THE MERGED UPPER MODEL: A LINGUISTIC ONTOLOGY FOR GERMAN AND ENGLISH

Renate Henschel and **John Bateman***

Project KOMET

GMD/Institut für Integrierte Publikations- und Informationssysteme

Dolivostr. 15, D-64293 Darmstadt, Germany

{henschel,bateman}@ipsi.darmstadt.gmd.de

Abstract

A detailed comparison of the Penman Upper Model and the KOMET German Upper Model has been carried out in order to construct a new *Merged Upper Model* capable of serving as the ideational basis for generation in both English and German. Previously proposed criteria for conducting such a merge are expanded on and evaluated. It is established that no (semi-)automatic merging of such knowledge sources can be expected to produce a reasonable result and that detailed comparison of the kind reported is essential. The result of the merge is now being used as the basis for sentence generation in English, German and Dutch.

1 INTRODUCTION: MULTILINGUAL LINGUISTIC 'ONTOLOGIES'

With the need to develop re-usable frameworks for organizing information (cf., e.g., the ARPA Knowledge Sharing Effort [Patil *et al.*, 1992]), workable proposals for 'ontologies' to provide such organization are increasing in importance. Such ontologies are now commonly applied in Natural Language Processing systems since there the representation of commonsense and domain knowledge is essential. Accordingly, a number of organizations of knowledge have been developed - some quite extensive. These organizations are evaluated by the extent to which they prove re-usable across distinct domains and applications. The consideration of the re-use of these organizations is, however, complicated by the range of differing design criteria that are employed in their construction; an extensive overview of approaches is given in [Bateman, 1992a]. One particular approach is to define *linguistically motivated* ontologies, where the criteria for organization rest on semantic distinctions that the grammar of a language needs to have drawn in order to motivate its deployment of grammatical distinctions.

Two sizeable linguistic ontologies constructed in this way are the Penman Upper Model developed for English text generation within the Penman project at USC/ISI [Penman Project, 1989] and the Upper Model for German developed similarly for text generation within the KOMET project at GMD/IPSI [Bateman *et al.*, 1991a]. The English Upper Model (*EUM*) is described in [Bateman *et al.*, 1990]; the concepts of the German Upper Model (*GUM*) go back to [Steiner *et al.*, 1988] and [Teich, 1992]. Both ontologies are individually examples of the most detailed such ontologies currently under development, each with over 200 domain and application independent concepts arranged in a subsumption lattice, spanning distinct types of processes (mental, communication, relational, actions), and diverse qualities and objects. Both ontologies have been used in a number of domains and show good re-usability characteristics mainly due to the fact that they are linguistically motivated. Thus, for example, if language generation is required there is typically 100% re-usability across domains in contrast to the 50% described by [Pirlein, 1993] for the, largely non-linguistically motivated, LILOG ontology.[1]

There are a number of suggestions for the evaluation of ontologies in terms of formal properties of consistence and coherence of the information those ontologies contain (e.g., [Horacek, 1989, Guarino, 1994]). With a restriction to linguistically motivated ontologies, we can now state further design principles concerning what is to be represented and how. Although these principles were originally developed in order to carry out a detailed comparison of the *EUM* and the *GUM*, they are generally applicable for all linguistically motivated ontologies; evaluating the concepts proposed within such an ontology according to the principles described in this paper should improve the status of that ontology overall.

The main result of the *EUM-GUM* comparison is a *Merged Upper Model* presently used within the KOMET project as the basis for multilingual sentence generation in English, German and Dutch.[2] This then also provides an early answer to a question concerning a different kind of re-usability of linguistically motivated ontologies, i.e., the extent to which they can be re-used across distinct *languages* rather than across distinct domains.

2 THE MERGING METHOD

2.1 Starting points

Merging distinct ontologies is a problem that will occur more frequently as new proposals are to be reconciled. [Hovy and Nirenburg, 1992] propose a general method for creating a merged ontology out of different ontologies where it does not matter whether

*Also on indefinite leave from the Penman Project, USC/Information Sciences Institute, Marina del Rey, Los Angeles.

[1] This is simply because a linguistically motivated ontology is bound to the semantics of a grammar and not to general, possibly domain-transcendent knowledge. The two kinds of ontologies should therefore be seen as performing different kinds of work. For extensive motivations for maintaining a linguistically motivated ontology, see [Halliday and Matthiessen, to appear, Bateman, 1992a].

[2] The comparison is based on the English Upper Model and German Upper Model data files from July 1992. Both are expressed in the knowledge representation language LOOM ([MacGregor and Brill, 1989]).

differences are language dependent or due to different linguistic theories. The commonalities and differences in two ontologies are classified according to Hovy and Nirenburg as follows:

1. **Identity**: The same concept is found in both ontologies.

2. **Extension**: There is a concept in one ontology which is missing in the other, but which specializes the latter ontology further.

3. **Cross classification**: The partitioning of identified concepts into subconcepts differs in the considered ontologies.

The merging procedure then keeps all concepts of cases (1) and (2) and resolves case (3) by exhaustive cross classification.

A simplified version of this procedure is proposed in [Hovy and Knight, 1993]. Here, the cross classification resulting from nonmatching partitions of identified concepts into subconcepts is replaced by parallel subordination of those subconcepts. This results in a substantial reduction in the concepts necessary, but leaves open the question of the mutual relation between concepts stemming from different source ontologies. Participating NLP components can be controlled well by such a shared ontology, but its adequacy as a point of communication in a joint MT system is less clear.

We have found that it is necessary to go beyond the original merging methodology in a number of ways. Nevertheless, as a consequence of the criteria for merging that we propose, not all existent concepts of the *EUM* and *GUM* need find their representation in the Merged Upper Model and cross classification is still significantly reduced without impairing inter-translatability across concepts arising from different source ontologies.

2.2 Problems with identity

The crucial point in [Hovy and Nirenburg, 1992] is the notion of 'identity'. The decision how to deal with different concepts (identification, extension, or cross classification) is based on the possibility of stating an identity between concepts of different language ontologies. This is somewhat problematic. In the comparison between the English and the German upper models, we took as identification criterion the equivalence of the sentences or phrases which can be generated by the concepts. This correspondence relies on the assumption that German and English sentences have a one-to-one-mapping and that translation is a totally information preserving relation. Although this is not true in general, we based our merging on the assumption that it may be true for simple sentences if we abstract out the textual and interpersonal—i.e., non-*experiential* (examples below)—dimensions of utterances, and the language distance is close. Hence, the whole construction has to be seen in the context of its own relativity.

2.3 Further Principles

Principle 1: Removal of non-experiential concept discrimination

Difficulties can arise when ontologies to be merged are themselves inherently problematic in some way. Internal problems should not be automatically transmitted to a merged ontology. Thus, during merging, the distinctions drawn in individual ontologies need always to be evaluated internally before being admitted. The exclusion of textual and interpersonal information in a merged upper model provides an additional important criterium for an 'extended identification' of concepts *within* ontologies to be merged. Two common kinds of non-experiential concept discriminations were found in the *GUM*. The first introduces distinct upper model concepts in order to motivate lexicogrammatical realization by differing types of grammatical units. The second introduces distinct upper model concepts to motivate the *selection* of semantic roles from a given semantic configuration that are to be lexicogrammatically expressed.

An example of the first kind is offered by the concepts *G-Relational* and *G-Relationship*[3]. These are both responsible for the generation of processes experientially classifiable as *relational*, but whereas *G-Relationship* causes an attributive or adverbial realization, *G-Relational* causes a clausal realization. Thus, phrases (1) and (2) can only be generated from different semantic input—expressed here in the form of the typed semantic assertions of the Penman Sentence Plan Language (SPL) [Kasper, 1989]:[4]

(1) Das Mädchen ist krank. (The girl is sick.)

```
(a / classificatory
   :attribuant (m / person :lex mädchen)
   :classifier (k / quality :lex krank))
```

(2) das kranke Mädchen (the sick girl)

```
(a / property-ascription
   :domain (m / person :lex mädchen)
   :range (k / quality :lex krank))
```

This problem does not surface so often in the *EUM*, although there are occasional violations—e.g., the inclusion of 'rhetorical relations' that are explicitly textual (see [Bateman *et al.*, 1990] for details).

Examples of the second kind are offered by sentences (3a-b).

(3)a. Der Lehrer antwortet, dass das Raumschiff zurück gekehrt ist.
(The teacher answers that the spaceship has returned.)

b. Der Lehrer antwortet den Schülern, dass das Raumschiff zurück gekehrt ist.
(The teacher answers the students that...)

The differences in (3) arise from differences in the number of semantic participants in the answering-event that are made grammatically explicit. Both (3a) and (3b) could be used to describe the same *experiential* event, the selection being made on non-experiential grounds (e.g., lack of relevance of a participant, being known from context, known from preceding text, etc.) specific to the text being created.

[3] Concepts from the English Upper Model and the German Upper Model will be differentiated where relevant by prefixing either 'E-' or 'G-' as appropriate.

[4] *Classificatory* is a subtype of *GUM* concept *Relational, property-ascription* a subtype of *relationship*. Further, in the SPL examples in this paper, lexical selection is specified directly by means of the keyword `:lex` to avoid complicating the discussion unnecessarily.

Both are, however, classified semantically underneath distinct *GUM* concepts. These distinct concepts have differing obligatory role configurations, which requires that the selection of semantic (experiential) type has to be made according to the participants that are to be expressed – a decision that is often made on textual grounds without a change in experiential perspective. In the *EUM* the only semantic distinction in this area of communication processes is between 'telling'-like events (*addressee-oriented*) and 'saying'-like events (*non-addressee-oriented*), which is a difference in experiential perspective.[5] In the *EUM*, concept discrimination is made with respect to differing possible *realizations* of roles, not straightforward absence/presence of roles as in the *GUM*. 'Missing' surface participants can be modelled more adequately by an upper model-grammar interface which allows defined semantic roles to have zero realization. This is an elegant way to deal with optional participants, passive, and impersonal constructions.

The net effect of both kinds of violations[6] of principle 1 is that the number of concepts is increased and necessary decisions concerning lexicogrammatical realizations are avoided. The proliferation of concepts—if allowed—would complicate considerably the task of 'identification' of similar concepts in ontologies to be merged.

Principle 2: Intelligent cross classification

Even following extended identification of concepts, it is not sufficient to provide cross products for those concepts that are not identifiable but which classify overlapping semantic areas. This is clarified by the following concrete, although much abbreviated, example of merging in the area of material (action) process types. The decisions that are required here are typical of the merging process as a whole.

The *E-Material-Process* hierarchy distinguishes processes more or less with regard to transitivity patterning. An *E-Nondirected-Action* is a process without external causation (mostly intransitive, although transitive sentences where the object is not affected or created by the action also fall into this class). *E-Ambient-Process* and *E-Motion-Process* are not exhaustive subconcepts of *E-Nondirected-Action*.

The vase broke.	*Nondirected-Action*
I play piano.	*Nondirected-Action*
The tourist ran.	*Motion-Process*
It rains here.	*Ambient-Process*

An *E-Directed-Action* in contrast is a process with an external causer as additional participant. *E-Directed-Actions* divide into *E-Creative-Material-Action* and *E-Dispositive-Material-Action*.

The child broke the vase.	*Dispositive-Material-Action*
The lion chased the tourist.	*Dispositive-Material-Action*
Mary baked a cake.	*Creative-Material-Action*

The *GUM* differentiates *G-Agent-centered*, *G-Affected-centered*, *G-Agent-only* and *G-Affected-only* as disjoint *G-Action* subtypes. Here, we have at first a classification with regard to kind and number of participants. Examples for the semantic representations of the intransitive process types are given in (4) and (5), again in SPL notation:

(4) Der Tourist rannte. (The tourist ran.)

```
(r / action :lex rennen
   :agent (t / tourist))
```

(5) Die Pflanze geht ein. (The plant is dying.)

```
(e / action :lex eingehen
   :affected (p / pflanze))
```

The transitive processes (with two participants) are further broken up into *G-Agent-centered* and *G-Affected-centered*. The *G-Affected-centered* process type is a very special case of a transitive process. The definition is given in [Steiner *et al.*, 1988] thus: "X *affected-centered-verb* Y iff X causes that Y *affected-centered-verb*". Examples are:

Das Kind zerbricht die Vase. ↔
Das Kind bewirkt, dass die Vase zerbricht.

The child breaks the vase. ↔
The child brings it about that the vase breaks.

Thus a process is called *G-Affected-centered* if the realizing verb is able to form an ergative pair. All *G-Affected-centered* processes have at least two participants, the *G-Agent* and the *G-Affected*.

The *G-Agent-centered* process is differentiated with regard to the different participant types for the second participant:

Affecting	Der Bauer fällt <u>den Baum</u>. (The farmer is felling the tree.) *G-Agent* *G-Affected*
Effecting	Die Mutter malt <u>ein Haus</u>. (The mother is painting a house.) *G-Agent* *G-Effected*
Ranging	Ich spiele <u>Klavier</u>. (I play piano.) *G-Agent* *G-Process-range*

At first sight, there are few commonalities between these two ontologies. Without deeper introspection, one can only state an identity

E-Ambient-Process = *G-Natural-Phenomenon*

and could mechanically build a cross classification as shown in Figure 1. Some created concepts should, however, be omitted from this 'cross product ontology'.

The first obvious argument is the number of participants. These are contradictory in the following cross concepts: *E-Directed-Action/G-Agent-only* and *E-Directed-Action/G-Affected-only*. A comparison of the low level concepts shows further that the

[5] The former necessarily involves an addressee semantically, someone who is intended to be listening, while the latter does not. This difference is grammaticized in English in the acceptability/non-acceptability of 'I told him that...'/'I said him that...'. In order to grammaticize an addressee in a saying-like event, it is necessary to respect its less central role and to use the form 'I said *to him* that...'.

[6] There are further similar cases; for example the *GUM* also includes *interpersonally* motivated concept discriminations such as *negative-feature-ascription* and *negative-quality*. These govern the generation of negative assertions, thus pre-empting a more appropriate speech function control of negation.

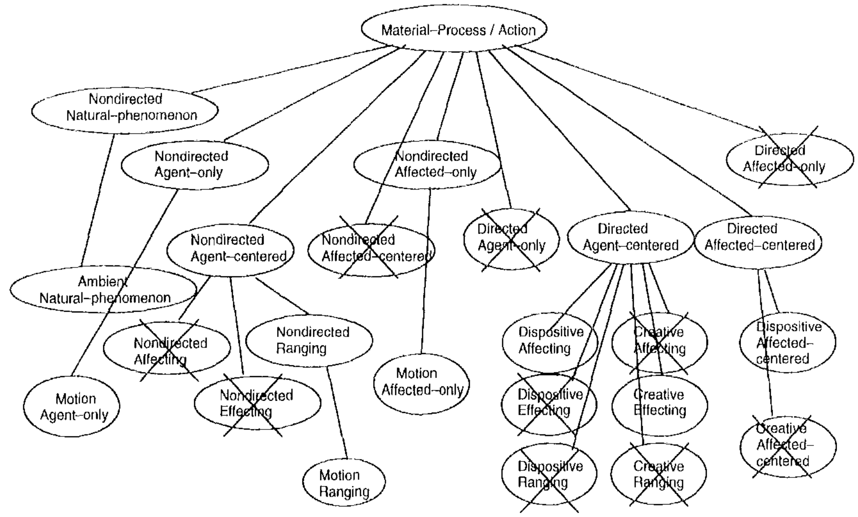

Figure 1: Mechanical merge of the material processes by cross classification

following *GUM* and *EUM* concepts can in fact be identified:

E-Dispositive-Material-Action =
G-Affecting + *G-Affected-centered*
E-Creative-Material-Action = *G-Effecting*.

This rules out the cross concepts:

E-Dispositive-Material-Action/ G-Effecting,
E-Dispositive-Material-Action/ G-Ranging,
E-Creative-Material-Action/ G-Affected-centered,
E-Creative-Material-Action/ G-Affecting,
E-Creative-Material-Action/ G-Ranging.

Furthermore, it is known from the definition of *E-Nondirected-Action* in [Bateman *et al.*, 1990] that such processes are either intransitive or they have a second participant which is in meaning nothing else than the *G-Process-range* participant. Hence, the cross concepts:

E-Nondirected-Action/ G-Affecting,
E-Nondirected-Action/ G-Effecting,
E-Nondirected-Action/ G-Affected-centered

as well as its subconcepts

E-Motion-Process/ G-Affecting,
E-Motion-Process/ G-Effecting

are ruled out. Finally, the exhaustive coverage of the low level subtypes in the *EUM* and *GUM* supports the following identities:

E-Nondirected-Action/ G-Natural-phenomenon
= *E-Ambient-Process/ G-Natural-phenomenon*
E-Nondirected-Action/ G-Agent-centered
= *E-Nondirected-Action/ G-Ranging*,
E-Directed-Action/ G-Affected-centered
= *E-Dispositive-Material-Action/ G-Affected-centered*,

E-Directed-Action/ G-Agent-centered
= *E-Dispositive-Material-Action/ G-Affecting*
+ *E-Creative-Material-Action/ G-Effecting*.

By these kinds of detailed considerations, we have filtered an intelligent merge out of the mechanical merge. Within the intelligent merge, we omit the German differences concerning the participant number (*G-Agent-only*, *G-Ranging*) since these violate principle 1, and do not establish the very subtle *G-Affected-centered* type. Preferring the English terminology the result is given in Figure 2.

This turns out to be mainly the *EUM* subhierarchy for material processes. To also cover the German requirements, the *Nondirected-Action* concept is differentiated into *Nondirected-Doing* and *Nondirected-Happening* according to the distinction between *Agent-only* and *Affected-only*. Therefore we do not need to preserve the German participant types *Agent* and *Affected*, and can infer the relevant information from the new *Nondirected-Action* subconcepts. The German SPL examples (4) and (5) then have the revised semantic form:

(4′)
```
(r / nondirected-doing :lex rennen
    :actor (t / tourist))
```

(5′)
```
(e / nondirected-happening :lex eingehen
    :actor (p / pflanze))
```

Because we have fixed the semantic differences between the *G-Agent* and the *G-Affected* participant in the process types we do not need this differentiation as participant roles again. Hence, we choose the English participant types *E-Actor* and *E-Actee*, the correspondence of which to the German *G-Agent*, *G-Affected*, *G-Effected* and *G-Process-range* differs with the process type (see Figure 2). For further details of the merging of all 12 top-level regions of the two ontologies, see [Henschel, 1993].

Principle 3: Flexible semantics-grammar interface

One peculiarity of the proposed merging is that we do not assume a straightforward correspondence between concepts (especially process types) and sets of surface sentences. That means, disjoint concepts

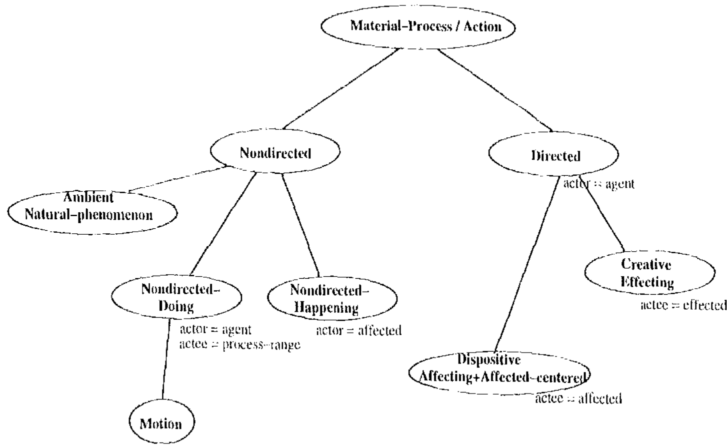

Figure 2: Merging proposal for the material process type

in the Merged Upper Model do not necessarily correspond to disjoint sets of surface sentences – only to disjoint semantic perspectives on them. The interface between the upper model and the grammar needs to be written in such a way that it is possible in some cases to generate the same sentence from different semantic input. This approach meets the differences between the process type partitioning in the *EUM* and the *GUM* without eliminating both perspectives and without creating new cross product types (as it would be the case in the simple merging strategy), but by giving the upper model-grammar interface more flexibility[7]. As a consequence, a sentence such as (6) can now have two distinct semantic representations (7a-b) according to the Merged Upper Model; the concept *destination* in (7b) is a subconcept of *relational-process*, which is disjoint to *material-process* in (7a).

(6) Der Sohn begleitet seinen Vater in die Stadt.
(The son accompanies his father to the city.)

```
(7)a. (b / material-process :lex begleiten
        :actor (p / person :lex sohn)
        :actee (v / person :lex vater)
        :destination (s / one-or-two-d-location
                         :lex stadt))

   b. (b / destination :lex begleiten
        :domain (v / person :lex vater)
        :range (s / one-or-two-d-location
                   :lex stadt)
        :third-party-agent
             (p / person :lex sohn))
```

The two semantic representations correspond to two genuinely alternative experiential perspectives on the event, one focusing more on its action-like nature, the other more on its relational-like nature.

3 RESULTS AND CONCLUSIONS

3.1 The Merged Upper Model

By applying the principles for merging set out above, it was possible to fully replace both upper models by a single merged upper model that differs very slightly from the *EUM*. The Merged Upper Model can in fact be obtained from the *EUM* by a small number of additions (6 new concepts and 1 change of role restrictions to an existing concept). This lack of difference supports the claims concerning multilinguality of functional descriptions made in [Bateman *et al.*, 1991b]. There it is argued that a functional grammar already goes beyond strictly language specific distinctions: the re-usability of the vast majority of the *EUM* organization for German demonstrates that that organization is not tied solely to English. This is further reinforced by the experience during the merge that where the *EUM* extended on distinctions made in the *GUM*, these extensions were generally equally applicable and useful for German (see [Henschel, 1993] for relevant examples).

The result of our merging procedure is an ontology fulfilling the construction ideas of [Hovy and Knight, 1993] in that the resulting ontology contains all concepts necessary for the operation of the PENMAN module and the KOMET module. However, it contradicts the merging theory of Hovy and Knight in that it states some theoretical principles for the merge construction which should be maintained by the source ontologies as well as in the merge.

3.2 Merging Statistics

Because of their questionable status, we leave the 'rhetorical relations' out of account in the statistical comparison. Without this RST-subhierarchy the *EUM* includes 252 concepts. The *GUM* makes no precise distinction between upper and domain model. For the comparison, 235 *GUM* concepts are considered. The Merged Upper Model contains 258 concepts.

Identity

We found 167 identical concept names (excluding the RST-relations), from which only 87 concepts can really be identified. Identical meaning can strongly be stated for 106 concepts (i.e. 19 have distinct names). The main identification areas are the object and the quality hierarchy as well as the temporal one. The

[7] Giving this interface this flexibility is in any case argued for on other grounds in [Bateman, 1992b].

precise distribution for strong identical meaning is shown by the numbers outside of brackets in Figure 3.

Union

If both considered ontologies are equally weighted as in [Hovy and Nirenburg, 1992], individual concepts in an ontology must be maintained in any merge. However, in our approach we have extensively made use of an ontology-internal concept union. This is a result of the general ontology design principles given in Section 2.3. The clause/PP distinction, for example, which is often a concept discrimination criterion in the *GUM* violates Principle 1 and so this discrimination is not preserved in the Merged Upper Model. Therefore, leaving out of account the clause/PP distinction, identical concepts then amount to 163. The number and distribution of concepts identical after union is shown by the numbers in brackets in Figure 3. 106 concepts are strongly identical and 57 merged concepts are identical with the unions of different *GUM* concepts.

Extension

Extension can be found in both directions. Because of the emphasis we have given to the *EUM*, most of the extensions are *EUM* concepts which extend the *GUM* further. These are 60 concepts, 11 for the *Mental-Process*, 11 *Participants* and 38 others from the *Relational-Process* hierarchy. On the other hand, only 4 German participant concepts have found their way into the Merged Upper Model.

Cross classification

An essential field for cross classification has been avoided by the relaxation of the upper model-grammar interface stated in Principle 3 in Section 2.3. For example, whereas the cross classification discussed for the *Material-Process/Action* hierarchy in Section 2.3 would have cross classified 2 English subconcepts with 5 German subconcepts and their subhierarchies respectively, resulting in 42 merged concepts, 9 concepts are sufficient to cover all distinctions expressed in the *EUM* and the *GUM*.

Summary

Summarizing the merging statistics, strong identity can be found for 41%. If we allow identification of unified concepts, identity can be stated for 63%. About 25% of the merged UM are created by extension, and only 3.6% by cross classification. Beside this, there is a small part of the Merged Upper Model (8%) where the concepts are not created by identification, extension and cross classification, but by preferring *EUM* concepts over *GUM* ones.

3.3 Future work

In the current merging process, we have only looked for identities and differences between the given English and German Upper Models. We did not try to improve the inherent consistency of both, although it became clear during the merge that certain distinctions should be removed and others further developed; these local improvements are detailed in [Henschel, 1993] and will be incorporated in future versions of the Merged Upper Model.

In addition, one of our aims with the Merged Upper Model is to provide a stable basis for further

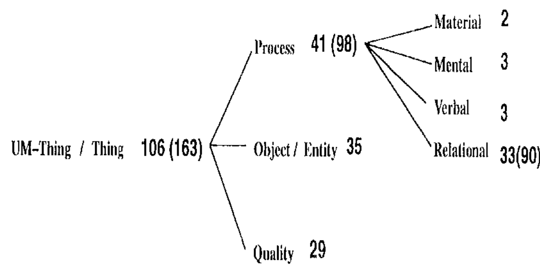

Figure 3: Identity statistics and distribution

extension—both to include further linguistic phenomena and to cover further languages. We expect that an organization of information based on the requirements of natural language grammars will provide a more stable and re-usable result than organizations based on the requirements of individual computational systems. We are already using the Merged Upper Model as the basis for sentence generation in Dutch and there is suggestion here that, again, few additional concepts appear necessary. Of more interest is the extension to rather different languages, some of which has already been begun. Detailed accounts of this work of extension and comparison are necessary since automatic merging will rarely be possible when these most general levels of information organization are considered.

Finally, extensions in future may also be made by comparison with other ontologies—although here it is necessary to be very careful concerning the kinds of ontologies considered. Since the Merged Upper Model is explicitly a linguistically motivated ontology, comparison with ontologies with differing motivations can be difficult. In considering the ontology of the LILOG project, for example, the mixture of linguistic and non-linguistic information criticized by [Lang, 1991] should not be carried over into the merge.

The evaluation of the resulting linguistic ontologies as potential semantic type hierarchies for representations in machine translation, analysis and multilingual generation is then a clear further step.

References

[Bateman *et al.*, 1990] John A. Bateman, Robert T. Kasper, Johanna D. Moore, and Richard A. Whitney. A general organization of knowledge for natural language processing: the PENMAN upper model. Technical report, USC/Information Sciences Institute, Marina del Rey, California, 1990.

[Bateman *et al.*, 1991a] John A. Bateman, Elisabeth A. Maier, Elke Teich, and Leo Wanner. Towards an architecture for situated text generation. In *International Conference on Current Issues in Computational Linguistics*, Penang, Malaysia, 1991. Also available as technical report of GMD/Institut für Integrierte Publikations- und Informationssysteme, Darmstadt, Germany.

[Bateman *et al.*, 1991b] John A. Bateman, Christian M.I.M. Matthiessen, Keizo Nanri, and Licheng Zeng. The re-use of linguistic resources across languages in multilingual generation components. In *Proceedings of the 1991 International Joint Conference on Artificial Intelligence, Sydney, Australia*, volume 2, pages 966 – 971. Morgan Kaufmann Publishers, 1991.

[Bateman, 1992a] John A. Bateman. The theoretical status of ontologies in natural language processing. In Susanne Preuß and Birte Schmitz, editors, *Text Representation and Domain Modelling – ideas from linguistics and AI*, pages 50 – 99. KIT-Report 97, Technische Universität Berlin, May 1992. (Papers from KIT-FAST Workshop, Technical University Berlin, October 9th - 11th 1991).

[Bateman, 1992b] John A. Bateman. Towards Meaning-Based Machine Translation: using abstractions from text generation for preserving meaning. *Machine Translation*, 6(1):1 - 37, 1992. (Special edition on the role of text generation in MT).

[Guarino, 1994] Nicola Guarino. The ontological level. In R. Casati, B. Smith, and G. White, editors, *Philosophy and the Cognitive Sciences*. Hölder-Pichler-Tempsky, Vienna, 1994.

[Halliday and Matthiessen, to appear] Michael A.K. Halliday and Christian M.I.M. Matthiessen. *Construing experience through meaning: a language-based approach to cognition*. de Gruyter, Berlin, to appear.

[Henschel, 1993] Renate Henschel. Merging the English and the German Upper Model. Technical report, GMD/Institut für Integrierte Publikations- und Informationssysteme, Darmstadt, Germany, 1993.

[Horacek, 1989] Helmut Horacek. Towards principles of ontology. In D. Metzing, editor, *Proceedings of the German Workshop on Artificial Intelligence: GWAI89*, pages 323 – 330. Springer-Verlag, Berlin, Heidelberg, New York, 1989.

[Hovy and Knight, 1993] Eduard Hovy and Kevin Knight. Motivating shared knowledge resources: an example from the Pangloss collaboration. In *Proceedings of IJCAI Workshop on Knowledge Sharing and Information Interchange*. International Joint Conference on Artificial Intelligence, 1993.

[Hovy and Nirenburg, 1992] E. Hovy and S. Nirenburg. Approximating an interlingua in a principled way. In *Proceedings of the DARPA Speech and Natural Language Workshop*. Arden House, New York, 1992. Also available from USC/Information Sciences Institute (Marina del Rey, Los Angeles) as Technical Report ISI/RR-93-345, Febuary 1993.

[Kasper, 1989] Robert T. Kasper. A flexible interface for linking applications to PENMAN's sentence generator. In *Proceedings of the DARPA Workshop on Speech and Natural Language*, 1989. Available from USC/Information Sciences Institute, Marina del Rey, CA.

[Lang, 1991] Ewald Lang. The LILOG ontology from a linguistic point of view. In O. Herzog and C.-R. Rollinger, editors, *Text understanding in LILOG: integrating computational linguistics and artificial intelligence, Final report on the IBM Germany LILOG-Project*, pages 464 – 481. Springer-Verlag, Berlin, 1991. Lecture notes in artificial intelligence, 546.

[MacGregor and Brill, 1989] Robert MacGregor and David Brill. The LOOM manual, 1989. USC/Information Sciences Institute, Marina del Rey, CA.

[Patil *et al.*, 1992] Ramesh S. Patil, Richard E. Fikes, Peter F. Patel-Schneider, Don McKay, Tim Finin, Thomas R. Gruber, and Robert Neches. The DARPA knowledge sharing effort: progress report. In Charles Rich, Bernhard Nebel, and William R. Swartout, editors, *Principles of knowledge representation and reasoning: proceedings of the third international conference*, Cambridge, MA, 1992. Morgan Kaufmann.

[Penman Project, 1989] Penman Project. PENMAN documentation: the Primer, the User Guide, the Reference Manual, and the Nigel manual. Technical report, USC/Information Sciences Institute, Marina del Rey, California, 1989.

[Pirlein, 1993] Thomas Pirlein. Reusing a large domain-independent knowledge base. In *Fifth International Conference on Software Engineering and Knowledge Engineering (SEKE'93)*, San Francisco, 1993.

[Steiner *et al.*, 1988] Erich H. Steiner, Ursula Eckert, Birgit Weck, and Jutta Winter. The development of the EUROTRA-D system of semantic relations. In Erich H. Steiner, Paul Schmidt, and Cornelia Zelinksy-Wibbelt, editors, *From Syntax to Semantics: insights from Machine Translation*. Frances Pinter, London, 1988.

[Teich, 1992] Elke Teich. Komet: grammar documentation. Technical report, GMD/Institut für Integrierte Publikations- und Informationssysteme, Darmstadt, West Germany, 1992.

Syntax