INTERNATIONAL COMPUTER SCIENCE INSTITUTE

1947 Center St. • Suite 600 • Berkeley, California 94704-1198 • (510) 666-2900 • FAX (510) 666-2956

Multilinguality and FrameNet

Birte Lönneker-Rodman*

TR-07-001

March 2007

Abstract

The report introduces some notions relevant to multilingual resources for Natural Language Processing and summarizes the model of FrameNet, a lexical resource for English. As the development of FrameNet resources for languages other than English advances, theoretical and practical issues related to multilingual representation and the processing of such resources become apparent. The aim of this report is to characterize current practices in creating multilingual FrameNets as well as to point out issues related to the status of the frame part of FrameNet as "interlingua". The details discussed and the conclusions drawn are intended to be helpful to both designers and users of multilingual FrameNets.

*Thanks to all FrameNet staff and visitors for discussion and insights into their work, especially to Collin F. Baker, Michael Ellsworth, Charles J. Fillmore, Kyoko H. Ohara, Miriam R. L. Petruck, and Carlos Subirats. – The author is supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD).

1 Introduction

With respect to Natural Language Processing, the term *multilingual* has been applied to both data (static or representational resources) and algorithms or tools for manipulating the data or helping users manipulate it (e.g. Francopoulo *et al.*, 2006b). This report concerns representational resources, or the encoding of linguistic data. In the most general terms, multilinguality of language data refers to the existence of such resources for more than one language. Practically, multilingual data can only be compared if they represent the same kind of linguistic information: e.g., the lexicon, orthography, phonology, or syntax.

This report focuses on multilinguality of lexica as static language resources, and in particular on FrameNet, which is a lexical database with many components. The potential of FrameNet for language comparisons and multilingual applications has been discussed earlier, for example by Boas (2002, 2005). However, new theoretical and practical issues related to multilinguality and FrameNet are expected to become apparent as soon as further FrameNet resources for languages other than English will be released (see Section 4 for such resources). The aim of this report is to summarize current practices in creating multilingual FrameNets as well as to point out issues related to the status of the frame part of FrameNet as "interlingua" (Boas, 2005). The details discussed and the conclusions drawn are intended to be helpful to both designers and users of multilingual FrameNets.

The report is structured as follows. Section 2 introduces notions and issues related to multilingual language resources. Section 3 describes the model of FrameNet as a lexical resource. Combining notions and insights from these two sections, Section 4 characterizes current practices in building multilingual FrameNet-like resources and points out possible problematic issues. Section 5 concludes by proposing future directions for both the English FrameNet as such and the linking of resources in other languages to it.

2 Multilinguality of Data Resources

This section discusses notions related to multilingual language resources, taking the level of the lexicon as an example. The expressions *building block*, *component* and *(structural) entity* will be used interchangeably to designate abstract information types individuated by the design of a given resource. The existence and function of a certain building block depends on the particular theory or approach underlying the resource. To allow for cross-lingual comparison, at least some of the building blocks within each resource of the multilingual set should have the same function.

In order to make the discussion easily accessible, a simple example resource is introduced in Subsection 2.1. It will be referred to throughout the remainder of the section, which includes a discussion of a descriptive scale from organizational similarity to interrelatedness of multilingual resources (2.2), ways of establishing (2.3) and representing (2.4) interrelatedness, and notions used to characterize interrelatedness (2.5).

2.1 Example Model of a Lexical Resource

A very simple lexicon model derived from the design of WordNet (Miller *et al.*, 1990) will illustrate the discussion of central notions. The building blocks of the example model are as follows:

- **synset:** a set of near-synonyms; for example, the set of nouns {`land:2, ground:7, soil:3`} with synset number 09335240 and the meaning (gloss) "material in the top layer of the surface of the earth in which plants can grow (especially with reference to its quality or use)";[1]

- **variant:** a word sense as an integral part of a synset; for example, the word sense `land:2` or `ground:7`, both of which are members of the above synset and share its meaning;
- **has_hyponym relation:** a directed semantic relation between two synsets indicating hyponymy; for example, the relation between {`land:2, ground:7, soil:3`} and {`bottomland:1, bottom:6`} (gloss: "low-lying alluvial land near a river") can be rendered as {`land:2, ground:7, soil:3`} `has_hyponym` {`bottomland:1, bottom:6`}.

These components will be represented by the graphical symbols shown in Figure 1.

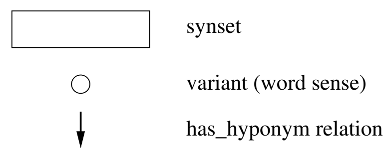

Figure 1: Components of the example lexicon model.

2.2 Two Ends of the Multilinguality Scale

At the particular level(s) of information each of them encode, a set of language resources can exhibit different degrees of relatedness. In what follows, two ways of relatedness are distinguished: organizational similarity (2.2.1) and interrelatedness (2.2.2). They should be understood as two ends of a scale rather than as excluding principles.

2.2.1 Organizational similarity

Organizational similarity is exhibited by multilingual language resources that use the same *organizational principle*. For example, multilingual WordNets exist which have been designed along the lines of the original Princeton WordNet (Miller *et al.*, 1990) and therefore recognize the same structural entities. Still, these resources are likely to differ from their model resource. First of all, the overall numbers (of instances) of synset nodes and relations are not necessarily the same across languages. Second, even if these numbers are equal, the distribution of the components can be different within a relational resource: The hierarchical position of a synset designating the same concept may not be the same cross-linguistically. Figure 2 illustrates this state of affairs by representing different structures of corresponding lexicon areas in Language A (left) and Language B (right). Due to the different relational arrangement of synsets, it is impossible to conclude which of them refer to the same concept and therefore correspond cross-linguistically, from inspecting their position alone – although certain equivalences are more likely than others.

As Pianta *et al.* (2002) point out, differences in multilingual lexicons (covering the same semantic area) are usually due to two factors: language specific differences in the structure of the semantic space, as given by the lexicon of the language, and individual design decisions during resource creation.

[1] This example has been drawn from the current WordNet database, 3.0, `http://wordnet.princeton.edu/` [28 February 2007].

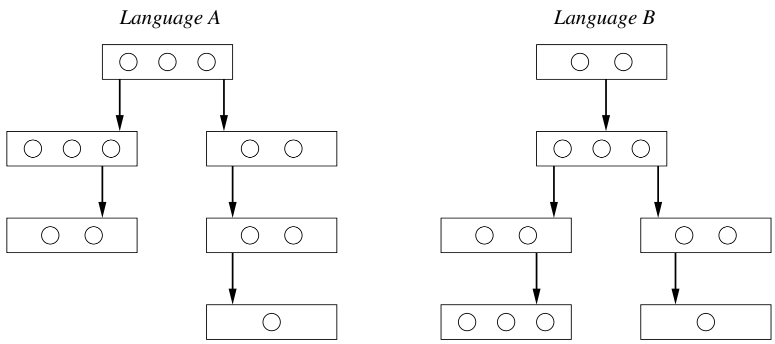

Figure 2: Organizational Similarity.

2.2.2 Interrelatedness

Resources covering the same kind of linguistic information and exhibiting the same organizational principle can further be interrelated. The interrelation, also called *linking*, *mapping*, or *alignment*, can be implemented (represented) at different levels and in a variety of ways.

We can think of two organizationally similar resources as being minimally interrelated by one global link between the resources as a whole. However, true *interrelatedness* involves cross-lingual links between internal components of the resources. Figure 3 displays organizationally similar resources that are also interrelated at the level of synsets (see dashed lines).

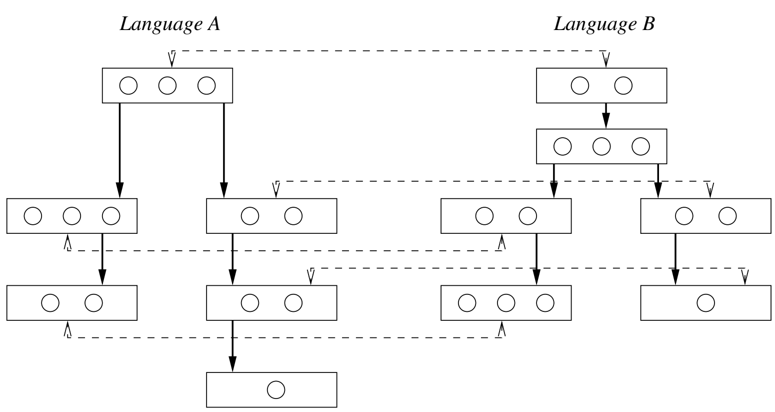

Figure 3: Interrelatedness.

2.3 Establishing Interrelatedness

Interrelated language resources can be subdivided with respect to the way in which links are established. In the course of the EuroWordNet project (Vossen, 1999), which aimed to build an interrelated multilingual WordNet for eight European languages, the project team distinguished two methods, which they called the Merge approach (2.3.1) and the Expand approach (2.3.2).

2.3.1 Merge approach

When the *Merge approach* is adopted, independent resources for different languages are first built from scratch. Later, links that relate selected types of components (e.g. synsets) cross-linguistically are added (Vossen, 1999). Figure 4 schematically illustrates this method.

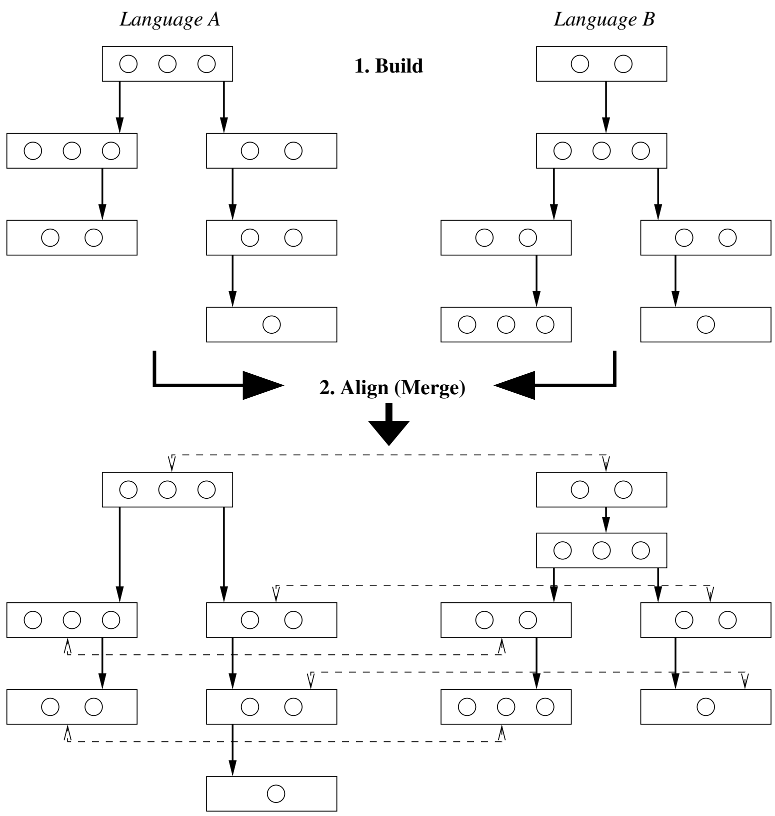

Figure 4: The Merge approach.

In the opinion of Pianta *et al.* (2002), this approach potentially involves difficulties during the alignment phase. They argue that independently built resources might be more divergent than linguistically necessary because resource designers are free to make different decisions at a more abstract (world-knowledge related) level.

2.3.2 Expand approach

With the *Expand approach*, a resource for one language, which is regarded as stable at that time, is transferred to another language. This approach has been called *Expand approach* (Vossen, 1999). This implies that initially, the overall structure of the resource is kept unchanged and only obviously language specific information is replaced. With respect to lexicons, it is assumed that conceptual building blocks stay in place whereas lexical building blocks change (are given new values, disappear or are added). For example, when a given WordNet is transferred into another language, the Expand approach at first maintains the original number of synsets and the semantic relations between them; it only replaces the members of a synset by word senses in the new language, wherever available, or adds new synset members where appropriate. Direct links between the building blocks (synsets) are created immediately at the time of this original transfer. (Step 1 in Figure 5.)

A problem for this method is the well-known fact that any given two languages might lexicalize different concepts, and therefore also the relations between valid building blocks vary from language to language. For this reason, the Expand approach requires a subsequent "cleaning" of the resource resulting from the initial transfer (see step 2, "adapt", in Figure 5). At least some areas of the new resource (Language B) are likely to be in need of restructuring, and restructured lexicon areas lose at least some of the initial direct equivalence links to the model resource.

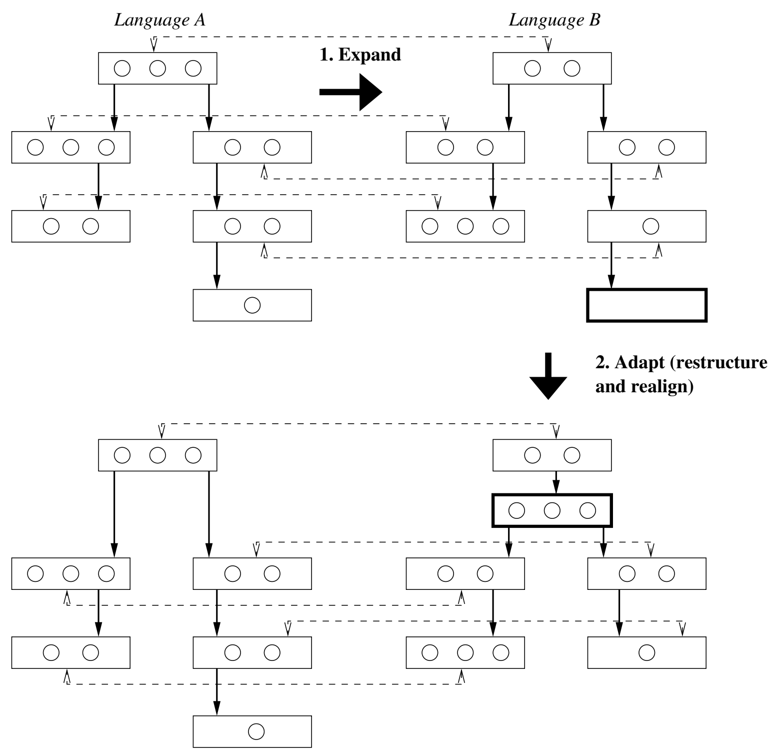

Figure 5: The Expand approach.

The Expand approach tends to produce structurally highly similar resources. A problem is that this can entail the risk of neglecting language-specific differences in lexicalization and therefore in the structure of the lexicon. In particular, as a result of the translation of lexicalizations proper to Language A, "artificial" lexical labels (such as infrequent multi-word units) might appear in Language B after the Expand step (cf. the empty synset in Language B in Figure 5, which might have been unappropriately labelled with an infrequent multi-word). Also, unless the adaptation step is supported by further monolingual evidence from Language B, concepts lexicalized in this language alone (but not in A) are likely to be missed (cf. the newly introduced synset without interrelation links in Figure 5). In summary, the Expand approach tends to produce structurally highly similar resources, at the risk of neglecting language-specific differences in lexicalization and, therefore, the structure of the lexicon.

2.4 Representing Interrelatedness

Interrelated multilingual language resources contain information on which monolingual instances of their building blocks correspond to each other. Different linking methods are available and choosing one is a design decision that depends on the individual properties and applications of the resource. Such design issues include the directness of the relation (2.4.1) and the available set of cross-lingual relation types (2.4.2).

2.4.1 Directness

The general nature of the relatedness between multilingual resources can be direct or indirect. A *direct mapping* relates two corresponding entities in language A and language B. An *indirect mapping* uses an *intermediate index*, also called inter-lingual index (Vossen, 1999) or shared axis (Francopoulo *et al.*, 2006a). The index is represented separately from any of the individual languages and the mapping contains information about which language-specific entities in languages A and B correspond to which entity in the intermediate index. Figure 6 represents the two approaches.

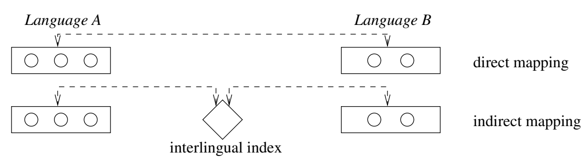

Figure 6: Directness of Mapping.

When more than two languages are to be interrelated, the intermediate index helps to keep the number of relations manageable. If one of the languages is selected as interlingua directly, there are two problems: lexical gaps in that language; and difficulties to keep the development aligned if the "interlingua-language" resource changes.

In EuroWordNet, the inter-lingual index (ILI) is based on the English WordNet 1.5. Concepts lexicalised in a given language but not directly represented in the ILI are defined by adding more than one semantic relation to the ILI, and/or by considering its relation to the ILI as well as its language-internal relations to other concepts. New versions of the English WordNet can still be related to the EuroWordNet ILI and therefore to the individual languages, based on incremental mappings to WordNet 1.5 that were automatically generated to align different versions of WordNet itself. However, current projects working on resources linked to WordNet usually align their data directly to the most current version of WordNet rather than to the ILI index (Piek Vossen, p.c., 4 March 2007).

Technically, both direct and indirect relatedness can be achieved in at least two ways: by assigning the same index (ID) to corresponding entities or by introducing an additional representational entity into the multilingual framework that explicitly represents the relation. The latter could be, for example, a mapping table containing the IDs of each corresponding entry. This solution also allows storage of additional information concerning the relation itself, further discussed in the next Subsection (2.4.2).

2.4.2 Cross-lingual relation types

Cross-linguistically as well as language-internally, relations between entities can be of different types. The most widely used relation across languages is Equivalence, by which the related entities are declared to have the same meaning, as in traditional dictionaries. Cross-lingual Equivalence of meaning is difficult to define in general, as each semantic theory might have a slightly different approach to this question.[2]

Because of different lexicalization patterns, there can be entities for which there is no direct equivalence, irrespective of the approach adopted to create the resource. If no relations besides Equivalence are available, such entities cannot be linked to anything outside the monolingual resource. The meaning of a cross-linguistically unrelated entity can then only be inferred from language-internal information, combined with equivalence information on entities that are internally related to the entity in question.

To allow for a more flexible cross-linguistic alignment, multilingual resources can define other types of relations across languages besides Equivalence. For example, EuroWordNet defines the cross-lingual hyponymy relation EQ_HAS_HYPONYM, among others (Vossen, 2004, p. 165). Figure 7 exemplifies this relation in a direct mapping environment, where it is used to relate a more general concept in language A to a slightly more specific concept, lexicalised in language B only. An instance of the same relation has also been used to relate a more general concept in Language B to a concept lexicalised in Language A only. The language-specific synsets are provided with stronger outlines of their boxes in the figure. – A further example of non-equivalent relations is that of typed links provided with tests representing different selectional restrictions within the Lexical Markup Framework, as explained in Francopoulo *et al.* (2006a).

2.5 Characterizing Interrelatedness

This section introduces the notions of mapping coverage (2.5.1) and mapping granularity (2.5.2), each characterizing an aspect of interrelatedness.

2.5.1 Mapping coverage

The term *mapping coverage* is used here to designate the ratio of instances of building blocks provided with a cross-lingual link, to all instances of this building block in a given resource. The building block must be in principle available for interrelation.

Mapping coverage cannot be complete (100%) if no appropriate link types are provided for components that do not have a direct equivalent, in the sense of equivalence adopted by the resource designers. The inclusion of different types of cross-language relations besides Equivalence and the actual implementation of relevant instances of these relations then increases mapping coverage. For example, the usage of the cross-linguistic hyponym relation allows the highlighted concepts in Figure 7 to have cross-linguistic links, which they would lack otherwise. In that example, Equivalence relations only account for a mapping coverage of 5 out of 6 synsets (83.3%) in both directions. Counting both Equivalent and cross-linguistic Hyponym relations, however, mapping coverage is 100%, as each synset is provided with at least one cross-lingual link.

[2] As far as the EuroWordNet model is concerned, cross-lingual equivalence between two synsets does not necessarily mean that they also display the same set of language-internal relations to other concepts, as some of those might not be lexicalised in all languages (cf. Vossen, 2004, p. 162).

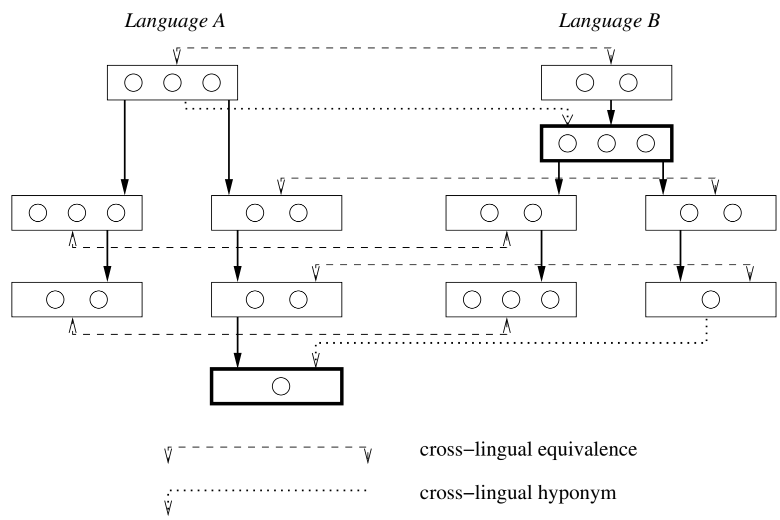

Figure 7: Non-equivalence cross-lingual relations.

2.5.2 Mapping granularity

Mapping granularity refers to the level and number of building block *types* that get matched.

For example, in the simple WordNet model discussed above, synsets, variants, and the semantic relation of hyponymy have been identified as building blocks. Within EuroWordNet and MultiWordNet (Pianta *et al.*, 2002), equivalences between multilingual WordNets are implemented at the synset level. In principle, however, Equivalence links could also be created between variants, which would add to the number of matched entity types (building block types) and thereby contribute to a fine-grained mapping. The degree of granularity of a mapping depends on the types of buildings blocks which are available as well as on the purpose of the mapping. For example, an alignment of variants in the WordNet model will be redundant to the synset mapping in many cases. But it could be helpful in other cases, because synonymy within a synset is not complete: for example, certain usage conditions such as register are disregarded in some WordNets.

Only those entities that are conceived as building blocks of the resource model are available for mapping in the first place. For example, it is not possible within the WordNet model to establish equivalence between morphemes, because they are not accessible individually.

Figure 8 illustrates cross-lingual relations at different levels of granularity, from the "trivial" interrelatedness between the resources themselves to the relatively fine-grained level of word senses (variants or word form-sense pairs).

3 The FrameNet Model

In the previous section, a very simple lexicon model has been used for illustration of multilinguality concepts. This section describes the FrameNet model, particular multilinguality aspects of which will be studied in Section 4 below.

In FrameNet, a word sense is represented as a Lexical Unit (LU). A Lexical Unit has a name and a short textual definition but is defined above all as a combination of semantico-conceptual and morpho-syntactic information. The latter comprises part of speech, word forms and order of units

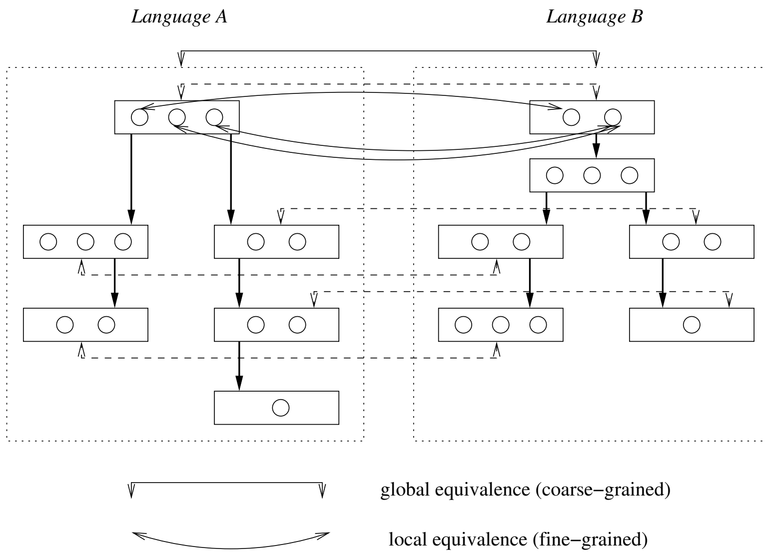

Figure 8: Relating fine-grained components.

within multi-word terms; this information has also been called the Lexical Database of FrameNet (Ruppenhofer *et al.*, 2006, p. 130) and will not be further discussed in this report. LUs are linked to semantico-conceptual information by membership in a particular *frame*, the background for understanding occurrences of the Lexical Unit in running text. FrameNet analyses and annotates occurrences of LUs with respect to their frame-semantic and syntactic behavior inside a sentence. A set of such analyses taken together show the semantic and syntactic combinatorial profile of the LU. The part of the database that stores the annotations has been referred to as the Annotation Database (Ruppenhofer *et al.*, 2006, p. 130). The internal structure of the annotations will not be explained in this report.

The part of the FrameNet database containing the semantico-conceptual information in the form of frames has been called the Frame Database (Ruppenhofer *et al.*, 2006, p. 130). It will be at the center of the model presented in the following, which comprises:

- **Frames** provided with a name, a definition, and a semantic type, and having frame elements (Subsection 3.1);
- **Frame elements (FEs)** provided with a name, a definition, a semantic type, and a coreness status (Subsection 3.2);
- Frame-internal **relations between frame elements** (internal FE-to-FE-relations, Subsection 3.3);
- Directed **relations between frames** and their frame elements, exemplified by the Inheritance relation (Subsection 3.4);
- **Lexical Units** with a name and a description, membership in a frame, and a relation to the Lexical Database not shown in this model.

The building blocks of the model are represented in UML style in Figure 9.

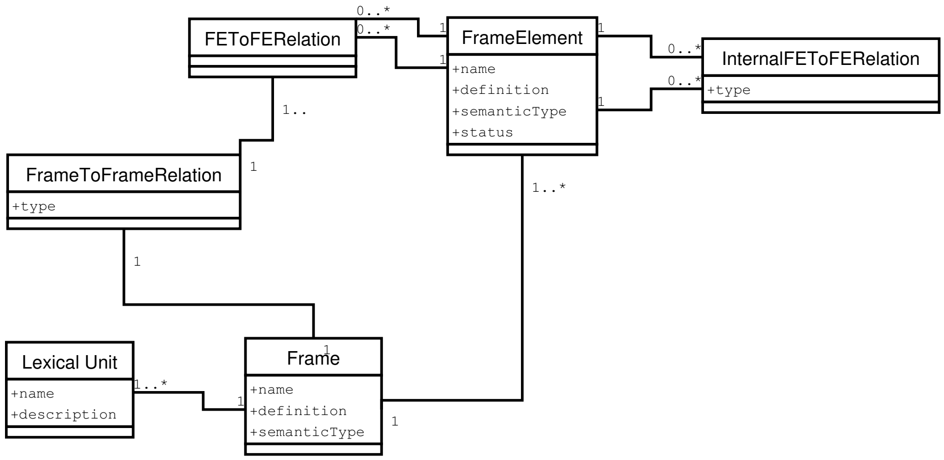

Figure 9: FrameNet model of frames

The FrameNet model as presented here does not include all features of the original FrameNet design. Semantic types are treated as attributes here whereas in FrameNet, they are themselves arranged into a hierachy; frame relations are limited to Inheritance, but FrameNet defines several other frame relations; and implicit frame elements, so-called null instantiations, are not taken into account in what follows. For a comprehensive description of FrameNet, the reader is referred to Ruppenhofer *et al.* (2006).

A slightly different and still more simplified version of the FrameNet model can be found in Francopoulo (2005). That work maps the FrameNet model onto the Lexical Markup Framework (LMF) model but ignores the semantic types and FE-to-FE-relations of FrameNet, which however play an important role in the definition of frame-to-frame-relations.

Quantifications provided in the subsequent sections indicate the pervasiveness of the described entity, attribute, or value. The numbers should be seen against the background figure of 831 frames in the FrameNet database, at the time of this investigation (a stage between data releases 1.3 and 1.4).

3.1 Frame

A frame is a conceptual structure that describes a particular type of situation, object, or event along with its participants and props, which are referred to as frame elements (Ruppenhofer *et al.*, 2006, p. 5). In spite of its conceptual nature, a frame cannot be defined without knowledge of the Lexical Units that are said to *evoke* the frame; for example, both *ache.v* and *itch.v* (among others) evoke the `Perception_body` frame. Frame definition is an iterative process that involves corpus work to find and study potential Lexical Units belonging to the frame.

An example of how the meaning of Lexical Units and frame definitions depend on each other is the restriction imposed on a frame, stating that all its LUs should entail the same set of stages and transitions (Ruppenhofer *et al.*, 2006, p. 13). For example, the situation of the `Communication` frame (containing LUs such as *communicate.v*) is defined as “a Communicator conveys a Message to an Addressee” and therefore includes a stage when the Addressee actually receives the Message. Only Lexical Units entailing this should belong to the `Communication` frame; others require a different frame. For example, the verb *e-mail* does *not* evoke `Communication` because it does not entail that the Addressee actually receives the Message (Ruppenhofer *et al.*, 2006, p. 17).

The following subsections describe the name (3.1.1), definition (3.1.2), and semantic type (3.1.3) of a frame.

3.1.1 Frame Name

A frame name is a mnemonic that stands for the frame, a convenient "handle" used besides a numeric ID. Frame names should be unique. Only some parts of frame names are conventionalized, such as the ending `_scenario` which is used for certain general frames not conveying perspective.

3.1.2 Frame Definition

A frame definition is a description of the situation or object represented by the frame. Usually, a frame definition mentions at least the most central frame elements and their relation to each other.

Some frame definitions also summarize the syntactic behavior of Lexical Units in the frame. For example, the `Scrutinity` frame definition states:

(1) Some words in this frame allow alternate expressions of the Ground and the Phenomenon.

Alternatively or additionally, some frame definitions generalize the syntactic realization pattern of certain frame elements across all LUs. Here is an example from the `Appearance` frame:

(2) In this class of perception words, a Phenomenon, *typically expressed as External Argument*, and its perceptual characteristics are given some description. [Italics added]

Scheffczyk (2006) states a number of consistency rules that allow the detection of frame definitions not mentioning core FEs, among other things. The rules are a first step towards the set-up of more formal restrictions on the format of frame definitions.

3.1.3 Semantic Type of Frames

There are two categories of semantic types available for frame labeling: general semantic types and framal types.

A general semantic type attached to a frame indicates that all LUs evoking that frame should have this or a more specific semantic type. For example, the semantic type Artifact is attached to the frames `Clothing` and `Containers`, among others. So, all the LUs in these frames should designate artifacts (cf. Ruppenhofer *et al.*, 2006, p. 111).

The so-called framal types are semantic types that can be attached to frames only, not to Lexical Units; currently, these are restricted to Non-lexical frame and Non-perspectivalized frame. They are regarded as meta information on the frame itself: A Non-lexical frame does not contain LUs[3] and a Non-perspectivalized frame does not convey perspective (cf. Ruppenhofer *et al.*, 2006, p. 112).

[3] This only superficially contradicts what has been said above in the introduction on frames (3.1). A Non-lexical frame can appear as Parent frame within the frame Interhiatnce relation (cf. Subsection 3.4 below); its direct or indirect Child frames should still have Lexical Units.

Frames are not required to have a semantic type. 119 frames in the current FrameNet database have been provided with a semantic type label, 92 of which are framal types. However, frames can accumulate semantic type labels; for example, the framal types can be combined.

3.2 Frame Element

A frame element is named (3.2.1) and defined (3.2.2) with respect to the general conceptual situation described by the frame. The behavior of Lexical Units evoking a frame influences the definition of frame elements of that frame. All LUs of a frame must have the same number of frame elements, and those must be of the same semantic type (3.2.3) (Ruppenhofer *et al.*, 2006, p. 11). Special treatment of frame elements can however be expressed by frame element relations, as described in Section 3.3 below. Frame elements are further classified according to their coreness status (3.2.4).

3.2.1 Frame Element Name

A frame element name is a mnemonic handle standing for the definition of a frame element, besides its unique ID. In contrast to frame names, frame element *names* are not unique. Rather, frame elements mapped onto each other in frame-to-frame relations such as Inheritance (Section 3.4 below) are often named the same, in order to reflect their similarity in function within the frame.

3.2.2 Frame Element Definition

A frame element definition describes the meaning of the frame element with respect to the frame it belongs to. Sometimes, the definition of an FE also refers to other frame elements of the same frame.

Each frame element definition should contain at least one example sentence. Inside the example sentence, the frame-evoking LU and the frame element being explained are tagged using idiosyncratic XML markup. Tools for viewing FrameNet data convert this markup into presentational devices; for example, a stretch of text corresponding to a frame element can be provided with a distinctive background color. Example 3 is an example sentence illustrating the Cognizer_agent frame element of the `Seeking` frame, as found in the definition of this FE.

(3) <fex name="Cognizer_agent">John</fex> <t>looked</t> for his pencil.

Some frame element definitions also generalize the syntactic realization pattern of the frame element across Lexical Units. Here is an example from the definition of Goal in the `Filling` frame: "Goal is generally the NP Object in this frame". This reflects the observation that LUs of the same frame often overlap in the way they realize their arguments syntactically (cf. Ruppenhofer *et al.*, 2006, p. 128), but each specific generalization of this kind must be based on corpus investigation or existing annotations.

Apart from the presence of examples, frame element definitions do not have a canonic internal structure. For example, there can be more than one example sentence, and there might be comparisons to other frame elements. As a contribution to the standardization of the format of frame element definitions, Scheffczyk (2006) states a number of consistency rules. Among other things, the rules can detect frame element definitions without example sentences.

3.2.3 Semantic Type of Frame Elements

A semantic type on a frame element constrains the type of the semantic head word of a syntactic phrase which realizes or "fills" the role of this frame element, in any annotated sentence. For example, the semantic type Sentient has been attached to many FEs called Agent, including the one in the `Communication` frame. Therefore, all occurrences of LUs evoking `Communication` should show an Agent in their annotations which is a Sentient. In Example 4, the noun phrase *the catechist* is annotated as Agent and the head noun *catechist* meets the semantic type requirement of being a Sentient.

(4) The catechist communicated with Todd by way of simple communication and a flannel board.

Frame elements are not formally required to have a semantic type but if they do, this also places certain restrictions on frame elements in other frames they are related to (see Section 3.4). Usually, there is no more than one semantic type on a frame element, but the FrameNet software allows and handles more. For example, by multiple Inheritance (Section 3.4), more than one semantic type on a frame element can be inherited and both of them could be kept in the Child frame. There is a rather limited set of semantic types for frame elements; currently, 28 distinct types are attached to 3,678 frame elements.

3.2.4 Coreness Status

Each Frame Element must be provided with a coreness status. This information pertains to the conceptual centrality of the entity the FE represents, with respect to the frame as such. There are three main coreness statuses: core, peripheral, and extra-thematic.

Core frame elements are defined as instantiating "a conceptually necessary component of a frame, while making the frame unique and different from other frames" (Ruppenhofer *et al.*, 2006, p. 26). Nevertheless, to make the notion of conceptual necessity practically applicable, some formal properties are considered as well. These properties are: frame LUs require overt specification of the frame element; when omitted, the frame element receives a definite interpretation; and the semantics of the frame element are unpredictable from its form, in particular from any marking prepositions (cf. Ruppenhofer *et al.*, 2006, pp. 26–27).

In contrast to core FEs, **peripheral** FEs do not uniquely characterize a frame. They mark such notions as Time, Place, Manner, Means, and Degree, which can be instantiated in any frame that is semantically appropriate. To distinguish peripheral Frame Elements from extra-thematic ones, it is important to know that peripheral FEs do not introduce additional, independent or distinct events from the main reported event, represented by their frame (cf. Ruppenhofer *et al.*, 2006, pp. 27–28).

Finally, **extra-thematic** frame elements introduce another state of affairs as a backdrop against which the frame is situated. Because of the existence of this additional state or event, which is actually a different frame, extra-thematic frame elements are understood not to conceptually belong to the frames they appear in. Unlike core and peripheral frame elements, extra-thematic frame elements do not have a frame-specific understanding (cf. Ruppenhofer *et al.*, 2006, pp. 27–28; 127). Extra-thematic frame elements are applicable to even more frames than peripheral frame elements and can be given fairly general definitions on their own (see Appendix A in Ruppenhofer *et al.* (2006)). Example 5 is the general definition of the extra-thematic frame element Beneficiary, as given by Ruppenhofer *et al.* (2006, p. 140). Beneficiary appears as an extra-thematic FE in the frames `Creating` and `Placing`, among others.

(5) [Beneficiary] applies to participants that derive a benefit from the occurrence of the event specified by the target predicate. Further, the target predicate should involve some sort of agent that intends that the benefit go to the Beneficiary.

Core-unexpressed is a fourth value to indicate coreness status. In the frame where this frame element appears, it is core and it is indeed also expressed. The "unexpressed" status is important when it comes to Inheritance; therefore, this status is further discussed in Section 3.4 on Inheritance.

3.3 Relations between Frame Elements

As mentioned above, all LUs of a frame must have the same frame elements (Ruppenhofer *et al.*, 2006, p. 11). But this requirement, even if reduced to equivalence of *core* frame elements only, is sometimes difficult to fulfil. In some cases, respecting it strictly would lead to the creation of numerous specific frames with only very few LUs. Therefore, frame-internal relations between frame elements are used to make the requirement more flexible. There are currently three types of such frame element relations: CoreSet, Requires, and Excludes.

A **CoreSet** or **coreness set** groups together two or more core frame elements such that sentences are informationally complete and pragmatically felicitous with only a subset of them expressed. Those frame elements have an informational and conceptual interdependence. For example, Source, Path, Goal, and Direction are grouped together as a CoreSet in the `Self_motion` frame. Any annotated sentence for any of the LUs in `Self_motion` exhibits between one and four of these CoreSet frame elements. That means that elements in a CoreSet are logically connected via "or" (Scheffczyk, 2006, p. 47). CoreSet relations appear in 228 frames.

The **Excludes** relation groups frame elements that are conceptually related such that if one of them appears in a sentence, the other(s) cannot; logically, this can be expressed by an "xor" connection between them (Scheffczyk, 2006, p. 48). Typically, but not always, an Excludes relation comes in pairs, which makes FE exclusion reciprocal. Currently, Excludes relations are found in 112 frames. Specific reasons for having an Excludes relation are:

- reciprocal vs. non-reciprocal construal of the frame situation, for example of `Forming_relationships`. The first alternative uses one frame element (e.g. Participants) while the second alternative uses two different frame elements instead (e.g. Participant_1 and Participant_2), see Examples 6–7:

  (6) [Her parents$_{Participants}$] separated and later divorced.

  (7) [Mandela$_{Participant_1}$] under pressure to separate [from "wayward" wife$_{Participant_2}$]

- situations such as `Killing` that can be brought about either by a sentient Agent or a non-sentient Cause, see Examples 8–9:

  (8) [A person$_{Killer}$] killed two patients by applying corrosive plasters to their chests.

  (9) [The fiery blast$_{Cause}$] killed everyone on deck instantly.

- other frame elements that can take the form of either an event or a participant in the event, such as Capture (event) vs. Pursuer (participant) in the `Evading` frame (Examples 10–11):

(10) They have been padlocked by owners fleeing [militant kidnappings$_{Capture}$].

(11) The imperial ambassador fled [an enraged crowd$_{Pursuer}$].

A **Requires** relation between two frame elements indicates that if one of them is realized in a sentence, the other one must also. Logically, they are connected via "implies" (Scheffczyk, 2006, p. 49). Usually, this relation involves frame elements that also have an Excludes relationship to others. Typically, but not always, Requires relations come in pairs, which makes the relationship reciprocal rather than unidirectional. Requires relations are found in 49 frames.

3.4 The Frame Inheritance Relation

Inheritance is the strongest relation between frames, linking a more general Parent frame to a more specific Child frame. Its semantics is laid out in Subsection 3.4.1. Special attention needs to be paid to core-unexpressed frame elements (3.4.2) and to the restrictions applied to frame-internal frame element relations (3.4.3). Inheritance can also be compared to LU membership in a frame (3.4.4). FrameNet allows multiple Inheritance: More than one Parent can be defined for a specific frame if this is considered necessary.

3.4.1 Inheritance Semantics

With the Inheritance relation, "anything which is strictly true about the semantics of the Parent must correspond to an equally or more specific fact about the Child," (Ruppenhofer *et al.*, 2006, p. 104–106). This explicitly includes the inheritance of all frame elements (except, possibly, for extra-thematic ones) along with their semantic type and coreness type (Scheffczyk, 2006, p. 55), and inheritance of the general semantic type of the frame itself. With respect to frames, Baker *et al.* (2003) have characterized this type of inheritance as monotonic. Frame Inheritance has also been compared to Is-a relations in ontologies (Ruppenhofer *et al.*, 2006, p. 8).

3.4.2 Core-Unexpressed Frame Elements

The coreness status **core-unexpressed** is an exception to the requirement that all frame elements should be inherited via the Inheritance relation between frames. FEs of coreness status core-unexpressed are not inherited. Rather, they are lexically incorporated into the meaning of the LUs evoking the Child frame. An example is provided by Ruppenhofer *et al.* (2006, pp. 28–29) which refers to the frame element Act of the very general `Intentionally_act` frame. This frame contains Lexical Units such as *do.v* and *perform.v*. The Act frame element is core with these Lexical Units, as it must be expressed (cf. Example 12).

(12) [I$_{Agent}$]'ll do [the vacuuming$_{Act}$].

In frames that inherit from `Intentionally_act`, the Act FE is not expressed separately from the Lexical Units. Rather, the LUs of those frames incorporate the Act FE into their semantics, as they themselves designate specific actions. Example 13 illustrates the occurrence of an LU from the `Intentionally_Create` frame, a direct Child of `Intentionally_Act`, where the Parent FE Agent is mapped onto the frame element Creator in the Child, but no overt Act FE is possible.

(13) [Tolkien$_{Creator}$] created [a mythical world$_{Created_entity}$].

A logically possible consequence of the core-unexpressed status might be the existence of Inheritance relations lacking associated frame element mappings altogether. The FrameNet database does not show any such cases currently.

3.4.3 Frame Element Relations within an Inheritance Situation

In frames linked to each other by Inheritance, some restrictions on frame-internal frame element relations (Section 3.3) follow logically from the definition of the Inheritance relation itself. The requirement that any fact about a Parent must be equally or more specifically represented in the Child implies special rules for frame elements that are frame-internally linked via frame element relations. Formal consistency conditions for the behavior of frame element relations under Inheritance are defined in Scheffczyk (2006).

Of all frame elements related via a CoreSet relation in the Parent, at least one should be present in the Child frame. Thus, a CoreSet relation in a Parent can also get removed in a Child. In fact, the removal of elements from a (disjunctive) CoreSet as well as the replacement of the CoreSet disjunction by the default conjunction of frame elements make the Child frame more constrained and therefore more specific, which complies with the definition of the Inheritance relation. Similarly, frame developers are not allowed to bind more elements into the CoreSet of a Child frame than there are in the corresponding CoreSet of the Parent frame: for each frame element in a Child CoreSet, there must be at least one corresponding frame element in a CoreSet of the parent frame.

If the Parent frame has mutual Excludes relations between two frame elements, the most important rule to observe is that if one of them is core, it must be inherited; further, frame elements that mutually exclude each other in a Child frame must also exclude each other in the Parent frame, *if they both exist*. This means that a Child can introduce a new Excludes relation only if at least one new frame element is introduced. However, the Child frame can drop an Excludes relation that exists in its Parent.

A frame element requiring another one by Requires in the Child frame must also require it in the Parent frame, if they both exist in the Parent. Requires relations in the Parent can be dropped in the Child frame, where the respective frame elements would then be connected via the default conjunction between FEs. Child frames can also add Requires relations to frame elements that they introduce (i.e. that do not exist in the Parent).

3.4.4 Inheritance and LU Membership

As has been explained in Section 3.1 above, facts about a frame are defined in such a way that they describe the behavior of all Lexical Units in it. Conversely, this can also been seen as a requirement for the Lexical Units evoking the frame to fulfil all its conditions in terms of frame elements, semantic types, and relations between frame elements. In some ways, the behavior of a Lexical Unit can be more restricted than the frame allows. For example, certain LUs in a frame might use exclusively one of two frame elements related to each other via Excludes. This makes LU membership conditions very similar to Child frame conditions: "one may [...] consider LU membership in a frame to be an identical relationship to Inheritance from a frame," (Ruppenhofer *et al.*, 2006, p. 127). Still, a Lexical Unit and a Child frame are fundamentally different insofar as an LU is an *instance* of the semantic class described by the frame, whereas a Child frame is itself a class and can have its own instances. Figure 10 illustrates this.

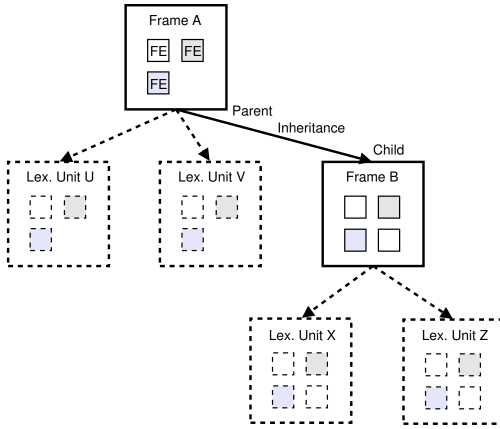

Figure 10: FrameNet model of Inheritance.

4 Multilingual FrameNets

This section discusses aspects of multilinguality of the FrameNet model and of non-English FrameNets under construction. At first, some consequences will be derived from the FrameNet model itself, with respect to the organizational similarity of multilingual FrameNets (4.1) and their potential for interrelatedness (4.2). A discussion of interrelatedness of FrameNets necessarily includes the issue of the status of frames as interlingua. This question will be dealt with in terms of the possible interpretation of an Equivalence relation between frames. In Subsection 4.3, the current practice of expanding the frame part of the FrameNet database and its consequences will be summarized.

Examples concern mainly Spanish (Subirats and Sato, 2004) and Japanese (Ohara *et al.*, 2003, 2004). Further examples have been provided by the SALSA project which produces FrameNet-style annotation of German texts (Burchardt *et al.*, 2006). A contrastive case study of German and Slovenian conducted by the author revealed similar types of examples, some of which are mentioned in what follows.

4.1 Organizational Similarity of FrameNets

In terms of the terminology introduced in Section 2, multilingual databases built along the FrameNet model are *organizationally similar*; this includes the lexicon and annotations (cf. Ruppenhofer *et al.*, 2006, p. 130). In principle, FrameNets allow for cross-lingual relations at many levels, including those of frames, lexical units, lemmas, or even word forms and annotated sentences. It is however generally understood that the frame part of the database (presented in Section 3 above) has the highest degree of inherent equivalences across languages and lends itself most easily to interrelatedness. Nevertheless, parts of the frame and frame element definitions are still language specific and therefore excluded from inherent equivalence (4.1.1).

4.1.1 Organizationally Similar Components of the Frame Database

The existence of inherently language specific parts within the frame and frame element definitions restricts the relationship between definitions to organizational similarity. Currently, only entire definition texts can be addressed as building blocks of the FrameNet model. Language-dependent parts thereof are not always marked as such; they are not building blocks of their own.

The language specific aspects of definitions include example sentences and generalizations of syntactic realization patterns. The English examples should be replaced in FrameNets for other languages by original examples from those languages that fulfil the same function; in other words, example sections are organizationally similar but not necessarily semantically equivalent.

Similarly, the summary of syntactic realization patterns provided in some frame and frame element definitions needs to be verified and possibly adapted to reflect syntactic facts in other languages. For example, the generalization in the English `Perception_body` frame that the possessor of the body part usually appears as a possessive determiner[4] does not hold for Slovenian, where the possessor of the body part is expressed as direct object (Example 14) or in a prepositional phrase, depending on the LU.

(14) Noge me bolijo.
Legs me hurt.
My legs hurt.

4.2 Equivalence and Non-equivalent Interrelatedness at Frame Level

Frames are presumably language independent to a fair degree (Ruppenhofer *et al.*, 2006, p. 130). Therefore, an alignment of FrameNets via frames as interlingua has been proposed by Boas (2005). Nevertheless, many building blocks of the frame model are defined at least partly with respect to the behavior of Lexical Units in the frame. This introduces the possibility of language specific differences also at frame level, which result in deviations from (cross-linguistic) Equivalence between frames. This section discusses possible reasons for non-equivalence, from the coarse-grained level of frames themselves (as identified by frame names; Subsection 4.2.1) to finer-grained levels: the list of frame elements of a frame (4.2.2), the coreness status and semantic type of frame elements (4.2.3), and FE relations (4.2.4).

4.2.1 Frames

There are at least two possible reasons for the definition of a completely new frame in a non-English FrameNet: **1.** inadequacy of frame definitions in the corresponding semantic domain or area of the English FrameNet; **2.** inadequate coverage of the domain in the English FrameNet, i.e. English Lexical Units and surrounding frames have not been defined yet.

In both the Spanish and the Japanese FrameNet projects, at least some frames were found to be inadequate for the respective target language. Spanish FrameNet data has not yet been officially released but can be viewed in a currently access-protected version of FrameSQL (Subirats

[4] The definition says: "The body part affected [...] is typically expressed by the noun heading the external argument, and this noun is typically accompanied by a possessive determiner that refers to the possessor of the body part: *My head hurts!*."

and Sato, 2004; Sato, 2005).[5] The data shows that several Spanish frames have been defined that lack an equivalent of the same name in the English resource. For example, the Spanish frame `Return` is not present in the current English FrameNet Release 1.3. For Japanese, it has been suggested that frames related to communication might be shaped differently than in English. For example, the English `Statement` frame seems to be realized by two different frames in Japanese: `Statement_Verbal_Act` and `Statement_Verbal_Transfer` (Ohara *et al.*, 2003).[6]

Spanish and Japanese FrameNet proceed by expanding the English data, thereby concentrating their work on those domains in which frames have already been defined for English (Subsection 4.3 describes the expansion process, including the restructuring and realignment steps that are necessary when frames within these domains differ cross-linguistically). However, if Lexical Units are not selected by frame expansion but in some other way, it is likely that frames are needed that have not yet been covered by the English FrameNet, or that doubts will arise as to the extension or adequacy of an existing frame with respect to certain new Lexical Units. For example, it remains unclear whether FrameNet contains a frame pertaining to certain word senses of English *be.v* or *do.v* as in the following examples:

(15) How are you?

(16) How are you doing?

The examples should be read as real inquiries expecting real answers, not as mere politeness or greeting formulae. The corresponding example sentences investigated in Slovenian and German are inquiries and statements about the general well-being of a person, not restricted to a purely emotional or a purely physical state. Currently, in FrameNet, the `Feeling` frame seems to be close to the intended meaning. It defines a situation in which "an Experiencer experiences an Emotion or is in an Emotional_state." But this definition and the frame element definitions do not make clear (enough) whether the Emotion or Emotional_state frame elements also comprise physical states; if they exclude them, this frame is not fully appropriate for the word senses under investigation and a new frame would have to be added.

The SALSA project produces exhaustive FrameNet-style annotation of German texts. Here, Lexical Units are "pre-selected" by the text to be annotated and do not necessarily pertain to domains for which frames exist in the English FrameNet. Therefore, the SALSA project is confronted with the problem of limited coverage in FrameNet, encountered when German word senses appearing in the text cannot be described in terms of original FrameNet frames. In this case, Burchardt *et al.* (2006, p. 971) create a so-called proto frame for each uncovered word sense. The proto frames are predicate-specific but do have definitions and frame elements. They can also be linked to other frames. Proto frames might either correspond to English frames not yet defined, or they might represent frames that are specific for German.

4.2.2 Frame Elements

It has been proposed that some English frames might be adjusted to accommodate data from other languages, by adding or splitting frame elements. For example, the Spanish `Motion` frame contains an extra-thematic FE Intention, which is not present in the English `Motion` frame (Subirats and

[5] Thanks to Carlos Subirats for making this preview of the Spanish FrameNet data accessible to ICSI staff and visitors.

[6] English FrameNet researchers plan to reorganize the `Statement` frame in the future.

Sato, 2004). For Japanese, it has been suggested that the frame element Path within the `Path_Shape` frame should be replaced by two new frame elements, Route and Boundary ("sub-categories of Path"); cf. Ohara *et al.* (2004). Finally, the frame element Message in communication-related frames might need to be split into finer distinctions for Japanese (Ohara *et al.*, 2003). Differences in frame elements resulting in FE splitting have also been observed by Burchardt *et al.* (2006, p. 971) when using FrameNet frames for German corpus annotation.

Strictly speaking, the language-specific addition of frame elements to a FrameNet frame results in a different, more restricted frame for that language, because all frame elements form a conjunctive set (they are logically related via "and"). This more specific non-English frame could be related to the English one by a cross-lingual Inheritance relation, whereby it would become a cross-lingual Child frame of the English frame; but it should not normally be related to it via Equivalence. A possible exception to this strict definition of Equivalence is the treatment of *extra-thematic* frame elements, based on the fact that these frame elements are exempt from being logically connected to the others via "and" (Scheffczyk, 2006). It can be argued that the definition of an Equivalence relation still complies with the FrameNet model if it disregards extra-thematic frame elements. Following a definition of cross-lingual frame Equivalence where all but the extra-thematic frame elements have to be the same cross-linguistically, the Spanish `Motion` frame that inserts an additional extra-thematic frame element not present in the English frame (as described above) still counts as a cross-lingual equivalent of the English frame.

In general, non-English FrameNets featuring Equivalence and other cross-lingual links to the English FrameNet should state explicitly which types of differences in frame elements are tolerated by the Equivalence relation, and if and how the other types of divergence are covered by other cross-lingual relations.

4.2.3 Coreness status and Semantic Type

Language-internally, differences in coreness status of frame elements cannot be accounted for by the Inheritance relation, as the rule definitions for consistency (Scheffczyk, 2006) show. Therefore, frames with FEs of different coreness status should also not be regarded as equivalent across languages. This is justified insofar as sentences with target LUs of the same frame should always display all core FEs of the frame (except for certain FEs related by frame-internal relations, cf. Section 3.3).

Formal and checkable conditions on coreness status such as obligatory overt specification for core FEs depend on the syntactic behavior of the Lexical Units in the frame. This can only be observed in corpus data and might turn out to be language specific.

Differences in coreness status have not been reported in the literature for Spanish or Japanese, but case studies on German and Slovenian suggest that some English frames are not reusable in these two languages due to differences in coreness status. As discussed above for frames (4.2.1), these observations might be explained by FrameNet's limited coverage, with the necessary frame containing all and only the necessary frame elements with the correct coreness status not yet defined for English either. Therefore, the two following examples should only be considered as an illustration of the *method* of comparing frames and Lexical Units with respect to FE coreness status.

Consider the `Topic` frame, of which German and Slovenian prepositional verbs corresponding to

English *be about* were tentatively considered members.[7] The frame has three core frame elements, Communicator, Topic, and Text. The Text FE is defined as "a set of propositions that is coherent in being about a Topic". However, the Lexical Units investigated for German and Slovenian rarely occur in combination with a Text FE. If a Text FE is realized, this is accommodated by an additional prepositional phrase. This suggests that based on the behavior of these verbs, only Topic and Communicator are core FEs, whereas Text is peripheral.

The English `Process_continue` frame holds Lexical Units such as *continue.v* and *go_on.v*. It has been considered as a possible host frame for Lexical Units roughly corresponding to the English verb *last*[8] (currently not covered by FrameNet). However, the Duration frame element of `Process_continue` is merely peripheral according to the FrameNet definition. The German and Slovenian LUs seem to require an overt realization of this frame element, or of an element standing for End-Point, which makes the `Process_continue` frame inadequate for their description.

Differences in coreness status and/or in semantic type should be considered as indicators of differences in frame elements, with a possible discrepancy in the number of frame elements. Ultimately, they introduce a cross-linguistically different frame, even if the differences are very subtle.

4.2.4 Frame Element Relations

Finally, as a function of the number, coreness status and semantic type of frame elements, the relations between FEs might also differ cross-linguistically. For example, some CoreSet relations (Subsection 3.3) in English reflect a kind of metonymic alternation in English. Ruppenhofer *et al.* (2006, p. 33) illustrate this with the constituent that fills the subject position in sentences involving the English verb *open* in the `Closure` frame. This position can be filled by either the FE Agent or the FE Instrument (Examples 17 and 18) which are further grouped into a CoreSet.

(17) John opened the door.

(18) The key opened the door.

In Japanese, however, "subject selection is much more restricted and Instruments and Means rarely, if ever, appear as subjects," (Ruppenhofer *et al.*, 2006, p. 33). The reluctance of the Japanese Instrument FE to fill a subject position can result in its loss of coreness status. A CoreSet including Instrument could thus not exist in such Japanese frames, because coreness sets are by definition restricted to core FEs. In other words, the English CoreSet relating Agent and Instrument cannot be preserved in those Japanese frames where Instrument is not core. Consequently, the English `Closure` frame, as it is currently defined, would be inappropriate for Japanese if Instrument loses core status.

The addition of new frame elements or a switch in coreness status from peripheral to core can also make it necessary to extend an existing FE CoreSet. This direction of CoreSet modification has been observed by Burchardt *et al.* (2006, p. 971), where a frame taken over from English was extended by a newly introduced FE in order to cover German data. At the same time, this new frame element in the German frame became an additional member of a CoreSet already imported as part of the English frame.

[7] The LUs under consideration were the German prepositional verb *gehen um*, and its Slovenian counterpart *iti za*.

[8] For example, Slovenian *trajati.v* as in *Do kdaj traja kriminalka?* – (literally) 'Until when does the crime film last?', meaning 'When does the crime film end?'.

4.3 Expanding FrameNet Frames

At the level of frames, all existing non-English FrameNets known to the author follow the Expand approach as presented in Section 2.3.2 above. This section reports mainly on Spanish FrameNet. The Expand approach has also been adopted by other projects such as Japanese FrameNet and German FrameNet[9], as well as by projection approaches. The latter include (Johansson and Nugues, 2006) dealing with Swedish LUs and example annotations and (Fung and Chen, 2006) producing Chinese LUs.

Expand. FrameNets for Spanish and Japanese are implicitly related to the English FrameNet at the level of frames, by importing frame names and definitions from the original Berkeley FrameNet database. For example, Subirats and Sato (2004) assume that frames that *share the same name* in English and Spanish should also have the same characteristics. This implies Equivalence relations between all frames of the same name. When the developers notice that the English frames are not appropriate for their respective language, the differences are dealt with in the restructuring and possibly the realignment phase.

Restructure. Explicit restructuring takes place when new frames with new frame names are introduced into the non-English resource (cf. Subsection 4.2.1 above). As mentioned in Subsection 4.2.1 above, the definition of new language-specific frames has been necessary for both Spanish and Japanese.

If changes are applied to attributes or elements of the frame (Subsections 4.2.2 to 4.2.4 above) while keeping the original frame name, there is currently no means to explicitly document the way in which the resulting resource differs from the English original. Nevertheless, most of these changes imply a violation of strict Equivalence conditions, as described in Section 4.2 above. Currently, such discrepancies can only be tracked down by comparing the monolingual resources at all levels.

Realign. Realignment is not documented in the resources themselves but is performed, to a certain degree, within the FrameSQL tool for Spanish and English (Sato, 2005). FrameSQL is an application for browsing and viewing FrameNet data.

Cross-lingual alignment is performed between frames of the same name, presupposing a degree of Equivalence sufficient to allow for a cross-linguistic inspection of frame elements and frame LUs. At the level of Lexical Units, alignment relies on Spanish-English translation equivalences from the Edic dictionary, and a disambiguation of Edic entries by verifying the host frames of the words in translation pairs: dictionary translations are retained for cross-lingual alignment only if they belong to the same frame.

Cross-lingual Inheritance relations between frames are established in post-hoc manner within FrameSQL based on language-internal frame-to-frame relations. Spanish LUs that belong to frames created specifically for Spanish will be shown if the translation of an English LU into Spanish proposed by Edic is found in the corresponding (equivalent) Spanish frame, or a Child frame thereof. Sato (2005) describes this as, for example, putting "a Frame-to-Frame relation between Eng *Self_motion* [the cross-lingual Parent frame] and Spa *Motion_manner* [the cross-lingual Child frame]."

[9]http://gframenet.gmc.utexas.edu/ [10 March 2007]

In some cases, Spanish frames are not linked to each other via Inheritance, but rather by the less restrictive Use relation. In this case, the cross-lingual realignment discussed above is not applied in FrameSQL. For example, the Spanish-only frame `Return` **Uses** the `Arriving` frame. The `Arriving` frame exists also in English. But currently, no cross-lingual relation between English *return.v* (in the `Arriving` frame) and Spanish *retornar.v* or *volver.v* (in the Spanish `Return` frame) has been established in FrameSQL.

5 Conclusion: Possible Future Directions

Multilingual FrameNets as currently under construction are organizationally similar. At the level of frames, they exhibit an implicit form of interrelatedness, based on identity of frame names. However, conditions for cross-linguistic Equivalence between frames have not been defined formally yet. This report has shown that such conditions might be established based on consistency rules already applied resource-internally to the Berkeley FrameNet. In light of those rules, cross-lingual relations between frames should be regarded as strict Equivalence only in such cases where all attributes and frame elements as well as frame element relations are the same in both languages. This includes attributes of frame elements as defined by the FrameNet model presented in Section 3. A possible exception to the Equivalence conditions is the set of extra-thematic frame elements of a frame. This set is ignored in previously defined consistency conditions on language-internal Inheritance and might therefore also be neglected when postulating cross-lingual Equivalence.

Non-English FrameNets featuring Equivalence and other cross-lingual links to the English FrameNet should state explicitly which types of differences in frame elements are tolerated by the Equivalence relation they use, and if and how other types of divergence are covered by other cross-lingual relations. A consensus among the developers of non-English FrameNets with respect to the exact definition of the cross-lingual relations used to align FrameNets would further facilitate multilingual processing based on FrameNets. The definition of a standardized set of cross-lingual FrameNet relations and their encoding within the resources should be based on analyses of current and future data from different languages.

Strict Equivalence between the textual definitions (as attributes of frames and frame elements) is currently not possible, due to the inclusion of language specific elements. A possible development of FrameNet and its counterparts in other languages might be to further structure the definitions. This structure could be inspired, for example, by the definitions of common extra-thematic frames in the Appendix of the FrameNet book (Ruppenhofer *et al.*, 2006), which systematically separate meaning from form descriptions. Whereas interrelatedness, possibly of the Equivalence type, should be applied to the meaning part, the form part can be considered as merely organizationally similar. Example sentences within the definition are language specific as well, and would be exempt from the Equivalence conditions on frames.

As long as attributes or frame elements of non-English frames might diverge from their English counterparts of the same frame name, algorithms using the data for bilingual or multilingual processing cannot presuppose strict cross-linguistic Equivalence between frames of the same name. Depending on the purpose of the algorithm, existence of all frame elements and equivalence of all attributes might thus have to be checked before further processing the data.

Whether and how diversion from strict Equivalence between frames should be indicated is a design question. It has been argued above that most changes to the original frame actually result in a different frame, which might still be closely related to the original one, by way of a cross-lingual frame relation such as cross-lingual Inheritance. To facilitate further realignment, frames could therefore be renamed if they are modified (for example, by adding a language code to the original name) or some other indication of this modification might be given (for example, in a further attribute). If multilingual resources are designed for multilingual processing tasks including cross-lingual document retrieval, machine translation, or computer aided language learning, more explicit information on the kind of relatedness might be desired within the resources, or as an extra module. A cross-lingual version of the frame relation editor which is part of the FrameNet software (cf. Petruck *et al.*, 2004) could be implemented and used for this purpose. Thereby, cross-lingual Inheritance and other types of cross-lingual frame relations could be indicated and would further be reflected by explicit cross-linguistic frame element relations. This would result in an increase in coverage and a finer granularity of the cross-lingual mapping.

It seems likely that FrameNets in languages other than English will continue to be mapped directly onto the English resource, without the intermediary of an index. This raises the issue of documenting the relations between different versions of the FrameNet resource itself. FrameNet will continue to grow and might therefore introduce as new frames some that, for one reason or the other, have already been defined in other languages. Also, some of the current frame definitions and frame relations in FrameNet might be revised in the future. A process known as Reframing (cf. Petruck *et al.*, 2004) describes cases where FrameNet frames are split and annotations are shifted from one frame to the other accordingly. These and all other revisions of the original resource that influence the validity of cross-lingual Equivalence relations are of interest when keeping multilingual resources synchronized. Mappings between different versions of FrameNet documenting the relationship between old (removed or replaced) and new (added or replacing) elements of the frame part of the database therefore facilitate the transition of cross-lingual alignments. In the current FrameNet release (1.3), such version-to-version changes in frames and frame elements, including coreness status and semantic type, are documented in the file `framesDiff.xml`. A similar file documenting change in relations between frames and frame elements might be developed in the future.

References

Baker, C. F., Fillmore, C. J., and Cronin, B. (2003). The structure of the FrameNet database. *International Journal of Lexicography*, **16**(3), 281–296.

Boas, H. C. (2002). Bilingual FrameNet dictionaries for machine translation. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC-02)*, volume IV, pages 1364–1371, Las Palmas, Spain.

Boas, H. C. (2005). Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography*, **18**(4), 445–478.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006). The salsa corpus: a german corpus resource for lexical semantics. In *Proceedings of LREC 2006*, pages 969–974, Genoa, Italy.

Francopoulo, G. (2005). Extended examples of lexicons using LMF (auxiliary working paper for LMF). Technical report, INRIA-Loria.

Francopoulo, G., Bel, N., George, M., Calzolari, N., Monachini, M., Pet, M., and Soria, C. (2006a). Lexical Markup Framework (LMF) for NLP multilingual resources. In *Proceedings of the Workshop on Multilingual Language Resources and Interoperability*, pages 1–8, Sydney.

Francopoulo, G., Bel, N., George, M., Calzolari, N., Monachini, M., Pet, M., and Soria, C., editors (2006b). *Proceedings of the Workshop on Multilingual Language Resources and Interoperability*. ACL, Sydney, Australia.

Fung, P. and Chen, B. (2006). Robust word sense translation by em learning of frame semantics. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 239–246, Sydney, Australia. Association for Computational Linguistics.

Johansson, R. and Nugues, P. (2006). A FrameNet-based Semantic Role Labeler for Swedish. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Session*, pages 436–443, Sydney.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. J. (1990). Introduction to WordNet: an on-line lexical database. *International Journal of Lexicography*, **3**(4), 235–244.

Ohara, K. H., Fujii, S., Saito, H., Ishizaki, S., Ohori, T., and Suzuki, R. (2003). The Japanese FrameNet project: A preliminary report. Poster presentation at the Pacific Association for Computational Linguistics (PACLING'03).

Ohara, K. H., Fujii, S., Ohori, T., Suzuki, R., Saito, H., and Ishizaki, S. (2004). The Japanese FrameNet project: An introduction. In *LREC 2004. The Fourth international conference on Language Resources and Evaluation. Proceedings of the Satellite Workshop "Building Lexical Resources from Semantically Annotated Corpora"*, pages 9–11, Lisbon, Portugal.

Petruck, M. R. L., Fillmore, C. J., Baker, C. F., Ellis, M., and Ruppenhofer, J. (2004). Reframing FrameNet data. In *Proceedings of the 11th EURALEX International Congress*, pages 405–416, Lorient, France.

Pianta, E., Bentivogli, L., and Girardi, C. (2002). MultiWordNet. Developing an aligned multilingual database. In *Proceedings of the 1st International WordNet Conference*, Mysore, India.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., and Scheffczyk, J. (2006). FrameNet II: Extended theory and practice.

Sato, H. (2005). A bilingual viewer for FrameNets. Presentation at the 2nd Cross-lingual FrameNet meeting. 1. April 2005, Saarbrücken, Germany.

Scheffczyk, J. (2006). FrameNet Consistency Management. Draft.

Subirats, C. and Sato, H. (2004). Spanish FrameNet and FrameSQL. In *4th International Conference on Language Resources and Evaluation. Workshop on Building Lexically Annotated Corpora*, Lisbon, Portugal.

Vossen, P. (1999). EuroWordNet General Document. Version 3. Technical report, University of Amsterdam.

Vossen, P. (2004). EuroWordNet: a multilingual database of autonomous and language-specific wordnets connected via an Inter-Linual-Index. *International Journal of Lexicography*, **17**(2), 161–173.