Merging the English and the German Upper Model

Renate Henschel

IPSI-GMD
Dolivostrasse 15
D-64293 Darmstadt
Germany

July 1992, July 1993

Abstract. A detailed comparison of the Penman Upper Model and the KOMET German Upper Model is carried out in order to construct a new *Merged Upper Model* capable of serving as the ideational basis for automatic text 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 within the KOMET Project as the basis for generation in English, German and Dutch.

Keywords. Ontology, natural language semantics, text generation, multilinguality.

1 Introduction

The text generation projects PENMAN for English ([Penman Project, 1989]) and KOMET for German ([Bateman *et al.*, 1991]) share the same architecture: Semantic specifications expressed in terms of a knowledge base control the decisions of a systemic generation grammar. The knowledge base consists of a general ontology - the Upper Model (UM) - and a certain number of domain models, the concepts of which are always subsumed by Upper Model concepts. Beside this terminological knowledge base (also called ideation base following Halliday's terminology), it is also planned to include a textual base and an interaction base which will be responsible for textual knowledge and user modelling, respectively. It is an advantage of this architecture that these three bases are disjoint modules.

The PENMAN project is now participating in the multilingual Machine Translation project PANGLOSS ([Frederking *et al.*, 1993]), which uses a knowledge base as interlingua. This requires that all participating analysis and generation components (ISI, CMU, University of New Mexico) are able to operate with one and the same ontology. To this end, [Hovy and Nirenburg, 1992] have proposed a general method for creating a merged ontology out of different ontologies where it does not matter, whether differences are language dependent or due to different linguistic theories. The commonalities and differences in two ontologies can be classified according to [Hovy and Nirenburg, 1992] 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 keeps all concepts of cases (1) and (2) and resolves case (3) by exhaustive cross classification.

In this paper, a detailed comparison of the English Upper Model (EUM) used in the PENMAN project and the German Upper Model (GUM) from the KOMET project is made, and from this a Merged Upper Model - called the *Merge* - is developed.[1] We include in our consideration some more principles concerning ontology design in general which go beyond the merging methods of [Hovy and Nirenburg, 1992] (identification, extension and cross classification). Thus not all existent concepts of the EUM and GUM need find their representation in the Merge. The Merged UM has been tested as knowledge source for

[1]The merged UM was developed in July 1992, the documentation follows now because a sweet little disturber was born in between.

the German, English and Dutch grammars. For that, the interface between the existing grammars and the new UM has been modified accordingly.

This paper is based on the English Upper Model and German Upper Model datafiles from July 1992. Both are expressed in the knowledge representation language LOOM ([MacGregor and Brill, 1989]). The English Upper Model is described in [Bateman *et al.*, 1990]; the concepts of the German Upper Model go back to [Steiner *et al.*, 1988] and [Teich, 1992].

2 Principles for the Merge construction

1. Starting from the paper of [Hovy and Nirenburg, 1992], all concepts and relations in both models are compared, looking for identity, extensions and differences.

2. In the following construction we have given more emphasis to the English Upper Model. The English UM has a longer development tradition and contains a lot of experiences from a team of senior scientists, whereas the German UM was developed in a smaller group, for shorter time and contains a number of more or less ad hoc concepts partly introduced as temporary fillers between the level of abstraction found in the EUM and that found in the available German grammar. These latter concepts have not been referred to in the merging process.

3. Going beyond the merging criteria of [Hovy and Nirenburg, 1992], we assume as a general guideline in constructing the merge that UM concepts should not reflect a particular surface realization of a concept. E.g., concepts are not distinguished with regard to their realization as a prepositional phrase versus a clause, attribute versus predicative clause, verb versus nominalization. As to these distinctions, the generation process should be controlled from a textplanner outside the ideational knowledge base. This is mainly observed in the EUM, whereas the GUM suffers from a overhead of such distinctions.

4. Another general design principle concerns the number of participants. We argue that the number of surface realized participants should not be the criterion for the discrimination of process concepts. Missing surface participants can be modelled more adequately by a UM-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. According to this principle, a lot of GUM concepts can be merged into one because they reflect only a different number of surface participants.

5. A 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 in the UM do not necessarily correspond to disjoint sets of surface sentences - only to disjoint semantic perspectives on them. The interface between the UM and the grammar should be written in such a way, that it is possible in some cases to generate the same sentence from different semantic input. This approach would meet 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 a merging strategy mentioned under 1.), but by giving the UM-grammar interface more flexibility[2].

3 Problems with Identity

The crucial point in [Hovy and Nirenburg, 1992] is the notion of "Identity". The decision how to deal with different concepts (extension, cross classification or identification) is based on the possibility to state an identity between concepts of different language ontologies. This is rather problematic, and we are aware of that. 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. This is not true in general. We based our merging on the assumption that it may be true for simple sentences if we let out the textual and interpersonal dimensions of utterances, and the language distance is close. Hence, the whole construction has to be seen in its own relativity.

4 Construction overview

The differences between the EUM and the GUM mainly concern the ontology of process types. The german UM is more relational-process committed, whereas the EUM more towards material processes. This means that a significant number of sentences which are analyzed within the GUM as *Relational-Process* should be given to the English grammar as *Material-Process*. This is not a language specific difference, but a theory dependent one and reflects the differences between the systemic approach of HALLIDAY [Halliday, 1985] which the EUM is based on and the approach of STEINER and FAWCETT (see [Fawcett, 1980]) which is the base for the GUM. There are four solutions to this problem:

1. Choosing the HALLIDAYan approach,
2. choosing the STEINER approach,
3. creating a lot of new process types by cross classification of both, or
4. making the grammar ambiguous.

The first two solutions would cause the necessity of severe changes in the English or the German grammar. The third solution would create a lot of ill-looking artificial concepts without natural evidence. By the last we mean, the demarcation between relational and material processes is relaxed so that for processes which are subsumed by the overlapping

[2]Giving the UM-grammar interface more flexibility is argued for on other grounds in [Bateman, 1992].

area of these process types the grammar can generate exactly the same sentence from a relational and a material SPL-input (see example 9 in section 5.2). With the following merged UM, we will propose this fourth solution. It has the advantage that the English and the German grammar can remain as they are, and the UM does not need additional concepts in the top level.

The quality and the object hierarchies are nearly identical and can be merged by identification leaving out some ad hoc German concepts.

5 Documentation of the concept merging

In the following discussion all examples are taken from [Bateman *et al.*, 1990, Steiner *et al.*, 1988, Teich, 1992]. Examples are given as sentences which can be generated by UM concepts, or as semantic specifications which function as input to the systemic generation grammar. Semantic specifications are expressed in the sentence planning language SPL defined in [Penman Project, 1989]. For those not familiar with SPL we give an extract from the BNF notation for a sentence plan:

plan := (variable / concept attribute*)
attribute := :relation plan

Concepts and relations are taken from the Upper Model or subordinated domain models.[3]

The UM concepts referred to in running text are emphasized by italics. Sometimes they are prefixed by G- or E- in order to make their affiliation unique.

5.1 Top level

Comparing the highest levels of the German and the English UM (see Figure 1), we see already that the GUM gives more emphasis to the relational process type because it occurs in fact twice, as *G-Relational* and *G-Relationship*. The intended difference here is as follows: The process *G-Relational* is reserved only for relations realized as a sentence, whereas *G-Relationship* stands for attributive realizations (see section 5.6).

We argue that such differences are textually determined and should not be reflected in an ideational semantic component (see merging principle 3). Hence, we prefer the English version here, and will conflate the German concepts *G-Relational* and *G-Relationship* into one concept *Relational*, which subsumes all subhierarchies of *G-Relational* and *G-Relationship* where - as we shall see below - some of these subhierarchies can be conflated too.

With this removal in mind, the correspondences between English and German top level concepts are as follows:

[3]Relations can also be selected from the small set of SPL special keywords, such as `lex` (used below), `theme`, etc.

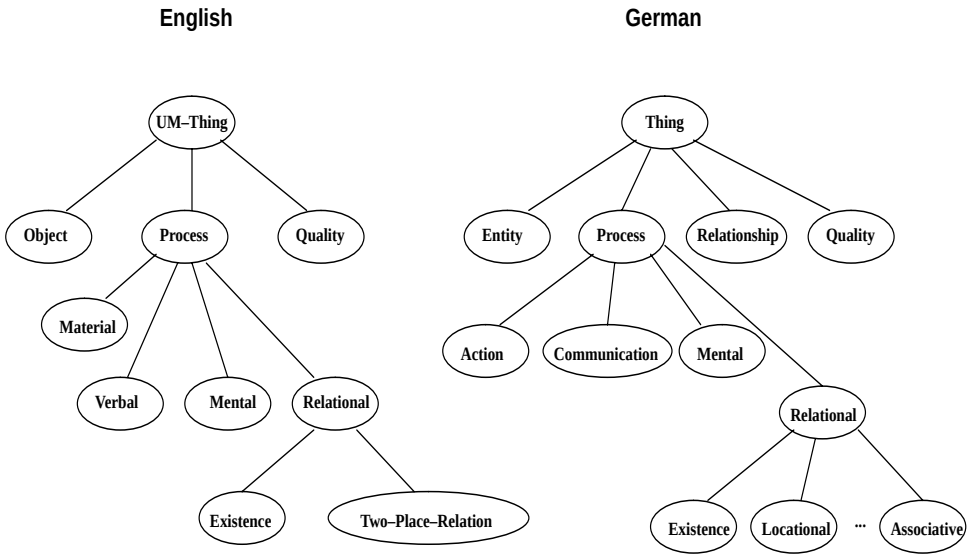


Figure 1: UM top level

English		German	Merge
E-Process	=	G-Process + G-Relationship	Process
E-Object	=	G-Entity	Object
E-Quality	=	G-Quality.	Quality

These identities are not self-evident. They are a consequence of the detailed comparison of the three subhierarchies given in the following subsections.

5.2 Process types

As mentioned above, the main difference between the EUM and the GUM concerns the ontology of process types. The partitioning of the concept *Process* into subconcepts is not identical. The german UM is more relational-process committed, whereas the EUM is more committed towards material processes. Figure 2 gives an overview.

The relations stated in figure 2 can be expressed in terms of set inclusion as follows:

(1) E-Material-Process $\supset$ G-Action

(2) E-Relational-Process $\subset$ G-Relational + G-Relationship

(3) E-Mental-Process = G-Mental

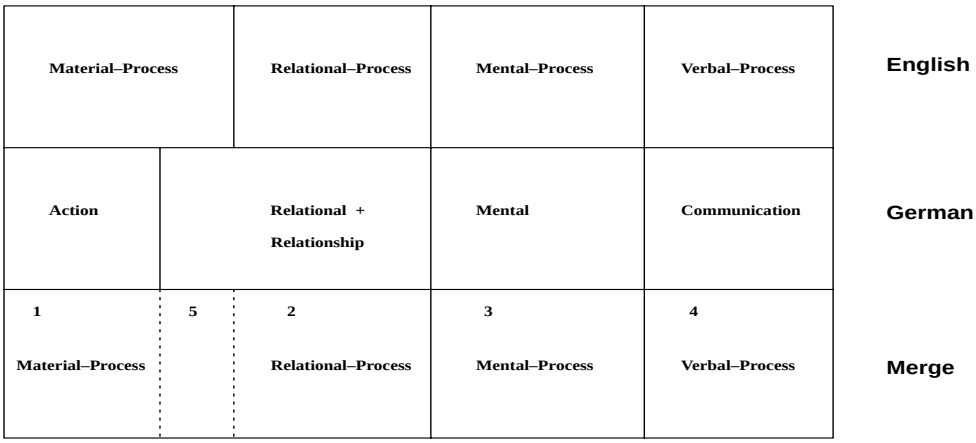


Figure 2: German and English process types

(4) E-Verbal-Process = G-Communication

(5) E-Material-Process $\cap$ G-Relational $\neq \emptyset$

Corresponding examples of the identical kernels of the four process types are:

(1) Der Bauer fällt den Baum.
(The farmer is felling the tree.)

(2) Die Regierung hat kein Geld.
(The government has no money.)

(3) Die Klasse gedachte des verstorbenen Schülers.
(The class remembered the dead classmate.)

(4) Der Präsident erzählte eine Geschichte.
(The president told a story.)

In the merged UM, we choose the EUM terms *Material-Process*, *Relational-Process*, *Mental-Process*, *Verbal-Process* for them.

The more difficult problem is the overlap between *G-Relational* and *E-Material-Process*. In the German grammar, sentences like

(5) Der Hund brachte dem Mann eine Verletzung bei.
(The dog caused the man an injury.)

(6) Der Meister nannte die Frau eine gute Arbeiterin.
(The master called the woman a good worker.)

(7) Der Lehrer bezeichnet den Schüler als fleissig.
(The teacher called the pupil hard-working.)

(8) Der Arzt wohnt in Leipzig.
(The doctor lives in Leipzig.)

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

are analyzed as *G-Relational*, which relates the second and the third participant (counted from the unmarked surface order):

- Associative

  (5') Der Mann hat eine Verletzung.
  (The man has an injury.)

- Classificatory

  (6') Die Frau ist eine gute Arbeiterin.
  (The woman is a good worker.)
  (7') Der Schüler ist fleissig.
  (The pupil is hard-working.)

- Locational

  (8') Der Arzt ist in Leipzig.
  (The doctor is in Leipzig.)
  (9') sein Vater — in die Stadt
  (The father – to the city)

The first participant (the subject in the examples) is modelled by the role *Third-party-agent*, the agent who brings the relation into existence.

The English grammar would analyze them as *E-Material-Process*.

This distinction is not a language dependent one. It reflects the different linguistic theories used in the German and English grammar. As already mentioned in section 4, we propose a flexible solution to this problem. We do not want to create a new fixed partitioning of the *Process* concept preferring one or the other theory, but relax the borders between the overlapping process types (solution four in section 4). That means, that the grammar is able to generate sentences of the case in question such as the examples (5) to (9) by giving them as *Relational-Process* or as *Material-Process* without any differences in the surface

realization. Hence, the grammar covers both perspectives - the HALLIDAYan and the STEINER approach - to a partitioning of sentences into process types. The process type a user chooses depends on what semantic perspective he prefers.

For example, the input SPL for sentence (9) in terms of the German UM would look like[4]:

```
(b / locational   ;--- as a subconcept of relational ---
       :lex begleiten
        :attribuant (v / person :lex vater)
        :location (st / one-or-two-d-location :lex stadt)
        :third-party-agent (s / person :lex sohn))
```

In terms of the Merged UM, sentence (9) can have the following two distinct SPL-representations:

Hallidayan approach

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

Fawcett/Steiner approach

```
(b / destination  ;--- as a subconcept of relational-process
        :lex begleiten
        :domain (v / person :lex vater)
        :range (st / one-or-two-d-location :lex stadt)
        :third-party-agent (s / person :lex sohn))
```

5.3 Material Process

The *E-Material-Process* hierarchy and the *G-Action* hierarchy differ a lot (see figure 3).

Roles

Let us consider at first the relevant participants more theoretically. In the analysis of the English grammar, [Halliday, 1985] distinguishes for material processes two roles with regard to ergativity (medium and agent) and two further roles with regard to the transitivity of a clause (actor and goal).

[4]Lexical selection is specified here directly by means of the :lex keyword to avoid complicating the discussion unnecessarily.

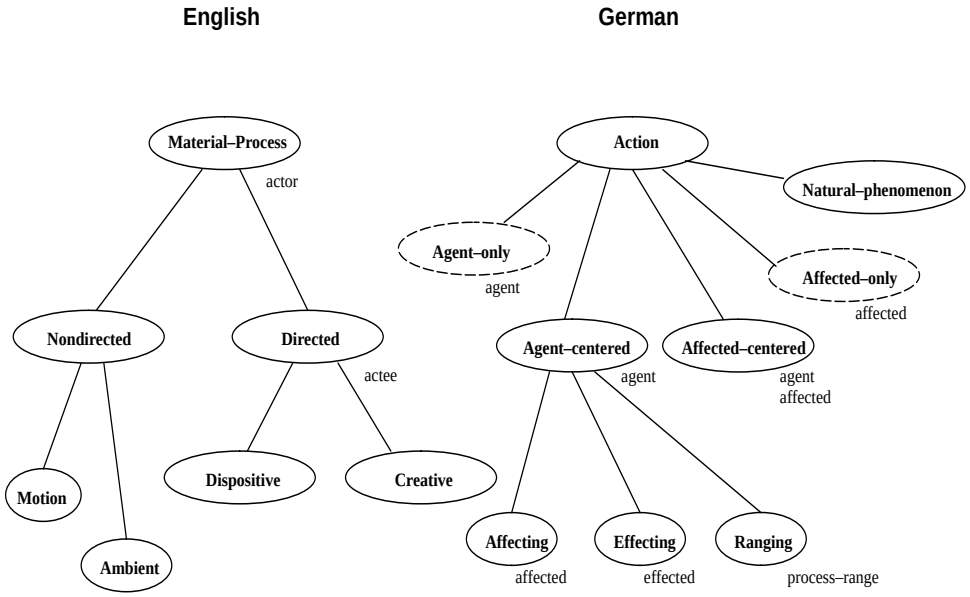


Figure 3: Material processes

The transitivity pattern is the traditional perspective on processes. Every process must have an actor, the participant that "does the deed". Usually the actor is in nominative case. The process can be extended to another participant, the goal. This constitutes the known intransitive - transitive classification.

The ergativity reflects causativity. The medium is the central role critically involved in the process which brings the process into existence. Medium and process form the kernel of the sentence. With the agent, an external causation of the process can be given. How these roles correspond to each other is shown in the examples below:

(10a) The child broke the vase.
actor *goal*
agent *medium*

(10b) The vase broke.
actor
medium

(11a) The lion chased the tourist.
actor *goal*
agent *medium*

(11b) The tourist ran.
actor
medium

The EUM uses the terms *E-Actor* and *E-Actee* for the participants actor and goal.

In the GUM the participant *G-Agent* has a different meaning. Agent is here the causer

too, but more literally the participant which intensionally performs an action. So, the not very evident parallelism of (10b) and (11b) disappears:

(10a) The child broke the vase.
agent *affected*

(10b) The vase broke.
affected

(11a) The lion chased the tourist.
agent *affected*

(11b) The tourist ran.
agent

The second participant can either be *G-Affected* or *G-Effected* or *G-Process-range* following the grammatical functions found in [Fawcett, 1980]. The *G-Affected* is the participant affected by the action, the *G-Effected* is created by the action, and the *G-Process-range* neither nor. Process-range is used for NPs which belong more to the process itself than playing the role of an participant. This participant type has an equivalent in the subtype of *E-Actee E-Process-Range*. The examples (21) to (23) below exemplify these participant types.

From the examples (10) and (11), it is already obvious that the English and German participants could not be identified. How to deal with this in the Merge will be clarified at the end of this section.

Crossing concepts

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

(12) The vase broke. *Nondirected-Action*
(13) I play piano. *Nondirected-Action*
(14) The tourist ran. *Motion-Process*
(15) It rains. *Ambient-Process*

An *E-Directed-Action* in contrast is a process with an external causer as additional participant (examples 10a, 11a). *E-Directed-Actions* divide into *E-Creative-Material-Action* and *E-Dispositive-Material-Action*.

(16) The child broke the vase. *Dispositive-Material-Action*

(17) The lion chased the tourist. *Dispositive-Material-Action*
(18) Mary baked a cake. *Creative-Material-Action*

Concerning the *G-Action* process type, the current GUM is not elaborated to the extent which the German grammar presupposes. So, in Figure 3 the concepts presupposed by the grammar and missing in the GUM are shown by broken ellipses. We will include them in our discussion. Also, the roles *G-Effected* and *G-Process-range* are missing in the GUM, but necessary to the grammar.

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 intransitive process types are given here in SPL notation:

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

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

(20) 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] as follows:

X *affected-centered-verb* Y iff X causes that Y *affected-centered-verb*

Das Kind zerbricht die Vase. <—> Das Kind bewirkt, dass die Vase zerbricht.
The child brakes the vase. <—> The child brings it about that the vase brakes.

So 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

  (21) Der Bauer fällt den Baum.
  (The farmer is felling the tree.)
  G-Agent *G-Affected*

- Effecting

  (22) Die Mutter malt ein Haus.
  (The mother is painting a house.)
  G-Agent *G-Effected*

- Ranging

  (23) Ich spiele Klavier.
  (I play piano.)
  G-Agent *G-Process-range*

We will not consider other phenomena which are addressed by Steiner's definition of range because they are not covered by both the German grammar and the English grammar so far.

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

$$E\text{-}Ambient\text{-}Process = G\text{-}Natural\text{-}Phenomenon,$$

and could mechanically build a cross-classification as shown in Figure 4. However, in this cross product ontology, some created concepts can be omitted. The first obvious argument is the number of participants which is contradictory in the following cross concepts:

$$E\text{-}Directed\text{-}Action/\, G\text{-}Agent\text{-}only \quad \text{and}$$
$$E\text{-}Directed\text{-}Action/\, G\text{-}Affected\text{-}only.$$

A comparison of the low level concepts shows that the following can be identified:

$$E\text{-}Dispositive\text{-}Material\text{-}Action = G\text{-}Affecting + G\text{-}Affected\text{-}centered \quad \text{and}$$
$$E\text{-}Creative\text{-}Material\text{-}Action = G\text{-}Effecting.$$

This rules out the cross concepts

$$E\text{-}Dispositive\text{-}Material\text{-}Action/\, G\text{-}Effecting,$$
$$E\text{-}Dispositive\text{-}Material\text{-}Action/\, G\text{-}Ranging,$$
$$E\text{-}Creative\text{-}Material\text{-}Action/\, G\text{-}Affected\text{-}centered,$$
$$E\text{-}Creative\text{-}Material\text{-}Action/\, G\text{-}Affecting,$$
$$E\text{-}Creative\text{-}Material\text{-}Action/\, G\text{-}Ranging\ .$$

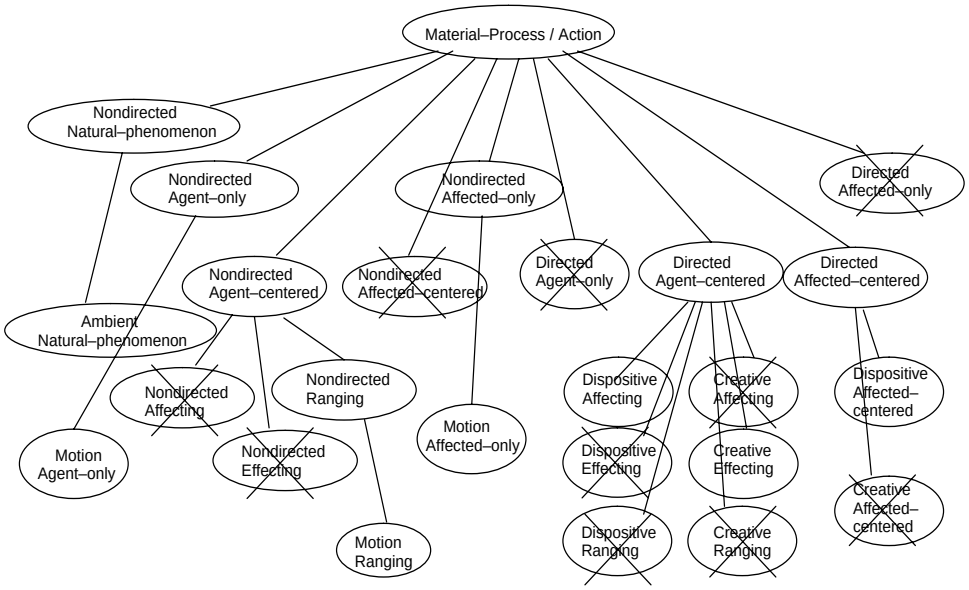


Figure 4: Mechanical merge of the material processes by cross classification

From the definition of *E-Nondirected-Action* in [Bateman *et al.*, 1990], it is known 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.

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* and

E-Nondirected-Action/ G-Agent-centered $=$ *E-Nondirected-Action/ G-Ranging*,

E-Directed-Action/ G-Affected-centered
$=$ *E-Dispositive-Material-Action/ G-Affected-centered*,

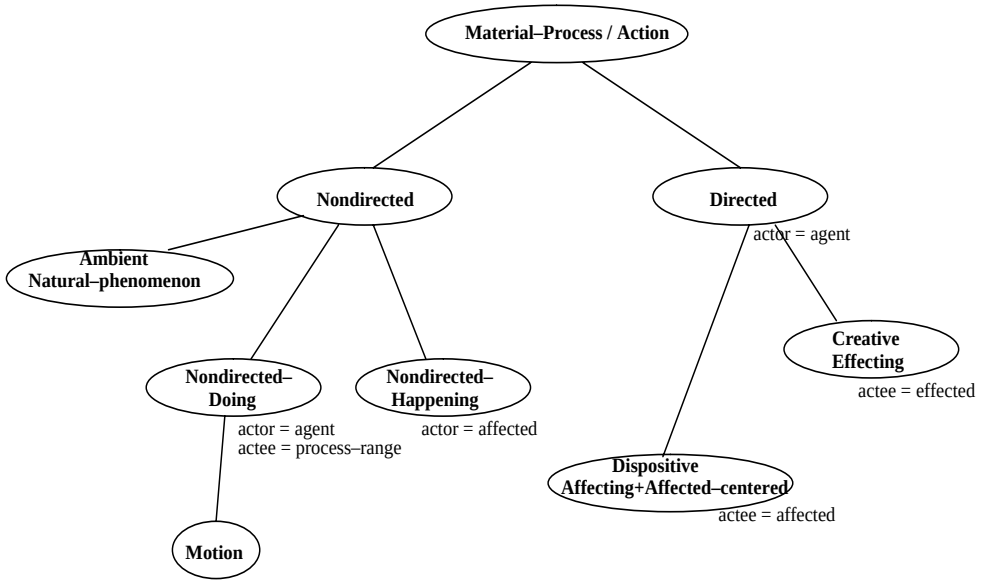


Figure 5: Merging proposal for the material process type

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

Thus we have filtered an intelligent merge out of the mechanical merge.

Within the intelligent merge, we will omit the german differences concerning the participant number (*G-Agent-only*, *G-Ranging*), and not establish the very subtle *G-Affected-centered* type. Preferring the english terminology the result is given in Figure 5.

This is mainly the EUM subhierarchy for material processes. To meet the German approach, the *Nondirected-Action* concept is differentiated into *Nondirected-Doing* and *Nondirected-Happening* according to the distinction between *Agent-only* and *Affected-only*. So 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 (19) and (20) than have the following form:

```
(19')  (r / nondirected-doing
          :lex rennen
          :actor (t / tourist))

(20')  (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

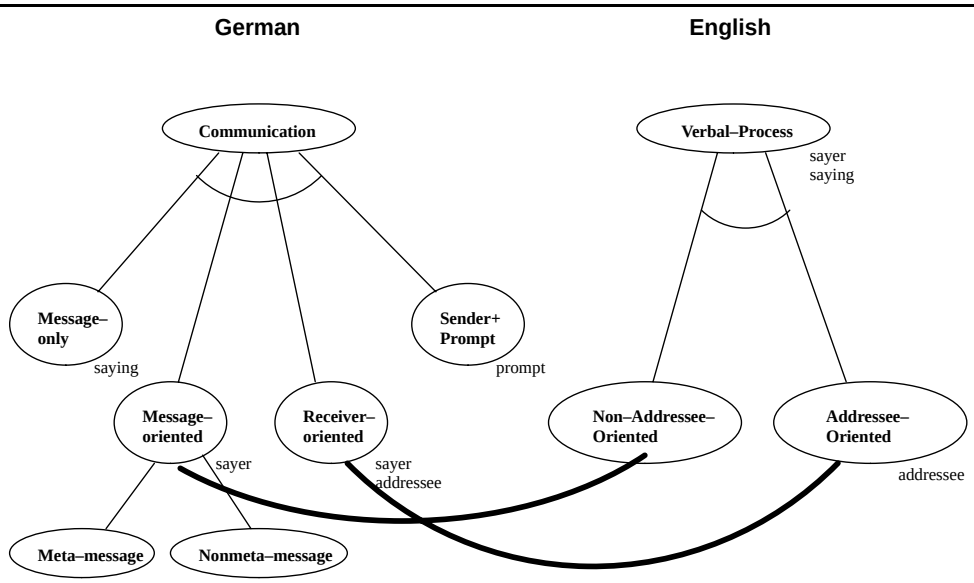


Figure 6: Verbal Process

again. Hence, we will 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 5).

5.4 Verbal Process

Crossing concepts

The main subtype differentiation criterion in both ontologies (see Figure 6) seems to be the presence of a receiver/addressee. More precisely, the possible surface realization of the addressee as a direct object is the base for considering a process as *Addressee-oriented* in English as well as in German. So, *E-Non-Addressee-Oriented* and *G-Message-Oriented*, and *E-Addressee-Oriented* and *G-Receiver-Oriented* mainly correspond to each other. As always, the GUM makes finer distinctions with respect to the number and kind of participants. We will remove these special types *G-Message-Only* and *G-Sender+Prompt* for the merged UM, for they can be treated as subtypes of *Non-Addressee-Oriented* and do not need an extra conceptual reflection in the UM (see principle 4).

Hence, the following correspondences are proposed:

English		German	Merge
E-Verbal-Process	=	G-Communication	Verbal-Process
E-Non-Addressee-Oriented	=	G-Message-Oriented + G-Message-Only + G-Sender+Prompt	Non-Addressee-Oriented

E-Addressee-Oriented = G-Receiver-Oriented Addressee-Oriented

The examples below show how the removed German concepts can be expressed in terms of the Merge:

- G-Message-only

  (24) Es heisst, dass das Raumschiff zurückgekehrt ist.
  (It is said that the spaceship has returned.)

```
(h / non-addressee-oriented
        :saying (z / motion-process
        :lex zurückkehren
        :actor (r / raumschiff)))
```

  The necessity of an impersonal construction can be inferred from the missing *Sayer* participant.

- G-Sender+Prompt

  (25) Der Arzt antwortete dem Patienten auf seine Frage.
  (The doctor answered the patient's question.)

```
(a / non-addressee-oriented
        :lex antworten
        :sayer (a /arzt)
        :addressee (p / patient)
        :prompt (f / frage))
```

  The existence of the *Prompt* role is enough evidence for the correct generation of this transitivity type.

Roles

Both UMs distinguish the participants *Sayer*, *Saying* and *Addressee* which can be stated as identical in terminology and meaning. The *E-Saying* subtype *E-Quote* for information being communicated by a direct quotation can extend the German UM in a suggestive way and is adopted for the merged UM.

The participant role *G-Prompt* covers an additional kind of participant in *G-Communication* which is not reflected in the EUM but is equally necessary for English. The following examples from [Steiner *et al.*, 1988] illustrate this participant:

(26) Der Arzt antwortete dem Patienten auf seine Frage.
sayer *addressee* *prompt*
(The doctor answered the patient's question.)

(27) Der Schüler entgegnete auf die Frage, dass er krank gewesen sei.
sayer *prompt* *saying*
(The school boy responded to the question that he was sick.)

The concepts *G-Meta-message* and *G-Nonmeta-message* are introduced to differentiate the *Saying* participant further.

- Meta-message

  (28) Der Präsident erzählt eine Geschichte.
  (The president told a story.)

  (29) Sie bezichtigt den Nachbarn der Lüge.
  (She accused the neighbour of lying.)

- Nonmeta-message

  (30) Der Messegast bedankt sich beim Polizisten, dass er ihm geholfen hat.
  (The exhibition visitor thanked the policeman for helping him.)

  (31) Luis Trenker schlug Hans den Aufstieg vor.
  (Luis Trenker suggested the climb to Hans.)

This differentiation holds for all *G-Communication* processes, not only for *G-Mesage-oriented* ones. So, it would be more adequate to put these concepts under *Saying* in the *Participant* ontology, and not as subconcepts of *G-Message-oriented*.

The English translations of the given examples demonstrate the usefulness of the German-only concepts within the English grammar, also. Hence, we extend on the EUM and include them in the Merge (see Figure 6).

5.5 Mental process

Figure 7 shows the German and the English ontologies for mental processes. The proposed merge is shown by cross connections between them.

Crossing concepts

There are two main subclasses *G-Processor-oriented* and *G-Phenomenon-oriented* on the German side and *E-Mental-inactive* and *E-Mental-active* which could nearly identified with each other. *Mental-inactive* is defined to describe so-called inactive mental processing as "see", "hear", "taste", "smell", "think", "believe", "know", "realize", "want", "fear", "desire", "like". *Mental-actives* are a kind of reverse *Mental-inactives*. Examples are "please", "frighten", "amaze", "attract". It turns out that in unmarked realizations of *Mental-inactives* the *Senser* role is realized as subject:

(32) Henry likes Bananas.

whereas in *Mental-Actives* as object:

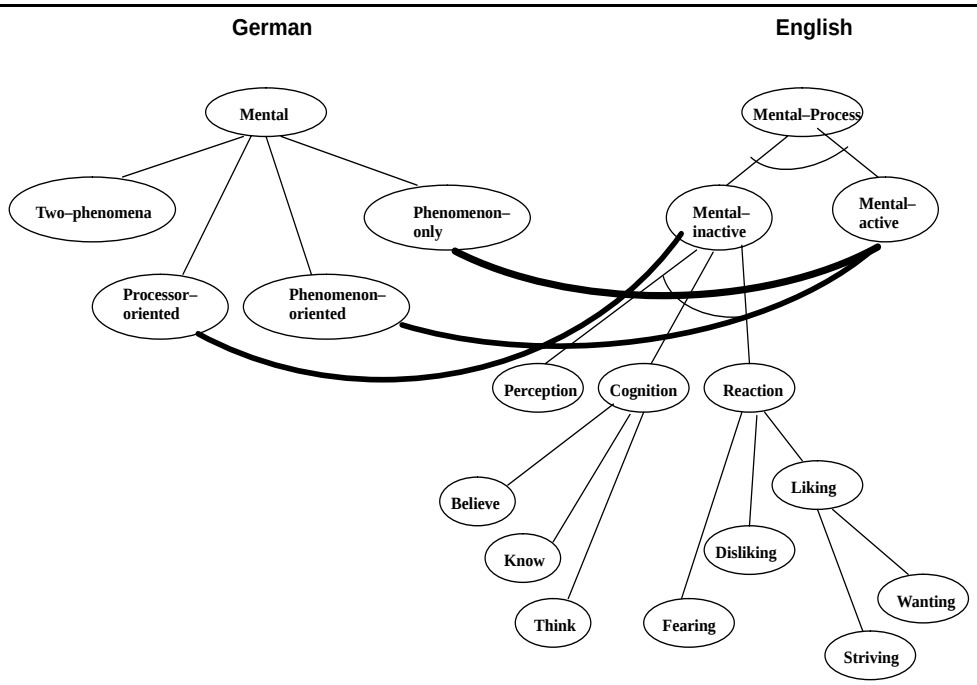


Figure 7: Mental Process

(33) Bananas please Henry.

This is exactly the difference between the German concepts *G-Processor-oriented* and *G-Phenomenon-oriented* as defined in [Steiner *et al.*, 1988]. The following examples from [Steiner *et al.*, 1988] and [Teich, 1992] exemplify them:

- Processor-oriented

  (34) Hans denkt nach.
  (Hans is thinking.)

  (35) Hans mag das Buch.
  (Hans likes the book.)

  (36) Die Klasse gedachte des verstorbenen Schülers.
  (The class remembered the dead classmate.)

- Phenomenon-oriented

  (37) Die Geschichte beeindruckte mich.
  (The story impressed me.)

The English *Mental-Process* has the roles *E-Senser* and *E-Phenomenon* which correspond exactly to the German mental process roles *G-Processor* and *G-Phenomenon*.

With regard of the number of roles, the German ontology differentiates between subtypes for a phenomenon-oriented process. *G-Phenomenon-oriented* stands for a mental process

with two expressed roles, the *Phenomenon* and the *Processor*, whereas *Phenomenon-only* is the semantic reflection of mental processes with only the *Phenomenon* role, e.g.

(38) Es geht um eine wichtige Frage.
(It concerns an important question.)

We have argued that role number distinctions should not be covered by the UM. Such information can easily be inferred from the actual SPL input expression. There is no need to cause an extra concept differentiation.

These considerations lead to the following concept merging:

English		German	Merge
E-Mental-Active	=	G-Phenomenon-oriented + G-Phenomenon-only	Mental-Active
E-Mental-Inactive	=	G-Processor-oriented	Mental-Inactive

English-only concepts

All the English concepts which elaborate the *Mental-Process* hierarchy further are useful for the German grammar too. Hence they should also be components of the merged UM.

English concept	German examples
E-Perception	sehen, hören, fühlen, schmecken, riechen
E-Cognition	
E-Think	nachdenken, denken
E-Know	wissen, verstehen
E-Believe	glauben, denken
E-Reaction	
E-Disliking	hassen
E-Fearing	fürchten, sich fürchten, befürchten
E-Liking	
E-Wanting	wollen, möchten
E-Striving	mögen, gern haben

German-only concepts

We will exclude the concept *G-Two-phenomena* from our discussion because it has no realization in the German grammar until now.

5.6 Relational Process

In this section, we consider the correspondences between the subhierarchies below *E-Relational-Process* on one side and below *G-Relational* and *G-Relationship* on the other

side. *G-Relational* and *G-Relationship* are both responsible for the generation of relational processes; they differ with regard to the realized rank. Roughly, *G-Relationship* causes an attributive or adverbial realization, *G-Relational* a clausal realization. So, the phrases (39a) and (39b) can only be generated from different SPL input:

(39a) Das Mädchen ist krank.
(The girl is sick.)

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

(39b) das kranke Mädchen
(the sick girl)

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

Because we reject reflecting the difference "clausal realization versus group realization" within our ontology, each *G-Relational* concept is in the following merged with a corresponding concept out of the *G-Relationship* hierarchy into one common concept which now can be identified with a corresponding concept in the EUM.

G-Relationship mirrors more or less the terminology of *E-Relational-Process* with some differences in the distribution of concepts within the *Logical* and *Causal* subhierarchies. *G-Relational* includes a number of special GUM relations for clause generation shown in Figure 9 which have to mapped to their group counterpart in the *G-Relationship* hierarchy.

The main one-to-one correspondences are given below and are shown graphically in Figure 8:

E-Existence = G-Existence

E-Intensive = G-Identifying + G-Classificatory + G-Feature-Ascription

E-Generalized-Possession = G-Associative + G-Generalized-Possession

E-Circumstantial = G-Cirumstantial + G-Locational[5]

E-Logical = G-Logical[6]

Roles

The assumed roles for relational processes in the EUM are *E-Domain* and *E-Range* representing the first and the second participant in every relation. This is mirrored by the

[5] + G-Purpose + G-Cause + G-Consequence + G-Scope + G-Generalized-Means
[6] without G-Purpose, G-Cause, G-Consequence, G-Scope + G-Generalized-Means

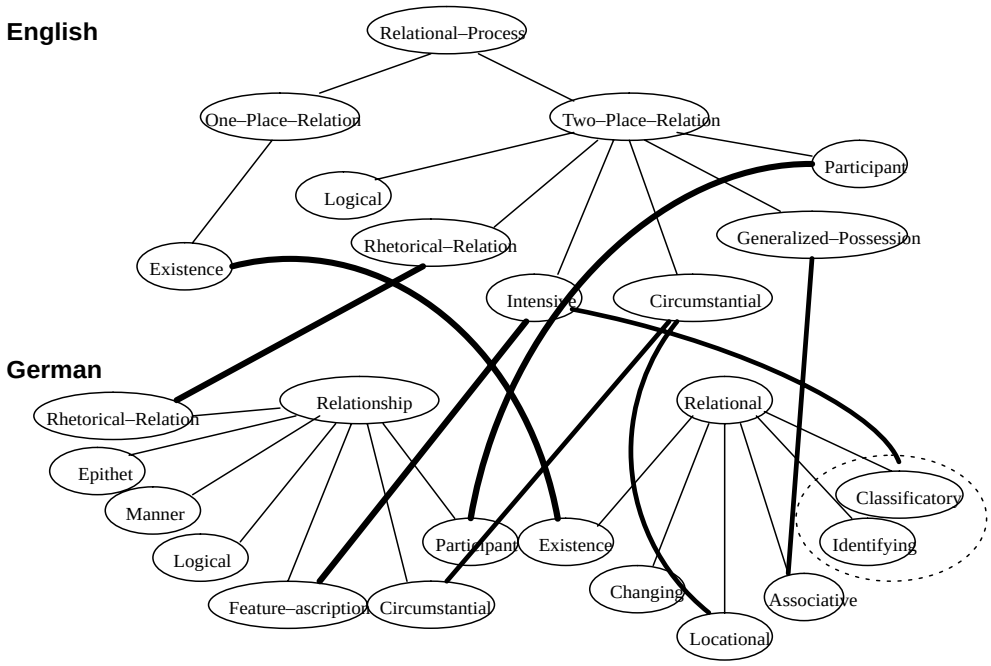


Figure 8: Relational Process

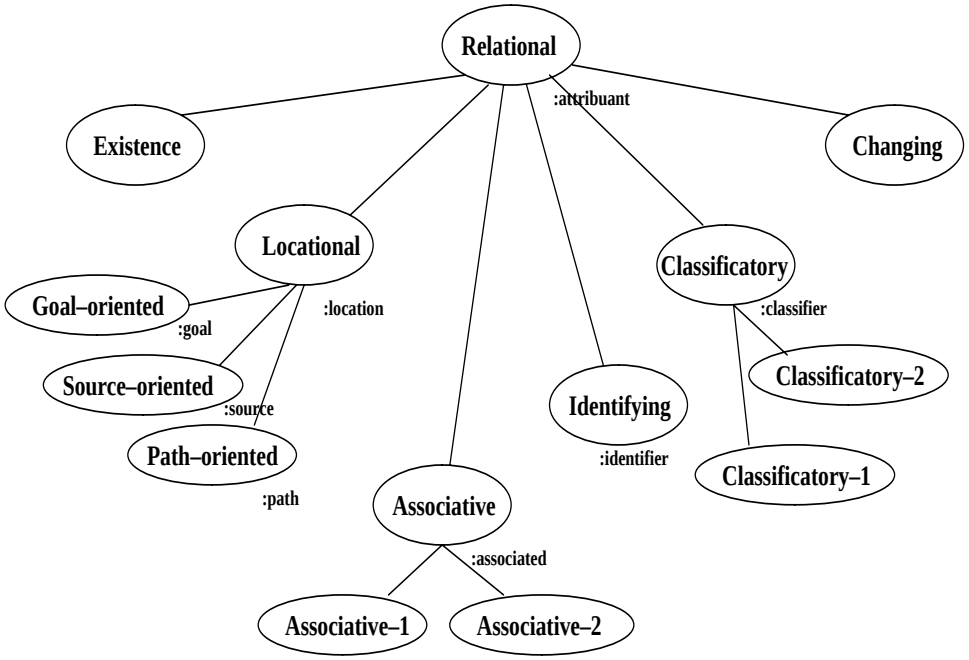


Figure 9: German Relational Process Types

process type *G-Relationship* in the GUM, whereas the *G-Relational* has the only-German participants *G-attribuant* for the first participant, a number of different roles for the second participant, and *G-Third-party-agent* for a possibly involved third participant. In putting together the *G-Relationship* with *G-Relational*, the first participants will be expressed only by *Domain* and the highly differentiated second participant roles will be summarized by *Range*. A preservation of the German differentiation is not necessary because it is already expressed by the different process subtypes. The precise correspondences can be seen in Figure 15.

Intensive Relation

The *E-Intensive* subhierarchy has been adapted for the *G-Feature-ascription* hierarchy and is therefore terminologically identical (see Figure 10). However, with the distinction between *G-Relational* and *G-Relationship* mentioned above, the *G-Feature-ascription* covers only the generation of attributive PPs and adjectives whereas feature ascribing clauses are described with the German process types *G-Identifying* and *G-Classificatory* . This is a kind of doubling, which can be removed in the merged UM. The following correspondences can be stated:

E-UM-Identity = G-Identifying + G-Identity[7]

E-Class-Ascription = G-Classificatory-1 + G-Class-Ascription

E-Property-Ascription = G-Classificatory-2 + G-Property-Ascription

The distinction between *E-Class-Ascription* and *E-Property-Ascription* is defined by the filler of the *E-Range*, the second participant. In *E-Class-Ascription* the range is restricted to be filled by an *E-Object*, in *E-Property-Ascription* by an *E-Quality*. The German UM reflects the same distinction with the subtypes of *G-Classificatory*. The examples below show how the German concepts can be substituted by English ones.

- Identifying

  (40) Der Lehrer ist der Chef.
  (The teacher is the boss.)

```
GERMAN    (b / Identifying
                  :attribuant (l / person :lex lehrer)
                  :identifier (ch / person :lex chef)


MERGE     (b / UM-Identity
                  :domain (l / person :lex lehrer)
                  :range (ch / person :lex chef)
```

[7]The *G-Identity* which denotes the same phenomenon as *G-Identifying* is not used by German the grammar.

- Classificatory-1

  (41) Hans ist ein Lehrer.
  (Hans is a teacher.)

```
GERMAN    (b / Classificatory-1
                  :attribuant (h / person :name hans)
                  :classifier (l / person :lex lehrer)


MERGE     (b / Class-Ascription
                  :domain (h / person :name hans)
                  :range (l / person :lex lehrer)
```

- Classificatory-2

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

```
GERMAN    (b / Classificatory-2
                  :attribuant (s / person :lex mädchen)
                  :classifier (k / quality :lex krank)


MERGE     (b / Property-Ascription
                  :domain (s / person :lex mädchen)
                  :range (k / quality :lex krank)
```

From the German only concepts – broken ellipses in Figure 10 – the *G-Negative-feature-ascription* can be removed as a ad hoc concept not used anymore in the grammar. *G-Quantity-selection* and *G-Measure-ascription* can also play a role for the English grammar and appear in SPL specifications. Without loss of generative power the same status can be given to them in the German grammar.

Possession

Possession in both broader and narrower senses is dealt with in the EUM with the help of the concepts *E-Generalized-Possession* and *E-Generalized-Possession-Inverse* and its subhierarchies (see Figure 11). The inverse concept represents the same semantic relation between possessor and possessed, only the domain and range fillers are reversed. So, *E-Generalized-Possession* represents processes like "to own / besitzen", "to consist of / bestehen aus". The *E-Generalized-Possession-Inverse* stands for "to be owned by / gehören" or "to be a component of / ein Teil sein von". The GUM provides the concept *G-Associative* for all processes dealing with possession and possession changes. The difference in the surface realization which in the EUM is represented by the standard and the inverse concept is covered in the GUM by the subtypes *G-Associative-1* and *G-Associative-2*. Hence,

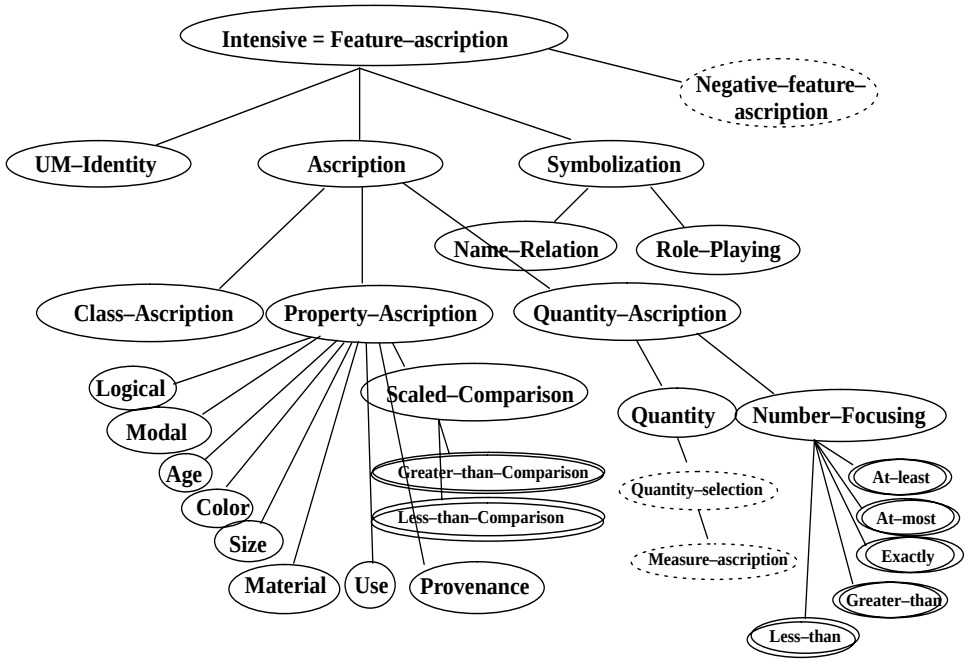


Figure 10: Intensive Relation

we have here clear one-to-one correspondences. Some of the finer concepts in the English possession hierarchy are missing in the GUM, but could be useful for the German grammar as the following examples show:

- E-Ownwership

  (43) Die Regierung hat kein Geld.
  (The government has no money.)

- E-Owned-By

  (44) Das Geld gehört der Regierung.
  (The money belongs to the government.)

- E-Name-Of

  (45) Das Schiff heisst Knox.
  (The ship is called Knox.)

- E-Part-Of

  (46) Schiffe sind Teile einer Flotte.
  (Ships are parts of a fleet.)

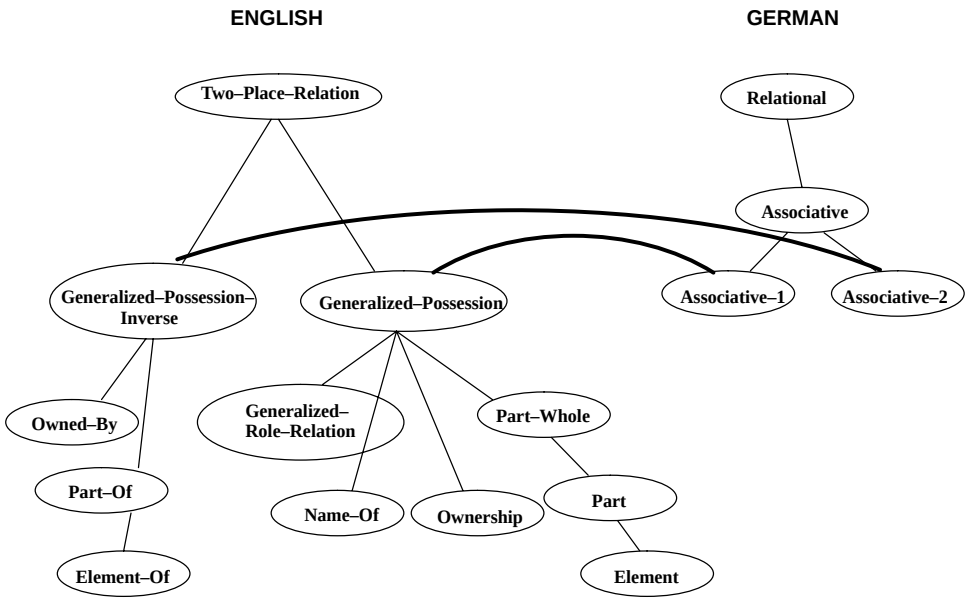


Figure 11: Possession Relations

Circumstances

The demarcation of the *Circumstantial* ontology and the *Logical* ontology is not very precise in both UMs. The *Circumstantial* hierarchy should contain relations between a process and an object, the *Logical* hierarchy between process and process. So it is open to question how the *Causal-Relation*s are to be classified. For the time being, we follow the English version which puts *Causal-Relation*s under *Circumstantial*. For a comparison it is necessary to compare the *Logical* and the *Causal-Relation* hierarchies simultaneously (see Figure 12). 7 concepts form the *E-Logical* ontology. They all have counterparts in the *G-Logical* hierarchy. 6 of them bear identical terms, and the *E-Logical-Condition* can easy be identified with *G-Condition*. The *G-Logical* hierarchy in addition to that contains a number of other relations which correspond to English concepts not from the *E-Logical* hierarchy. Direct identity can be stated for

G-Generalized-Means = E-Generalized-Means,

G-Scope = E-Specific-Matter.

The concept *G-Consequence* was originally defined to control the generation of purpose subclauses. Here again a textual distinction, the distinction between purpose adverbial nominal groups and adverbial subclauses, was taken as the motivation of the distinction between *G-Purpose* and *G-Consequence*. In the merged UM, both realizations are covered by the *Purpose* concept. The same is the case with *G-Cause* and *G-Reason* which are therefore conflated into one concept. *G-Cause* controls the clausal realization of a cause, *G-Reason* the realization as prepositional group.[8]

[8] An analogous problem seems to be the English concept *Cause-Effect* with its RST subtypes.

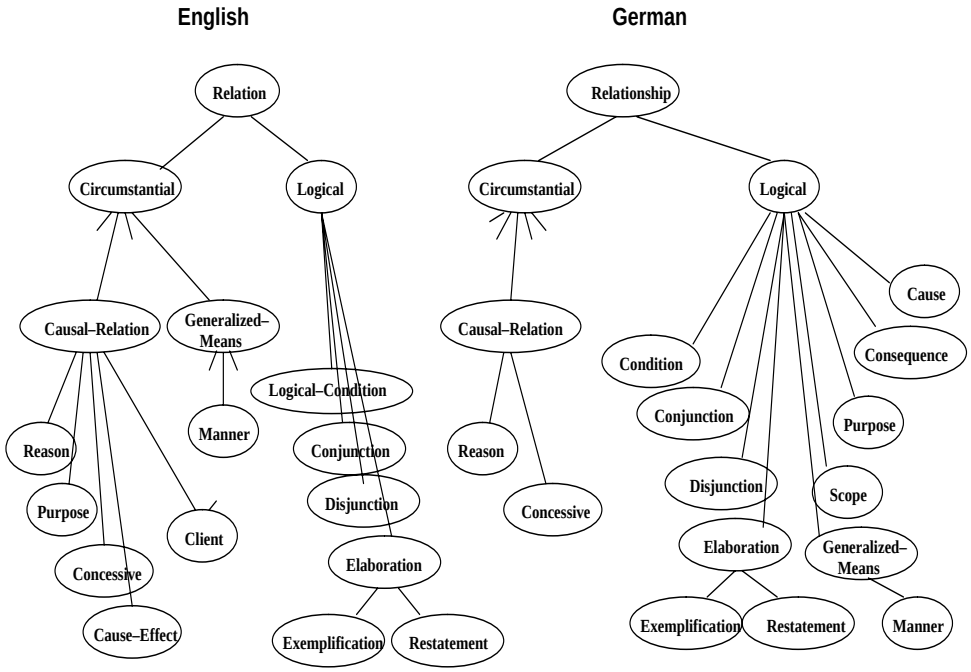


Figure 12: Causal and Logical Relations

We obtain for the shared *Causal-Relation*s

E-Reason = G-Reason + G-Cause,
E-Concessive = G-Concessive,
E-Purpose = G-Purpose + G-Consequence.

Spatio-Temporal Ontology

Within the *Circumstantial* hierarchy, the main subhierarchy is the spatio-temporal complex. It is present with nearly the same structure in both UMs. 39 concepts have identical names and meaning. In Figure 13 the 11 English-only concepts are marked with double circles, the 8 German-only concepts with broken circles.

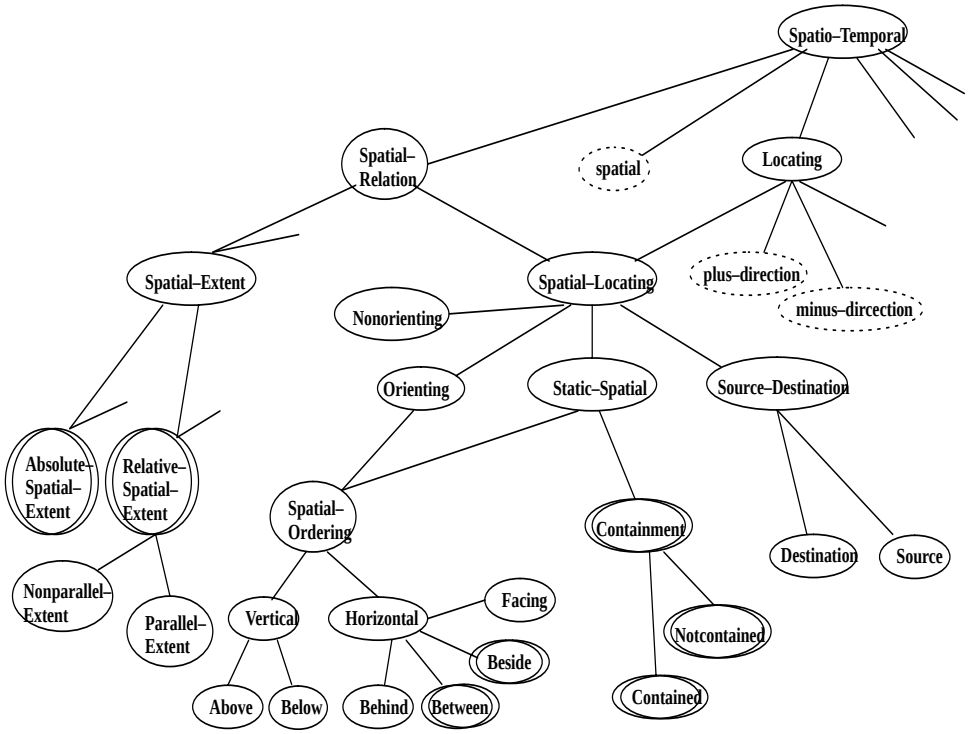
English-only concepts

The English-only concepts

E-Absolute-Spatial-Extent,
E-Absolute-Temporal-Extent,
E-Relative-Spatial-Extent,
E-Relative-Temporal-Extent,
E-Anterior-Extremal

are concepts which inherit from two different superconcepts. They can be included into

(a) Spatial Ontology



(b) Temporal Ontology

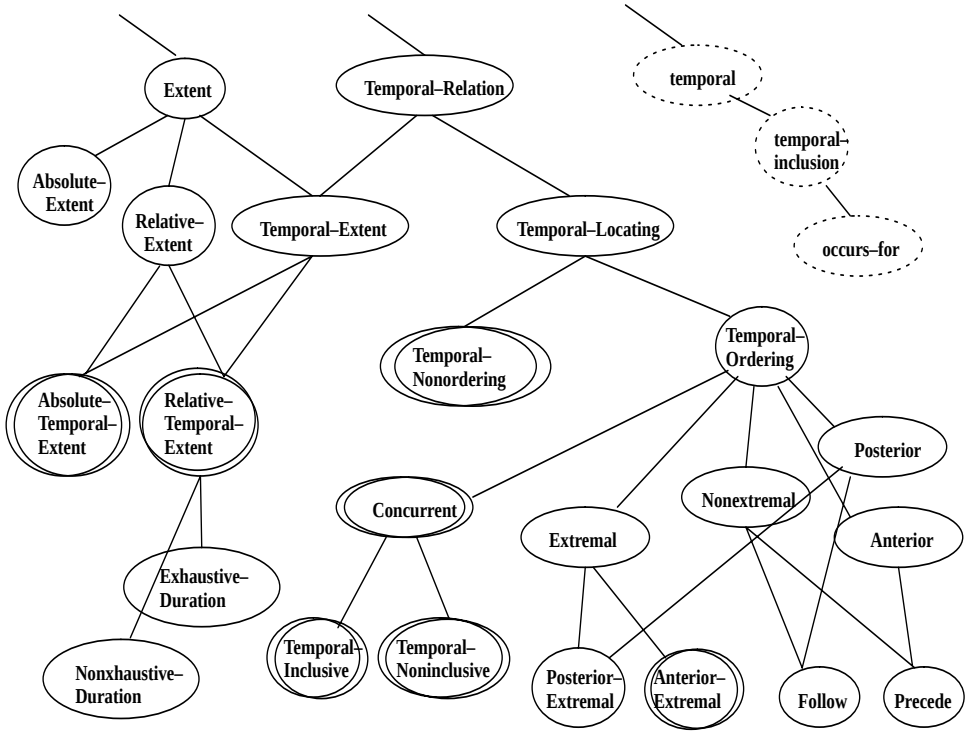


Figure 13: Spatio-Temporal Ontology

the merged UM without difficulties, and even without specializing it further since they represent nothing else than the conjunction of their superconcepts.

The concepts *E-Between* and *E-Beside* are missing in the GUM, as well as *E-Concurrent*, *E-Temporal-Inclusive* and *E-Temporal-Noninclusive*. *E-Temporal-Inclusive* has an identical counterpart in the concept *G-Temporal-inclusion*. The other concepts can be included into the merged UM by making the shared spatio-temporal hierarchy more delicate. They are mostly useful for generation in German also.

E-Between	(47) zwischen den Türmen (between the towers)
E-Beside	(48) neben dem Turm (beside the tower)

German-only concepts

The following German concepts

G-Spatial,
G-Temporal,
G-Temporal-inclusion,
G-Occurs-for,
G-Follows,
G-Precedes

seem to be ad hoc concepts introduced for the moment and not used by the current German grammar. This gives us reason to leave them unrepresented in the merged UM.

G-Spatial and *G-Temporal* are even defined twice in the GUM, one time in the relational hierarchy, another time in the object hierarchy. We suggest to leave them out here, and to keep them in the object hierarchy. *G-Follows* and *G-Precedes* may correspond to the more deeply embedded English concepts *E-Follow* and *E-Precede*.

The intension of *G-Temporal-inclusion* is represented in the EUM by the concept *E-Temporal-Inclusive*.

Extended participant relations

Beside the causal and spatio-temporal relations, some other concepts for wider participants have their place in the circumstantial ontology. Figure 14 shows these concepts for the EUM and their correspondences in the GUM. The corresponding concepts all bear identical names except for *E-Specific-Matter* which is equivalent to *G-Scope* placed under G-Logical. *E-Manner* has its correspondence in *G-Manner* which is placed in the GUM under G-Logical. For the Merge, we prefer in both cases the EUM version.

The English only concepts generally represent useful extensions of the GUM. The following examples give some illustrations.

- Inclusive

  (49) Johann ging mit Maria spazieren.
  (John went for a walk with Mary.)

- Exclusive

  (50) Johann ist ohne Maria spazieren gegangen.
  (John went for a walk without Mary.)

- Alternative

  (51) Johann trug seine blauen Schuhe anstelle seiner weissen.
  (John went with his blue shoes instead of his white ones.)

- Additive

  (52) Johann ging spazieren wie Maria auch.
  (John went for a walk as well as Mary.)

- Agentive

  (53) Der Vater ist von seinem Sohn in die Stadt begleitet worden.
  The father was accompanied by his son to the city.)

- Enablement

  (54) Heinrich löste das Problem durch harte Arbeit.
  (Henry solved the problem by hard work.)

- Instrumental

  (55) Johann schlug den Nagel mit einem Hammer.
  (John hit the hammer with a nail.)

- Similarity

  (56) Heinrich ähnelt Johann.
  (Henry resembles John.)

- Difference

  (57) Heinrich unterscheidet sich von Johann.
  (Henry differs from John.)

5.7 Participants

The concepts for participants reflected in the EUM and the GUM have already been considered in detail in the corresponding sections about the different process types. The identities found are summarized in Figure 15.

Some more marginal participants from the EUM (not discussed until now) make sense for German too:

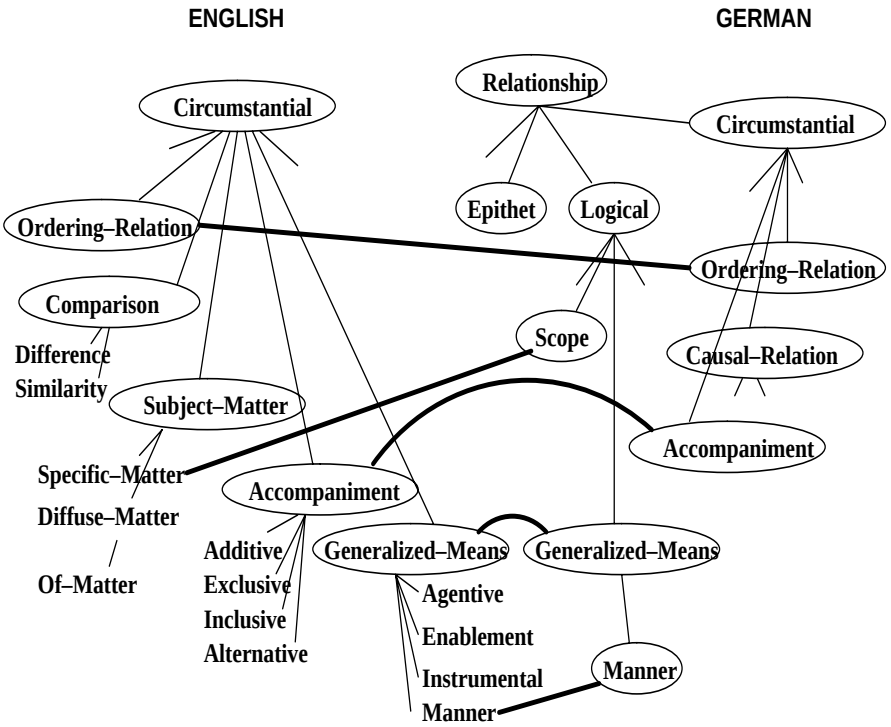


Figure 14: Extended Participant Relations

- Beneficiary

  (58) Meine Tante gab dem Farmer eine Ente.
  (My aunt gave the farmer a duck.)

- Client

  (59) Schenk mir ein kühles Bier ein!
  (Pour me out a cold beer.)

- Material

  (60) Er baute das Haus aus Holz.
  (He built his house of wood.)

- Standard – Attribuend – Compare-Quality

  (61) Hans ist zwei Jahre älter als Thomas.
  Attribuend Compare-Quality Standard
  (John is two years older than Tom.)

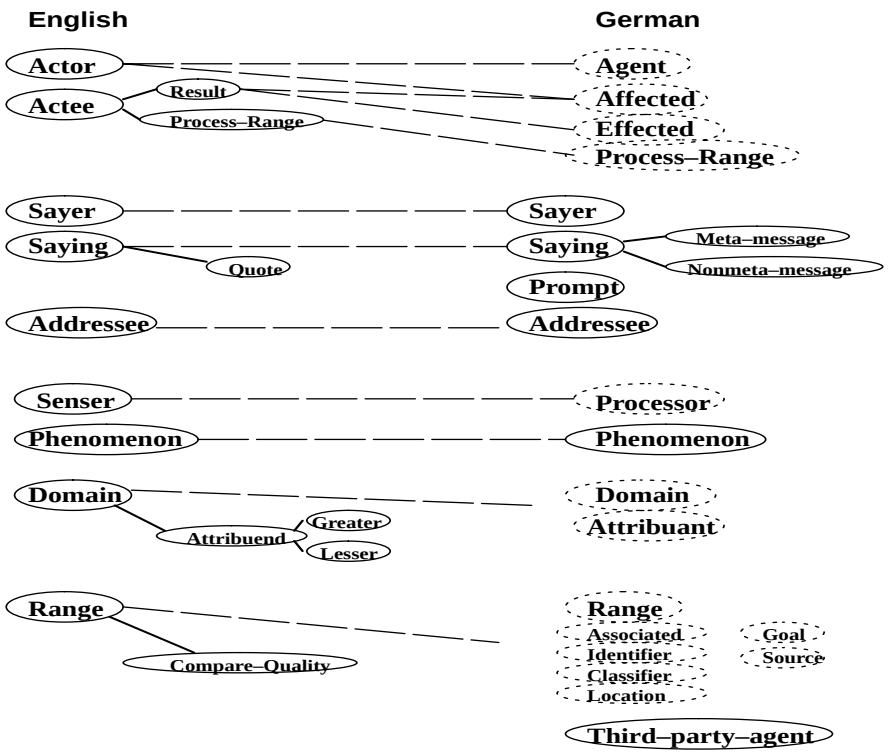


Figure 15: Correspondences for the Participants

5.8 Object hierarchy

The essential part of the Object taxonomy in the EUM and the GUM can be identified replacing the GUM name *Entity* by the EUM name *Object*. The comparison is given in Figure 16. The english-only concepts (marked by double ellipsis)

Set, Disjunctive-Set, Ordered-Set

extend the identical kernel taxonomy.

The German-only concepts *absolute* and *relative* are meant to distinguish the cardinal and ordinal usage of numbers. *G-measure* should assist in the generation of measure units. However, the implementations which use these concepts are not working. So, the concepts have not yet enough evidence in their favour to be included in the merged UM.

The concept *G-Generic-person* triggers the generation of pronouns in the German grammar. This is a textual decision which should be omitted from the ideational knowledge base. The concept *G-Time-of-speaking* is removed for the same reason.

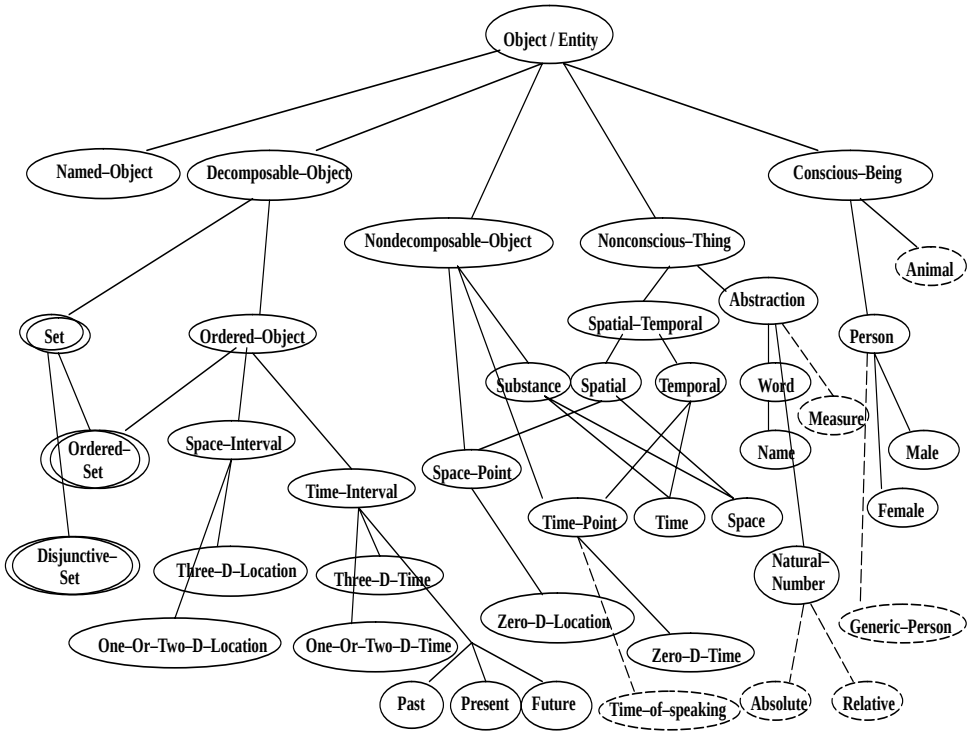


Figure 16: Object Hierarchy

5.9 Qualities

The *Quality* subhierarchies of the EUM and the GUM are identical in meaning and terminology with the exception of two additional German concepts, namely *G-Cardinal* and *G-Negative-quality* which are not present in the English UM. Because these concepts are recently not used by the German grammar, we propose to remove them entirely. Figure 17 shows the resulting quality subhierarchy of the merged UM which is in fact the one from the EUM.

6 Merging statistics

Because of their questionable status, we leave the RST-Relations out of account in the statistical comparison. Without the RST hierarchy 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 UM contains 258 concepts.

Identity

We have 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

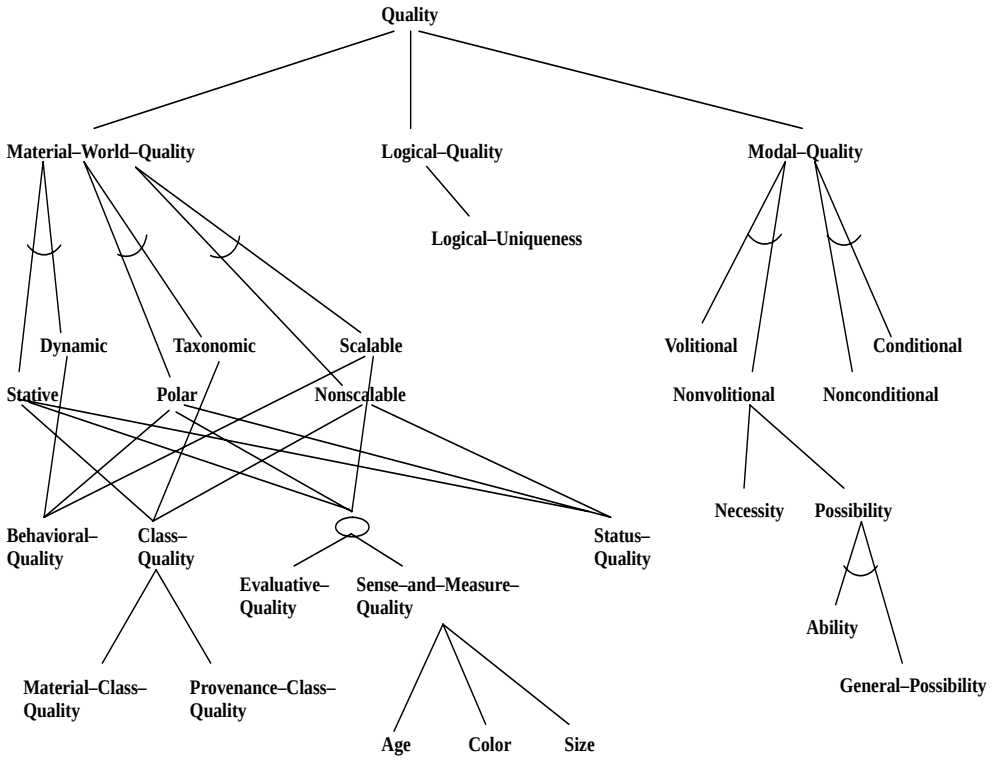


Figure 17: Quality Hierarchy

and the quality hierarchy as well as the temporal one. The precise distribution for strong identical meaning is shown by the numbers outside of brackets in Figure refmerge-stats.

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. The clause/PP distinction which is a concept discrimination criterion in the GUM violates design principle 3. Therefore, this discrimination is not preserved in the Merged UM as described in section 5.6. Therefore, leaving out of account the clause/PP distinction, identical concepts then amount to 163 (see the numbers in brackets in Figure 18). 106 concepts are strongly identical and 57 merged concepts are identical with the union 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 UM.

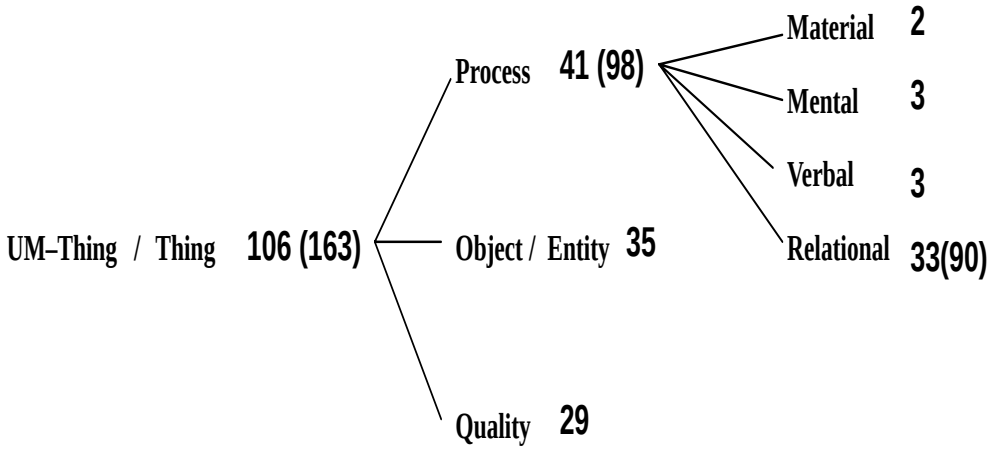


Figure 18: Merging statistics

Cross classification

An essential field for cross classification has been avoided by the relaxation of the UM-grammar-interface described in section 5.2. Thus cross classification is only necessary in the *Material-Process/Action* hierarchy where 2 English subconcepts have to be cross classified with 5 German subconcepts and their subhierarchies respectively. This would generate 42 merged concepts. An extract of that was given in Figure 4. However as shown in section 5.3, 9 concepts are sufficient to cover all distinctions expressed in the EUM and the GUM (see Figure 5).

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 UM (8%) where the concepts are not created by identification, extension and cross classification, but by preferring EUM concepts over GUM ones.

7 Table of concept correspondences

This section provides a list of all concept correspondences discussed in the preceding sections. The lack of a corresponding concept is denoted by —. The Merged UM concepts are mainly the English ones, underlined concepts are the extensions of the EUM which are introduced due to the GUM. German concepts which seem to have the status of ad hoc concepts (see discussion above) are not taken into account. The *Quality*, the *Object* and the *Rhetorical-Relation* hierarchy are not listed because of their nearly overall identity.

ENGLISH	GERMAN	MERGE
Top level		
UM-Thing	Thing	UM-Thing
Process	Process + Relationship	Process
Material-Process	Action + Relational [9]	Material-Process
Mental-Process	Mental	Mental-Process
Verbal-Process	Communication	Verbal-Process
Relational-Process	Relational + Relationship	Relational-Process
Object	Entity	Object
Quality	Quality	Quality
Material		
Nondirected-Action	—	Nondirected-Action
—	Agent-only + Ranging	Nondirected-Doing
Motion-Process	—	Motion-Process
—	Affected-only	Nondirected-Happening
Ambient-Process	Natural-Phenomenon	Ambient-Process
Directed-Action	—	Directed-Action
Dispositive-Material-Action	Affecting + Affected-centered	Dispositive-Material-Action
Name-Event	—	Name-Event
Creative-Material-Action	Effecting	Creative-Material-Action
Mental		
Mental-Active	Phenomenon-oriented + Phenomenon-only	Mental-Active
		Mental-Active
Mental-Inactive	Processor-oriented	Mental-Inactive
Perception	—	Perception
Cognition	—	Cognition
Believe	—	Believe
Know	—	Know
Think	—	Think
Reaction	—	Reaction
Fearing	—	Fearing
Disliking	—	Disliking
Liking	—	Liking
Striving	—	Striving
Wanting	—	Wanting
Verbal		
Non-Addressee-Oriented-Verbal-Process	Message-oriented + Message-only + Sender+Prompt	Non-Addressee-Oriented-Verbal-Process
Addressee-Oriented-Verbal-Process	Receiver-oriented	Addressee-Oriented-Verbal-Process
Relational		
One-Place-Relation	—	One-Place-Relation

[9] The part of Relational which integrates a Third-party-agent can be classified as Material, too

Existence	Existence	Existence
Two-Place-Relation	—	Two-Place-Relation
Intensive	Identifying + Classificatory + Feature-Ascription	Intensive
Circumstantial	Circumstantial + Locational + Purpose + Cause + Consequence + Generalized-Means + Scope	Circumstantial
Generalized-Possession	Associative-1 + + Generalized-Possession	Generalized-Possession
Generalized-Possession-Inverse	Associative-2	Generalized-Possession-Inverse
Logical	Logical [10]	Logical
Rhetorical-Relation	Rhetorical-Relation	Rhetorical-Relation
Participant + Prompt	Participant	Participant

Intensive

UM-Identity	Identifying + Identity	UM-Identity
Ascription	Classificatory + Ascription	Ascription
Class-Ascription	Classificatory-1 + Class-Ascription	Class-Ascription
Property-Ascription	Classificatory-2 + Property-Ascription	Property-Ascription
Logical-Property-Ascription	Logical-Property-Ascription	Logical-Property-Ascription
Modal-Property-Ascription	Modal-Property-Ascription	Modal-Property-Ascription
Age-Property-Ascription	Age-Property-Ascription	Age-Property-Ascription
Color-Property-Ascription	Color-Property-Ascription	Color-Property-Ascription
Size-Property-Ascription	Size-Property-Ascription	Size-Property-Ascription
Material-Property-Ascription	Material-Property-Ascription	Material-Property-Ascription
Provenance-Property-Ascription	Provenance-Property-Ascription	Provenance-Property-Ascription
Use-Property-Ascription	Use-Property-Ascription	Use-Property-Ascription
Scaled-Comparison	—	Scaled-Comparison
Greater-than-Comparison	—	Greater-than-Comparison
Less-than-Comparison	—	Less-than-Comparison
Quantity-Ascription	Quantity-Ascription	Quantity-Ascription
Quantity	Quantity	Quantity
Number-Focusing	Number-Focusing	Number-Focusing
At-Least	At-Least	At-Least
At-Most	At-Most	At-Most
Exactly	Exactly	Exactly
Greater-than	Greater-than	Greater-than
Less-than	Less-than	Less-than
Ascription-Inverse	—	Ascription-Inverse
Property-Of	—	Property-Of
Symbolization	Symbolization	Symbolization
Name-Relation	Name-Relation	Name-Relation
Role-Playing	Role-Playing	Role-Playing

Generalized-Possession

Generalized-Role-Relation	—	Generalized-Role-Relation

[10] without Cause, Consequence, Purpose, Scope

Part-Whole	Part-Whole	Part-Whole
Part	Part	Part
Element	Element	Element
Part-Of	—	Part-Of
Element-Of	Element-Of	Element-Of
Ownership	—	Ownership
Owned-By	—	Owned-By
Name-Of	—	Name-Of

Circumstantial

Causal-Relation	Causal-Relation + Cause + Consequence + Purpose	Causal-Relation
Reason	Reason + Cause	Reason
Purpose	Purpose +Consequence	Purpose
Concessive	Concessive	Concessive
Cause-Effect	—	Cause-Effect
Ordering-Relation	Ordering-Relation	Ordering-Relation
Subject-Matter	—	Subject-Matter
Specific-Matter	Scope	Specific-Matter
Diffuse-Matter	—	Diffuse-Matter
Of-Matter	—	Of-Matter
Generalized-Means	Generalized-Means	Generalized-Means
Manner	Manner	Manner
Instrumental	—	Instrumental
Enablement	—	Enablement
Agentive	—	Agentive
Accompaniment	Accompaniment	Accompaniment
Inclusive	—	Inclusive
Exclusive	—	Exclusive
Alternative	—	Alternative
Additive	—	Additive
Comparison	—	Comparison
Similarity	—	Similarity
Difference	—	Difference
Spatio-Temporal	Spatio-Temporal + + Locational	Spatio-Temporal
Spatial-Relation	Spatial-Relation + Locational	Spatial-Relation
Temporal-Relation	Temporal-Relation	Temporal-Relation
Locating	Locating	Locating
Extent	Extent	Extent
Absolute-Extent	Absolute-Extent	Absolute-Extent
Relative-Extent	Relative-Extent	Relative-Extent

Spatial-Relation

Spatial-Extent	Spatial-Extent	Spatial-Extent
Relative-Spatial-Extent	—	Relative-Spatial-Extent
Absolute-Spatial-Extent	—	Absolute-Spatial-Extent
Nonparallel-Extent	Nonparallel-Extent	Nonparallel-Extent
Parallel-Extent	Parallel-Extent	Parallel-Extent
Spatial-Locating	Spatial-Locating	Spatial-Locating
Nonorienting	Nonorienting	Nonorienting
Orienting	Orienting	Orienting

Static-Spatial	Static-Spatial	Static-Spatial
Spatial-Ordering	Spatial-Ordering	Spatial-Ordering
Vertical	Vertical	Vertical
Above	Above	Above
Below	Below	Below
Horizontal	Horizontal	Horizontal
Behind	Behind	Behind
Between	—	Between
Beside	—	Beside
Facing	Facing	Facing
Containment	—	Containment
Contained	—	Contained
Notcontained	—	Notcontained
Source-Destination	Source-Destination	Source-Destination
Destination	Destination + Goal	Destination
Source	Source	Source

Temporal-Relation

Temporal-Extent	Temporal-Extent	Temporal-Extent
Absolute-Temporal-Extent	—	Absolute-Temporal-Extent
Relative-Temporal-Extent	—	Relative-Temporal-Extent
Exhaustive-Duration	Exhaustive-Duration	Exhaustive-Duration
Nonexhaustive-Duration	Nonexhaustive-Duration	Nonexhaustive-Duration
Temporal-Locating	Temporal-Locating	Temporal-Locating
Temporal-Nonordering	Temporal-Nonordering	Temporal-Nonordering
Temporal-Ordering	Temporal-Ordering	Temporal-Ordering
Extremal	Extremal	Extremal
Nonextremal	Nonextremal	Nonextremal
Posterior	Posterior	Posterior
Anterior	Anterior	Anterior
Posterior-Extremal	Posterior-Extremal	Posterior-Extremal
Anterior-Extremal	—	Anterior-Extremal
Follow	Follow	Follow
Precede	Precede	Precede
Concurrent	—	Concurrent
Temporal-Inclusive	Temporal-Inclusion	Temporal-Inclusive
Temporal-Noninclusive	—	Temporal-Noninclusive

Logical

Logical-Condition	Condition	Logical-Condition
Conjunction	Conjunction	Conjunction
Disjunction	Disjunction	Disjunction
Elaboration	Elaboration	Elaboration
Exemplification	Exemplification	Exemplification
Restatement	Restatement	Restatement

Participants

Material

Actor	in Directed-Action	Agent	Actor
	in Nondirected-Happening	Affected	Actor
Actee	in Nondirected-Doing	Affected	Actee
	in Dispositive-Material-Action	Affected	Actee
	in Creative-Material-Action	Effected	Actee
	in Nondirected-Action	Process-range	Actee
Process-Range		Process-Range	Process-range
Result	in Dispositive-Material-Action	Affected	Result
	in Creative-Material-Action	Effected	Result

Verbal

Sayer	Sayer	Sayer
Saying	Saying	Saying
—	Meta-message	Meta-Message
—	Nonmeta-message	Nonmeta-Message
Quote	—	Quote
Addressee	Addressee	Addressee
—	Prompt	Prompt

Mental

Senser	Processor	Senser
Phenomenon	Phenomenon	Phenomenon

Relational

Domain	Attribuant + Domain	Domain
Attribuend	—	Attribuend
Greater	—	Greater
Lesser	—	Lesser
Range	Identifier + Classifier + Associated Goal + Source + Location + + Location + Range	Range
Compare-Quality	—	Compare-Quality
—	Third-party-agent	Third-Party-Agent
Beneficiary	—	Beneficiary
Recipient	—	Recipient
Client	—	Client
Addressee	Addressee	Addressee
Agentive	—	Agentive
Material	—	Material
Standard	—	Standard

8 Future work

In the current merging process, we have only looked for identities and differencies between the given English and German UM. We did not try to improve the inherent consistency of both. But as a by-product of the considerations carried out, we have found some unclear edges, which should be cleared in future work. Here, we can only mention the relevant topics for that:

- All RST-relations should be removed from the UM and put into a text base, because an ideational knowledge base (as the UM) should not contain any textual relations. This contradicts the construction criterion 3 (see also footnote 17 in [Bateman *et al.*, 1990]).
- The *domain* and *range* concept as participants of a *two-place-relation* must be clearly distinguished from the LOOM-built-in concepts domain and range.
- The area with *accompaniment*, *generalized means*, *causal-relation*, *ordering-relation*, *comparison* should be thought over. These concepts could be ordered with respect to domain and range restrictions.
- It is open to question, whether the special status of participants is adequately reflected by placing them under *two-place-relations* in parallel to *intensive*, *generalized-possession* and *circumstantial*.

Acknowledgement

I was only able to perform the detailed comparison of both UM's thank to personal communication with Elke Teich concerning the GUM and John Bateman concerning the EUM whom I would like to express many thanks.

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.*, 1991] 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, 1992] 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).

[Fawcett, 1980] Robin P. Fawcett. *Cognitive Linguistics and Social Interaction*. Exeter University and Julius Groos Verlag, Exeter and Heidelberg, 1980. Exeter Linguistic Studies 3.

[Frederking *et al.*, 1993] Robert Frederking, Ariel Cohen, Dean Grannes, Peter Cousseau, and Sergei Nirenburg. The PANGLOSS MARK I MAT system. In *Sixth conference of the European Chapter of the Association for Computational Linguistics*, page 468. Association for Computational Linguistics, 1993.

[Halliday, 1985] Michael A.K. Halliday. *An Introduction to Functional Grammar*. Edward Arnold, London, 1985.

[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.

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

[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.

[Steiner *et al.*, 1988] Erich H. Steiner, Paul Schmidt, and Cornelia Zelinsky-Wibbelt. *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.

Appendix – The Merged UM in LOOM

As shown above, the Merged UM can be obtained out of the EUM by a small number of additions. The necessary extensions to the datafile of the EUM which make it possible to serve for the German generation grammar, too, are listed below:

```
(defconcept Nondirected-Doing
  :is (:and Nondirected-Action :primitive))


(defconcept Nondirected-Happening
  :is (:and Nondirected-Action :primitive))


(defreified-relation Meta-Message
  :is (:and Saying :primitive))


(defreified-relation Nonmeta-Message
  :is (:and Saying :primitive))


(defreified-relation Prompt
  :is (:and Participant :primitive))


(defreified-relation Two-Place-Relation
  :is (:and (:concept Relational-Process) :primitive)
  :constraints (:at-least 1 Domain)
  :constraints (:at-least 1 Range)
  :constraints (:at-most 1 Third-party-agent))


(defreified-relation Third-party-agent
  :is (:and Participant :primitive))
```