3. Semantic frames as interlingual representations for multilingual lexical databases

Hans C. Boas

1. Introduction[1]

Globalization and its effects on many areas of life requires a previously unforeseen level of detail of cross-linguistic information without which it is difficult, if not impossible, to provide accurate resources for efficient communication across language boundaries. Over the past decade, research in computational lexicography has thus focused on streamlining the creation of multilingual lexical databases in order to meet the ever-increasing demand for tools supporting human and machine translation, information retrieval, and foreign language education. However, creating multilingual lexical databases poses a number of problems that are more numerous and more complicated than those encountered in the creation of monolingual lexical databases.

One of the main problems that arises in the creation of multilingual lexical databases (henceforth MLLDs) is the development of an architecture capable of handling a wide spectrum of linguistic issues such as diverging polysemy structures (cf. Boas 2001, Viberg 2002), detailed valence information (cf. Fillmore and Atkins 2000), differences in lexicalization patterns (cf. Talmy 2000), and translation equivalents (cf. Sinclair 1996, Salkie 2002). A closely related question is whether MLLDs should employ an interlingua to map between different languages. If one decides in favor of an interlingua for mapping purposes, a choice needs to be made between using an unstructured interlingua as in EuroWordNet (Vossen

1. This paper was first published in 2005 in the *International Journal of Lexicography* Vol. 18.4: 445–478. I am grateful to Charles Fillmore, Collin Baker, Carlos Subirats, Kyoko Hirose Ohara, Hans U. Boas, Jonathan Slocum, Inge De Bleecker, Jana Thompson, and three anonymous referees for very helpful comments on the material discussed in the article.

1998, 2004), or a structured interlingua as in ULTRA (Farwell et al. 1993) or SIMuLLDA (Janssen 2004).

Another problem underlying the creation of adequate MLLDs concerns the sources of information used for constructing them. Whereas most MLLDs primarily rely on machine-readable versions of existing print dictionaries, very few take advantage of the multitude of information contained in electronic corpora that have become available for increasing numbers of languages over the past decade.[2]

This paper addresses these important issues by demonstrating how the English FrameNet database (Fillmore et al. 2003a) provides a solid basis for conducting cross-linguistic research, thereby facilitating the creation of MLLDs capable of overcoming a number of important linguistic problems.

As we will see, semantic frames as well as the underlying framework of Frame Semantics (Fillmore 1982, Fillmore and Atkins 1994) have been successfully employed by a number of FrameNet-type projects for languages other than English. In these projects, semantic frames play a central role in the building and connection of lexicon fragments across languages such as English, German, Spanish, and Japanese.

The remainder of the paper is structured as follows. Section 2 describes in detail some of the cross-linguistic problems that the architecture of any MLLD needs to address. Section 3 provides a brief survey of Frame Semantics. Section 4 discusses the architecture of FrameNet, which forms the basis for the creation of parallel lexicon fragments described in Section 5. This architecture, which employs semantic frames as an interlingual representation for connecting the various lexicon fragments differs in important ways from other types of interlingua approaches. Instead of using traditional lexical-semantic concepts such as synonymy, antonymy, and meronymy in combination with conceptual ontological information, the complementary approach proposed in this paper aims at linking parallel lexicon fragments by means of semantic frames. Section 6 compares the structure of MLLDs created on frame semantic principles with the architecture of other MLLDs. Finally, Section 7 provides a summary and gives an overview of open research questions.

2. See Atkins et al. (2002) for a recent approach to the design of multilingual lexical entries within the ISLE framework.

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 60)

2. Linguistic problems for multilingual lexical databases

2.1. Polysemy

Whereas polysemy is seldom a serious problem in human communication, lexicographers have traditionally been concerned with how to best account for the fact that one word can carry several different meanings (cf. Leacock and Ravin 2000). Over time, lexicographic procedures have been established that have resulted in the listing of multiple dictionary senses for polysemous words where sub-senses are grouped together with their respective definitions (cf. Béjoint 2000: 227–234). However, dictionaries often vary in their organization of word senses, which makes it difficult to compare definitions across different dictionaries (cf. Atkins 1994, Goddard 2000). For example, in their discussion of the verb *risk*, Fillmore and Atkins (1994) compare the definitions found in ten different print dictionaries and come to the conclusion that "all the dictionaries agree on the clear stand-alone existence of Sense 1 (*risk your life*), but cannot agree on Sense 2 (*risk falling/a fall*) and Sense 3 (*risk climbing the cliff*)" (Fillmore and Atkins 1994: 353).

Looking beyond the well-known issues surrounding the treatment of polysemy in a single language, we find even greater problems when it comes to accounting for polysemy across languages. Overcoming these problems is not only important for the design of traditional lexicons, but also crucial for the successful implementation of MLLDs. In other words, without a satisfactory account of cross-linguistic polysemy, it is difficult, if not impossible, to construct adequate MLLDs. For example, Altenberg and Granger (2002) distinguish between three different types of cross-linguistic polysemy patterns that can be located along a continuum, where complete overlap of word senses is on one end of the continuum, and no correspondence among word senses across languages is found at the other end of the continuum. On one end of the continuum we find "overlapping polysemy" which refers to cases in which items in two languages have roughly the same meaning extensions (Altenberg and Granger 2002: 22). An example of overlapping polysemy is provided by Alsina and DeCesaris' (2002) comparison of the adjective *cold* with its Spanish and Catalan counterparts *frío* and *fred*. The authors discuss the varying degrees of polysemy exhibited by the three adjectives and come to the conclusion that the three adjectives exhibit "almost complete" overlapping polysemy patterns. Overlapping polysemy poses relatively few problems for multilingual dictionaries, but it is unfortunately very rare.

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 61)

In contrast, diverging polysemy structures are very common. In their contrastive study of English *to crawl* and French *ramper*, Fillmore and Atkins (2000) demonstrate that the two verbs exhibit semantic overlap when it comes to the basic senses describing "the primary motion of insects and invertebrates, and the deliberate crouching movement of humans" (2000: 104). However, they differ widely in their meaning extensions when it comes to more specialized senses. For example, whereas English *crawl* can be used to describe slow-moving vehicles, French requires *rouler au pas* (literally: move at walking pace, or slowly) instead of *ramper*. Similarly, whereas *crawl* exhibits a meaning extension describing "creatures teeming" (*You got little brown insects crawling about all over you.* (2000: 96)), French requires *grouiller* instead of *ramper* to express the same concept (Fillmore and Atkins 2000: 107). Examples such as these show that adequate MLLDs must not only take into consideration the multitude of different senses of words across languages, but also have to include effective mechanisms that allow for the linking of extended word senses in diverging polysemy patterns.[3]

The third type of cross-linguistic phenomenon posing problems for MLLDs are cases in which there are no clear equivalents in the target language. As Altenberg and Granger (2002: 25) point out, these cases may lead to two types of problems: "either the lack of a clear translation equivalent in the target language results in a large number of zero translations, indicating that the translators have great difficulties finding a suitable target item, or in a wide range of translations, indicating that the translators find it necessary to render the source item in some way but, in the absence of a single prototypical equivalent, vary their renderings according to context." However problematic it may be to find proper equivalences for "difficult" lexical items cross-linguistically, it is necessary to account for them within MLLDs. Without their inclusion, neither humans nor machines will be able to successfully employ MLLDs for translation purposes. With this brief overview of problems surrounding cross-linguistic polysemy patterns, we now turn to another linguistic issue that needs to be accounted for when designing MLLDs, namely the accuracy of syntactic and semantic valence patterns.

3. For examples of diverging polysemy patterns among nouns, see Svensen (1993) on *wood* and *forest* and their French and German equivalents. See Chodkiewicz et al. (2002: 264) on the various meanings of *proceedings* and their French equivalents.

2.2. Syntactic and semantic valence patterns

Besides providing information about a word's different senses, any MLLD should provide detailed syntactic information illustrating the various ways in which meanings can be realized. To illustrate, consider the following examples.

(1) a. The mother cured the child.
    b. The mother cured the measles.
    c. The mother cured {the child/the measles} with pills.

(2) a. The mother cured the ham.
    b. The mother cured the ham with hickory smoke.

(3) a. [NP, V, NP]
    b. [NP, V, NP, PP_with]

The sentences in (1) exemplify some of the syntactic valence patterns associated with one sense of *to cure*, namely the healing sense. In contrast, the examples in (2) illustrate some of the syntactic valence patterns found with the preserving food sense of *cure*. The syntactic frames in (3) summarize the syntactic commonalities among the two different senses of *cure*. That is, whereas the syntactic frame in (3a) represents the valence pattern exhibited by (1a), (1b), and (2a), the syntactic frame in (3b) summarizes the valence patterns of (1c) and (2b). From the perspective of a human user the information in (1)–(3) is readily interpretable because humans have already stored the representation that makes the link between the underlying meaning of the senses and their different syntactic realizations.

However, NLP-applications face a much harder task when trying to identify the different meanings of *cure* because they are typically trying to establish the meanings based on syntactic information of the type in (3) alone. That is, without having access to information about the different semantic types of Noun Phrases or Prepositional Phrases that may occur with the different senses in postverbal position, it is difficult to decide what sense of *cure* is expressed. This example illustrates that lexical databases should contain adequate information not only about a word's different senses, but also how a single sense of a word may be realized in different ways at the syntactic level.[4]

4. Note that resources such as WordNet (cf. Fellbaum 1998) provide important information that can be used to determine the semantic type of complements.

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 63)

Similar issues arise in multilingual environments. Discussing the various Swedish counterparts for *get*, Viberg (2002: 139) reviews the "large number of senses which are both lexical and grammatical." As Table 1 shows, the multitude of syntactic frames associated with *get* are relevant for the identification of the appropriate sense.

Table 1. The major meanings of *get* (cf. Viberg 2002: 140)

Meaning	*Frame*	*Example*
Possession	get + NP have + got + NP	Peter got a book Peter has got a book
Modal: Obligation	have got to + $VP_{infinitive}$ gotta + $VP_{infinitive}$	Peter has got to come Peter has gotta come
Inchoative	get + ADJ/Participle	Peter got angry
Passive	get + PastPart (by NP)	Peter got killed (by a gunman)
Causative Motion:	get + NP + to $VP_{infinitive}$	Peter got Harry to leave
Subject-centered	get + Particle get + PP	Peter got up/in/out . . . Peter got to Berlin
Object-centered	get + NP + PP	Peter got the buns out of the oven

Similar to our discussion of *cure* above, it is clear that any lexical database must contain fine-grained valence information of the kind contained in Table 1 in order to successfully identify the different senses of *get*. At the next step, MLLDs should also provide information about translation equivalents in other languages. Table 2 lists the most frequent Swedish equivalents of *get*.

Table 2. The most frequent Swedish equivalents of English *get* (cf. Viberg 2002: 141)

Possession		*Motion*		*Inchoative*	
få	'get'	komma	'come'	bli	'become'
ha	'have'	gå	'go'		
ta	'take'	stiga	'step'		
ge	'give'	kliva	'stride'		
skaffa	'acquire'	resa sig	'rise'		
hämta	'fetch'				

The Swedish data demonstrate that the identification of Swedish equivalents of *get* require detailed information about the specific sense of *get* in English source texts. Any MLLD aimed at providing useful information for humans and machines will therefore have to include detailed syntactic and semantic valence information showing how to map specific sub-senses of a word from one language into another language. The following section discusses a related problem, namely different types of lexicalization patterns across languages.

2.3. Differences in lexicalization patterns

As Talmy (1985, 2000) points out, languages show strong preferences as to what kinds of semantic components they lexicalize. This property, in turn, has a number of important implications for the design of MLLDs. For example, Japanese motion verbs differ from English motion verbs in how they realize various types of paths (Ohara et al. 2004). The verbs *wataru* ('go across') and *koeru* ('go beyond, go over') "describe motion in terms of the shape of the path traversed by the theme that moves" (Ohara et al. 2004: 10). As examples (4a) and (4b) show, *wataru* ('go across') is used with an accusative-marked direct object NP describing a path. Ohara et al. point out that *kawa* ('river') in (4a) "denotes an area that lies between two points in space", whereas *hasi* ('bridge') "refers to a medium or a passage that is constructed between the two points."

(4) a. nanminga kawa o watatta
refugees NOM river ACC went.across
'The refugees went across (crossed, traversed) the river.'

b. nanminga hasi o watatta
refugees NOM bridge ACC went.across
'The refugees crossed the bridge.' (Ohara et al. 2004: 10)

Differences arise when we look at semantically related verbs such as *koeru* ('go beyond') which takes an accusative marked direct object NP such as *kawa* ('river') in (5a). However, *koeru* does not allow *hasi* ('bridge') as its direct object as is illustrated by (5b).

(5) a. nanminga kawa o koeta
refugees NOM river ACC went.beyond
'The refugees went beyond (passed) the river.'

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 65)

b. *nanminga hasi o koeta
refugees NOM bridge ACC went.beyond
(Intended meaning) 'The refugees passed the bridge.'
(Ohara et al. 2004: 10)

According to Ohara et al. (2004), the differences between these verbs illustrate the necessity to identify and include in lexical descriptions the subcategories of different types of paths that can occur with motion verbs in Japanese. They point out that *wataru* ('go across') may be described as taking an accusative-marked route, while *koeru* ('go beyond') may be characterized as taking an accusative-marked boundary as the direct object (2004: 10).[5] These examples demonstrate that Japanese makes a more fine-grained distinction between different types of path expressions than English. In other words, whereas in English the type of path is typically unimportant in terms of lexical selection, Japanese verbs exhibit a larger variety of lexicalization patterns with respect to path expressions.

While these systematic differences in lexicalization patterns pose relatively few problems to bilingual speakers, it is far from clear as to how these differences between languages should be encoded in MLLDs. That is, in order to successfully "mirror the expertise of bilingual humans" (Sinclair 1996: 174), it is first necessary to determine how to systematically account for differences in lexicalization patterns in the design of MLLDs. We return to this issue in Section 5.

2.4. Measuring paraphrase relations and translation equivalents

Another linguistic problem requiring attention in the design of MLLDs concerns two related issues, namely dealing with paraphrase relations and measuring translation equivalents across languages. When accounting for paraphrase relations, lexical databases should include information about the fact that certain words and multi word expressions are paraphrases of each other, i.e., they may be substituted for each other and still express the same meaning. Compare the following examples.

(6) Jana argued with Inge about the theory.

(7) Jana had an argument with Inge about the theory.

5. For a discussion of different lexicalization patterns posing similar types of problems, see Talmy (1985) for motion verbs in English and Atsugewi, and Subirats & Petruck (2003) for emotion verbs in English and Spanish.

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 66)

Both sentences express the same type of situation. However, the two examples differ in how the situation is expressed syntactically. In (6) it is the verb *argue* which takes *Jana* as a subject, and *with Inge* and *about the theory* as prepositional complements. In (7), it is the multi word expression *to have an argument*, which occurs with *Jana* as its subject, and *with Inge* and *about the theory* as its prepositional complements. This example shows that the number of words evoking a given meaning may differ across sentences. Any lexical database that is used for translation purposes must not only take into account paraphrase relations within a single language, but it should also include a description of how to map such paraphrases cross-linguistically.

In other words, when it comes to translation equivalents, the question is not only how to "measure" them cross-linguistically, but also how to match them from different paraphrases in the source language to different types of paraphrases in the target language. Consider the following examples from German, which are translation equivalents of (6) and (7).

(8) a. Jana stritt mit Inge über die Theorie.
Jana argued with Inge about the theory
'Jana argued with Inge about the theory.'

b. Jana stritt sich mit Inge über die Theorie.
Jana argued self with Inge about the theory
'Jana argued with Inge about the theory.'

(9) Jana hatte einen Streit mit Inge über die Theorie.
Jana had an argument with Inge about the theory
'Jana had an argument with Inge about the theory.'

In (8a) and (8b), we find the verb *streiten* ('to argue') and its counterpart *sich streiten* ('to argue'), respectively. In this context, there is no obvious difference in meaning that would be caused by choosing one verb over the other. Similarly, the multi word expression *einen Streit haben mit* ('to have an argument with') in (9) expresses the same type of situation as the sentences in (8). These three sentences are important because they exemplify the difficulty of identifying paraphrase relations within one language, and translation equivalents across languages.[6] In contrast to bilingual

6. An anonymous reviewer points out that another way of capturing such paraphrase relations would be to apply Mel'čuk's Meaning-Text Theory (Mel'čuk et al. 1988) and its Explanatory Combinatory Dictionaries. On this view, a

human speakers, who possess what Chesterman (1998: 39) calls translation competence ("the ability to relate two things"), multi-lingual NLP applications have to rely on MLLDs to supply information about translation equivalents. Without the inclusion of paraphrase relations and the different numbers and combinations of word senses across languages it will be difficult to solve problems such as those discussed above. With this overview, we now turn to a discussion of Frame Semantics and the structure of the English FrameNet database. In Section 5, we return to the linguistic issues discussed in this section and demonstrate how they can be tackled by MLLDs that employ semantic frames as an interlingua.

3. Frame Semantics

Frame Semantics, as developed by Fillmore and his associates over the past three decades (Fillmore 1970, 1975, 1982, Fillmore and Atkins 1992, 1994, 2000), is a semantic theory that refers to semantic "frames" as a common background of knowledge against which the meanings of words are interpreted (cf. Fillmore and Atkins 1992: 76–77).[7] An example is the `Compliance` frame, which involves several semantically related words such as *adhere*, *adherence*, *comply*, *compliant*, and *violate*, among many others (Johnson et al. 2003). The `Compliance` frame represents a kind of situation in which different types of relationships hold between so-called "Frame Elements" (FEs), which are defined as situation-specific semantic roles.[8] This frame concerns ACTS and STATES_OF_AFFAIRS for which PRO-

---

lexical function is a meaning relation between a keyword and other words or phraseological combinations of words. Using paraphrase mechanisms, we can link such paraphrases as *streiten* and *einen Streit haben* (cf. (8) and (9)) with lexical functions:

V0(argument) = argue
Oper1(argument) = have

See Mel'čuk & Wanner (2001) for a lexical transfer model using Meaning-Text Theory for machine translation.

7. For a detailed overview of Frame Semantics, see Petruck (1996).
8. Names of Frame Elements (FEs) are capitalized. Frame Elements differ from traditional universal semantic (or thematic) roles such as Agent or Patient in that they are specific to the frame in which they are used to describe participants in certain types of scenarios. "Tgt" stands for target word, which is the word that evokes the semantic frame.

TAGONISTS are responsible and which violate some NORM(S). The FE ACT identifies the act that is judged to be in or out of compliance with the norms. The FE NORM identifies the rules or norms that ought to guide a person's behavior. The FE PROTAGONIST refers to the person whose behavior is in or out of compliance with norms. Finally, the FE STATE_OF_AFFAIRS refers to the situation that may violate a law or rule (see Johnson et al. 2003).

With the frame as a semantic structuring device, it becomes possible to describe how different FEs are realized syntactically by different parts of speech. The unit of description in Frame Semantics is the lexical unit (henceforth LU), which stands for a word in one of its senses (cf. Cruse 1986). Consider the following sentences in which the LUs (the targets) *adhere*, *compliance*, *compliant*, *follow*, and *violation* evoke the `Compliance` frame. FEs are marked in square brackets, their respective names are given in subscript.[9]

(10) [$_{\text{<Protagonist>}}$ Women] take more time, talk easily and still *adhere*$^{\text{Tgt}}$ [$_{\text{<Norm>}}$ to the strict rules of manners].

(11) It is also likely to improve [$_{\text{<Protagonist>}}$ patient] *compliance*$^{\text{Tgt}}$ [$_{\text{<Norm>}}$ in taking the daily quota of bile acid].

(12) [$_{\text{<Protagonist>}}$ Patients] were$^{\text{Supp}}$ [$_{\text{<Act>}}$ *compliant*$^{\text{Tgt}}$] [$_{\text{<Norm>}}$ with their assigned treatments].

(13) So now the Commission and other countryside conservation groups, have produced [$_{\text{<Norm>}}$ a series of guidelines] [$_{\text{<Protagonist>}}$ for the private landowners] to *follow*$^{\text{Tgt}}$.

(14) [$_{\text{<Act>}}$ Using a couple of minutes for private imperatives] was$^{\text{Supp}}$ a [$_{\text{<Degree>}}$ serious] *violation*$^{\text{Tgt}}$ [$_{\text{<Norm>}}$ of property rights].

The examples show that FEs may occur in different syntactic positions, and that they may fulfill different types of grammatical functions (subject, object, etc.). One of the major advantages of describing LUs in frame semantic terms is that it allows the lexicographer to use the same underlying semantic frame to describe different words belonging to different parts of speech. The design of the FrameNet database, to which we now turn, is influenced by and structured along frame-semantic principles.

---

9. Support verbs (Supp) such as *to be* or *to take* do not introduce any particular semantics of their own. Instead, they create a verbal predicate "allowing arguments of the verb to serve as frame elements of the frame evoked by the noun". (Johnson et al. 2003)

4. FrameNet

The FrameNet database developed at the International Computer Science Institute in Berkeley, California, is an on-line lexicon of English lexical units (LUs) described in terms of Frame Semantics. Between 1997 and 2003, the FrameNet team collected and analyzed lexical descriptions for more than 7,000 LUs based on more than 130,000 annotated corpus sentences (Baker et al. 1998, Fillmore et al. 2003a). The process underlying the creation of lexical entries in FrameNet involves several steps. First, frame descriptions for the words or word families targeted for analysis are devised. This procedure consists roughly of the following phases:

> (1) characterizing schematically the kind of entity or situation represented by the frame, (2) choosing mnemonics for labeling the entities or components of the frame, and (3) constructing a working list of words that appear to belong to the frame, where membership in the same frame will mean that the phrases that contain the LUs will all permit comparable semantic analyses. (Fillmore et al. 2003b: 297)

The second step in the FrameNet workflow concentrates on identifying corpus sentences in the British National Corpus exhibiting typical uses of the target words in specific frames. Next, these corpus sentences are extracted mechanically and annotated manually by tagging the Frame Elements realized in them. Finally, lexical entries are automatically prepared and stored in the database. An important feature of the FrameNet workflow is that it is not completely linear. That is, at each stage of the workflow, FrameNet lexicographers may discover new corpus data that might force them to re-write frame descriptions because of the need to include or exclude certain LUs in the frame. Similarly, if frames are found to include LUs whose semantics are too divergent, frames have to be "re-framed" (see Petruck et al. 2004), i.e., they have to be split up into separate frames (for a full overview of the FrameNet process, see Fillmore et al. (2003a) and Fillmore et al. (2003b)).

The FrameNet database (http://framenet.icsi.berkeley.edu) offers a wealth of semantic and syntactic information for several thousand English verbs, nouns, and adjectives. Each lexical entry in FrameNet is structured as follows: It provides a link to the definition of the frame to which the LU belongs, including FE definitions, example sentences exemplifying prototypical instances of FEs (For more information on the structure of the FrameNet database, please see Baker et al. (2003)). In addition, it offers information about various frame-to-frame relations (e.g., child-

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 70)

parent relation and sub-frame relation (see Fillmore et al. 2003b and Petruck et al. 2004)) and includes a list of LUs that evoke the frame.

The central component of a lexical entry in FrameNet consists of three parts. The first provides the Frame Element Table (a list of all FEs found within the frame) and corresponding annotated corpus sentences demonstrating how FEs are realized syntactically (see Fillmore et al. 2003b). In this part, words or phrases instantiating certain FEs in the annotated corpus sentences are highlighted with the same color as the FEs in the FE table above them. This type of display allows users to identify the variety of different FE instantiations across a broad spectrum of words and phrases. The Realization Table is the second part of a FrameNet entry. Besides providing a dictionary definition of the relevant LU, it summarizes the different syntactic realizations of the frame elements. The third part of the Lexical Entry Report summarizes the valence patterns found with a LU, that is, "the various combinations of frame elements and their syntactic realizations which might be present in a given sentence" (Fillmore et al. (2003a: 330)). As the first row in the valence table for *comply* in Figure 1 shows, the FE NORM may be realized in terms of two different types of external arguments: either as an external noun phrase argument, or as a prepositional phrase headed by *with*. Clicking on the link in the column to the left of the valence patterns leads the user to a display of annotated example sentences illustrating the valence pattern.[10]

Accessing the Lexical Entry Report for a given LU not only allows the user to get detailed information about its syntactic and semantic distribution. It also facilitates a comparison of the comprehensive lexical descriptions and their manually annotated corpus-based example sentences with those of other LUs (also of other parts of speech) belonging to the same frame. Another advantage of the FrameNet architecture lies in the way lexical descriptions are related to each other in terms of semantic frames. Using detailed semantic frames which capture the full background knowledge that is evoked by all LUs of that frame makes it possible to systematically compare and contrast their numerous syntactic valency patterns.

Our discussion of FrameNet shows that it is different from traditional (print) dictionaries, thesauri, and lexical databases in that it is organized

10. Frame Elements which are conceptually salient but do not occur as overt lexical or phrasal material are marked as null instantiations. There are three different types of null instantiation: Constructional Null Instantiation (CNI), Definite Null Instantiation (DNI), and Indefinite Null Instantiation (INI). See Fillmore et al. (2003b: 320–321) for more details.

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 71)

around highly specific semantic frames capturing the background knowledge necessary to understand the meaning of LUs. By employing semantic frames as structuring devices, FrameNet thus differs from other approaches to lexical description (e.g. ULTRA (Farwell et al. 1993), WordNet (Fellbaum (1998), or SIMuLLDA (Janssen 2004)) in that it makes use of independent organizational units that are larger than words, i.e., semantic frames (see also Ohara et al. 2003, Boas 2005). In the following sections I show how the inventory of semantic frames can be utilized for the construction of MLLDs. Drawing on data from Spanish, Japanese, and German I demonstrate the individual steps necessary for the construction of parallel FrameNets.

Number Annotated	Patterns		
3 TOTAL	Act	Norm	
(3)	NP Ext	PP[with] Dep	
1 TOTAL	Norm	Norm	Protagonist
(1)	NP Ext	PP[with] Dep	CNI --
17 TOTAL	Norm	Protagonist	
(2)	PP[with] Dep	CNI --	
(15)	PP[with] Dep	NP Ext	
2 TOTAL	Norm	State_of_Affairs	
(1)	DNI --	NP Ext	
(1)	PP[to] Dep	NP Ext	

Figure 1. FrameNet entry for *comply*, Valence Table

5. Using semantic frames for creating multilingual lexicon fragments

5.1. Producing FrameNet-type descriptions for other languages

In order to construct a non-English FrameNet, we first download the English FrameNet MySQL database (see Baker et al. 2003 for a detailed description of the FN database structure). Next, all English-specific information is removed from the language-specific database tables. This includes, for example, all information about Lexical Units in the top left

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 72)

part of the original FrameNet database tables in Figure 2 (e.g. Lemma, Part of Speech, Lexeme, Lexeme Entry, Word Form), as well as all information relating to annotated corpus example sentences in the lower left part of the original FrameNet database tables in Figure 2 (e.g. Corpus, Sub-corpus, Document, Genre, Paragraph).

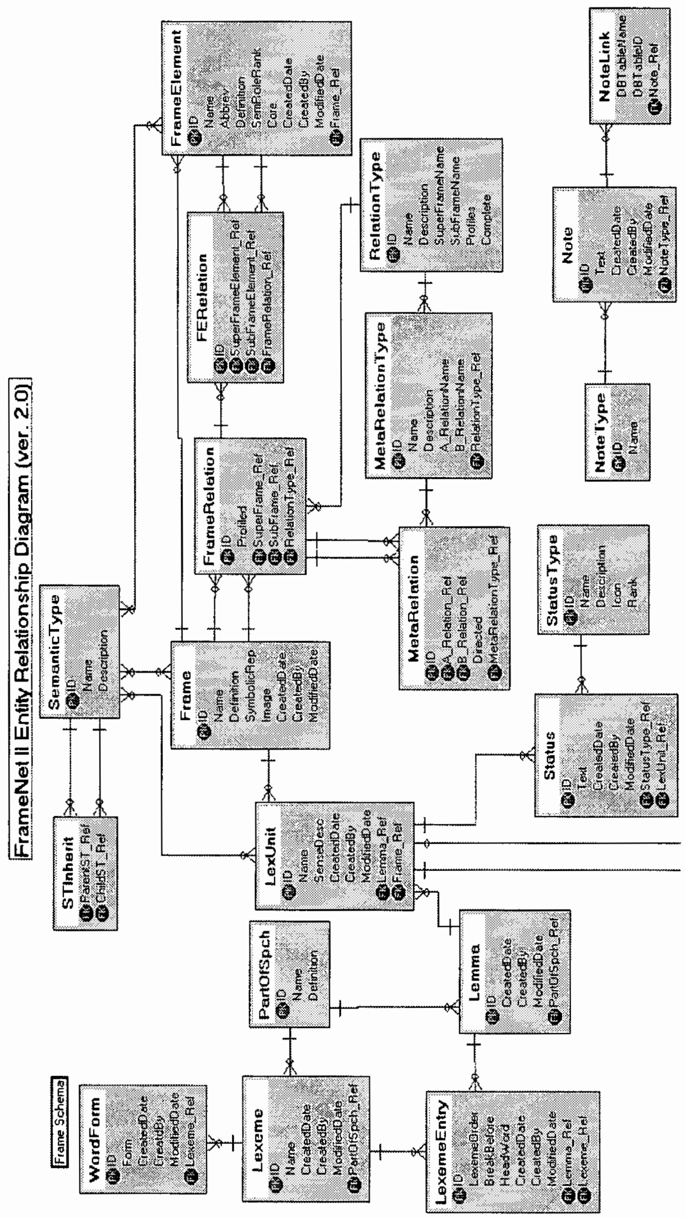
Once all English-specific information is removed, only information not specific to English remains in the database tables. This includes conceptual information in the upper right of the FrameNet database diagram in Figure 2, such as the Frames table, the FrameRelation table, the FERelation table, the FrameElements table, among other information. Once the FrameNet database has been stripped of its English-specific lexical descriptions and accompanying information, work begins on the second stage, namely re-populating the database with non-English lexical descriptions.

The first step consists of choosing a semantic frame from the stripped-down original database. For example, one might choose the `Communication_response` frame, which deals with communicating a reply or response to some prior communication or action (Johnson et al. 2003). English LUs belonging to this frame include the verbs *to answer, to counter*, and *to rejoin*, as well as the nouns *answer*, *response*, and *reply*, among others. In the FrameNet database we learn from the FrameElement table that this frame contains the FEs ADDRESSEE, MESSAGE, SPEAKER, TOPIC, and TRIGGER.

The second step in re-populating the database to arrive at a full-fledged non-English FrameNet is to identify with the help of dictionaries and parallel corpora lists of LUs in other languages that evoke the same semantic frame. This process is similar to the initial stages of English FrameNet (see Fillmore et al. 2003a), except for the fact that it is easier to compile lists of LUs because one already has access to existing frame descriptions and frame relations.[11] Our compilation of LUs for the `Communication_response` frame yields a list that includes German verbs and nouns such as *beantworten* ('to answer'), *entgegnen* ('to reply'), *die Ant-*

11. The availability of a stripped-down FN database with existing frames and FEs means that non-English FrameNets do not have to go through the entire process of frame creation (Fillmore et al. 2003: 304–313). It is important to keep in mind that at present FrameNet covers about 8900 lexical units in more than 600 frames. This means that its coverage of the English lexicon is somewhat limited when compared with other resources such as WordNet. Similarly, FrameNets for other languages will exhibit comparable limitations until FrameNet covers much larger areas of the English lexicon (or, even full coverage).

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 73)



(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 74)

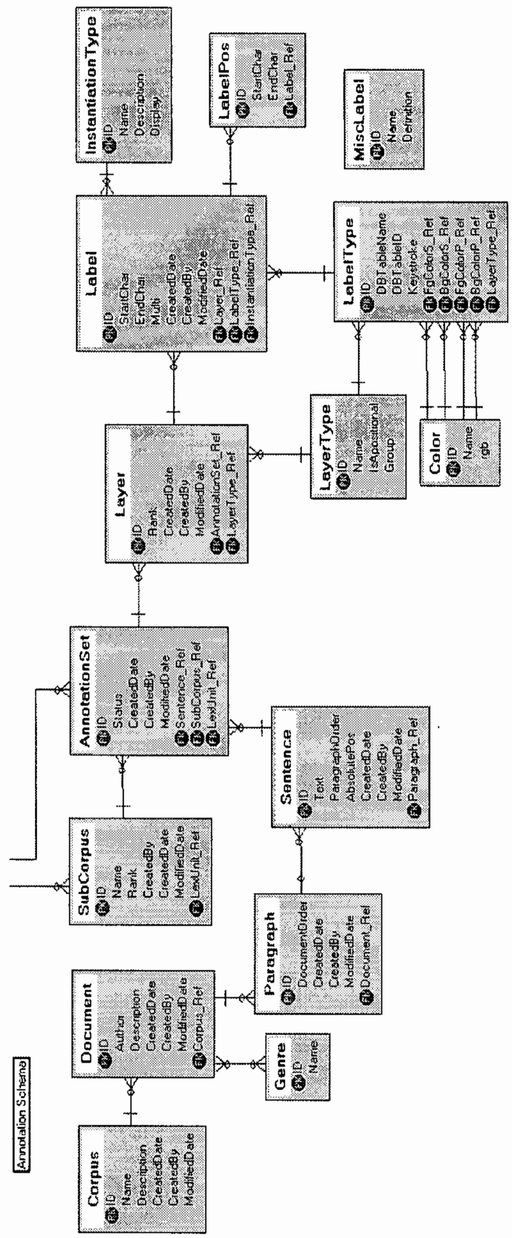


Figure 2. Structure of the FrameNet database (cf. Baker et al. 2003)

wort ('answer'), and *die Entgegnung* ('reply'). For Japanese, we find verbs such as *uke-kotae suru* ('to answer') and *ootoo suru* ('to reply') and nouns such as *kotae* ('answer'), which evoke the `Communication_response` frame. Similarly, in Spanish we find verbs such as *desmentir* ('deny') and *responder* ('to respond') and nouns such as *respuesta* ('response').

At this point it is necessary to briefly mention some similarities and differences among non-English FrameNets. Between the Spanish, Japanese, and German FrameNets there are differences in software setup and data sources used. Whereas Spanish FrameNet uses all of the original English FrameNet software (and has compiled its own corpus) (see Subirats and Petruck 2003), Japanese FrameNet is developing its own set of software tools to augment the tools provided by English FrameNet (see Ohara et al 2003). There are two projects concerned with developing FrameNet-type descriptions for German. The SALSA project at the University of the Saarland (Saarbrücken, Germany) (Erk et al. 2003) has developed its own annotation software and set of tools to annotate the entire TIGER corpus (König and Lezius 2003) with semantic frames. Its goal is to apply English-based frames to the TIGER corpus data, inventing new frames where necessary. In contrast, German FrameNet (Boas 2002), currently under construction at the University of Texas at Austin, is adapting the original FrameNet tools and aims to provide parallel lexical entries that are comparable in breadth and depth to those of English FrameNet. Another project, BiFrameNet (Fung and Chen 2004) focuses on the lexical description of Chinese and English for machine translation purposes. It differs from other FrameNets in that it takes a statistically-based approach to producing bilingual lexicon fragments.

To illustrate the process by which the stripped-down FrameNet database is repopulated with non-English data, the remainder of this section focuses primarily on the workflow of the Spanish FrameNet project (Subirats and Petruck 2003).[12] Once the appropriate lists of LUs evoking the frame are compiled for Spanish, they are added to the database using FrameNet's Lexical Unit Editor (cf. Fillmore et al. 2003b: 313–315). More specifically, for each LU information is stored about "(1) its name,

12. Spanish FrameNet currently contains about 80 annotated frames (with about 480 lexical units) as well as 500 frames that have not yet been annotated. Currently, SALSA has annotated approximately 540 lexical units, totaling more than 25,000 verb instances in the TIGER corpus. As both Japanese FrameNet and German FrameNet are currently in their beginning stages, no data have yet been made public.

(2) its part of speech, (3) its meaning, and (4) information about its formal composition" (Fillmore et al. 2003: 313). After adding all of the relevant information about each LU belonging to a frame to the database, a search is conducted in a very large corpus in order find sentences that illustrate the use of each of the LUs in the frame. This approach is parallel to the procedure employed by the original Berkeley FrameNet. Spanish FrameNet uses a 300 million-word corpus, which includes a variety of both New World and European Spanish texts from different genres such as newspapers, book reviews, and humanities essays (Subirats and Petruck 2003). To search the corpus and to create different subcorpora of sentences for annotation, the Spanish FrameNet project employs the Corpus Workbench software from the Institut für Maschinelle Sprachverarbeitung ('Institute for Natural Language Processing') at the University of Stuttgart (Christ 1994). Using an electronic dictionary of 600,000 word forms and a set of deterministic automata, a number of automatic processes select relevant example sentences from the corpus and subsequently compile subcorpora for each syntactic frame with which an LU may occur (cf. Subirats and Ortega 2000 and Ortega 2002). As in the creation of the original FrameNet, the subcorpora are then manually annotated with frame semantic information in order to arrive at clear example sentences illustrating all the different ways in which frame elements are realized syntactically. For annotation and database creation, Spanish FrameNet (SFN) employs the software developed by the original Berkeley FrameNet project. Figure 3 illustrates how the FrameNet Desktop Software is used by SFN to annotate part of an example sentence in the `Communication_response` frame.

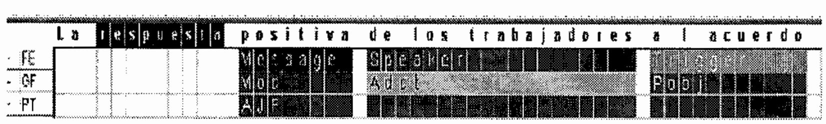


Figure 3. Annotation of a Spanish sentence in the `Communication_response` frame (Subirats and Petruck 2003)

The top line shows the example sentence *La respuesta positiva de los trabajadores al acuerdo* with the target noun *respuesta* ('response'), which evokes the `Communication_response` frame. Underneath the top line are three separate layers, one each for information pertaining to frame element names (FE), grammatical functions (GF), and phrase types (PT). After having become familiar with the frame and frame element defini-

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 77)

tions, annotators mark whole constituents with the appropriate colored tags representing the different frame elements of the `Communication_response` frame. In Figure 3, *positiva* ('positive') is tagged with the FE MESSAGE, *de los trabajadores* ('by the workers') is tagged with the FE SPEAKER, and *al acuerdo* ('to the accord') is marked with the FE TRIGGER. Once example sentences are marked with semantic tags, syntactic information about grammatical functions (GF) and phrase types (PT) is added semi-automatically and hand-corrected if necessary. Figure 4 shows only a small part of the software used for semantic annotation by members of the Spanish FrameNet team. Recall that manual semantic annotation covers the full range of examples of sentences illustrating each possible syntactic configuration in which a lexical item may occur. As such, Figure 4 gives a more complete illustration of the FrameNetDesktop Annotator software graphical user interface.

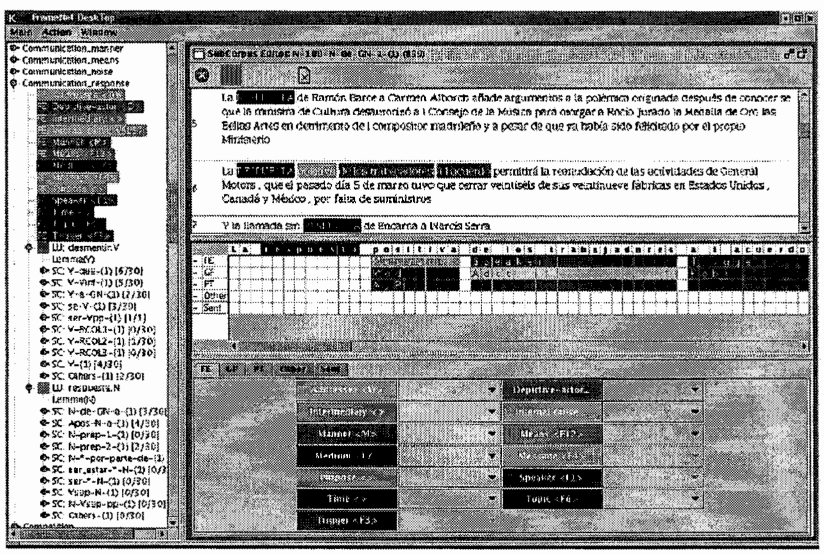


Figure 4. Annotation of a Spanish sentence using the FrameNet Annotator (Subirats and Petruck 2003)

The FrameNet Annotator window is divided into four main parts. The left part is the navigation frame that allows annotators to directly access all frames as well as their respective frame elements and lexical units contained in the MySQL database. The navigation frame shows different communication frames (`Communication_manner` and `Communication_noise` among others), where `Communication_response` is highlighted by an annotator to reveal the frame's FEs (ADDRESSEE, MEDIUM, and SPEAKER, among others). Clicking on a frame name reveals a list of LUs evoking the frame, in this case *desmentir* ('deny') and *respuesta* ('response') with their corresponding subcorpora containing example sentences previously extracted from the 300 million-word corpus (Subirats and Petruck 2003).

Selecting a lexical unit's subcorpus displays its respective example sentences in the top right part of the FrameNet Annotator window, in this case three example sentences with the target noun *respuesta*, which is highlighted in black. Clicking on one of the corpus sentences allows annotators to view it with the full set of layers in the middle part on the right of the Annotator window (see also Figure 3). The fourth part on the bottom right of the Annotator window displays the content space with the specifications for the different frame elements of the `Communication_Response` frame.[13]

Using the Annotator tool, members of the Spanish FrameNet team annotate a set of relevant corpus sentences in each subcorpus (see description above), thereby arriving at an extensive set of annotated subcorpora for each LU. As with the original FrameNet, the resulting annotated sentences represent an exhaustive list of the ways in which frame elements may be realized syntactically with a given target word. Once annotation is completed, the lexical units are stored with their annotated example sentences in the FrameNet MySQL database, which at the end of the workflow described in this section has evolved from a FrameNet database whose tables have been stripped of all of their English-specific data into a corresponding Spanish FrameNet database. Thus, Spanish FrameNet (and, to some degree, the corresponding Japanese and German FrameNets) is comparable in structure with that of the original English FrameNet database in that it contains the same set of frames and frame relations. It differs from English FrameNet in that the entries for argument taking nouns, verbs, and adjectives are in Spanish. Users may access the Spanish FrameNet database by the same set of web-based reports as for the original English FrameNet, i.e., for each LU in the database it is possible to display an Annotation Report, a Lexical Entry Report, and the corresponding valence tables. With this overview in mind, we now look at

13. Frame Elements are automatically annotated with grammatical function (GF) and phrase type (PT) information.

how semantic frames may be used to connect parallel lexicon fragments. More specifically, I show that the frame-semantic approach to MLLDs overcomes many of the problems faced by other MLLDs discussed in Section 2.

5.2. Linking parallel lexicon fragments via semantic frames

With FrameNets for multiple languages in place, the next step towards the creation of MLLDs on frame-semantic principles consists of linking the parallel lexicon fragments via semantic frames in order to be able to map lexical information of frame-evoking words from one language to another language (see also Heid and Krüger 1996, Fontenelle 2000, Boas 2002). Since the MySQL databases representing each of the non-English FrameNets are similar in structure to the English MySQL database in that they share the same type of conceptual backbone (i.e., the semantic frames and frame relations), this step involves determining which English lexical units are equivalent to corresponding non-English lexical units.

Table 3. Partial Realization Table for the verb *answer*

FE Name	*Syntactic Realizations*
Speaker	NP.Ext, PP_by_Comp, CNI
Message	INI, NP.Obj, PP_with.Comp, QUO.Comp, Sfin.Comp
Addressee	DNI
Depictive	PP_with.Comp
Manner	AVP.Comp, PPing_without.Comp
Means	PPing_by.Comp
Medium	PP_by.Comp, PP_in.Comp, PP_over.Comp
Trigger	NP.Ext, DNI, NP.Obj, Swh.Comp

To exemplify, consider the `Communication_response` frame discussed in the previous section. Suppose this frame, along with its frame elements and frame relations is contained in multiple FrameNets, where each individual database contains language-specific entries for all of the lexical units that evoke the frame in that language. Once we identify with the help of bilingual dictionaries a lexical unit whose entry we want to connect to a corresponding lexical unit in another language, we have to carefully consider the full range of valence patterns. This is a rather lengthy and complicated process because it is necessary that the different syntactic frames associated with the two lexical units represent translation equivalents in context. This procedure is facilitated by the use of parallel-aligned corpora, which allow a comparison between the LUs when they are embedded in different types of context (see, e.g. Wu 2000, Salkie 2002).[14] Consider, for example, the verb *answer*, whose individual frame elements may be realized syntactically in many different ways.[15] The realization table (in Table 3) is an excerpt from the FrameNet lexical entry for *answer*, which contains an excerpt from the valence tables as well as the corresponding annotated corpus sentences.

The column on the left contains the names of Frame Elements belonging to the `Communication_Response` frame, the column on the right lists their different types of syntactic realizations. For example, the FE SPEAKER may be realized either as an external noun phrase or a prepositional phrase complement headed by *by*. Alternatively, the FE SPEAKER does not have to be realized at all as in imperative sentences such as *Never answer this question with a straight no.*

Table 4. Excerpt from the Valence Table for *answer*

	Speaker	*TARGET*	*Message*	*Trigger*	*Addressee*
a.	NP.Ext	answer.v	NP.Obj	DNI	DNI
b.	NP.Ext	answer.v	PP_with.Comp	DNI	DNI
c.	NP.Ext	answer.v	QUO.Comp	DNI	DNI
d.	NP.Ext	answer.v	Sfin.Comp	DNI	DNI

Recall from Section 4 that each lexical entry also gives a full valence table illustrating the various combinations of frame elements and their syntactic realizations, which might be present in a given sentence. The valence table for the verb *answer* lists a total of 22 different linear sequences of Frame Elements, totaling 32 different combinations in which these sequences may be realized syntactically. As the full valence table for *answer* is rather long, we focus on only one linear sequence of Frame

14. We are currently looking into the possibility of automating this process by using a script that matches non-English examples expressing a specific constellation of FEs with their corresponding English examples expressing the same constellation of FEs.
15. We focus on verbs here, but similar procedures are followed for nouns and adjectives.

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 81)

Elements, namely the one in which the FE SPEAKER is followed by the target LU *answer* and the FE MESSAGE. The annotated example sentences in (15) correspond to the valence table excerpt in Table 4.

(15) a. Every time [$_{<\text{Speaker}>}$ you] *answer*$^{\text{Tgt}}$ [$_{<\text{Message}>}$ no], I shall adorn you with these pegs. [$_{<\text{Trigger}>}$ DNI] [$_{<\text{Addressee}>}$ DNI]

b. [$_{<\text{Speaker}>}$ She] *answered*$^{\text{Tgt}}$ [$_{<\text{Message}>}$ with another question]. [$_{<\text{Trigger}>}$ DNI] [$_{<\text{Addressee}>}$ INI]

c. [$_{<\text{Speaker}>}$ He] *answered*$^{\text{Tgt}}$, [$_{<\text{Message}>}$ This beer is expensive] [$_{<\text{Trigger}>}$ DNI] [$_{<\text{Addressee}>}$ DNI]

d. [$_{<\text{Speaker}>}$ He] *answered*$^{\text{Tgt}}$ [$_{<\text{Message}>}$ that he had gone too far now and that the country expected a dissolution]. [$_{<\text{Trigger}>}$ DNI] [$_{<\text{Addressee}>}$ DNI]

Table 4 is an excerpt from the full valence table for the verb *answer* and shows how one of the 22 different linear sequences of FEs may be realized in four different ways at the syntactic level. That is, besides sharing the same linear order of Frame Elements with respect to the position of the target LU *answer*, all four valence patterns have the FE SPEAKER realized as an external noun phrase, and the FEs TRIGGER and ADDRESSEE not realized overtly at the syntactic level, but null instantiated as Definite Null Instantiations (DNI). In other words, in sentences such as *He answered with another question* the FEs TRIGGER and ADDRESSEE are understood in context although they are not realized syntactically.

With both the language-specific as well as the language-independent conceptual frame information in place, we are now in a position to link this part of the lexical entry for *answer* to its counterparts in other languages. Taking a look at the lexical entry of *responder* ('to answer') provided by Spanish FrameNet, we find a list of Frame Elements and their syntactic realizations that is comparable in structure to that of its English counterpart in Table 4.

Spanish FrameNet also offers a valence table that includes for *responder* a total of 23 different linear sequences of Frame Elements and their syntactic realizations. Among these, we find a combination of Frame Elements and their syntactic realizations that is comparable to the English in Table 4 above. For example, the Frame Element MESSAGE may be realized as an adverbial phrase functioning as an object (AVP.AObj), a direct object quotation phrase (QUO.DObj), or a direct object phrase headed by *que* (queSind.DObj). Alternatively, it may not be realized syntactically, and therefore be understood as a definite null instantiation (DNI) based

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 82)

Table 5. Partial Realization Table for the verb *responder*

FE Name	*Syntactic Realizations*
Speaker	NP.Ext, NP.Dobj, CNI, PP_por.COMP
Message	AVP.AObj, DNI, QUO.Dobj, queSind.DObj, queSind.Ext
Addressee	NP.Ext, NP.IObj, PP_a.IObj, DNI, INI
Depictive	AJP.Comp
Manner	AVP.AObj, PP_de.AObj
Means	VPndo.AObj
Medium	PP_en.AObj
Trigger	PP_a.PObj, PP_de.PObj, DNI

Table 6. Excerpt from the Valence Table for *responder*

	Speaker	*TARGET*	*Message*	*Trigger*	*Addressee*
a.	NP.Ext	responder.v	QUO.DObj	DNI	DNI
b.	NP.Ext	responder.v	QueSind.DObj	DNI	DNI

on the context. Because of space limitations, we cannot discuss here all 23 linear sequences of Frame Elements and their syntactic realizations. Instead, we focus on only the one linear sequence that corresponds to the English counterpart(s), namely sentence (a) in Table 4. Consider the excerpt from the valence table of *responder* in Table 6.

Comparing Tables 4 and 6, we see that *answer* and *responder* exhibit comparable valence combinations with the Frame Elements SPEAKER and MESSAGE realized at the syntactic level, and the Frame Elements TRIGGER and ADDRESSEE not realized syntactically, but implicitly understood (they are both definite null instantiations). Having identified corresponding semantic frames, lexical units, and their semantic and syntactic combinatorial possibilities, it is now possible to link the parallel English and Spanish lexicon fragments by establishing correspondence links between the parts of the entries of the two lexical units shown it Tables 3–6 via semantic frames.

It is important to keep in mind that at this stage it is not yet possible to automatically connect lexical entries of the source and target languages. For example, although bilingual lexicon fragments might match in terms

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 83)

of their syntactic and syntactic valences, they might differ in terms of domain, frequency, connotation, and collocation in the two languages. This means that one must carefully compare each individual part of the valence table of a lexical unit in the source language with each individual part of the valence table of a lexical unit in the source language with each individual part of the valence table of a lexical unit in the target language. This effort requires at the first stage a detailed comparison using bilingual dictionaries and mono-lingual as well as parallel corpora in order to ensure matching translation equivalents (cf. also Boas 2001, Teubert 2002, Subirats and Petruck 2003, Ohara et al. 2004).[16] Once the translation equivalents are identified, it is possible to link the parallel lexicon fragments. As Figure 5 illustrates, the semantic frame serves as an inter-lingual representation between the valence and realization tables of the LUs in English and Spanish, thereby effectively establishing links between translation equivalents (annotated corpus sentences are not included).

In Figure 5, *answer* and *responder* are indexed with 'a'. This index points to the respective first lines in the valence tables of the two verbs and identifies the two syntactic frames as being translation equivalents of each other. At the top of the box in Figure 5 we see the verb *answer* with one of its 22 linear sequences of Frame Elements, namely SPEAKER, TRIGGER, MESSAGE, and ADDRESSEE (cf. Table 4 above). For this linear sequence, Figure 5 shows one possible set of syntactic realizations of these Frame Elements, that given in row (a) in Table 4 above. The 9a-designation following *answer* indicates that this lexicon fragment is the ninth linear configuration of Frame Elements out of a total of 22 linear sequences. Of the ninth linear sequence of Frame Elements 'a' indicates that it is the first of a list of various possible syntactic realizations of these Frame Elements (there are a total of four, cf. Table 4 above). As pointed out above, SPEAKER is realized syntactically as an external noun phrase, MESSAGE as an object noun phrase, and both TRIGGER and ADDRESSEE are null instantiated. The bottom of Figure 5 shows *responder* with the first of the 17 lin-

---

16. An anonymous reviewer has pointed out that bilingual dictionaries may not include all the necessary information. This suggests that in order to find appropriate translation equivalents it is necessary to rely on multiple resources simultaneously (dictionaries, corpora, intuitions of bilingual speakers, etc.). At the same time it is important to keep in mind that any of the individual resources used for creating bilingual lexicon fragments may have particular shortcomings (e.g. coverage).

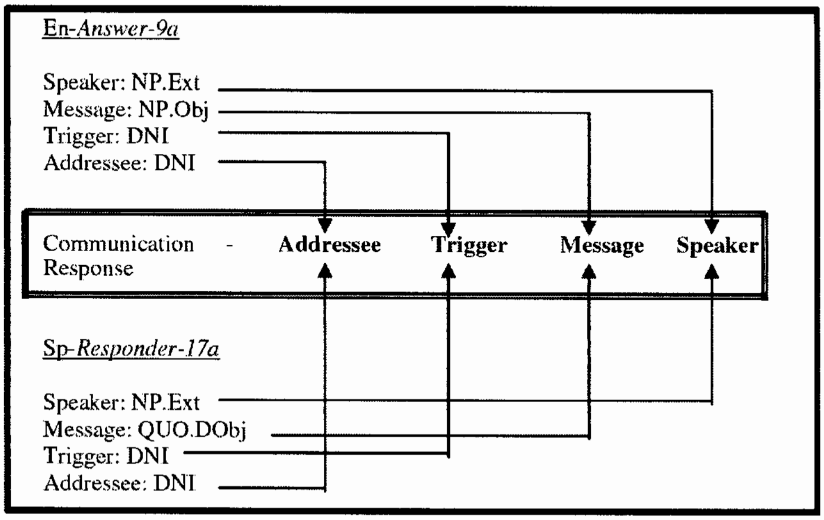


Figure 5. Linking partial English and Spanish lexicon fragments via semantic frames

ear sequences of Frame Elements (recall that there are a total of 23 linear sequences). For one of these linear sequences, we see one subset of syntactic realizations of these Frame Elements, namely the first row catalogued by Spanish FrameNet for this configuration (see row (a) in Table 6).

We can now link the two independently existing partial lexical entries at the top and bottom of Figure 5 by indexing their specific semantic and syntactic configurations as equivalents within the `Communication_Response` frame. This linking is indicated by the arrows pointing from the top and the bottom of the partial lexical entries to the mid-section in Figure 5, which symbolizes the `Communication_Response` frame at the conceptual level, i.e. without any language-specific specifications. The linking of parallel lexicon fragments is achieved formally by employing Typed Feature Structures (Emele 1994) that allow us to co-index the corresponding entries in a systemized fashion (see, e.g. Heid and Krüger 1996).

It is important to keep in mind that the English and Spanish data discussed in this section represent only a very small set of the full lexical entries of *answer* and *responder* in the `Communication_Response` frame. As such, these examples serve to illustrate how to systematically link parallel English and Spanish FrameNet fragments.[17] More specifically, in Figure 5 we have only looked at one possible syntactic realization out of one set of Frame Elements in a specific linear order. For the same order of Frame Elements there are four additional syntactic configurations (cf. Tables 4 and 6 above). For each of these sets, similar entries are needed in order to link them to each other. Recall that FrameNet provides for *answer* in the `Communication_Response` frame a total of 22 linear sequences of Frame Elements, totaling 32 different combinations in which these sequences may be realized syntactically. In order to arrive at a complete parallel lexicon fragment for *answer* and *responder*, it is necessary to create entries for each of the 32 combinations of *answer* and subsequently linking them to their corresponding Spanish counterparts. The same process is applied to link other lexical units across multilingual FrameNets.[18]

Clearly, the procedure outlined here appears to be very time intensive as currently the translation equivalents for each Frame Element Configuration (FEC) are largely determined manually, with the help of parallel corpora and bilingual dictionaries. Demanding though this procedure may be, it provides a solid basis for overcoming the types of linguistic problems typically encountered in the creation of multilingual lexical databases.

---

17. The current architecture of German FrameNet is based on identical (i.e., translation equivalent) texts. Using multilingual corpora such as the Europarl corpus (Koehn 2002), frame-evoking words are identified and subsequently explored in monolingual corpora in order to determine the full range of their uses. Then, other words in the same frame are explored (see Boas 2002). One problem not addressed in this paper (and currently under investigation) concerns translation mismatches where a single semantic frame or Frame Element may not be sufficient as an interlingual representation to map from one language to another language (see Section 2.3 for an example). Clearly, this is an important issue that needs to be addressed in future work. EuroWordNet (Vossen 2004) has developed a set of equivalence relations in combination with an Inter-Lingual-Index (ILI) in order to address mismatches between languages.
18. As this process is very time and labor intensive, efforts are currently under way to arrive at different ways for extracting parallel lexicon fragments automatically. A first step is to use parallel corpora to automatically identify translation equivalents in context in order to determine frame membership of lexical units across languages. For approaches incorporating automatic acquisition of lexical information from parallel corpora see Wu (2000), Farwell et al. (2004), Green et al. (2004), and Mitamura et al. (2004).

Another important point to keep in mind is that in this paper semantic frames do not serve as a true interlingua in which a concept is realized independently of a source language. However, the model presented here is neither a purely transfer-based system, because semantic frames are understood as an independently existing conceptual system that is not tied to any particular language. At this early point, semantic frames have been developed primarily on the basis of English, so it may appear as if they can only be used to describe the semantics of English LUs and one or two other languages. However, this is not the case. Because at this point semantic frames are best characterized as entities that combine aspects of true interlinguas and of transfer-based systems, I am using the term 'interlingual representation.' Once more languages are described using the FrameNet approach we may arrive at true universal semantic frames (e.g. communication, motion, etc.), which may then serve as a true interlingua. The remaining culture-specific frames (e.g. calendric unit frame; see Petruck and Boas 2003) will then have to be modeled using a transfer-based approach (see also Mel'čuk and Wanner (2001: 28), who propose the inclusion of transfer-mechanisms for systems that utilize true interlinguas).

5.3. Advantages of MLLDs based on Frame Semantics

Applying frame semantic principles to the design of MLLDs overcomes a number of theoretical and practical issues outlined in Section 2. With regard to polysemy we have seen that assigning different senses of words to individual semantic frames allows us to capture their syntactic and semantic distribution in great detail. This step shifts issues surrounding polysemy from the level of words to the level of semantic frames and FEs. As such, it is not only possible to describe overlapping polysemy effectively, but also diverging polysemy.

Table 7. Syntactic frames highlighting different parts of the `Communication_Statement` frame (Boas 2002: 1370)

1	[$_{\langle\text{speaker}\rangle}$ They] *announced*$^{\text{Tgt}}$ [$_{\langle\text{message}\rangle}$ the birth of their child].
2	[$_{\langle\text{medium}\rangle}$ The document] *announced*$^{\text{Tgt}}$ [$_{\langle\text{message}\rangle}$ that the war had begun].
3	[$_{\langle\text{speaker}\rangle}$ The conductor] *announced*$^{\text{Tgt}}$ [$_{\langle\text{message}\rangle}$ the train's departure] [$_{\langle\text{medium}\rangle}$ over the intercom].

For example, consider the `Communication_Statement` frame, which describes situations such as the following: the SPEAKER produces a (spoken or written) message, the ADDRESSEE is the person to whom the message is communicated, the MESSAGE identifies the content of what the SPEAKER is communicating to the ADDRESSEE, the MEDIUM is how the message is communicated, and the TOPIC is the subject matter to which the MESSAGE pertains. The verb *announce* is extremely flexible with respect to different types of perspectives it may take on a communication statement event.

Consider the examples in Table 8 discussed by Boas (2002). In each of the sentences, *announce* highlights different Frame Elements and their relations to each other. In German, each of the different uses of *announce* requires a different verb as a translation equivalent depending on the Frame Element Configuration and the type of perspective it takes on the communication statement scenario.

When *announce* occurs with only the SPEAKER and the MESSAGE frame elements, German prefers the use of *bekanntgeben*, *bekanntmachen*, *ankündigen*, and *anzeigen*, but not *ansagen* and *durchsagen*.[19] This is because the latter two verbs are primarily used in cases in which a MEDIUM frame element represents some sort of (electronic) equipment used to communicate

Table 8. Different syntactic frames of *announce* and corresponding German verbs (Boas 2002: 1370)

1	**speaker**	**TARGET**	**message**	
	NP.Ext	announce.v	NP.Obj	
	bekanntgeben, bekanntmachen, ankündigen, anzeigen			
2	**medium**	**TARGET**	**message**	
	NP.Ext	announce.v	Sfin_that.Comp	
	bekanntgeben, ankündigen, anzeigen			
3	**speaker**	**TARGET**	**message**	**medium**
	NP.Ext	announce.v	NP.Obj	PP_over.Comp
	ankündigen, ansagen, durchsagen			

19. In reality, a much finer-grained distinction (including contextual background information) is needed to formally distinguish between the semantics of individual verbs. E.g., *anzeigen* is used in a much more formal sense than the other verbs. In contrast, *ankündigen* is primarily used to refer to an event that will occur in the future (see Boas 2002).

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 88)

the MESSAGE to the ADDRESSEE such as in the third sentence in Table 7. This demonstrates that it is not sufficient to simply generalize over senses of words that may be used as synonyms of each other. Instead, it is necessary for MLLDs to capture the full range of possible translation equivalents before arriving at decisions about which German verbs may serve as possible equivalents to a specific syntactic frame listed in an entry for an English lexical unit.[20]

MLLDs based on frame semantic principles may also help with overcoming problems surrounding word sense disambiguation caused by analogous valence patterns. Our discussion of *cure* and *get* in Section 2 illustrated that the proper identification of verb senses occurring with multiple syntactic frames is often difficult. By detailing how different types of syntactic frames are used to express diverse semantic concepts represented by semantic frames it becomes possible to correctly identify a word sense not only within a single language, but also mapping that sense to appropriate translation equivalents across languages.[21] For example, when *cure* occurs with the [NP, V, NP] syntactic frame, it may express either the preservation sense (*The mother cured the ham*), or the healing sense (*The mother cured the child*), depending on the choice of semantic object. Explicitly stating the different semantics of the postverbal object and other constituents in frame semantic terms as part of the lexical entry not only allows us to disambiguate the two senses straightforwardly. It also enables us to identify the proper translation equivalent for other languages by

---

20. Note that it will not suffice to only map a lexical unit's equivalents to German. Instead, a MLLD based on frame semantic principles has to map each syntactic frame of a German lexical unit back to a syntactic frame of an English lexical unit in order to ensure that the two are capable of expressing the same semantic space. Whenever there are discrepancies, a revision of mappings between lexical entries will be necessary. This example illustrates that although parallel corpora may be helpful for the automatic acquisition of bilingual lexicon fragments, it is still necessary to manually check the translation equivalents before finalizing any parallel lexicon fragments (see Boas 2001, 2002).
21. Syntactic frames alone are not sufficient for identifying the correct word sense. Instead, it is necessary to first determine the semantic types of the verb's arguments (using other lexical resources such as WordNet). Once we have information about the semantic types of the verb's arguments, it then becomes possible to link the syntactic frame to specific semantic frames, thereby correctly identifying word senses. For details about the linking of semantic and syntactic information for each of a word's multiple senses, see Goldberg (1995), Rappaport Hovav & Levin (1998), and Boas (2001).

using semantic frames to map the senses across languages. For German, we thus find *pökeln* for the preservation sense of *cure*, and *heilen* for the healing sense of *cure*.

Another advantage of employing semantic frames for the structuring of MLLDs is that knowledge about different lexicalization patterns can be accounted for systematically at the level of Frame Elements. The differences in lexicalization patterns between English and Japanese motion verbs discussed in Section 2.3 have shown that the two languages vary in the types of PATH Frame Elements. Whereas English exhibits only one general PATH FE, Japanese makes a more fine-grained distinction into ROUTE and BOUNDARY (cf. Ohara et al. 2004). To account for these differences, it is necessary to introduce the notion of Frame Element sub-categories that identify ROUTE and BOUNDARY as subtypes of the more general PATH FE. When mapping a PATH FE from English to Japanese it is thus important to rely on the valence patterns to determine the subtype of PATH FE for Japanese. For example, in English *the bridge* and *the river* may appear as a PATH FE with verbs such as *go*, *pass*, and *traverse*. As we have seen in Section 2.3, *wataru* ('go across') behaves similarly to English in that it may occur with *hasi* ('the bridge') and *kawa* ('the river'). In contrast, *koeru* ('go beyond') only occurs with *kawa*, but not with *hasi*. In a frame-based MLLD this difference is accounted for in terms of lexical entries that specify for each lexical unit the different combinations of FEs with which it occurs. Using the mapping and numerical indexing mechanisms outlined in the previous section, we can then link English and Japanese lexicon fragments according to the equivalent Frame Element Configurations. It is at this level that the fine-grained differences between the ROUTE and BOUNDARY subcategories of Japanese path FEs and their English PATH counterpart are encoded.

6. Differences to other MLLDs

Frame-based MLLDs differ from other MLLDs in a number of significant ways. The first difference is in their overall architecture. For example, EuroWordNet (Peters et al. 1998, Vossen 2004) consists of individual databases for eight European languages structured along the original Princeton WordNet for English (Fellbaum 1998). As such, EuroWordNet relies on decontextualized concepts for lexical descriptions. The sense relations between semantically related words (synsets) such as hyponymy, antonymy, meronymy, etc. differ from semantic frames in that they repre-

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 90)

sent ontological relations holding between synsets. These sense relations are internal to the conceptual architecture of EuroWordNet. In contrast, frame-based MLLDs are based on linguistically motivated concepts (semantic frames) that are external to the units of analysis. As such, frame-based MLLDs and MLLDs based on WordNet such as EuroWordNet offer complementary types of information.

The second difference between frame-based MLLDs and other MLLDs is the combination of syntactic and semantic information. Some lexical databases provide detailed conceptual ontologies representing hierarchies of different lexical relations. For example, SIMuLLDA (Janssen 2004) provides a fine-grained formal concept analysis for nouns in English and French. But it does not offer any significant information about their syntactic distribution such as different types of modification. EuroWordNet (Vossen 2001, 2004) offers a detailed semantic analysis of lexical semantic relations between synsets, but it only contains partial syntactic information in the form of one or two example sentences illustrating how a word is used in context. In contrast, other lexical resources such as SIMuLLDA and EuroWordNet differ from frame-based MLLDs in that they provide different types of conceptual information as well as access to ontological information which is not currently available in frame-based dictionaries. Moreover, WordNet and its multilingual counterpart EuroWordNet offer a much broader coverage than FrameNet and its multilingual extensions.

Another difference concerns the methodology used to create and link MLLDs. In EuroWordNet, each language-specific WordNet is an autonomous language-specific ontology where each language has its own set of concepts and lexical-semantic relations based on the lexicalization patterns of that language (cf. Vossen 2004).[22] EuroWordNet differentiates between language specific and language-independent modules. The language-independent modules consist of a top concept ontology and an unstructured Inter-Lingual-Index (ILI) that provides mappings across individual language WordNet structures and consists of a condensed universal index of meaning (so far, 1024 fundamental concepts) (Vossen 2001, 2004). Each ILI record consists of a synset and an English gloss specifying its meaning and source. Although most concepts in each WordNet are

22. In EuroWordNet, there are no concepts for which there are not words or expressions in a language. In contrast, GermaNet (Hamp & Feldweg 1997, Kunze & Lemnitzer 2002), which is a spin-off from the German EuroWordNet consortium, uses non-lexicalized, so-called artificial concepts for creating well-balanced taxonomies.

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 91)

ideally related to the closest concepts in the ILI, there is a set of equivalence relations that map between individual WordNets and the ILI (cf. Vossen 2004: 164–167).

Identifying equivalents across languages with EuroWordNet requires three steps. First, one must identify the correct synset to which the sense of a word belongs in the source language. Next, using an equivalence relation (e.g. EQ_HAS_HYPERONYM (when a meaning is more specific than any available ILI record), Vossen 2004: 164) the synset meaning is mapped to the ILI (which is linked to a top-level ontology). Finally, the corresponding counterpart is identified in the target language by mapping from the ILI to a synset in the target language.

Frame-based MLLDs differ from the EuroWordNet architecture in that all meanings are described directly with respect to the same semantic frame. Differences between the languages are thus to be found in the various ways in which the conceptual semantics of a frame are realized syntactically. On this approach, semantic frames are only used to identify and link meaning equivalents (Frame Elements). As we have seen in Section 5.2, the linking of the syntactic valence patterns is established by directly identifying the translation equivalents (on the basis of parallel corpora) and indexing them with each other.[23] Differences between languages are thus to be found in the various ways in which the conceptual semantics of a frame are realized syntactically.

It is important to keep in mind that at this early stage FrameNets for Spanish, German and Japanese are only linking their entries to existing English FrameNet entries, but not to entries across all the languages. The next step involves linking lexical entries across languages in order to test the applicability of semantic frames as a cross-linguistic metalanguage. Extending the FrameNet approach to different languages is in its preliminary stages. Clearly, much research on frame-based MLLDs remains to be done. One of the open questions concerns the description and mapping of adjectives and nouns across languages that differ in lexicalization patterns. This question has already been addressed by other MLLDs such as EuroWordNet. Another important issue concerns mismatches between languages. That is, we need to carefully consider the different strategies

23. Our approach differs from Fontenelle's (2000) analysis in that Fontenelle primarily relies on data from existing bilingual dictionaries to establish parallel lexicon fragments. Another difference is that Fontenelle augments his approach with additional semantic layers from Mel'čuk's Meaning-Text Theory in order to establish lexical functions.

that should be employed when encountering translation mismatches. Here, too, frame-based MLLDs may benefit from a variety of other resources to solve these problems: the detailed conceptual information contained in other resources such as EuroWordNet (Vossen 2004), information about complex translation mismatches provided by Acquilex (Copestake et al. 1995), statistical information on translation matches and mismatches provided by BiFrameNet (Fung and Chen 2004), or paraphrase relations as proposed by Mel'čuk's Meaning-Text Theory (Mel'čuk et al. 1988; see also Fontenelle 2000).

7. Conclusions and outlook

This paper has outlined the methodology underlying the design and construction of frame-based MLLDs. Starting with a discussion of the Berkeley FrameNet for English, I have shown how its semantic frames can be systematically employed to create parallel lexicon fragments for Spanish, Japanese, and German. In discussing the individual steps necessary for the creation of multilingual FrameNets, I have demonstrated how the use of semantic frames overcomes a number of linguistic problems traditionally encountered in cross-linguistic analyses. These include diverging polysemy structures, lexicalization patterns, and identifying and measuring paraphrase relations and translation equivalents.

At the center of the work-flow in the creation of frame-based MLLDs are the following three steps: (1) identification of translation equivalents based on existing English FrameNet entries, parallel corpora, and bilingual dictionaries; (2) attestation and semantic annotation of translation equivalents based on examples in both parallel corpora and large monolingual corpora; (3) creation of parallel lexical entries that are linked to English FrameNet entries on the basis of semantic frames. Since not all steps can be automated, this process is rather time and labor intensive.

The construction of frame-based MLLDs is only in its first phase. Clearly, future work will have to be extended to domains beyond those discussed in this paper to achieve broader coverage (i.e. beyond the 8,900 Lexical Units currently offered by FrameNet). Other multi-lingual resources such as EuroWordNet not only provide much broader coverage, but also contain useful conceptual information not currently encoded by FrameNet that may support this effort. Another important point will be to determine the feasibility of a truly independent metalanguage based on semantic frames for connecting multiple FrameNets. The idiosyncratic

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 93)

syntactic realizations of Frame Elements in the communication domain discussed in this paper for English and Spanish has shown that this is not an easy task. The fact that the large number of idiosyncratic valence patterns of verbs may evoke the same frame (or only certain aspects of a frame) suggests that it might be necessary to distinguish between truly universal frames and language-specific frames. The former would be modeled by linking the syntactic valence patterns of a lexical unit directly to a semantic frame. In this case semantic frames would serve as an interlingua as outlined in Section 5.3 above. The latter would be modeled by employing transfer rules between language pairs where specific transfer rules would have to specify how specific frames (or parts of frames) are mapped from one language to another. However, at this point it is too early to provide a definite answer to this problematic issue. It can only be addressed thoroughly once coverage has been extended significantly (both in terms of Lexical Units and of languages analyzed).

Future efforts will have to concentrate on finding mechanisms that allow for greater automation of the processes described in this paper, in particular the identification of translation equivalents in parallel corpora. Finally, it must be seen how multi-lingual FrameNets can be used to improve current and future machine translation systems.

References

Alsina, V. and J. DeCesaris
2002 Bilingual lexicography, overlapping polysemy, and corpus use. In: B. Altenberg and S. Granger (eds.), *Lexis in Contrast*, 215–230. Amsterdam/Philadelphia: Benjamins.

Altenberg, B. and S. Granger
2002 Recent trends in cross-linguistic lexical studies. In: B. Altenberg and S. Granger (eds.), *Lexis in Contrast*, 3–50. Amsterdam/Philadelphia: Benjamins.

Atkins, B.T.S.
1994 Analyzing the verbs of seeing: A frame semantic approach to corpus lexicography. In: C. Johnson et al. (eds.), *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*, 42–56. Berkeley: Berkeley Linguistics Society.

Atkins, B.T.S., N. Bel, F. Bertagne, P. Bouillon, N. Calzolari, C. Fellbaum, R. Grishman, A. Lenci, C. MacLeod, M. Palmer, G. Thurmair, M. Villegas, and A. Zampolli
2002 From resources to applications. Designing the multilingual ISLE lexical entry. In: *Proceedings of LREC 2002*, 687–693, Gran Canaria, Spain.

Baker, C.F., C.J. Fillmore, and J.B. Lowe
1998 The Berkeley FrameNet Project. In: *COLING-ACL'98: Proceedings of the Conference*, 86–90.

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

Béjoint, H.
2000 *Modern Lexicography*. Oxford: Oxford University Press.

Boas, Hans C.
2001 Frame Semantics as a framework for describing polysemy and syntactic structures of English and German motion verbs in contrastive computational lexicography. In: P. Rayson, A. Wilson, T. McEnery, A. Hardie, and S. Khoja (eds.), *Proceedings of Corpus Linguistics 2001*, 64–73.

Boas, Hans C.
2002 Bilingual FrameNet dictionaries for machine translation. In: M. González Rodríguez and C. Paz Suárez Araujo (eds.), *Proceedings of the Third International Conference on Language Resources and Evaluation*, 1364–1371. Las Palmas, Spain.

Boas, Hans C.
2005 From theory to practice: Frame Semantics and the design of FrameNet. In: S. Langer and D. Schnorbusch (eds.), *Semantisches Wissen im Lexikon*, 129–160. Tübingen: Narr.

Chesterman, A.
1998 *Contrastive Functional Analysis*. Amsterdam/Philadelphia: John Benjamins.

Chodkiewicz, C., D. Bourigault, and J. Humbley
2002 Making a workable glossary out of a specialized corpus: Term extraction and expert knowledge. In: B. Altenberg and S. Granger (eds.), *Lexis in Contrast*, 249–270. Amsterdam/Philadelphia: Benjamins.

Christ, O.
1994 A modular and flexible architecture for an integrated corpus query system. In: *COMPLEX'94*, Budapest, 1994.

Copestake, A., T. Briscoe, P. Vossen, A. Ageno, I. Castellon, F. Ribas, G. Rigau, H. Rodriguez, and A. Samiotou
1995 Acquisition of lexical translation relations from MRDs. *Machine Translation* 9: 183–219.

Cruse, A.
1986 *Lexical Semantics*. Cambridge: Cambridge University Press.

Emele, M.
1994 TFS – The typed feature structure representation formalism. In: *Proceedings of the International Workshop on Sharable Natural Language Resources (SNLR)*, Nara, Japan, 1994.

Erk, K., A. Kowalski, and S. Padó
2003 Towards a resource for lexical semantics: A large German cor-

pus with extensive semantic annotation. In: *Proceedings of ACL 2003*, Sapporo.

Farwell, D., L. Guthrie, and Y. Wilks
1993 Automatically creating lexical entries for ULTRA, a multilingual MT system. *Machine Translation* 8: 183–219.

Farwell, D., S. Helmreich, B. Dorr, N. Habash, F. Reeder, K. Miller, L. Levin, T. Mitamura, E. Hovy, O. Rambow, and A. Siddharthan
2004 Interlingual annotation of multilingual text corpora. In: *Proceedings of the North American Chapter of the Association for Computational Linguistics Workshop on Frontiers in Corpus Annotation*, 55–62. Boston, MA.

Fellbaum, C.
1998 *WordNet: An Electronic Lexical Database*. Cambridge, Mass.: MIT Press.

Fillmore, C.J.
1970 The grammar of hitting and breaking. In: R.A. Jacobs and P.S. Rosenbaum (eds.), *Readings in English Transformational Grammar*, 120–133. Ginn and Company.

Fillmore, C.J.
1975 An alternative to checklist theories of meaning. In: C. Cogen et al. (eds.), *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 123–131. Berkeley: Berkeley Linguistics Society.

Fillmore, C.J.
1982 Frame Semantics. In: Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, 111–138. Seoul: Hanshin.

Fillmore, C.J. and B.T.S. Atkins
1992 Toward a frame-based lexicon: The semantics of RISK and its neighbors. In: A. Lehrer and E. Kittay (eds.), *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*, 75–102. Hillsdale: Erlbaum.

Fillmore, C.J. and B.T.S. Atkins
1994 Starting where the dictionaries stop: The challenge for computational lexicography. In: B.T.S. Atkins and A. Zampolli (eds.), *Computational Approaches to the Lexicon*, 349–393. Oxford: Oxford University Press.

Fillmore, C.J. and B.T.S. Atkins
2000 Describing polysemy: The case of 'crawl'. In: Y. Ravin and C. Leacock (eds.), *Polysemy*, 91–110. Oxford: Oxford University Press.

Fillmore, C.J., C.R. Johnson, and M.R.L. Petruck
2003a Background to FrameNet. *International Journal of Lexicography* 16: 235–251.

Fillmore, C.J., M.R.L. Petruck, J. Ruppenhofer, and A. Wright
2003b FrameNet in action. The case of attaching. *International Journal of Lexicography* 16.3: 297–333.

Fontenelle, T.
2000 A bilingual lexical database for frame semantics. *International Journal of Lexicography* 14.4: 232–248.

Fung, P. and B. Chen
2004 BiFrameNet: Bilingual Frame Semantics resource construction by cross-lingual induction. In: *Proceedings of COLING 2004*. Geneva, Switzerland.

Goddard, C.
2000 Polysemy: A problem of definition. In: Y. Ravin and C. Leacock (eds.), *Polysemy*, 129–151. Oxford: Oxford University Press.

Goldberg, A.
1995 *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.

Green, R., B. Dorr, and P. Resnik
2004 Inducing frame semantic verb classes from WordNet and LDOCE. In: *Proceedings of the Workshop on Text Meaning and Interpretation*, Association for Computational Linguistics, Barcelona, Spain.

Hamp, B. and H. Feldweg
1997 GermaNet: A lexical-semantic net for German. In: P. Vossen, N. Calzolari, G. Adriaens, A. Sanfilippo, and Y. Wilks (eds.), *Proceedings of the ACL/EACL-97 Workshop on automatic information extraction and building of lexical semantic resources for NLP applications*, 9–15. Madrid.

Heid, U. and K. Krüger
1996 Multilingual lexicon based on Frame Semantics. In: *Proceedings of the AISB Workshop on Multilinguality in the Lexicon*. Brighton.

Janssen, M.
2004. Multilinguallexical databases, lexical gaps, and SIMuLLDA. *International Journal of Lexicography* 17.2: 137–154.

Johnson, C.R., M.R.L. Petruck, C.F. Baker, M. Ellsworth, J. Ruppenhofer, and C.J. Fillmore
2003 *FrameNet: Theory and Practice*. Technical Report. Berkeley: International Computer Science Institute.

Koehn, P.
2002 *Europarl: A multilingual corpus for evaluation of machine translation*. Ms., University of Southern California.

König, E. and W. Lezius
2003 *The TIGER language – A description language for syntax graphs, formal definition*. Technical report Institut für Maschinelle Sprachverarbeitung, University of Stuttgart.

Kunze, C. and L. Lemnitzer
2002 GermaNet – representation, visualization, application. In: *LREC 2002 Proceedings* Vol. V: 1465–1491.

Laecock, C. and Y. Ravin
2000 *Polysemy*. Oxford: Oxford University Press.

Mel'čuk, I., N. Arbatchewsky-Jumarie, L. Dagenais, L. Elnitsky, L. Iordanskaja, M.-N. Lefebvre, and S. Mantha
1988 *Dictionnaire explicatif et combinatoire du Français contemporain. Recherches lexico-sémantiques*. Montréal: Les Presses de l'Université de Montréal.

Mel'čuk, I. T. and Wanner
2001 Toward a lexicographic approach to lexical transfer in machine translation (Illustrated by the German-Russian Language Pair). In: *Machine Translation* 16: 21–87.

Mitamura, T., K. Miller, B. Dorr, D. Farwell, N. Habash, S. Helmreich, E. Hovy, L. Levin, O. Rambow, F. Reeder, and A. Siddharthan
2004 Semantic annotation for interlingual representation of multilingual texts. In: *Proceedings of the Workshop on Beyond Named Entity Recognition: Semantic Labeling for NLP Tasks*, LREC.

Ohara, K., S. Fujii, H. Saito, S. Ishizaki, T. Ohori, and R. Suzuki
2003 The Japanese FrameNet Project: A preliminary report. In: *Proceedings of the Pacific Association for Computational Linguistics (PACLING03)*, 249–254.

Ohara, K., S. Fujii, H. Saito, S. Ishizaki, T. Ohori, and R. Suzuki
2004 The Japanese FrameNet Project. An introduction. In: *Proceedings of the satellite workshop on building lexical resources from semantically annotated corpora*, 9–11. Fourth international Conference on Language Resources and Evaluation (LREC) 2004.

Ortega, M.
2002 *Intersección de autómatas y transductores en el análisis sintáctico de un texto*. MA Thesis, Polytechnic University of Catalonia, Spain.

Peters, W., I. Peters, and P. Vossen
1998 The reduction of semantic ambiguity in linguistic resources. In: A. Rubio, N. Gallardo, R. Catro, and A. Tejada (eds.), *Proceedings of the First International Conference on Language Resources and Evaluation*, 409–416. Granada.

Petruck, M.R.L.
1996 Frame Semantics. In: J. Verschueren, J-O Östman, J. Blommaert and C. Bulcaen (eds.), *Handbook of Pragmatics*, 1–13. Amsterdam/Philadelphia: Benjamins.

Petruck, M.R.L. and H.C. Boas
2003 All in a day's week. In: E. Hajičová, A. Kotěšovcová, and J. Mírovský (eds.), *Proceedings of the 17th International Congress of Linguists*, CD-ROM. Prague: Matfyzpress.

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

Rappaport Hovav, M. and B. Levin
1998 Building verb meaning. In: M. Butt and W. Geuder (eds.), *The Projection of Arguments*, 97–134. Stanford: CSLI Publications.

Salkie, R.
2002 Two types of translation equivalence. In: B. Altenberg and S. Granger (eds.), *Lexis in Contrast*, 51–72. Amsterdam/Philadelphia: Benjamins.

Sinclair, J.
1996 An international project in multilingual lexicography. In: J. Sinclair, J. Payne, and P. Hernández (eds.), Corpus to corpus: A study of translation equivalence. Special issue of the *International Journal of Lexicography* 9: 179–196.

Subirats, C. and M. Ortega
2000 Tratamiento automático de la información textual en español mediante bases de información lingüistica y transductores. *Estudios de Lingüistica del Espanol* 10.

Subirats, C. and M. Petruck
2003 *Surprise: Spanish FrameNet*. Presentation at the workshop on Frame Semantics, International Congress of Linguists, July 29th, 2003, Prague.

Svensén, B.
1993 *Practical lexicography. Principles and methods of dictionary-making*. Oxford: Oxford University Press.

Talmy, L.
1985 Lexicalization patterns: semantic structures in lexical forms. In: T. Shopen (ed.), *Language Typology and Syntactic Description*, 57–149. Cambridge: Cambridge University Press.

Talmy, L.
2000 *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press.

Teubert, W.
2002 The role of parallel corpora in translation and multilingual lexicography. In: B. Altenberg and S. Granger (eds.), *Lexis in Contrast*, 189–214. Amsterdam/Philadelphia: Benjamins.

Viberg, Å.
2002 Polysemy and disambiguation cues across languages: The case of Swedish *få* and English *get*. In: B. Altenberg and S. Granger (eds.), *Lexis in Contrast*, 119–150. Amsterdam/Philadelphia: Benjamins.

Vossen, P.
1998 Introduction to EuroWordNet. In: N. Ide, D. Greenstein, and P. Vossen (eds.), Special Issue on EuroWordNet. *Computers and the Humanities* 32: 73–89.

Vossen, P.
2001 Condensed meaning in EuroWordNet. In: P. Bouillon and F. Busa (eds.), *The language of word meaning*, 363–383. Cambridge: Cambridge University Press.

(V9 20/5/09 02:49) WDG (155mm×230mm) TimesNRMT 1114 Boas (AC1) pp. 59–100 1114 Boas_03_Ch03 (p. 99)

Vossen, P.
2004 EuroWordnet: A multilingual database of autonomous and language specific wordnets connected via an inter-lingual-index. *International Journal of Lexicography* 17.2: 161–173.

Wu, D.
2000 Bracketing and aligning words and constituents in parallel text using stochastic inversion transduction crammars. In: J. Veronis (ed.), *Parallel Text Processing: Alignment and Use of Translation Corpora*. Dordrecht: Kluwer.