

# ON THE ACQUISITION OF CONCEPTUAL DEFINITIONS VIA TEXTUAL MODELLING OF MEANING PARAPHRASES

BEATE FIRZLAFF and KARIN HAENELT

Integrated Publication and Information Systems Institute (IPSI) of GMD

Dolivostraße 15, D - 6100 Darmstadt, Germany

firzloff@ darmstadt.gmd.de, haenelt@darmstadt.gmd.de

phone ++49/(0)6151/869-811, fax -818

## 1 INTRODUCTION

Conceptual knowledge bases are core components of current natural language semantic information processing systems. Their content, structure and size crucially determines the applicability of these systems, and still their development and maintenance is a serious bottleneck. Taking into account the enormous amount of entries required for practical purposes, manual modelling procedures are hardly practicable. Much effort is therefore devoted to achieving a controlled automatic handling of these tasks.

In this paper a proposal is made to use a meaning dictionary as a source of conceptual knowledge, and to apply text understanding methods for processing this information. These methods include the automatic acquisition of conceptual definitions from meaning paraphrases, their transformation into a suitable representation for automatic processing, and the automatic update of conceptual definitions by means of text analysis.

A meaning dictionary explains the use of natural language expressions with respect to the access they provide to world knowledge. In principle, the knowledge accessed and thus represented in a meaning dictionary includes many of the different kinds of knowledge which are required for semantic information processing (i.e. "how nature works, practices, institutions, facts of nature, regularities, ways of doing" [Searle 80:227]).

Current research work on extracting information from machine-readable dictionaries (cf. e.g. [Briscoe/Copestake/Boguraev 90], [Walker/ Zampolli/ Calzolari 88], [Alshawi 87], [Nakamura/Nagao 88]) has mainly concentrated on semantic relationships of the vocabulary and has proposed methods for semi-automatic or partial extraction of this information. These methods basically rely on syntactic structures of meaning paraphrases or database-like coded fea-

tures and values (see also [Briscoe/Copestake/Boguraev 90]). The task of acquiring conceptual information, however, faces the problem that this information in traditional meaning dictionaries is itself conveyed in terms of the means natural language provides for communicating conceptual information, namely natural language texts (meaning paraphrases, definitions, even longer explanatory texts or abbreviated textual forms). Thus understanding conceptual definitions of meaning dictionaries requires textual mechanisms to be recognized and handled in order to analyse the information provided and to construct a semantic representation.

In this paper a new method of textually modelling meaning paraphrases is proposed for the acquisition of a formal representation of the conceptual information conveyed in the meaning paraphrase. It is described how these paraphrases are subjected to a manual text analysis process in accordance with the KONTEXT model [Haenelt/Könyves-Tóth 91]. 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). The conceptual modelling is based on the model of Semantic Emphasis [Kunze 91], which so far has been worked out for verbs. In order to be able to cover the meaning paraphrase as a whole, and thus to provide a basis for automatic bootstrapping, this approach has to be extended to modelling other word classes. A proposal for this is made in this paper.

First of all this paper introduces the approach to textually modelling meaning paraphrases, then an example is given. The concept and text structure related information in the system dictionary which is used as a basis of acquisition as well as for representing the newly acquired entries is presented in detail. The application of the dictionary to the analysis of a meaning paraphrase is explained.

## 2 TEXTUAL MODELLING OF MEANING PARAPHRASES

### 2.1 APPROACH

The KONTEXT model [Haenel/Könyves-Tóth 91] assumes that natural language texts do not communicate complex concepts immediately and at once, but rather provide sequential instructions concerning how to construct them incrementally. The instructions are established by the natural language expressions, which, used in texts, not only have the function of denoting concepts, but rather can be seen as operators which contribute to the construction of thematic structures and reference objects. Thematic structure and reference objects then guide the access to concepts and the incremental construction of new concept definitions communicated by the text. The KONTEXT model structures the information conveyed in a text and the information describing its contextual organization into five layers of text representation: sentence structure, thematic structure, view (on background knowledge), and background knowledge.

In order to construct a conceptual definition in accordance with a meaning paraphrase, the paraphrase must be analyzed step by step under textual aspects. Each natural language expression contributes information towards the construction of at least one of

the four upper layers of text representation. The information of the expressions of the meaning paraphrase is analyzed and a five-layered multi-state text representation is generated. This generation includes operations which incrementally construct (new) concept(s) and establish the relationship between natural language expressions, reference objects and concepts via the thematic structure.

### 2.2 EXAMPLE

As an example the following meaning paraphrase of "blank" is used: "A blank piece of paper has no writing on it." [COBUILD 87]. A textual representation of this paraphrase is shown in figure 1. The figure illustrates how the conceptual definition of "blank" is constructed step by step. It shows a text representation with four layers and five states. The 'sentence structure' contains the lexemes, their dependency structure (and further syntactic features not shown here). The information of the 'sentence structure' basically follows the PLAIN-grammar [Hellwig 80, 88]. The 'thematic structure' traces the discourse development. It shows contexts (cf. boxes 1-5 in figure 1; the numbers refer to the discourse states of their creation), and references to contexts (represented by the lines between the boxes) which correspond to discourse state transitions. The 'reference structure' contains the reference objects, their

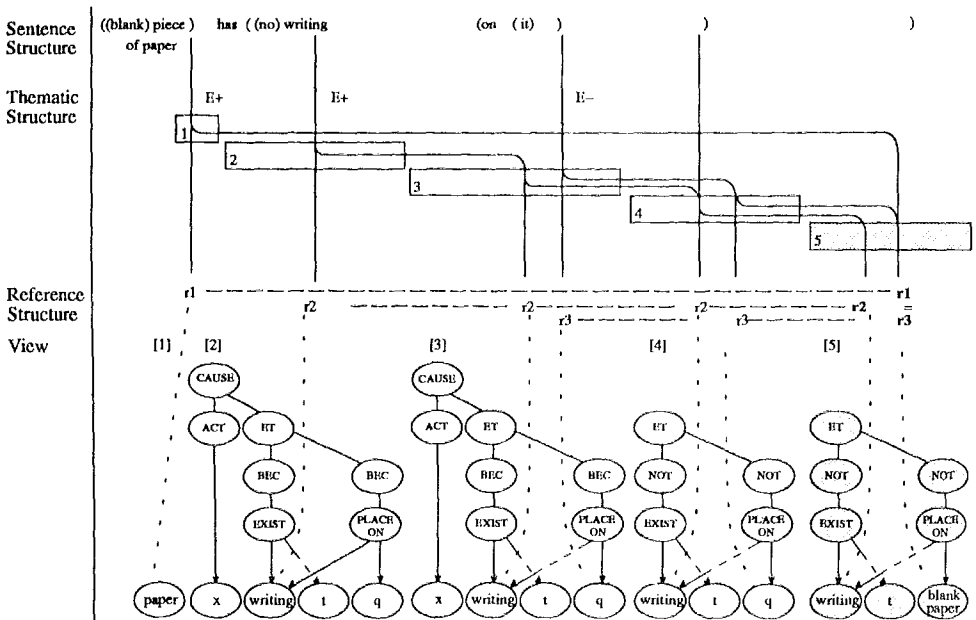


Figure 1: Textual modelling of the meaning paraphrase: "A blank piece of paper has no writing on it" (COBUILD 87)

development and relationship. The layer of 'view' shows views on the 'background knowledge' which have been established by state transitions. The 'background knowledge' (which is shown in figures 2 and 3) contains conceptual descriptions of prototypical situations in accordance with the model of Semantic Emphasis [Kunze 91]. These descriptions are explained in more detail in section 2.3.1.

The states trace the incremental construction of the final view ([5]) of the meaning paraphrase and the (re-)definition of the relationship between natural language expressions and reference objects. 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 by analyzing the meaning paraphrase under textual aspects.

## 2.3 THE KONTEXT DICTIONARY

The contribution natural language expressions make towards the layers of the text representation is described as a lexicalized text grammar. The information of all the five layers is modelled in feature structures which are processed with the CFS-system [Könyves-Tóth 91] [Bötcher 91]. Due to the formalism chosen the basic operation is unification.

### 2.3.1 CONCEPTUAL INFORMATION

The conceptual modelling of the KONTEXT dictionary is based on the model of Semantic Emphasis [Kunze 91]. The core components of this model are 'basic semantic forms' which describe prototypical situations commonly referred to by the verbs of a verb field. A 'basic semantic form' is a proposition consisting of predicates and (propositional or elementary) arguments. So, e.g. the propositional description of a 'write'-situation is: CAUSE ( ACT (x), ET (BEC (EXIST(w,t)), BEC (PLACE-ON (q,w))))). This description can be paraphrased in the following way: An action of 'x' causes a 'y' to become to exist as 'w' and this 'w' to become placed somewhere such that a relation 'ON' holds true between 'q' and 'w'. The scheme of 'write' is based on the general scheme of a produce-situation.

Situation descriptions provide the basic schemes of the layer of background knowledge of dictionary entries. They are used for the construction of the layer of 'view' during text analysis. They can immediately be applied to the modelling of verbs. For modelling further word classes the following extensions are made: Similarly to Conceptual Dependency approaches (e.g. [Schank 75]) in the KONTEXT approach 'basic semantic forms' are the backbone of conceptual modelling. Verbs and deverbative abstract nouns are modelled as referring to prototypical situa-

tions following the descriptions of the model of Semantic Emphasis. Prepositions are modelled as denoting parts of those situations which have actants which in the surface form can be realized with the preposition in question. Nouns are modelled as denoting participants of situations, and adjectives as denoting situations focusing on a particular participant. The defining situations of nouns and adjectives are determined in accordance with the meaning paraphrases of [COBUILD 87] following the textual conditions of applying the situations concerned. COBUILD definitions are also used in order to describe field external distinctions of verbs.

The formal components of the conceptual representation are the feature structure representation of 'basic semantic forms', instantiation rules and role derivation rules following the approach of [Kunze 91].

The propositional description of the 'write'-situation is modelled in feature structures in the following way:

```
[ CAUSE@<
  prop:[ a1:[ ACT@<
    prop:a1:ref: 'x' ]
    a2:[ ET@<
      prop:[ a1:[ BEC@<
        prop:a1:[ EXIST@<
          prop:[ a1:ref: 'w'
                a2:ref: 't' ]]]]
        a2:[ BEC@<
          prop:a1:[ PLACE-ON@<
            prop:[ a1:ref: 'q'
                  a2:ref: 'w' ]]]]]]]]
```

Figure 2: Propositional description of a 'write'-situation

Predicate names are written in capital letters. @< indicates reference to a type, where < enclose path descriptions. The arguments of predicates are values of the feature 'prop'. Elementary arguments correspond to possible reference objects.

Depending on the predicates used, instantiation rules are applied in order to get a more refined description of the situation. E.g. a proposition BEC(A) instantiates a monotonic path such that initially NOT(A) holds true, and finally (A) holds true. NOT(A) is also called the 'presupposition' of the situation, and (A) is (one of) its 'assertion(s)' [Jung/Küstner 90]. The predicate ET has two propositional arguments and requires its arguments to be instantiated at the same time. Thus, either the presupposition or the assertion of both arguments is selected.

For elementary arguments case relations are derived in accordance with a purely formal definition scheme (Kunze's approach (91) provides a theoretical description and formalization of Fillmore's (68) deep cases). The resulting case relations of e.g. the



```

place-on: [ ref: =<refer 1>
           emph: minus ]]
to-obj: [ exist: [ ref: [ ]
                 emph: plus ]
         place-on: [ ref: [ ]
                   emph: minus ]]]]
background:prop: write@<prop> ]

```

Figure 3: Feature structure representation of the result

The value of the feature 'inst' denotes the part of the situation description that corresponds to "no writing on it". It has been derived by applying instantiation rules to the propositional part of the conceptual description and by selecting the presupposition. Relationships to other layers of text representation are denoted as path specifications (<>).

### 3 EXPLANATION OF THE RESULTING CONCEPTUAL DEFINITION

The value of the feature 'inst' can be regarded as the conceptual definition of "blank". The meaning paraphrase, however, does not really explain "blank". It rather explains what it means for something to be blank. So, the conceptual definition acquired can be regarded as a rule for composing the adjective's meaning and a noun's meaning, if one takes meaning paraphrases as rule descriptions. The rule covers conventionalized readings of "blank" (this supports an observation described in [Boguraev/Pustejovsky 90]). Used in texts, however, these readings undergo further changes (cf. [Haenelt 92]). So, e.g. a word can stepwise be related to a complex concept in a text and then be used as a term.

### 4 CONCLUSION

In this paper a method has been described for transforming the conceptual information of a traditional meaning dictionary by means of textual modelling into a suitable representation for automatic semantic text processing. The concept definitions acquired can be used for content and text structure based retrieval (i.e. access to concepts or relevant text passages) or for the update of the dictionary (cf. [Firzlaff/Haenelt 92]). The method proposed is also used for automatic text analysis, and although it requires a basic set of dictionary entries to be transformed by manual analysis, its goal is to allow for an automatic continuation of this transformation. So, if the basic vocabulary of meaning paraphrases has been acquired it should be possible to analyze the majority of meaning paraphrases automatically. The conceptual definitions acquired from meaning paraphrases then can serve as a basis for analyzing further descriptive texts such as e.g. encyclopedias. In this sense the method described in this paper is meant to be a contribu-

tion towards the automatic bootstrapping of knowledge base construction.

### 5 REFERENCES

- [Alshawi 87] Alshawi, Hiyan: Processing Dictionary Definitions with Phrasal Pattern Hierarchies. In: Computational Linguistics, vol. 13, 1987.
- [Boguraev/Pustejovsky 90] Boguraev, Bran; Pustejovsky, James: Lexical Ambiguity and the Role of Knowledge Representation in Lexicon Design. In: Proc. COLING 1990.
- [Böttcher 91] Böttcher, Martin: The CFS System User Manual. Technical report IPSI 1990.
- [Briscoe/Copestake/Boguraev 90] Briscoe, Ted; Copestake, Ann; Boguraev, Bran: Enjoy the Paper: Lexical Semantics via Lexicology. In: Proc. COLING 1990).
- [COBUILD 87] Sinclair, John (ed. in chief): Collins COBUILD English Language Dictionary. London, Stuttgart: 1987.
- [Fillmore 68] Fillmore, Charles: The Case for Case. In: Bach, E.; Harms, R.T. (eds.): Universals in Linguistic Theory. New York, 1968, pp. 1-90
- [Firzlaff/Haenelt 92] Firzlaff, Beate; Haenelt, Karin: Applying Text Linguistic Principles to Modelling Meaning Paraphrases. In: Proc. EURALEX 1992.
- [Haenelt 92] Haenelt, Karin: Towards a Quality Improvement of Machine Translation: Modelling Discourse Structure and Including Discourse Development in the Determination of Translation Equivalents. In: Proc. 4th Int. Conf. on Theoretical and Methodological Issues in Machine Translation. 1992.
- [Haenelt/Könyves-Tóth 91] Haenelt, Karin; Könyves-Tóth, Michael: The Textual Development of Non-Stereotypic Concepts. In: Proc. 5th conference of European ACL, Berlin 1991. pp. 263-268
- [Hellwig 80] Hellwig, Peter: Bausteine des Deutschen. Germanistisches Seminar, Universität Heidelberg, 1980.
- [Hellwig 88] Hellwig, Peter: Chart Parsing according to the Slot and Filler Principle. In: Proc. COLING 1988.
- [Jung/Küstner 90] Jung, Uwe; Küstner, Herbert: Semantische Mechanismen der Negation. Berlin: Akademie-Verlag, 1990.
- [Könyves-Tóth 91] Könyves-Tóth, Michael: Incremental Evaluation of Disjunctive Feature Terms. Arbeitspapiere der GMD 1991. (technical report)
- [Kunze 91] Kunze, Jürgen: Kasusrelationen und Semantische Emphase. Berlin: Akademie-Verlag, 1991.
- [Nakamura/Nagao 88] Nakamura, Jun-ichi; Nagao, Makoto: Extraction of Semantic Information from an Ordinary English Dictionary and its Evaluation. In: Proc. COLING 1988, pp. 459-464
- [Schank 75] Schank, Roger C. (ed.): Conceptual Information Processing. Amsterdam: North Holland, 1975.
- [Searle 80] Searle, John R.: The Background of Meaning. In: Kiefer, F.; Bierwisch, M. (eds.): Speech Act Theory and Pragmatics. Dordrecht / Boston / London: D.Reidel Publishing Company, 1980.
- [Walker/Zampolli/Calzolari 88] Walker, Don; Zampolli, Antonio; Calzolari, Nicoletta (eds.): Automating the Lexicon: Research and Practice in a Multilingual Environment. Cambridge University Press, Cambridge 1988.