diamond_L \Box_L A') \otimes \Diamond^U(\Diamond_M \Box_M B') \quad \to \quad \Diamond_L \Box_L A' \otimes \Diamond_M \Box_M B', \quad \to \quad A' \otimes B'.$$

Actually, the coding of the domains is $\Diamond_L \Box^{\mathcal{L}} \Box_L A$ rather than $\Diamond_L \Box_L A$, where the intermediate $\Box^{\mathcal{L}}$ is needed to start a checking of LP constraints on the L- and M-domain.

The separation between reasoning about word order and reasoning about functor argument structure seems very clear in the system proposed in Versmissen's contribution. A price in the form of additional complexity in the lexical types and in the derivations has to be paid, and although it looks cheap, it would be good to have a careful calculation of the costs. For example, how well can bounded discontinuity be parsed with a multimodal categorial grammar using Versmissen's translation, compared to parsing with the classical CFG-parsers as modified by M.Reape[4]?

Versmissen's translation is given in somewhat intuititve terms; one would like to see a precise version of Reape's theory formulated in HPSG and then a theorem stating the correctness of the translation to multimodal categorial grammar. This would be a nice piece of formal comparison between HPSG and CG extensions.

Treating Auxiliaries and Negation in 2nd Order Lambek Calculus

The contribution by Martin Emms on "Treating Auxiliaries and Negation in a Second Order Lambek Calculus" first shows that an ordering constraint commonly imposed —modal < perfect < progressive < passive— can actually be derived in traditional Lambek calculus, and then applies type quantifiers to handle auxiliaries and negation.

To derive the auxiliary ordering constraint in Lambek grammar, a list of necessary assumptions is given; these for example specify the form of the verb subcategorized by the auxiliary, or the possible forms of the aux+verb combinations. While almost all violations of the ordering constraints already follow from the types of the auxiliaries with these verb form dependencies built in, one remaining violation can be excluded only by introducing a particular verb feature and omitting a corresponding type of the auxiliaries.

A second question on auxiliaries discussed in this contribution is whether they subcategorize for verbs or verb phrases, a subject of recent debate in the

HPSG community (see Netter e.a.[3]). If one takes the position that auxiliaries categorize for verbs, and the verb's subcategorization list is inherited to the auxiliary (resp. the aux+verb combination), then an auxiliary needs different lexical entries for verbs with different subcategorization lists.

While this kind of polymorphism is proposed in several HPSG analyses, it can most naturally be expressed in a second order Lambek calculus: since product types are definable, a universal quantifier can express (argument inheritence and) the schematic type of auxiliaries; for the type of the passive auxiliary being, this would be

$$\forall X.\,(\mathrm{VP}[+ing]/X)/(\mathrm{VP}[+p.part]/X).$$

However, in the second order Lambek calculus the difference between monomorphic VP complements for auxiliaries and polymorphic V-complement is a non-issue: the first implies the second, and with a slight additional assumption, the converse holds as well. This reveals that the difference is due to the combinatorial restrictions of the schemata for phrasal signs in HPSG and vanishes if combined with the hypothetical reasoning of Lambek's calculus.

Finally, there is a short discussion of scope ambiguities arising with auxiliaries and negation, with proposals for further use of second order quantifiers.

Emms' contribution gives some applications of second order properties in grammatical reasoning. In an additional contribution to Dyana-2 (see [1]), he showed that derivability of sequents in the second order Lambek calculus is generally undecidable. Since it seems unlikely that natural language grammars pose undecidable questions, a fragment of second order Lambek calculus is asked for that is decidable and sufficient for application to grammatical reasoning, but still provides some of the power of second order logic.

References

[1] M. Emms. The undecidability of a second order lambek calculus. DYANA-2, deliverable R1.1.C, September 1995.

[2] G. Morrill. *Type Logical Grammar.* Kluwer, 1994.

[3] J. Nerbonne, K. Netter, and C. Pollard, editors. *German in Head-Driven Phrase Structure Grammar*, volume 46 of *CSLI Lecture Notes.* Center for the Study of Language and Information, Stanford, CA, 1994.

[4] M. Reape. Word order in germanic and parsing. ESPRIT Basic Research Action 3175, DYANA, Deliverable R1.1.C, August 1991.

München, September 1995
Hans Leiß

Task 1.3, subtask 3

Polymorphic Treatments

Treating Modals and Negation in Second Order Lambek Calculus

Martin Emms

(Universität München)

Task 1.3, subtask 2

Parametric Variation in Categorial Perspective

French Object Clitics: A Multimodal Analysis

Esther Kraak
(Universiteit Utrecht)

Task 1.3, subtask 2

Parametric Variation in Categorial Perspective

Word Order Domains in Categorial Grammar

Koen Versmissen
(Universiteit Utrecht)