The NIBA[1]-Approach to Quantity Settings and Conceptual Predesign

Günther Fliedl*, Christian Kop[1], Willi Mayerthaler*,
Heinrich C. Mayr[1], Christian Winkler*

[1]Department of Business Informatics, *Department of Computational Linguistics
University of Klagenfurt, Austria

Abstract: This paper reports on a further step of the NIBA[2] project. We will distinguish some primary types of natural language (NL) quantification by means of linguistic analysis. Moreover, we are going to present some recent results in the field of NP-analysis in terms of our linguistic model (NTS Naturalness Theoretical Syntax [MFW98]) as far as quantification is concerned. Based on a typology of quantifiers we are going to discuss allquantification, existential quantification, numeral quantification (vs. indefinite articles), and some variants of fuzzy quantification. Furthermore, we will describe some basic aspects of the Klagenfurt Conceptual Predesign Model (KCPM [Fl96], [KM98]) and the relevance of quantification with respect to glossary entries.

1 Introduction

Similarly to COLOR-X [BR95], RADD [Al95], NL-OOPS [MG99]), in NIBA we are engaged in the derivation of conceptual schemata out of natural language based requirements specifications. Our approach focuses an intermediate step between requirements elicitation and conceptual design, the so-called KCPM (Klagenfurt Conceptual Predesign Model). KCPM is intended to harmonize the developer's and the end-user's view of a given universe of discouse (UoD), thus providing an interface for their mutual understanding. As opposed to conventional conceptual modeling, where input texts are formalized, KCPM suggests the idea of collecting and cataloguing UoD information in a non-textual form immediately after a linguistic analysis. Linguistic analysis is done by means of the NT(M)S-parser implemented in IF-PROLOG. The tool for KCPM is implemented in Visual Basic and ACCESS.

In this paper, we are especially interested in the illustration of different quantity-types and their equivalents in the glossary entries of KCPM. Together with transaction profiles, the type and the frequency of queries etc. these quantity-types influence decisions on data base tuning.

---

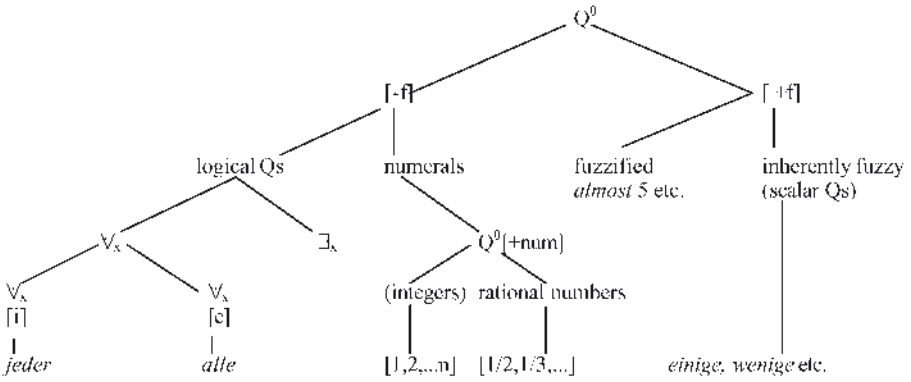
[1] The NIBA project is founded to a large extent by the Klaus Tschira Stiftung Heidelberg and by the *Jubiläumsfonds* of the Austrian National Bank.

[2] NIBA is an acronym for Natürlichsprachliche Informationsbedarfsanalyse (Natural Language Requirements Analysis). The NIBA project is an interdisciplinary cooperation of linguists and computer scientists at the University of Klagenfurt.

2 Classification of Quantifiers

One of the problems in natural language engineering is the simple fact that quantificational systems of natural languages are much richer than those provided by standard formal languages (e.g. predicate calculus, crisp mathematics). It is quite normal e.g. that end users speak in terms of allquantification and (fuzzy) numbers and/or fuzzy intervals (e.g. *about five, pretty much, very few, something in between five and ten* etc.). Let us, however, first present a very sketchy overview of quantifiers (Q) in natural languages:

TYPOLOGY OF QUANTIFIERS[3]



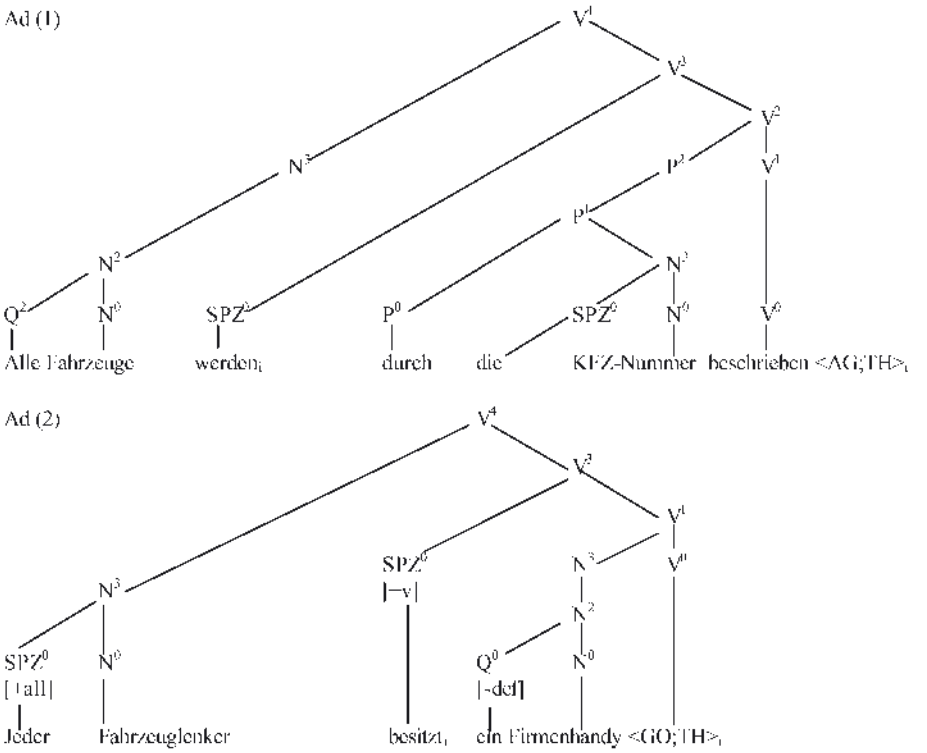
3 Allquantification

Allquantification in German is normally coded by means of *alle* (all) and *jeder* (every). In the following structures, we illustrate the difference in agreement (*jeder* implies singular, *alle* implies plural) by indicating different categories (Q^2 for *alle*, SPZ^0 [+all] for *jeder*).

(1) *Alle Fahrzeuge sind durch die KFZ-Nummer beschrieben.* (All cars are identified by their number.)

(2) *Jeder Fahrzeuglenker besitzt ein Firmenhandy.* (Every driver has a cellphone belonging to the firm.)

[3] [-f] = crisp, [+f] = fuzzy; in $\forall_x$[i] individuals are foregrounded, in $\forall_x$[e] ensembles.
Of course there are other quantifiers not taken into account in this sketchy tree, e.g. negated existential quantification (usually with NEG-incorporation) as in *nobody, none, never, nowhere* etc. It stands to remark, by the way, that human languages have at least two all-quantifiers (e.g. Engl. *all, every, each, any* or Spanish *todos, cada uno* etc.)



For the scope of quantification thing-types and connection-types regarding quantification are of significance. Thing-types in KCPM correspond to the unification of classes and value-types of attributes in UML [BRJ98]. A thing-type models relevant notions or technical terms occurring in the user's domain. Normally, they are expressed by noun phrases. Typical things (instances of thing-types) are natural or juristic persons, material or immaterial objects, abstract notions. Connection-types are employed to describe relations between thing-types. They can be perceived of as generalizations of the notions *association* and *attribute* (relationship and attribute) occurring in conceptual design. Fig.1 and 2 outline the way information is presented in KCPM. Since nouns and noun groups are treated as *thing-types*, they are represented in the *thing-type glossary*. Sentences (1) and (2) describe a part of a *connection-type*, namely the perspective from the subject of the sentence. In the column "min, max" the minimal cardinality is set to 1 because of the allquantifiers *alle, jeder*.

Fig.1: Thing-type glossary

Id	Name	....	Quantity Descriptor	....
T11	Fahrzeuglenker			
T12	Firmenhandy			
T13	Fahrzeug			
T14	Kfz-Nummer			
....	....	....	....	....

Fig.2: Connection-type glossary for sentences (1) and (2)[4]

Id	Name	Perspective				....
		Perspective#	Involved Thingtype	Name	Min,Max	
C04	....	P41	Fahrzeuglenker	Besitzt	1,?	
		P42	Firmenhandy		?,?	
C05	....	P51	Fahrzeuge	Beschrieben	1,?	
		P52	Kfz-Nummer		?,?	
....	....	....	....	....	....	....

4 Numeral quantification

In the default case, numeral quantification is done by means of simple and unequivocal expressions as shown in the following sentences:

(3) *Die Firma betreut maximal 1000 Kunden.* (The firm serves/attends to 1000 customers at the maximum.)

(4) *20 Mitarbeiter der Firma X verfügen über ein Notebook.* (20 employees of the firm X have a notebook.)

(5) *Die 20 Notebooks gehören der Firma.* (These 20 notebooks belong to the firm.)

In terms of linguistic categories the quantifieres applied in the sentences above correspond to numeralia. We classify them as Q^2 Phrases with the feature [+num]. As for the glossary entries, numeral quantifications are always possible candidates for cardinalities (see [Fl96] for further detail). In the given example, one object (*Firma* respectively *Notebook*) could be related to 20 respectively 1000 other objects (*Mitarbeiter*, *Notebook* respectively *Kunde*). The glossary in figure 4 represents this case.

Fig.3: Thing-type glossary entries for sentence (3), (4)

Id	Name	....	Quantity Descriptor	....
T21	Kunde		Min 0; Max 1000; type "stock"; period --	
T22	Firma			
T23	Mitarbeiter			
T24	Notebook			
....	....	....	....	....

Normally, if *Firma* is the organization the planned information systems is developed for, it would not be a part of the KCPM scheme (glossary). Consequently, there is no connection-type referring to *Firma* either. Thus cardinality would not be the appropriate concept to model the fact that the *Firma* serves 1000 clients. Instead, quantity description of a thing type (e.g. 1000 quantities of a customer or 20 notebooks) is more likely (see fig.3). The same holds for sentence (5). Here, the default interpretation would be: There are 20 employees (Mitarbeiter) owning a Notebook.

There are situations that might be interpreted differently. Discussing them, however, means to go beyond the scope of this paper and thus is deferred to a consecutive one.

[4] The question mark (?) in the column 'min, max' (cardinality) indicates: 'not yet decided'.