

Beate Firzlaff and Karin Haenelt, Integrated Publication and Information Systems Institute (IPSI), GMD-Darmstadt

# Applying Text Linguistic Principles to Modelling Meaning Paraphrases

*ABSTRACT: This paper proposes the application of text Linguistic principles to modelling meaning paraphrases. Text understanding methods are used in order to construct a formal representation of the conceptual information conveyed in the paraphrase, and text grammatical information is added to the lemma in order to provide information for automatic access in text understanding. The method of textually modelling meaning paraphrases is introduced by an example and discussed in detail. The relevance for text analysis and automatic dictionary update is explained.*

## 1. Introduction

Human users make use of meaning dictionaries in order to find out the meaning(s) of words or phrases they do not know. The meaning explanation supports the integration of new concepts into previous knowledge. The application of the meaning paraphrase basically is a process of natural language understanding.

For human users a long tradition of labour-intensive and highly experienced work of lexicographers has been producing voluminous meaning dictionaries which are continuously improved, enlarged and actualized. Thus much information concerning the meaning of words has been provided already. In order to save resources increasing effort is devoted to making these results also available for application in natural language systems (cf. e.g. (Briscoe/Copestake/Boguraev 1990), (Walker/Zampolli/Calzolari 1988), (Alshawi 1987), (Nakamura/Nagao 1988)).

Current research work basically has aimed at the construction of semantic taxonomies or the generation and completion of lexical entries in accordance with a lexical semantics theory (e.g. (Briscoe/Copestake/Boguraev 1990)). Methods have been provided for semi-automatic or partial extraction of semantic information from machine-readable dictionaries. The information extracted includes semantic taxonomies, grammar codes and lexical cooccurrences. The methods used are mainly syntactic parsing and retrieving database-like features of machine-readable dictionaries (for a short discussion cf. (Briscoe/Copestake/Boguraev 1990)).

Our contribution towards modelling meaning paraphrases differs from these approaches with respect to its goal, the type of information dealt with and the method used for analysis. We view meaning dictionaries under the aspect of their function in natural language knowledge communication. A meaning dictionary explains the use of natural language expressions with respect to the access they provide to world knowledge. These explanations are given in terms of the means natural language provides for communicat-

ing conceptual information, namely in terms of natural language texts (meaning paraphrases, definitions, even longer explanatory texts or abbreviated textual forms). Like other descriptive texts meaning explanations serve to introduce concepts by unfolding their structure and establishing (new) relationships between conceptual information and natural language expressions. A meaning dictionary is part of the natural language knowledge communication process in that it reflects a state achieved by previous communication, and in that it also can be used as a starting point of further communication.

As far as the type of information is concerned, we do not offer an approach for the extraction of certain types of information. We rather propose a method of constructing a textual representation of a meaning paraphrase which among further text related information contains a representation of concepts which have been communicated by the meaning paraphrase.

The method used for analysis is a process of text analysis. In this paper we describe how meaning paraphrases are subjected to a manual text analysis process in accordance with the KONTEXT model (Haenelt/Könyves-Tóth 1991) and we mainly focus on the conceptual information. The principles shown are general principles of text analysis and are also applied in the KONTEXT system to textual analysis of corpora (provided a basic set of dictionary entries is available). First of all a survey of the KONTEXT dictionary modelling is given, then the application of text linguistic principles to modelling meaning paraphrases is described on this background, and finally the relevance of this approach for lexicography is explained.

## 2. The KONTEXT dictionary

### 2.1 Structure of the KONTEXT dictionary

The KONTEXT model (Haenelt/Könyves-Tóth 1991) structures textual information into five layers of text representation:

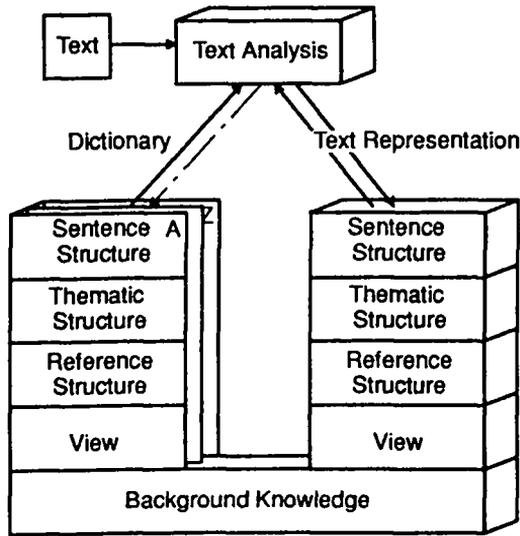
- sentence structure,
- thematic structure,
- referential structure,
- view (on background knowledge),
- background knowledge.

The two lower layers (view and background knowledge) model the conceptual information conveyed in a text, and the three upper layers (sentence structure, thematic structure, referential structure) describe the contextual organization of these concepts.

This structuring is also applied to modelling the entries of the KONTEXT dictionary, which can be regarded as lexicalized text grammar. Textual mechanisms are described in terms of the contribution linguistic means make towards the layers of the text representation.

The information of all the five layers is modelled in feature structures (Könyves-Tóth 1991), (Böttcher 1991). During the process of text analysis these structures provide "building blocks" for the construction of a text representation in accordance with the "instructions" (thematic structure and reference objects) of a particular text (cf. fig. 1). This paper, however, focuses on the conceptual layers (view and background knowledge).

**Figure 1:**  
The KONTEXT  
Text Analysis  
System



## 2.2 Conceptual information in the KONTEXT dictionary

The current modelling of the conceptual layers of the KONTEXT dictionary is based on COBUILD (1987). The method used for determining conceptual definitions is based on the hypothesis that meaning paraphrases (like other texts) provide access to concepts in a non-stereotypic way, i.e. they provide a sequence of instructions (sequence of natural-language expressions in the context of other expressions) of how to establish thematic and referential structures, which in turn guide the access to concepts and the construction of concepts. Currently, there is no way of automatic bootstrapping known. Therefore, in order to gain the basic set of dictionary entries, the meaning paraphrase is subjected to a manual text analysis process in accordance with the KONTEXT model (an example is given below). The concepts are expressed in terms of the model of Semantic Emphasis (Kunze 1991). The layer of background knowledge contains the 'basic semantic forms' identified in (Kunze 1991) via verb fields, i.e. concepts, which are commonly referred to by situationally equivalent verbs. The 'basic semantic form' is a proposition consisting of predicates and elementary arguments.

The model of Semantic Emphasis has been worked out in detail for verbs so far. For the conceptual modelling of adjectives (e.g. "blank") and of nouns (e.g. "writing", "paper"), however, basic principles of verb modelling can also be applied. Since it is not the purpose of this paper to introduce methods of adjective and noun modelling, only a short outline of some principles is given in order to support the explanation of the example given below.

Similarly to Conceptual Dependency approaches (Schank 1975), the KONTEXT approach uses 'basic semantic forms' as the backbone of conceptual modelling. Verbs and abstract nouns are modeled as denoting events or states, and concrete nouns as denoting participants of events and states, i.e. they are defined by their function. Further proposals for noun-based modellings (e.g. (Mel'čuk 1984), (Pustejovsky/Anick 1988), (Wierzbicka 1988), (Boguraev/Pustejovsky 1990)) are to be explored further and are to be integrated into the model. Adjectives also refer to conceptual structures, but in this case these structures are interpreted as embedding rules.

### 3. Application of text linguistic principles to the modelling of meaning paraphrases

Our method of textually modelling meaning paraphrases is introduced by an example, which is discussed in detail. The modelling is described step by step and modelling principles are explained.

#### 3.1 Approach

As an example the following meaning paraphrase of "blank" is used: "A blank piece of paper has no writing or other marks on it." (COBUILD 1987). In order to construct the concept 'blank' in accordance with this explanation, and in order to gain a valid relationship between reference expression, reference object and conceptual definition, a five-layered text representation of the meaning paraphrase must be constructed discourse state by discourse state, i.e. by sequentially interpreting the expressions textually and incrementally integrating the conceptual definitions. Initially the term in the phrase which is to be explained ("blank piece of paper") can be regarded as conceptually unspecified. It is the function of the meaning paraphrase to establish a concept structure and a thematic structure, which defines the relationship of reference expressions and concept structure. A graphical overview of the result is shown in figure 2.

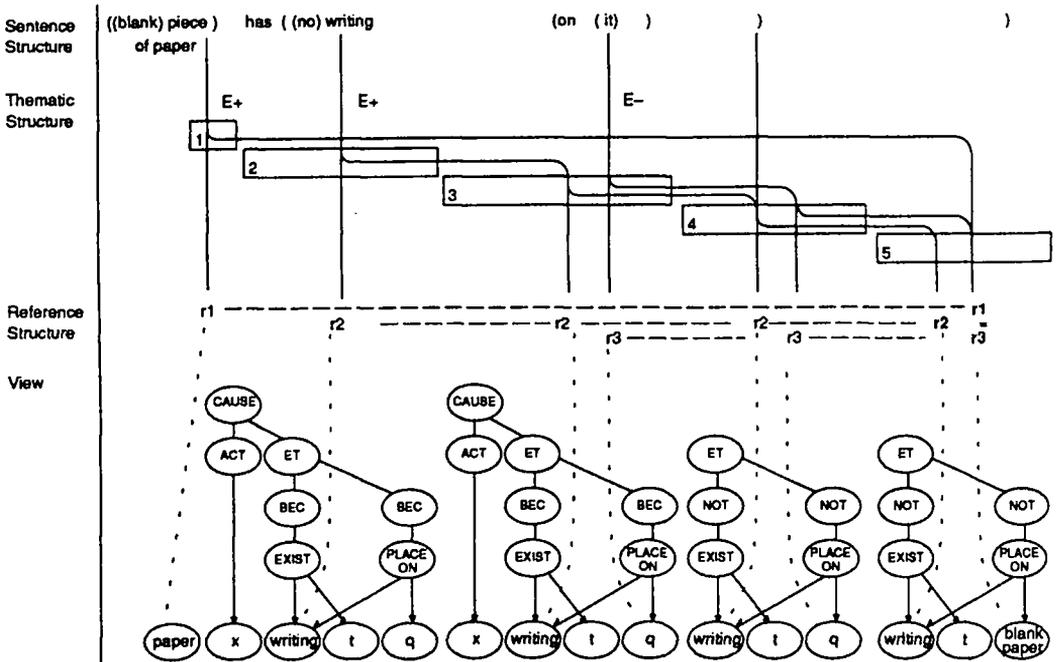


Figure 2: Textual modelling of the meaning paraphrase: "A blank piece of paper has no writing on it" (COBUILD 87)

The figure shows a text representation with five discourse states and four layers. The 'sentence structure' contains the lexemes, their dependency structure (and further syntactic features not shown here). The 'thematic structure' traces the discourse development. It shows discourse states which correspond to contexts (cf. boxes 1-5 in fig.2), and discourse state transitions which correspond to creations, closings and references to contexts (represented by the lines between the boxes). The 'referential structure' contains the reference objects, their development and relationship. The layer of 'view' shows views on the 'background knowledge' which have been established by the intermediate discourse states (1-4) and by the final discourse state (5). The 'background knowledge' is not shown here. In terms of predicate-argument structure the resulting conceptual definition of our example (cf. view state 5) is:

ET (NOT (EXIST (writing, t)), NOT (PLACE-ON (blank paper, writing))).

### 3.2 Textual principles

The result has been achieved by applying e.g. the following principles:

#### a) *determination of the defining situation*

The defining situation is determined in accordance with the dependency tree of the syntactic structure. The dominating verb introduces a situation with its complements as participants. In this case, however, the appropriate reading of "have", namely "have 7" (COBUILD 1987, 667), only "provides a verb for the structure" and "emphasizes that an action or event has a definite beginning or end". It indicates the abstractum "writing" which as a noun is a participant of a situation, but as a deverbative also introduces the situation of which it is a participant. In terms of the model of Semantic Emphasis (Kunze 1991) the situation introduced by "write" is expressed by the proposition:

CAUSE (ACT (x), ET (BEC (EXIST (v, t)), BEC(PLACE-ON (q,v))).

This can be paraphrased as: an action of 'x' causes a 't' to become to exist as 'v', and this 'v' to be placed on a 'q'. This situation is shown in state 2.

#### b) *negation*

the word "no" syntactically is a quantor of "writing". Conceptually it does not make a contribution of its own. It, however, has the effect, that the presupposition of the "write"-situation (which defines "writing") must be selected Jung/ Küstner 1990). According to the general instantiation rule (Kunze 1991) for predicates this is:

ET ( NOT (EXIST(v, t)), NOT (PLACE-ON (q, v))

- which is the initial state of a monotonous path denoted by BEC (A). According to the instantiation rule of BEC (A), at the beginning of the path NOT (A) holds true, and at the end A. The result is shown in state 4.

#### c) *anaphora resolution*

The pronoun "it" in its own context (state 3) denotes the participant 'q' of the "write"-situation. As a pronoun it also effects a referential equation with an antecedent of a

previous situation which had been realized emphatically (E+) (here: a surface form of an actant which is not a prepositional case). In this paraphrase it resumes (state 5) the theme "a blank piece of paper".

So, whilst the definiendum "blank" had no conceptual definition in the initial state (1), in the final state (5) a conceptual definition has been constructed and the relationship between natural language expression and reference object (with its conceptual definition) has been established via the thematic and referential structure of the meaning paraphrase.

### 3.3 Explanation of the conceptual definition

The meaning paraphrase does not really explain "blank". It rather explains "blank X". By doing this it gives an example of the situation and perspective the adjective "blank" imposes on a noun it modifies. The conceptual definition acquired can be regarded as a rule for composing an integrated meaning representation of the adjective and the noun. A similar idea of adjectives modifying nouns under a particular aspect has been proposed by Boguraev/Pustejovsky (1990). The approach shown here is a proposal of how to conceptually model this phenomenon. The rule covers conventionalized readings of "blank": The modelling of the meaning paraphrases of other readings of "blank" in COBUILD (1987) essentially leads to the same structure. They differ from this with respect to further participants of the situation. E.g. "blank cheque" involves a further situation "sign", and the "write"-situation is further specified with respect to what is being written ("the amount of money"). This further definition makes a "cheque" a more specialized "piece of paper". This generality also supports an observation described in (Boguraev/Pustejovsky 1990). Used in texts, however, words with initially conflicting conceptual definition may be connected syntactically. The readings – of the noun as well as of the adjective – then may undergo further changes.

### 3.4 Information on text structure in the KONTEXT dictionary

For the KONTEXT dictionary information of the 'referential structure' layer and of the 'thematic structure' layer are derived from this conceptual definition in accordance with rules which have been formulated for the model of Semantic Emphasis. This includes information like which participant of the situation can constitute a reference object, and which participant can be realized emphatically in which surface form (sentence structure). The description of the sentence structure basically follows the PLAIN-grammar (Hellwig 1980), which is extended by the notion of conditionally obligatory actants (Kunze 1991). In automatic text analysis this information is used for determining the sentence structure, the thematic structure (cf. (Daneš 1970), (Firbas 1971), (Hajičová/Sgall 1988), (Hajičová/Vrbová 1982)) and the reference structure (cf. Habel 1986)).

### 3.5 Coding example

In accordance with the KONTEXT system, feature structures are also used as representation formalism of the coding criteria for the manual analysis which is necessary for acquiring the basic set of dictionary entries and which is presented in this paper. As an example we provide some relevant parts of the entry for "write":



## 5. Bibliography

- ALSHAWI, Hiyan 1987: "Processing Dictionary Definitions with Phrasal Pattern Hierarchies." In: *Comp. Ling.* vol. 13.
- BOGURAEV, Bran; PUSTEJOVSKY, James 1990: "Lexical Ambiguity and the Role of Knowledge Representation in Lexicon Design." *COLING*
- BÖTTCHER Martin 1991: *CFS-System User Manual*. IPSI-report
- BRISCOE, Ted; COPESTAKE, Ann; BOGURAEV, Bran 1990: "Enjoy the Paper: Lexical Semantics via Lexicology." In: *Proc. COLING*
- DANEŠ, František 1970: "Zur linguistischen Analyse der Textstruktur." In: *Folia Linguistica* 4, pp.72-78
- FIRBAS, Jan 1971: "On the Concept of Communicative Dynamism in the Theory of Functional Sentence Perspective." In: *Sborník Praci Filosofické Faculty Brnské Uni.* A19 pp.135-144
- HABEL, Christopher 1986: *Prinzipien der Referentialität. Untersuchungen zur propositionalen Repräsentation von Wissen*. Heidelberg.
- HAENELT, Karin; KÖNYVES-TÓTH, Michael 1991: "The Textual Development of Non-Stereotypic Concepts." In: *Proc. of the 5th Conf. of European ACL*, pp. 263-268
- HAJIČOVÁ, Eva; SGALL, Petr 1988: "Topic and Focus of a Sentence and the Patterning of a Text." In: Petöfi, J. S.(ed.): *Text and Discourse Constitution*. Berlin, pp.70-96
- HAJIČOVÁ, Eva; VRBOVÁ, Jarka 1982: "On the Role of the Hierarchy of Activation in the Process of Natural Language Understanding." In: *Proc. COLING*
- HELLWIG, Peter 1980: *Bausteine des Deutschen. Germanistisches Seminar, Universität Heidelberg*.
- JUNG, Uwe; KÜSTNER, Herbert 1990: *Semantische Mechanismen der Negation*. Berlin. KÖNYVES-TÓTH, Michael 1991: *Incremental Evaluation of Disjunctive Feature Terms*. GMD technical report.
- KUNZE, Jürgen 1991: *Kasusrelationen und Semantische Emphase*. Berlin.
- MEL'ČUK, Igor A. 1984: *Dictionnaire explicatif combinatoire du français contemporain. Recherches lexico-semanticques*. Montreal.
- NAKAMURA, Jun-ichi; NAGAO, Makoto 1988: "Extraction of Semantic Information from an Ordinary English Dictionary and its Evaluation." In: *Proc. COLING*
- PUSTEJOVSKY, James P. G., ANICK, P. G. 1988: "On the Semantic Interpretation of Nominals." in: *Proc. COLING*
- SCHANK, Roger C. (ed.) 1975: *Conceptual Information Processing*. Amsterdam.
- SINCLAIR, John (ed. in chief) 1987: *Collins COBUILD English Language Dictionary*. London, Stuttgart.
- WALKER, Don; ZAMPOLLI, Antonio; CALZOLARI, Nicoletta (eds.) 1988: *Automating the Lexicon: Research and Practice in a Multilingual Environment*. Cambridge.
- WIERZBICKA, Anna 1988: *The Semantics of Grammar*. Amsterdam.

---

KEYWORDS: meaning dictionary, meaning paraphrases, electronic dictionary, text analysis, automatic dictionary generation, automatic dictionary update, knowledge acquisition