Compositional dialogue referents in phrase structure grammar

Dols, F.J.H.

Publication date:
1992

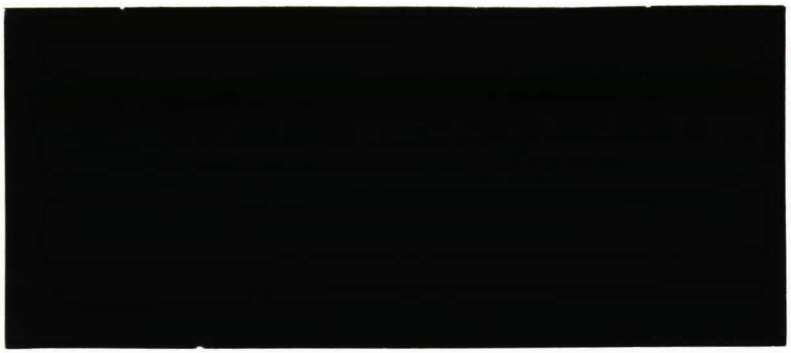
Document Version
Publisher's PDF, also known as Version of record

[redacted]

ITK

RESEARCH
REPORT

NSTITUTE FOR LANGUAGE TECHNOLOGY AND ARTIFICIAL INTELLIGENCE INSTITUUT VOOR TAAL- EN KENNISTECHNOLOGIE



ITK Research Report
march 1992

Compositional Dialogue Referents in Phrase Structure Grammar

Frens J.H. Dols

No. 21

This paper is a revised version of the one presented at The Second International Pragmatics Conference, Szczyrk, December 5-8, 1989, Poland. It will appear in Dols (1992).

ISSN 0924-7807

1 Introduction

Although semantics, implicatures and dialogue referents are established areas of natural language research, the notion of 'grammar' is still often restricted to the syntactic aspect, even in the context of natural language processing (NLP). An example is Zlatev et al. (1989), who discuss the merits of several frameworks for linguistic analysis. They give four meanings for 'grammar': (a) a theory, (b) a formalism, (c) a set of syntactic rules and (d) an analysis of a certain language. They write:

> ... the goal of NLP is that of creating an **NLP system**, i.e. a computational environment for representing and processing linguistic knowledge *and* a description of this knowledge for one or more languages. *(Zlatev, Eriksson, Kälgren, 1989, 2)*

They do not explicate the meaning of 'linguistic knowledge', but the subsequent discussion of frameworks (like GB, GPSG, TAG, LG, CG, DT, and PP) shows that they mean 'syntactic knowledge'. They do not mention the assertive or implicational semantic aspect in their comparison of these frameworks, nor do they mention aspects that pertain to the linguistic context, like those related to dialogue referents.

The role of dialogue referents in natural language has been investigated extensively in the frame-work of DRT (Kamp, 1982), in which individual variables ('discourse markers') represent introduced referents, which are subsequently used in formulae describing their properties.

In this paper I propose to extend phrase structure grammars with a component for the generation of formal descriptions of referents that are introduced by an utterance. These descriptions of dialogue referents derive from a compositional process parallel to the semantic analysis.

This paper aims at showing the feasibility of the proposed method by exploring simple examples introducing one or more referents at the

sentence level. I show how the rules generating the referents incorporate and extend the semantic representations produced by the phrase structure analysis. I will give examples involving referents introduced by the use of definite and indefinite phrases, and show also that the analysis easily extends to sentences with several definite or indefinite descriptions, which introduce more than one referent.

This work is closely related to work reported elsewhere (Dols 1989; Dols 1990). The general tenor of these reports is that the notion of 'grammar' should include (in addition to the syntactic component) not only a formal semantic part for representing the assertive content, but also a component for the representation of the conventionally implied content, and one for the representation of dialogue referent descriptions.

2 General background

This research was carried out in the context of the development of the Tendum dialogue system (Bunt et al. 1984). The grammar module in version 2.5 consists of a chart parser for an augmented phrase structure grammar. The grammar formalism includes a syntactic and a semantic component cooperating in the phrase structure analysis of sentences for e.g. quantified mass terms and discontinuous constituents (Bunt 1985; Bunt et al. 1987). In Dols (1990) a proposal was formulated to include a component for the representation of meaning aspects that are not part of the assertive content. Examples of such semantic aspects are postdeterminers and non-restrictive modifications. The grammar rules in this paper include in addition a component for generating dialogue referent descriptions on the basis of representations for the constituent phrases. The output of a parser for this extended grammar formalism therefore consists of one or more units consisting of the following four parts:

1. a bundle of feature-value pairs indicating pragmatic feature information derived from the sentence. In the Tendum system this aspect is called a "surface speech act" representation.

2. an expression of a higher-order representation language (extended lambda-calculus, Bunt 1985) that represents the assertive content of the sentence;

3. a formal representation of those semantic aspects that do not belong to the assertive content.

4. one or more formal representations denoting referents introduced by the sentence

In the next section I discuss the technique of generating the elements of the fourth part, the dialogue referent part.

3 Definite descriptions of dialogue referents

It has often been suggested that a dialogue system should accumulate information concerning discourse objects (Karttunen 1976; Kamp 1981; Heim 1982). This information may then be accessed for resolving anaphoric references and for reasoning about focus and topic. Heim cites Karttunen in this respect:

> " (...) one particular feature a text interpreter must have: that it must be able to recognize when a novel individual is mentioned in the input text and to store it along with its characterization for future reference." *(Karttunen 1976, 364, in: Heim 1982, 281)*

In the context of building a natural language processing system the formal descriptions for dialogue referents should be considered as building stones for the construction and maintenance of a model of the discourse, to be used for reasoning about topic and focus, resolving anaphora and ellipsis.

Definite descriptions seem appropriate for describing objects that may

occur as dialogue referents. In Dols (1989) a (bounded) uniqueness operator '!' defined in the context of the extended lambda-calculus mentioned in the previous section, was introduced to construct such definite descriptions. Using this operator, the semantic representation of the sentence 'The plane is due.' consists of an application of the predicate 'DUE' to the definite description !(PLANES, Cr) representing 'The plane':

DUE(!(PLANES, Cr))

The uniqueness operator '!' takes as its arguments a set (here: the set of planes) and a predicate (here: a function Cr, that determines the contextually relevant objects). It denotes the one object from the set of planes that is contextually relevant. The representation of dialogue referents presented in the following sections is in terms of such definite descriptions. The uniqueness operator enables us to generate complex definite descriptions for dialogue referents during the phrase structure analysis, paralleling the generation of semantic representations. The generation of these descriptions is defined in much the same way the semantic representation is generated. For certain phrases a representation is introduced in the dialogue referent part, in which the representations in the dialogue referent parts and the assertive content parts of the constituents may both play a role in the construction of the dialogue referent representations of the current phrase. The set of descriptions generated in this way represents the set of most recently introduced dialogue referents. In order to be able to identify in a simple way the introduced referents, unique keys for dialogue referents are also introduced (playing the same role as the individual variables for discourse markers in DRT).

4 Indefinite noun phrases and dialogue referents

An indefinite noun phrase like 'a plane' is often analysed as being ambiguous between a referential and a non-referential reading (Strawson 1950; Chastain 1975; Heim 1982). *Referring* indefinites can be represented like definite descriptions by means of the uniqueness operator, while *non-referring* indefinites are represented in terms of an existential quantification.

The following example shows that non-referring noun phrases may play a central role in the generation of dialogue referents. Imagine a dialogue situation, created by the following exchange:

A: A plane from Montreal is due.
B: Will it arrive at 12:00?

In this example, the indefinite noun phrase 'a plane' is not referential: it is not intended to refer to a particular plane. However, t is correct to continue the dialogue with the second sentence in which 'it' *does* refer to a particular plane. The point is that a referent was introduced by the first sentence, namely the plane that is due (according to the speaker). The generation of a definite description representing a dialogue referent means that *from that point in the dialogue* a specific referent is available.

It is interesting to see that for indefinite noun phrases the representation of the dialogue referents is exactly the same as the *semantic* representation of definite wh-descriptions. This is because the dialogue referents introduced by indefinite phrases like 'a plane' have wide scope, just like wh-phases. A sentence like 'A dog barks at the cat.' introduces a unique referent identified by the description 'The dog that barks at the cat.'. This is the same as the semantic representation for 'Which dog barks at the cat.', at least if we take the meaning of this wh-question to be a representation of the *way* the answer is to be computed.

De concrete details for the sentence 'A plane from Montreal is due.'

with respect to the dialogue referent part are as follows. Each grammar rule includes a semantic part SEM and a dialogue referent part REF, that builds descriptions of referents in terms of the SEM and REF structures of its constituents. Consider the next example. The noun phrase (NP) 'A plane from Montreal' consists of a determiner and a noun phrase consisting of a noun plus a (prepositional phrase as a) restrictive postmodifier. The resulting representations for the assertive content (SEM) and the dialogue referent (REF) of this NP are:

SEM NP: λ **P:** $\exists$**(SELECT(Planes,** λ **x: FROM(x, Montreal)),** λ **x: P(x))**
REF NP: #1, λ **P:** $!$**(SELECT(Planes,** λ **x: FROM(x, Montreal)),** λ **x: P(x))**

When combined with the representation of the verb phrase, the resulting representation for the sentence (S) 'A plane from Montreal is due.' after lambda-conversion is:

SEM S: $\exists$**(SELECT(Planes,** λ **x: FROM(x, Montreal)),** λ **x: Due(x)).**
REF S: #1, $!$**(SELECT(Planes,** λ **x: FROM(x, Montreal)),** λ **x: Due(x))**

In this example the analysis of the indefinite description 'a plane from Montreal' triggered the generation of a definite description 'the plane from Montreal that is due' for dialogue referent #1.

Suppose a subsequent utterance refers to this plane by an personal pronoun 'it'. The representation of this utterance should contain an (potentially ambiguous) anaphoric constant representing 'it', and the mechanism responsible for the resolution of intersentential anaphoric relations should be able to access the introduced referents. In this way it is not only possible to relate 'it' to the key '#1', but also to the formal representation 'the plane from Montreal that is due' that *describes* it in terms extracted directly from the introducing utterances.

The generation of dialogue referents during the phrase structure

analysis is not always straightforward, as the following example shows. The sentence 'I have a dog and a cat.' introduces two definite descriptions, one triggered by 'a dog' and one by 'a cat'. The referent component of the rule combining these two phrases will generate the descriptions for both introduced referents. However, in the negated sentence 'I do not have a dog and a cat' the rule that constructs the negation must cancel these descriptions, as this sentence does not introduce such referents. Consequently, the extended grammar rules must also include 'projection' conditions that restrict and regulate the process of generating referent descriptions, but this in not within the scope of this paper.

5 Noun phrase sequences and dialogue referents

In the following, I make use of so-called noun-phrase sequences (NPS; Scha 1981, pp. 147-148; Bunt 1985, pp. 148-149, 183, 195; Bunt et al. 1987; Dols 1989). The NPS structure represents the sequence of all arguments of a verb, which is useful in generating scope variants. Semantically, the arguments of a verb phrase have in common that they take part in the same relationship, specified by the semantics of the verb phrase. Unusual as this structure may be from a syntactic point of view, they are very useful and motivated by acceptable semantic reasons.

An example involving a definite and an indefinite noun phrase is given below, and one involving two indefinite descriptions. We may expect that a dialogue referent which is introduced by a multiple noun-phrases sentence, involves semantic aspects from more than one noun-phrase. The multi-noun phrase sentence 'A dog barks at the cat.' is represented by:

SEM: $\exists$ (Dog, λ x: Bark(x, !(Cat, CR)))

This sentence introduces a dialogue referent 'The dog that barks at the cat.', for which a suitable representation must be generated on the basis of the constituents of the sentence. The representations for the first noun phrase constituent 'a dog' are:

SEM NP1: λ **P:** $\exists$ **(SELECT(Dog, Cr),** λ **x: P(x))**
REF NP1: #1, λ **P: !(Dog,** λ **x: P(x))**

and the representations for the second noun phrase 'the cat' are:

SEM NP2: λ **P: P(!(Cat, CR))**
REF NP2: λ **P: !(Cat,** λ **x: P(x))**

(Note, that no key is introduced for 'REF NP2', as no new referent is involved.) The representations in the rule forming the noun phrase sequence for these noun phrases 'the dog' and 'the cat' are:

SEM NPS: λ **R: SEM NP1(** λ **x1: SEM NP2 (**λ **x2:**
R(x1, x2)))
REF NPS: #1, λ **R: REF NP1 (** λ **x1: SEM NP2 (**λ **x2:**
R(x1, x2)))

Note, that the dialogue referent part 'REF NPS' of the rule forming noun phrase sequences incorporates 'REF NP1' and 'SEM NP2'. This means that from the first noun phrase the structure of the dialogue referent is used, and from the second noun phrase the semantic structure. This is natural, as the dialogue referents introduced by a sentence are related to what the sentence expresses, which is represented by the semantic structures pertaining to the phrases. A similar involvement of semantic structures in non-standard new structures can be observed in proposals for generating conventional implicatures (compare the third grammar component mentioned in section 2.) in addition to the assertive content (see also Karttunen and Peters 1979; Dols 1990).

The development of the lambda conversions for the NPS representations are (first substituting SEM NP1):

SEM NPS: λ **R:** λ **P:** ∃ (**SELECT(Dog, Cr),** λ **x: P(x))**
[(λ **x1: SEM NP2 (**λ **x2: R(x1, x2)))]**

SEM NPS: λ **R:** ∃ (**SELECT(Dog, Cr),** λ **x: (** λ **x1:**
SEM NP2 (λ **x2: R(x1, x2)))** [(x)])

SEM NPS: λ **R:** ∃ (**SELECT(Dog, Cr),** λ **x: SEM NP2 (**
λ **x2: R(x, x2))))**

A similar development for SEM NP2 in this scheme gives finally:

SEM NPS: λ **R:** ∃ (**SELECT(Dog, Cr),** λ **x: R(x,**
!(Cat, Cr)))

If we develop REF NPS by substituting first REF NP1, followed by substituting SEM NP2, we obtain after lambda conversion:

REF NPS: #1, λ **R: !(Dog,** λ **x: R(x, !(Cat, Cr)))**

Finally, the representations for the verb phrase rule and for the rule forming a sentence are simply:

SEM VP: **BARK**
REF VP:

and

SEM S: **SEM NPS (SEM VP)**
REF S: #1, **REF NPS (SEM VP)**

If we apply the rule forming a sentence, we obtain finally:

SEM S: ∃ (**Dog,** λ **x: Bark(x,** !(**Cat, CR)**))
REF S: #1, !(**Dog,** λ **x: Bark(x,** !(**Cat, Cr**)))

The structure SEM S is a correct representation for the sentence 'a dog barks at the cat.' and REF S is a correct representation of the dialogue referent introduced by this sentence: 'The dog that barks the cat.'.

In the next example two indefinite noun phrases introduce two dialogue referents that refer to each other.

The semantic representation for the sentence 'A dog barks at a cat.' is:

SEM S: ∃ (**Dog,** λ **x1:** ∃ (**Cat,** λ **x2: Bark(x1, x2)**))

The dialogue referent part of the noun phrase sequence contains the following structure (the representation of the verb ('Bark') has already been incorporated by reduction):

REF NPS: **REF** NP_i (λ **x1: SEM** NP_j (λ **x2: Bark(x1, x2)**))

The dialogue referents structures pertaining to the noun phrase constituents are:

REF NP1: #1, (λ **P:** !(**Dog,** λ **x: P(x)**)
REF NP2: #2, λ **P:** !(**Cat,** λ **x: P(x)**)

As both noun phrase constituents are supposed to introduce a referent, the dialogue referent part of the NPS structure must be applied once for each constituent. This complicates the application of the NPS rule in that once for each NP its REF part must be substituted for REF NP_i and its SEM part for SEM NP_j.
Accordingly, the first time 'REF NP1' and 'SEM NP2' are substi-

tuted in REF NPS, the second time 'REF NP2' and 'SEM NP1'. This is accomplished by permuting all relevant noun phrase representations that are involved in the representation of the dialogue referents *and* permuting the arguments of the abstraction variable of the noun phrase sequence under construction. This may be implemented by a procedure in the parser that is triggered for each recognised NPS structure. As this would make the grammar formalism dependent upon the parser, I prefer a declarative approach using a permutation operator applied to the constituents of the rule and to the arguments of the predicate 'Bark'. The use of a permutation operator PERM requires that the constituent NP's are represented as a pair, of which each element can be addressed by a projection operator Π_i:

REF NPS: Π_1(**PERM(REF NP1, REF NP2))(λ x1:**
Π_2**(PERM(SEM NP1, SEM NP2))**
(λ x2: Bark(PERM(x1, x2))))

After conversion, application of this rule for NPS and those for S and VP explained above yields the following two dialogue referent representations:

REF S: #1, !(**Dog, λ x1:** $\exists$ **(Cat, λ x2: Bark(x1, x2))**
REF S: #2, !(**Cat, λ x1:** $\exists$ **(Dog, λ x2: Bark(x2, x1))**

These formula denote *the dog such that it barks a cat* and *the cat such that a dog barks it*, which are suitable descriptions of the newly introduced dialogue referents #1 and #2.

6 References

Bunt, H.C. (1985) *Mass Terms and Model-theoretic Semantics.* Cambridge University Press.

Bunt, H.C., Beun, R.J., Dols, F.J.H., van der Linden, J.A., and thoe Schwarzenberg, G.O. (1984) The TENDUM Dia-

logue System and its Theoretical Basis, in: *IPO Annual Progress Report 19.*

Bunt, H.C., Thesing, J.C., Sloot, K. van der (1987) Discontinuous constituents in trees, rules and parsing, in: *Proceedings of the Third ACL/Europe Conference.* Copenhagen.

Chastain, C. (1975) Reference and Context, in: Gunderson, K. (ed.) *Minnesota Studies in the Philosophy of Science, Vol. VII - Language Mind and Knowledge*, 194-269. University of Minnesota.

Dols, F.J.H. (1989) *The Representation of Definite Descriptions*, ITK Research Report 12, Institute for Language Technology and Artificial Intelligence (ITK). Tilburg University, The Netherlands.

Dols, F.J.H. (1990) Pragmatics of Postdeterminers, Non-restric- tive Modifications and Wh-phrases. In: Y. Wilks & P. Mc Kevitt. (eds) *Proceedings of the 5th Rocky Mountains Conference*, Las Cruces, New Mexico.

Dols, F.J.H. (1992) (ed.) *Pragmatic Grammar Components*, Tilburg University Press, TUP, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

Heim, I. (1982) *The semantics of definite and indefinite noun phrases.* Dissertation, University of Massachusetts.

Hilbert, D. and Bernays, P. (1970) *Grundlagen de Mathematik II.* Springer Verlag.

Kamp, J.A.W. (1981) A Theory of Truth and Semantic Representation, in: Groenendijk, J. & Stokhof, M. (eds.) *Formal Methods in the Study of Language.* Mathematisch Centrum, Amsterdam.

Karttunen, L. (1976) Discourse Referents, in: McCawley, J. (ed.) *Syntax and Semantics 7*, 363-385. Academic Press, New York.

Scha, R.J.H. (1981) Distributive, collective, and cumulative quantification, in: Groenendijk, J.A.G., Janssen, T.M.V., and Stokhof, M.B.J. (eds.) *Formal Methods in the Study of Language.* Proc. of the Third Amsterdam Colloquium, Mathematical Center, Amsterdam. Reprinted in: Groenendijk, J.A.G., Janssen, T.M.V., and Stokhof, M.B.J. (eds.) *Truth, interpretation and information.* 1984, Floris publications, Dordrecht.

Strawson, P.F. (1950) On Referring, in: *Mind 59, January 1950, 330-344.*

OVERVIEW OF ITK RESEARCH REPORTS

No	Author	Title
1	H.C. Bunt	On-line Interpretation in Speech Understanding and Dialogue Sytems
2	P.A. Flach	Concept Learning from Examples Theoretical Foundations
3	O. De Troyer	RIDL*: A Tool for the Computer-Assisted Engineering of Large Databases in the Presence of Integrity Constraints
4	E. Thijsse	Something you might want to know about "wanting to know"
5	H.C. Bunt	A Model-theoretic Approach to Multi-Database Knowledge Representation
6	E.J. v.d. Linden	Lambek theorem proving and feature unification
7	H.C. Bunt	DPSG and its use in sentence generation from meaning representations
8	R. Berndsen en H. Daniels	Qualitative Economics in Prolog
9	P.A. Flach	A simple concept learner and its implementation
10	P.A. Flach	Second-order inductive learning
11	E. Thijsse	Partical logic and modal logic: a systematic survey
12	F. Dols	The Representation of Definite Description
13	R.J. Beun	The recognition of Declarative Questions in Information Dialogues
14	H.C. Bunt	Language Understanding by Computer: Developments on the Theoretical Side
15	H.C. Bunt	DIT Dynamic Interpretation in Text and dialogue
16	R. Ahn en H.P. Kolb	Discourse Representation meets Constructive Mathematics

No	Author	Title
17	G. Minnen en E.J. v.d. Linden	Algorithmen for generation in lambek theorem proving
18	H.C. Bunt	DPSG and its use in parsing
19	H.P. Kolb	Levels and Empty? Categories in a Principles and Parameters Approach to Parsing
20	H.C. Bunt	Modular Incremental Modelling Belief and Intention
21	F. Dols	Compositional Dialogue Referents in Prase Structure Grammar (nog niet verschenen)
22	F. Dols	Pragmatics of Postdeterminers, Non-restrictive Modifiers and WH-phrases (nog niet verschenen)
23	P.A. Flach	Inductive characterisation of database relations
24	E. Thijsse H. Daniels	Definability in partial logic: the propositional part
25	H. Weigand	Modelling Documents
26	O. De Troyer	Object Oriented methods in data engineering
27	O. De Troyer	The O-O Binary Relationship Model
28	E. Thijsse	On total awareness logics
29	E. Aarts	Recognition for Acyclic Context Sensitive Grammars is NP-complete
30	P.A. Flach	The role of explanations in inductive learning
31	W. Daelemans, K. De Smedt en J. de Graaf	Default inheritance in an object-oriented representation of linguistic categories
32	E. Bertino H. Weigand	An Approach to Authorization Modeling in Object-Oriented Database Systems
33	D.M.W. Powers	Multi-Modal Modelling with Multi-Module Mechanisms: Autonomy in a Computational Model of Language

ITK: P.O. BOX 90153 5000 LE TILBURG THE NETHERLANDS