Building Concept Frames based on Text Corpora

Birte Lönneker

Institute of Romance Languages
Hamburg University
Von-Melle-Park 6, 20146 Hamburg, Germany
birte.loenneker@uni-hamburg.de

Abstract

Linguists have been using different kinds of frame representation since the emergence of the notion "frame". The main goal of the annotation system described in this paper is to provide an interactive and easy-to-use tool for structuring concept-specific information in linguistic frames for discourse analysis or cultural studies. These frames take into account background or "world" knowledge associated with the concepts, which is not necessarily present in lexicographic frames. A *frame hierarchy* providing default information, *example texts* containing specific information on a concept, and the *annotations* made by a user are combined together in one database. All frames have a predefined structure, and the information they contain is represented in natural language. The collected information can also be used as input to knowledge bases, or for defining patterns for Information Extraction.

1. Introduction

Linguists have been using different kinds of frame representation since the emergence of the notion "frame" in Minsky (1975). The main fields in which linguistic frames play a role are *discourse analysis* (van Dijk, 1987; Klein & Meißner, 1999), *etymology and metonymy* (Blank, 1999; Koch, 1999) and *lexicography* (Konerding, 1993; Fillmore, Wooters & Baker, 2001). The main goal of the annotation system described in this paper is to provide a tool for structuring concept-specific information in frames for discourse analysis and metonymy or metaphor studies. These frames take into account background or "world" knowledge associated with the concepts, which is not necessarily present in lexicographic studies like FrameNet (cf. Fillmore, Wooters & Baker (2001)).

The propositional frames produced with our tool can be used in sociological or cultural studies, or as input to knowledge bases and text generation systems. Working with corpora enables the users to encode knowledge that does not reflect their own culture. Furthermore, the combination of a frame hierarchy, the annotations made, and the underlying example texts in one database offers new possibilities of detecting patterns for Information Extraction. The combination of annotated propositions with the example texts allows to detect co-occurrences of frame information. The tool we developed can be used without having to learn a logical language. Annotations are made in natural language, according to a special frame structure.

The remainder of the paper is structured as follows: Section 2. presents the frame structure and the top hierarchy of frames used by the tool. In section 3., the annotation tool itself is explained and situated in the process of corpus collection and information gathering. Section 4. describes related work, and section 5. is the conclusion.

2. Prerequisites: structure and hierarchy of concept frames

The concept frames in our system are propositional frames. They are centered around a concept name, a noun, which can refer to "things" or "events" ("continuants" or "occurrents", cf. Guarino (1997)). In a simplistic view, the propositions are formed of the frame name as the subject, the subslot name as the predicate, and the filler name as the object of a sentence. In subsection 2.1., we give a more detailed presentation of the structure of the concept frames used in the system. Subsections 2.2. and 2.3. explain, respectively, the default information contained in the system, and how frames are arranged into a hierarchy.

2.1. A structure for concept frames

Since Minsky's (1975) first proposal of frames, the frame idea has been used and further developed for various areas (cf. Lönneker (2001)). Fillmore's *sentence frames* (cf. e.g. Fillmore (1968)), whose center of interest are the verb and its "case roles", have further been developed to more complex and abstract frames or "scenarios" (Fillmore, Wooters & Baker, 2001) being a background for diverse lexicalizations focusing on different aspects of the frame. In addition to Fillmore's frames, *propositional and graphical frames* are used in linguistics. Propositional frames consist of a list of propositions about a concept, graphical frames display frame elements and relationships or networks of frames in diagrams. In some graphical frames, all relations (those between "frame elements" or fillers as well as those between a frame and its "elements") are supposed to be "contiguity relations", and are not further differentiated (Koch, 1999; Blank, 1999).

In order to have a common and clear structure for propositional frames, we extended the frame structure of the Generic Frame Protocol (Karp, Myers & Gruber, 1995) used for knowledge representation tasks. Our concept frames consist of the following parts:

- a *frame name*, which is a noun or a nominal phrase (e.g., "entity");
- *slots* ("aspects" of the concept), named by nouns or nominal phrases (e.g. "existence"), grouping
- *subslots*, which consist of verbal phrases (a language-independent boolean marker, a verb, and—if needed—a preposition; e.g. "+ originate from" and "+ finish one's existence, because of"); and

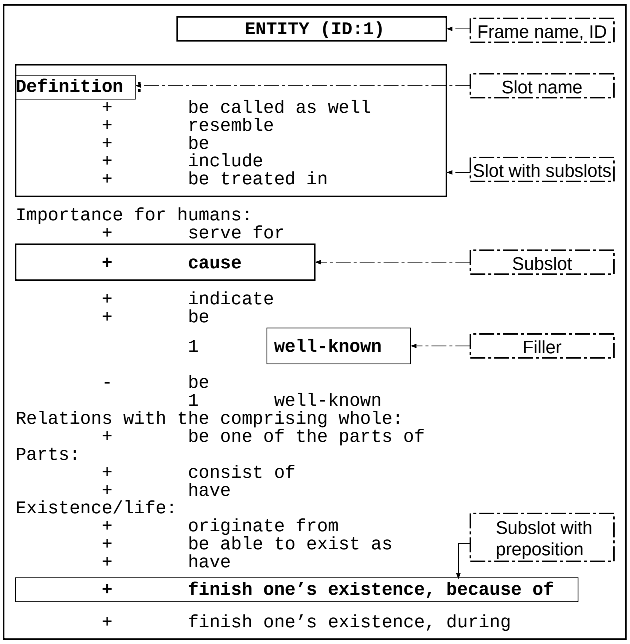

Figure 1: Structure of the ENTITY frame.

- *fillers*, which are named by adjectives (e.g. "well-known" as a filler of the subslot "+ be") or nouns, possibly modified. The nouns are the grammatical objects of the proposition, or the heads of noun phrases in the prepositional phrase introduced by the subslot. Modifiers of fillers are adjectives, prepositional phrases or relative clauses.

The resulting frame structure is illustrated in figure 1. Linguistic frames of this format should be easier to reuse in Knowledge Representation and Language Engineering tasks than the existing propositional and graphical frames. Propositional frames in our structure could also be used as an input to graphical frames and would provide them with named relations.

2.2. Default information in concept frames

In figure 1, the ENTITY frame illustrates our frame structure. At the same time, however, it contains some *information*: Slot names and subslot names as well as one filler (in two different subslots) are already contained in the frame. They are natural language entities.[1]

As can be seen from figure 1, most of the predefined or default information "known" by the frame system concerns frame names, slot names and subslot names. Only very few fillers are provided. This is due to the fact that the system contains definitions of concept frames at a high level of abstraction. These will be used as superframes for more individual concept frames that inherit the default information and are defined in more detail by the user.

[1] In our system, the natural language information used in the frames is provided for the languages French and German. The frame in figure 1 has been translated into English for this paper. Subsequent examples and screenshots will show French data.

The default information contained in the frame system is based on an adaptation of Konerding's (Konerding, 1993) so-called "matrix frames", the topmost concepts except "thing" or "entity" found in conventional dictionary definitions. The linguistic information collected concerns twelve different superframes ("matrix frames"). Konerding's frames consist of blocks of questions that might plausibly be asked about the concept represented by the superframe (e.g. possible questions concerning organisms). The verbs of these questions are taken over as default subslots into our frame system. This means that the combination of the frame name as a subject and the subslot fields as components of a verb phrase in a sentence will be semantically correct. The question blocks in Konerding (1993) have headlines indicating the aspects of the entity they describe, which can be taken over as default slots. However, as there are no answers to the questions in Konerding's work, it does not provide us with fillers.

The answers to the questions, i.e. the fillers, and more subslots for more specific concepts, can be found analyzing texts that deal with the concept in question. This is exactly what the annotation tool is used for.

2.3. A hierarchy of frames

The information concerning frame names, slot names, and subslot names of the twelve frames that are treated in Konerding (1993) shows some similarities from frame to frame. For this reason, we extracted all slot-subslot combinations occurring in every frame and inserted them into a new frame, the ENTITY-frame (cf. figure 1). This frame has the highest level of abstraction in our hierarchy. As all other frames are defined as subframes (at different levels) of the ENTITY-frame, they inherit the information provided by that frame. By this means, we ensure that there is some regularity in the frame information: Everything that is defined by the highest frame is also available in the subframes. This allows the user, for example, to compare all his frames regarding the fillers for some subslots provided by the highest frame, without getting an error.

Further analysis of the frame information in Konerding (1993) led to some more intermediate levels in the frame hierarchy, so that in the end we have four additional frames with a higher level of abstraction: ENTITY, CONTINUANT, PRIMARY OBJECT and ROLE/VIEWPOINT ON AN ENTITY. The twelve frames treated by Konerding are subframes of these, as shown in figure 2. Concerning the status of the hierarchy, we would like to add three remarks:

1. The hierarchy does not provide suitable frames for all entities of the world. For example, TIME and SPACE do not have their own frames; Konerding (1993, 184–185) proposes to treat them as "WHOLES of PARTS".

2. The hierarchy can be regarded as an informal top level ontology. The fact that it does not comply with formal ontology standards (cf. Guarino (1997)) shows once more that linguistic and ontological research on the same problem do not end up with the same solution, due to their different methods.

3. The location of the frames OBJECT, PERSON WITH

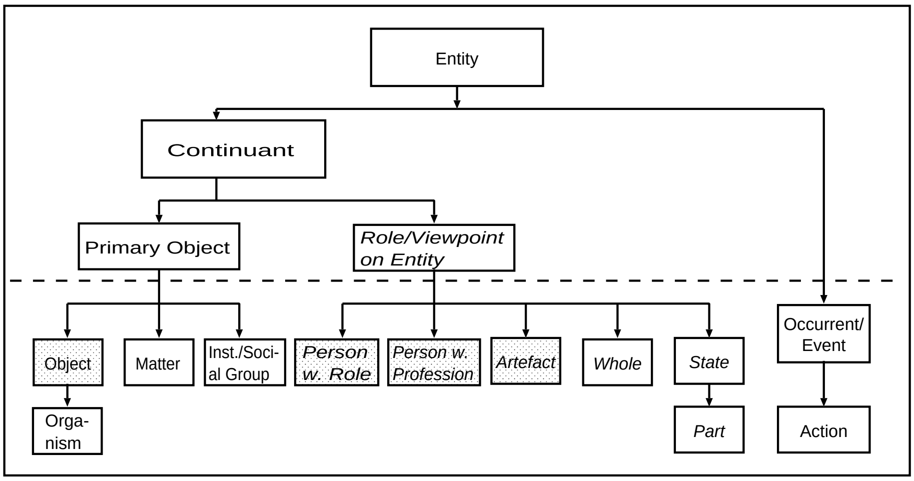

Figure 2: Frame hierarchy based on Konerding (1993).

A ROLE, PERSON WITH A PROFESSION and ARTEFACT in the hierarchy is problematic. The word "object" is highly polysemous and is also used in the names of higher-level frames. The names of the other problematic frames are lexicalized only in very specific areas, if at all. In our hierarchy, these frames are at the same level as other, lexicalized items (cf. the dotted line in figure 2). The annotation of some of their subframes will show whether they have to be "lifted up" in the hierarchy, melted together or replaced.

The frames of our hierarchy are potential superframes for subframes to be annotated. During the annotation of a new subframe, all the lexical information contained in the superframe is available to the user: He can reuse all subslots and fillers defined in the superframe. This advantage mainly concerns subslots up to now. The number of predefined subslots (inherited and new ones) in each frame can be found in table 1. For the less abstract frames, which are most likely to be chosen as superframes for concepts to annotate, the predefined subslots range between 40 and 60. It is of course also possible to directly annotate the lexicalized frames of the hierarchy; however, it seems more useful to derive the most often used fillers from their subframes.

Concluding this section, we can summarize that we have set up a detailed structure for propositional frames, as well as a hierarchy of frames containing default information, which is represented in natural language entities. The default information is available for inheritance and annotation. The hierarchy itself relies on this natural language information.

ID	Frame name	Super-frame ID	Inherited subslots	New sub-slots	Sum sub-slots
1	Entity	–	–	18	18
2	Constant object	1	18	12	30
4	Primary object	2	30	6	36
12	Role/Viewpoint on entity	2	30	7	37
5	Object	4	36	5	41
6	Matter	4	36	9	45
7	Institution/ Social Group	4	36	25	61
9	Person with role	12	37	17	54
10	Person with profession	12	37	10	47
11	Artefact	12	37	4	41
13	Whole	12	37	2	39
14	State	12	37	4	41
8	Organism	5	41	8	49
15	Part	14	41	3	44
3	Event	1	18	26	44
16	Action	3	44	18	62

Table 1: Number of default subslots per frame.

3. Using the annotation tool

Our annotation tool provides the user with an interactive interface for gathering concept-specific information that will be stored in a database. All information about a concept is arranged into the propositional frame structure explained in section 2.1. In our current work, the information relies on text resources that are collected on the internet; but, of course, any corpus providing information about the analyzed concept could be used. The text corpus is split into small portions (e.g. paragraphs, list entries) from which it is easy to extract propositions about the concept. The superframes containing information on concepts at a higher level of abstraction, as explained above in section 2.3., are provided by the system. From these frames, the user can choose a superframe for the concept he is working on. While annotating, he will thus be able to reuse the frame information inherited from that superframe. The context in which we use the annotation tool is briefly explained in subsection 3.1. Subsection 3.2. deals with the individual steps of annotating an example, subsection 3.3. discusses some example annotations, and subsection 3.4. presents two different ways of displaying the collected in-

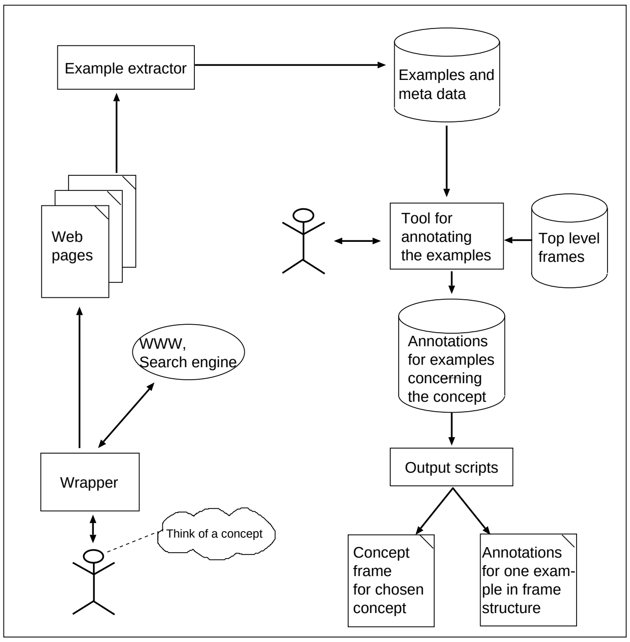

Figure 3: Context of the annotation tool.

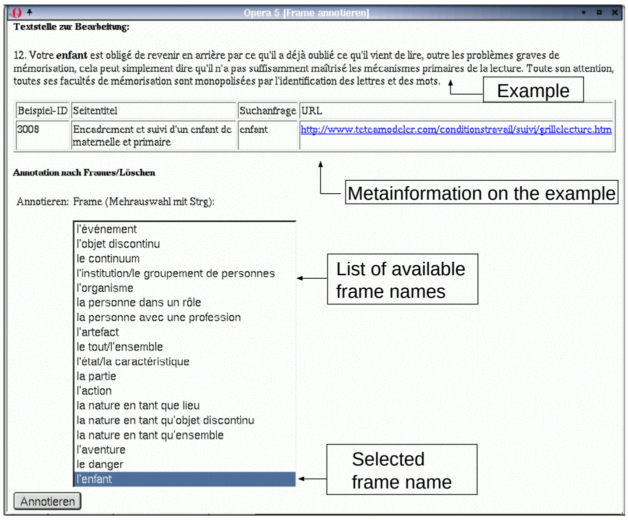

Figure 4: Selecting a frame for an example text.

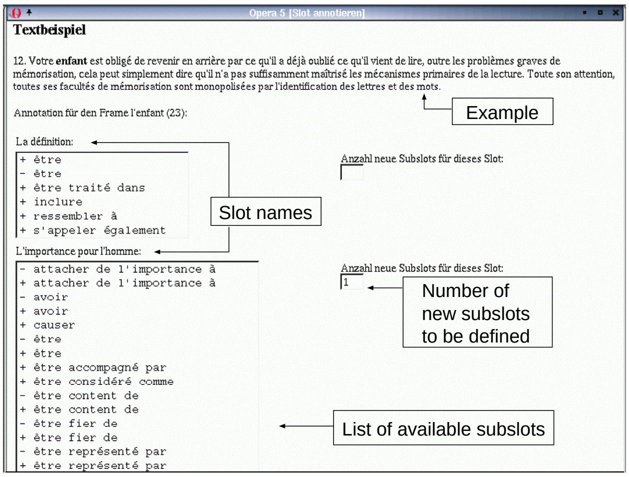

Figure 5: Annotating subslots.

formation.

3.1. Context

Figure 3 shows the context in which the annotation tool is used.

The text examples to be annotated are first collected from the internet (see left hand side of the figure) and written into a database. The concept for which a frame is to be established has to be defined by the user at the beginning of the collection phase. Among other parameters, he has to define at least one search string or regular expression which is likely to refer to the chosen concept. The example extractor splits the web pages collected by the wrapper into small pieces (e.g. paragraphs, list entries), each of them containing at least one occurrence of the specified search string. From these text portions, further referred to as *examples*, it will be easy for the user to extract propositions about the concept.

The annotation tool (see right hand side of figure 3) displays one example after another and asks the user to annotate them. It uses the information contained in the top level frames. The user can extend the hierarchy. Annotations are done in various steps using a web browser as graphical user interface (cf. subsection 3.2.). The annotations are written into the database as new frame information and can be displayed in various layouts (cf. subsection 3.4.).

3.2. Steps of the annotation process

During the annotation process, the user is confronted with an example (a portion of text) in which—due to the way in which the examples are collected—some information on the analyzed concept can be found. The annotation is performed in the following steps:

1. The example, together with some meta-information on where and when it has been collected, is displayed, as shown in figure 4. In the example, some words are written in bold characters. Those words matched the search string or regular expression of the collecting phase and are most likely to refer to the concept for which the user wants to build a frame.

   The frame name is selected from a list of known frames. It is possible to select more than one frame name, if the same example is to be used for the annotation of different frames. If the name of the needed concept frame is missing from the list, the frame can be created by the user as a subframe of one of the existing frames. It has to be given a name, and its superframe has to be chosen from the list of known frames.

2. The relations expressed in the example or inferred from it can be chosen for annotation from a concept-specific set of subslots (cf. figure 5). This list contains only those relations, in the guise of subslots, that have been defined as suitable for the concept frame or one of its superframes.

   New subslots can be defined. For this purpose, the user

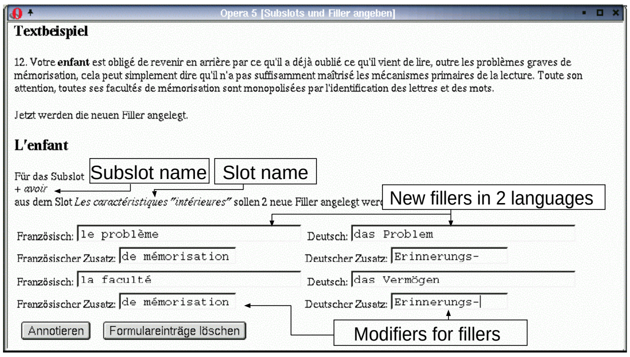

Figure 6: Defining new fillers.

enters the number of subslots per slot that he wants to create, as shown in figure 5. He will then be presented a form in which he defines a boolean value, a verb, and possibly a preposition for the new subslot(s). From then on, they are available in the subslot list for annotations of that frame.

3. In the last step, the user is shown the example together with the frame name(s) and the subslot(s) he chose for annotation. For each subslot, he can choose from a list the filler names for the connected concepts that he finds in the text. The available fillers are those which are known to the system as possible fillers for this subslot in this frame or in one of its superframes.

   If a filler name is missing from the list, the filler can be added analogously to step 2. There are two fields for each filler, the first one for the noun and the second one for modifiers of this noun, like adjectives or prepositional phrases. Figure 6 shows how two new fillers are defined in two languages in parallel.

All information once defined by the user will automatically be available for direct selection during the annotation of further examples. This decreases the time needed for annotation of examples belonging to the same frame.

It has to be stressed that the information to be annotated might be contained quite implicitly in the example text, so that it is extremely helpful for the user to have some "default information" to choose from. This is especially the case for the relations. In many cases, it is easy to detect that the text relates the concept frame to another concept in some way, but a verb or verbal phrase that expresses the relation is not explicitly present in the text. This is why the slots and subslots inherited by the top level frames are the most important default information in our system. Among them, the user might find the relation that is implicit in the example text. This issue is illustrated by some example annotations in the next subsection.

3.3. Examples and discussion

The examples in table 2 (see next page) illustrate that relations are expressed more or less explicitly in the text, and rarely by a verb. The French example texts have been taken from annotations concerning the CHILD frame, where they had been annotated with the subslot "+ être l'une des parties de" ("+ be a part of"). In the first row, the example text extracted from the internet is shown. The second row contains a translation of the example into English. The third row indicates the filler of the subslot that is expressed in the example. In the first two examples, the relation is expressed by the (more specialized) verb-preposition combinations "appartenir à" ("belong to") and "grandir au sein de" ("grow up in"). These verbs are more specialized than the verb phrase "être l'une des parties de" used in the annotation. The other examples show: an *enumeration* of the "subgroup" (children) and the group they belong to (example 3), an *explicit specification* ("en particulier") of the group to which children can belong (example 4), and a *preposition* introducing the name of the comprising group (example 5) as means for expressing the relation.

We annotated 50 examples for this frame and this subslot, in which the "+ be a part of"-relation is very rarely (< 10%) expressed by verbs. Other means like enumerations and prepositions (in various forms) are not specific to this relation, but can also indicate other subslots like "+ avoir" ("+ have"), "+ inclure" ("+ comprise"). Furthermore, the constructions indicating the subslot "+ être l'une des parties de" in the examples in table 2 are not the only ones encountered. These are the reasons why we do not try to annotate the examples automatically. Our annotations can, however, be useful for defining rules or patterns to automate some of the annotations in the future. In order to do so, possible relations or subslots for the analyzed concept will have to be known, and rules or patterns have to be defined for preprocessed texts, as usual in Information Extraction. Definitions for rules have to be based on previous annotations. Because of the "overlap" of some constructions just mentioned, the automatic extraction will reach neither the precision nor the recall of the human annotator. As a first step, we are concentrating on revising and refining the valid subslots for the superframes in our frame hierarchy on the basis of our annotations.

Concerning the annotations shown in table 2, a remark has to be added on the status of the fillers. We encode every proposition about the concept, also those that do not seem to express "typical" information. The typicality of the fillers and propositions, or their prominence in discourse, can only be quantified after a large number of annotations. This is why the example number 1 of table 2 ("Children belong to minority groups") has to be annotated in spite of its obvious or apparent marginality.

Until now, the tool has been used to annotate frames for the following concepts: NATURE, ADVENTURE, DANGER, and CHILD. This is a chain of frame and filler names (ADVENTURE is one of the fillers of the NATURE frame, and so on). The concept NATURE has been annotated with three different superframes ("roles"). Table 3 shows the number of examples collected for these concepts, and the number of propositions encoded in the database.

3.4. Displaying annotations and frames

The information contained in each frame can be displayed by a command line script. The output will be similar to figure 1 for all frames of the upper frame hierarchy. However, for the frames that have been annotated, there will be

	Example Text	Translation	Filler, Modifier
1	[...] besoins linguistiques des enfants autochtones ou appartenant à un groupe minoritaire	[...] linguistic needs of autochthone children or of children belonging to a minority group	le groupe, minoritaire
2	aide au développement des enfants qui grandissent principalement au sein de familles monoparentales [...]	help for the development of children who mainly grow up in monoparental families [...]	la famille, monoparental
3	Enfant & famille Canada	Child & family Canada	la famille
4	Faire en sorte que tous les groupes de la société, en particulier les parents et les enfants, reçoivent [...]	Proceed in such a way that all groups of the society, especially the parents and the children, get [...]	la société
5	Le mode de vie imposé aux enfants dans les sociétés développées, [...]	The lifestyle imposed on children in the developed societies, [...]	la société, développé

Table 2: Examples and fillers for the subslot "+ être l'une des parties de" of the CHILD frame, in French

Concept name	Number of examples	Annotated propositions
NATURE	457	252
ADVENTURE	239	202
DANGER	197	132
CHILD	965	2022

Table 3: Number of examples and annotations.

as many different fillers as found in the examples. These frames get so rich that it is impossible in this paper to show an example of an entire frame based on annotations. Figure 7 (see next page) therefore shows only small parts of the CHILD frame that we have annotated (in French). The number in front of a filler shows how often we chose this filler in this subslot. So, if we want to show only the most "typical" propositions about the concept that were found in the corpus, we could fix a minimum number or percentage of occurrences.

Another script, displaying annotations regarding individual examples, is integrated into the graphical user interface. It can be used in order to verify the annotations, or to have an overview of all annotations for one example. Its output is displayed in figure 8 (see next page).

Many other scripts can easily be written using the SQL query language of the database. For example, we used SQL scripts displaying all fillers and examples of a given subslot (similar to the information contained in table 2), in order to analyze by which means the relation encoded in the subslot was originally expressed in the text examples.

We can also imagine that a program displaying co-occurring fillers or subslot-filler combinations is useful for finding out internal relations between fillers. For example, if one of the fillers is "victim", in the example there might also be mentioned "of what". So there should be a filler for the slot "to be afflicted by". Do co-occurring subslot-filler combinations in the *External characteristics* or *Internal characteristics* slot in these cases refer to causes or effects of the victimhood? Another example is that of contradicting fillers or subslot-filler combinations: In this case, we could have a look at other fillers of the examples in order to find out why this predication is subject to discussion.

4. Related work

This section deals with two categories of related work: Ontology-orientated approaches (4.1.) and lexicographic resources (4.2.).

4.1. Ontological approaches

The large *ontology and knowledge base Cyc*[2] aims at representing "all" the background knowledge people need in order to communicate. In addition, it contains more specialized encyclopaedic knowledge. Much of the knowledge represented in Cyc was indirectly built on the basis of corpora. The knowledge enterers figured out what a reader of a desktop encyclopaedia or a newspaper had to know in order to understand the texts, and encoded this background knowledge in Cyc (Lenat & Guha (1990)). However, they did not link the encoded concepts to these corpora.

Automatic extraction of ontologies or other information from unstructured or semistructured texts requires preprocessing (syntactic parsing or Part-of-Speech-tagging). Rules for discovering a hierarchy or concepts and relationships based on the input texts have to be defined. *ASIUM* (Faure & Nédellec, 1998) and *WebOntEx* (Han & Elmasri, 2001) are systems that try to build ontologies from texts.

ASIUM (Faure & Nédellec, 1998) learns from syntactically parsed domain-specific texts which concepts are allowed to take which position in a verb frame. Nouns occurring in the same syntactical role after a given verb are grouped into basic clusters; these clusters are fused with similar clusters. In parallel, the subcategorization frames of the verbs are generalized. The new general cluster of nouns is validated and labeled by the user, which results in a domain relevant hierarchy of user-specified concept names subsuming the extracted nouns. No specific concept frames are built.

WebOntEx (Han & Elmasri, 2001) is supposed to semi-automatically extract domain ontologies from manually chosen web pages. It uses the tagging of the HTML source in order to identify the most relevant concepts appearing in the page. After Part-of-Speech-tagging, candidate concepts

[2] http://www.cyc.com

```
L'ENFANT (ID:23)

La définition:
    + ressembler à
        11 l'adolescent
        10 le jeune
        3 le nouveau-né
        [...]
    + être
        4 l'être
        3 l'être humain
        [...]
    - être
        15 l'adulte
    [...]
L'importance pour l'homme:
    + avoir
        52 les parents
        40 le droit
        30 la mère
        [...]
    + être représenté par
        7 la photo
      [...]
    [...]
Les relations avec le tout englobant:
    + être l'une des parties de
        29 la famille
        5 le groupe -- minoritaire
        5 la société
        4 la communauté
        2 la famille -- élargie
        2 le groupe -- de classe
        [...]
    [...]
[...]
Les relations avec les événements englobants:
    + jouer le rôle de
        68 le travailleur
        6 le délinquant
        5 le prostitué
        4 le réfugié
        4 le victime
        [...]
    + pratiquer
        14 le jeu
        13 le dessin
        8 l'apprentissage
        6 la lecture
        [...]
    [...]
```

Figure 7: Small parts of the CHILD frame, in French.

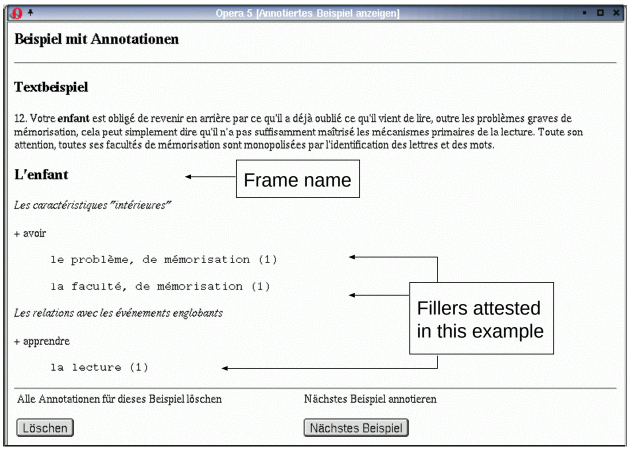

Figure 8: Overview of an annotated example.

are extracted taking nouns as candidates for types, verbs for relations, and adjectives for attributes, following Chen (1983). The extracted information has thus to be explicitly present in the texts.

Protégé (Grosso, 1999) and *OntoEdit*[3] are *tools* that can be used to build knowledge bases. The knowledge to be entered is supposed to be "known" by an expert. If the user is not an expert, he can annotate a concept with our tool and choose the most relevant information for entering into one of the professional tools. An easy way to do so would be to define a certain threshold of number of occurrences for the fillers.

4.2. Lexical resources

Lexical resources mainly do not aim at representing "typical" or "world" knowledge. For example, in the *French EuroWordNet*[4] (Vossen, 1999), most of the relations are hypernym relations. When annotating with our tool, a user is not bound to more or less conventional semantic relations between words, but he can define virtually any relation between concepts using natural language.

FrameNet[5] (Fillmore, Wooters & Baker, 2001) is structured according to cognitive principles and relies on a large general corpus. In the FrameNet approach, frames are traditionally centered around a verb. Frame elements (corresponding mainly to "case roles" of that verb) are marked in the corpus and annotated. They have to be explicitly present in a sentence. The main aim of the example sentences is to illustrate the grammatical constructions in which the frame elements can be used, and not to quantify their occurrences: not every matching example in the corpus gets annotated.

The *explanatory-combinatorial dictionary of French* by Mel'cuk et al. (1984) lists words that can be used to relate one word to another. The relations are represented as various "lexical functions". For example, the value of the lexical function $\mathbf{Oper}_1(\mathit{word})$ indicates the verb that correctly relates the first agent to the situation denominated by *word*, where the agent is the grammatical subject of a sentence and the situation is the direct object. As one of the values of the function $\mathbf{Oper}_1(\mathit{envie})$ is *avoir*, we know that we are allowed to define "envie" ("desire") as one of the fillers of a subslot with the verb "avoir" ("have"), attached to a frame which refers to the agent. Information of this "functional" kind is thus also contained in our frame system (as well as in the FrameNet data). A frame with some annotated subslot-filler entries, however, contains even more information, as for example it could say *who* typically has a desire, or even *who* typically has the desire *for what*.

5. Conclusions and future work

We have explained the background and use of an annotation tool for concept frames based on corpora. In order to store and provide reusable information, the frames have a common structure and are arranged in a hierarchy. Some default information is associated with the upper levels of this hierarchy and can be used for annotating their

[3] http://www.ontoprise.de
[4] http://www.hum.uva.nl/~ewn
[5] http://www.icsi.berkeley.edu/~framenet

subframes. The user can enter new information, which will then be available for further annotations. The names of frames, slots, subslots and fillers are defined in natural language, according to the given frame structure.

The tool is easy to use because it does not expect the user to have any knowledge of logic or formal ontologies. The resulting frames contain typical knowledge about the analyzed concepts and can be further used in discourse analysis and cultural studies. After a transformation which has to filter out the slot level of the frame structure, they can also be used as extension to existing knowledge bases. Until now, the tool has been used to annotate frames for the concepts NATURE (in different "roles"), ADVENTURE, DANGER, and CHILD.

The encoded propositions represented in the frames are inferred by the user from text examples collected from the internet. In the collection phase, the user has to define some values for parameters in order to get a corpus of examples containing information on the concept. The examples are stored in the same database together with the default frame information and the annotations. This combination allows of two further research possibilities:

1. Find out internal regularities in filler- or subslot-filler-combinations of frames ("co-occurrences" of fillers) and on which cognitive principle they might be built.

2. Look at the examples in order to find heuristic rules for automatically extracting some of the subslot-filler combinations. This task is extremely difficult if you want to infer the subslot from a text in which it is not explicitly mentioned as a verb.

We plan to build subframes for each of the lower frames in our frame hierarchy in order to verify and refine the default slots and subslots of the upper frames. We are also working on the two research issues mentioned above.

6. Acknowledgements

This work is funded by a Hamburg University grant and was repeatedly supported by the Hansische Universitätsstiftung.

7. References

A. Blank. 1999. Co-presence and Succession. A Cognitive Typology of Metonymy. In K.-U. Panther and G. Radden (eds.), *Metonymy in Language and Thought* (= Human Cognitive Processing, V. 4), pp. 169–191. Amsterdam/Philadelphia: John Benjamins.

P. P. Chen. 1983. English Sentence Structure and Entity-Relationship Diagrams. *Information Sciences* 29:127–149.

T. A. van Dijk. 1987. *Communicating Racism.* Newbury Park: Sage.

D. Faure & C. Nédellec. 1998. A Corpus-based Conceptual Clustering Method for Verb Frames and Ontology Acquisition. *Proceedings of the LREC workshop on Adapting lexical and corpus ressources to sublanguages and applications*, 5–12.

C. Fillmore. 1968. The Case for Case. In E. Back and R. T. Harms (eds.), *Universals in Linguistic Theory*, pp. 1-90. New York: Holt, Rinehart & Winston.

C. J. Fillmore, C. Wooters & C. F. Baker. 2001. Building a Large Lexical Databank Which Provides Deep Semantics. *Proceedings of the Pacific Asian Conference on Language, Information and Computation*. Available http://framenet.icsi.berkeley.edu/˜ framenet/papers/dsemlex16.pdf.

W. E. Grosso. 1999. *Knowledge Modeling at the Millennium. The Design and Evolution of Protégé-2000.* Stanford University, Stanford Medical Informatics. Available http://smi-web.stanford.edu/pubs/SMI_Abstracts/SMI-1999-0801.html.

N. Guarino. 1997. Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In M. T. Pazienza (ed.), *Information Extraction*, pp. 139–170. Berlin: Springer.

H. Han & R. Elmasri. 2001. Architecture of WebOntEx: A System For Semi-Automatic Extraction Of Ontologies From Web Pages. *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics*, X:81–86.

P. D. Karp, K. L. Myers & T. Gruber. 1995. The Generic Frame Protocol. *Proceedings of the 1995 International Joint Conference on Artificial Intelligence*, 1:768–774.

J. Klein & Iris Meißner. 1999. *Wirtschaft im Kopf.* Frankfurt a.M.: Lang.

P. Koch. 1999. Frame and Contiguity. On the Cognitive Basis of Metonymy and Certain Types of Word Formation. In K.-U. Panther and G. Radden (eds.), *Metonymy in Language and Thought* (= Human Cognitive Processing, V. 4), pp. 139–167. Amsterdam/Philadelphia: John Benjamins.

K.-P. Konerding. 1993. *Frames und lexikalisches Bedeutungswissen*. Tübingen: Niemeyer.

D. B. Lenat & R. V. Guha 1990. *Building large knowledge-based systems*. Reading, MA: Addison-Wesley.

B. Lönneker. 2001. The Frame Idea in Semantics and Knowledge Representation. *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics*, XIV:490–495.

I. A. Mel'cuk et al. 1984. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques 1.* Montréal: PUM.

M. Minsky. 1975. A framework for representing knowledge. In P. H. Winston (ed.), *The psychology of computer vision*, pp. 211–277. New York et al.: McGraw-Hill.

P. Vossen. 1999. *EuroWordNet General Document.* Version 3, Final. University of Amsterdam. Available http://www.hum.uva.nl/˜ ewn/docs/GeneralDocPS.zip.