Workshop Program

09:00	W E L C O M E	
09:15	Charles J. Fillmore and Collin F. Baker	The Evolution of FrameNet Annotation Practices
10:00	Kyoko Hirose Ohara, Seiko Fujii, Shun Ishizaki, Toshio Ohori, Hiroaki Saito, and Ryoko Suzuki	The Japanese FrameNet Project: An Introduction
10:30	Carlos Subirats and Hiroaki Sato	Spanish FrameNet and FrameSQL
11:00	B R E A K	
11:30	Michael Ellsworth, Katrin Erk, Paul Kingsbury, and Sebastian Padó	PropBank, SALSA and FrameNet: How Design Determines Product
12:00	Jan Hajič, Martin Holub, Marie Hučínová, Martin Pavlík, Pavel Pecina, Pavel Straňák, and Pavel Martin Šidák	Validating and Improving Czech WordNet via Lexico-Semantic Annotation of the Prague Dependency TreeBank
12:30 - 14:00	L U N C H	
14:00	Annette Frank	Generalizations over Corpus-Induced Frame Assignment Rules
14:30	Lea Cyrus, Hendrik Feddes, and Frank Schumacher	Annotating Predicate-Argument Structure for a Parallel TreeBank
15:00	Bonnie Dorr, Lori Levin, Owen Rambow, David Farwell, Rebecca Green, Nizar Habash, Stephen Helmreich, Eduard Hovy, Keith J. Miller, Teruko Mitamura, Florence Reeder, and Advaith Siddharthan	Semantic Annotation and Lexico-syntactic Paraphrase
15:30	B R E A K	
16:00	Myroslava Dzikovska, Mary Swift, and James Allen	Building a computational lexicon and ontology with FrameNet
16:30	Gerhard Fliedner	Using FrameNet for Question Answering
17:00	(General Discussion)	Crosslinguistic FrameNets

Workshop Organizers

Charles J. Fillmore, International Computer Science Institute, Berkeley, CA, U.S.A.
Manfred Pinkal, Computational Linguistics, Universität Saarlandes, Saarbrücken, Germany
Collin F. Baker, International Computer Science Institute, Berkeley, CA, U.S.A.
Katrin Erk, Computational Linguistics, Universität Saarlandes, Saarbrücken, Germany

Workshop Program Committee

Collin F. Baker
Katrin Erk
Charles J. Fillmore
Eva Hajičová, Center for Computational Linguistics, Charles University, Prague, Czech Republic
Ulrich Heid, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Germany
Mirella Lapata, Dept. of Computer Science, Univ. of Sheffield, Sheffield, U.K.
Martha Palmer, Dept. of Computer and Information Science, Univ. of Pennsylvania, PA, U.S.A.
Manfred Pinkal

Table of Contents

Author Index

The Evolution of FrameNet Anotation Practices*

Charles J. Fillmore and Collin F. Baker

International Computer Science Institute
Berkeley, California, U.S.A.
{fillmore,collinb}@icsi.berkeley.edu

Abstract

This paper traces the evolution of the annotation principles and practices of the Berkeley FrameNet project from 1997 to the present. Beginning with a straightforward way of building valence descriptions on the basis of annotated corpus sentences, represented originally in SGML tags around words and word strings, the project quickly saw the need to include many more kinds of information. With the switch to stand-off annotation represented in a relational database, convenient mechanisms for representing such added information were developed, leading ultimately to the ability to accomplish full-text semantic analysis.

1. The Origins of FrameNet

1.1. Frame Semantics

In beginning of the FrameNet project, we took as our theoretical and descriptive framework a set of concepts and assumptions from frame semantics: **frames** are the situation types against which lexical meanings are understood and interpreted; aspects and components of given frames are called **frame elements (FEs)**, roughly equivalent to deep cases or thematic roles, but defined relative to specific frames; a **lexical unit (LU)** is a word in a single sense, a pairing of a word with a sense. [1]

For any **target** LU (the LU being analyzed) our job has been to find good example sentences, to label the constituents of the sentence or phrase built up around that word with labels standing for the FEs that belong to the frame evoked by the target.

Although the idea of thematic roles broadly defined is widely accepted in linguistics, in FrameNet, we chose to define frame-specific role names (the FEs), distinct from the "standard" thematic roles for several reasons: first, there are many frames in which the participants do not nicely map into the usual thematic roles[2]; second, even in certain fairly long lists of general-purpose semantic semantic roles, we felt that we needed more specific names both for the sake of semantic transparency for the annotator and user, and for the imagined inference-machine that we hoped some day to build for frame-annotated texts[3].

In contrast to projects devoted exclusively to verbs[4], we were committed from the beginning to including words of all parts of speech in our frames. This means, for example, that both *devastate* and *devastation* appear in the Destroying frame, both *classify* and *classification* in Categorization, and while the structures built around verbs and nouns will differ syntactically, the semantic annotations will reflect their common frame membership.[5] And in contrast to WordNet (Miller et al., 1990; Fellbaum, 1998; Kohl et al., to appear), which in its earlier incarnations, showed relationships only between words of the same part of speech categories (in its **synsets**), in FrameNet, the sets of words belong to a single frame can be of multiple parts of speech.

1.2. Precursors

The pre-history of the FrameNet project has much to do with Pisa and the late Antonio Zampolli. The earliest paper on frame semantics that anyone noticed (Fillmore 1977) was delivered at one of Antonio's Pisa summer schools, and the paper that first expressed the ambition to create a frame-based lexicon was begun when Fillmore and Sue Atkins were both participants in another Pisa summer school (Fillmore and Atkins 1992). Later Fillmore (as outside consultant), Atkins, and the Pisa group were all participants in the EU-sponsored DELIS project directed by Ulrich Heid of IMS/Stuttgart, 1993-1995.

* We are grateful to the U.S. National Science Foundation for funding the FrameNet project through two three-year grants. The first, IRI #9618838 covered the period 1997-2000 and the second, ITR/HCI #0086132 covered the period 2000-2003; we refer to these as FrameNet I and II. We have also received a supplemental grants from NSF for a short extension of FrameNet II and another from DARPA for the exploration of some future courses of development.

[1]Thus a word with two different meanings counts as two LUs. Typically, different LUs belong to different frames, though occasionally a noun can name both a situation type and a component of such a situation, as, e.g., the two uses of *replacement* Ex. (i) or the two uses of *possession* (ii).

(i) a. His replacement was a necessary step.
    b. His replacement was even worse than he was.

(ii) a. He lost his most precious possession last year.
     b. He lost possession of the farm last year.

[2]For example, in the Try_defendant frame, the FEs include the DEFENDANT, the JUDGE, the JURY, the GOVERNING_AUTHORITY, and the CHARGES. It is difficult to see how more than one or two of these can reasonably be equated with any of the standard roles

[3]Information that could support inference can of course be built into defining phrases that separately explain the precise participation of more abstractly identified–or perhaps only numbered–argument types, but FrameNet is hoping to be able to create inference mechanisms that build on patterns of frame names and frame-element names rather than depending on cross-linguistically varying formulations of a defining language.

[4]Such as PropBank (Kingsbury and Palmer, 2002; Palmer et al., to appear), VerbNet (Kipper et al., 2000), and Levin (1993).

[5]It is true, however, that FrameNet has done more annotation on verbs than on other word classes.

The DELIS project was an ambitious corpus-based effort to discover and document the meanings and valences of verbs of communication and sensory perception, in Danish, Dutch, English, French and Italian, and summarize the annotations in a typed-feature structure formalism. Most of the papers derived from the project were on verbs of perception in these languages (Atkins, 1994; Braasch, 1994; Emele and Heid, 1994; Fillmore, 1993; Fillmore, 1995; Heid, 1994; Heid and Krüger, 1996; Ostler, 1995).

2. The FrameNet I Period

FN started as a purely lexicographic project, concentrating at first on predicates or frame-bearing words[6]. Our objective was to provide for each predicate an account of its valence possibilities, this to be presented in terms of semantic functions, linked to the grammatical functions (GF) and phrase types (PT) in which they are realized. Later, but still within FrameNet I, we developed a method of annotating frame-filling words, mainly for collocation-collecting purposes, which consisted of identifying the **governor** of the target and the boundaries of the phrase within which the target word participated in the frame evoked by the governor. Our annotations were attached to sentences extracted from a very large corpus[7] and were chosen so as to exemplify each valence possibility for each target.

In the data structures used in FrameNet I, the values for FE, GF and PT were represented as attributes on an SGML element called "C" (for "constituent"), as shown in Ex. (1)

(1)

```
<C FE="Agent" GF= "ext" PT="NP">Couriers
</C> <C Target="yes"> carried</C>
<C FE="Theme" GF="Obj" PT="NP">drugs</C>
<C FE="Source" GF="Obl" PT="PP">from Rome</C>
<C FE="Goal" GF="Obl" PT="PP"> to Paris</C>
```

The boundaries were determined by an annotator, who selected a phrase with the cursor and assigned it a relevant FE name; once the phrases were identified, automatic processes–equipped with information about the part of speech categories of the constituent words and the phrase's position relative to the target–"guessed" the GF and the PT, guesses which the annotator could check and correct if necessary. The resulting annotation looks something like that shown in Table 1.

Further automatic processes would group the annotations and identify patterns in the configurations of GF, FE and PT information that characterized each target. These **valence descriptions**, then, were presented as generalizations over the permitted combinations of {FE, GF, PT} triples found for each word in the corpus. Annotators' sampling of sentences to include frequently occurring collocations gave some indication of relevant lexical collocations, which can be thought of as preferred "fillers" of FE constituents associated with given frame evokers.

	Couriers	CARRIED	drugs	from Rome	to Paris
FE	Agent		Theme	Source	Goal
GF	External		Object	Comp	Comp
PT	NP		NP	PP	PP

Table 1: Layered Annotation of a Sentence with *carried* as the Target

For various reasons we needed to develop a streamlined roster of GFs that could cover the grammatical relations that would show relevant facts about the LUs. Since for our purposes, about a strict distinction between arguments and adjuncts was not important, we used the GF name "Complement" to cover all non-nuclear relations with verbs[8]. The table also shows the GF name "External" where "Subject" might be expected; this choice is explained below.

2.1. Changes during FN1

2.2. Expansion of Annotated Sentence Types

The original expectation was that, given the very large size of the corpus we were working with, we would be able to find so many example sentences for each verb LU that we could limit ourselves to the syntactic patterns appearing in "basic" sentences—i.e. in relatively short, active voice, indicative mood sentences, where the target verb appeared in the main clause. Our rationale was that passivization, extraction, heavy NP shift, subject-auxiliary inversion, gapping, conjunction reduction, tough-movement, etc. are not lexically governed, but are general syntactic phenomena, and thus the concern of syntacticians, rather than lexicographers. Other sorts of syntactic variation, such as the dative alternation, **are** lexically specific (e.g. *tell him the answer/ *explain him the answer*) and thus proper objects for FrameNet annotation and analysis.

Furthermore, we had originally conceived of limiting ourselves to annotating those FEs which were in direct syntactic construction with the target word, including the subjects of finite VPs. We wanted to treat event nominals like the corresponding verbs so far as possible, so that their arguments would be instantiations of the same frame element, in the same grammatical relation to the frame-evoking word. It seemed odd to refer to the "subject" of a noun, so we chose the term "External", for "external argument", i.e. external to the phrase containing elements immediately governed by the target word[9], and use it for corresponding arguments of verbs, noun and adjectives. Thus in Ex. (2-a), *Watt* would have the GF "External" with respect to the target noun *stroll*, just as in Ex. (2-b) it has the GF "External" with respect to the verb *strolled*.

(2) a. Watt, out for a sabbath STROLL past the Golf-

[6] Elsewhere in this collection the preferred name is **FEE**, frame-evoking element and **autosemantic words**.

[7] For most of the project, and for all of FrameNet I, we worked almost exclusively with the British National Corpus (Burnard and Aston, 1998), provided to us courtesy of Oxford University Press, containing more than 100 million words, balanced across genres. We have recently added a roughly equal amount of text from the LDC North American Newswire, which we mainly use for vocabulary that is different in American English, such as the terminology of the criminal justice system.

[8] Additional GF names are called for in the case of noun targets.

[9] But not including "extracted" elements.

House . . .

b. Watt STROLLED past the Golf-House.

In fact, there is still a general preference for annotating structurally simple sentences[10]. Unfortunately, we have found that for many LUs there aren't enough examples with such simple structure to achieve our goals. Thus, it has been necessary to include in our samples (a) verbs embedded in control situations, where the controller NP satisfies one of the arguments of the target, assigning it the GF "External"; while it has a GF with respect to its own governor, its relation to the controlled target is simply external to the constituent headed by that verb; and (b) sentences with "extracted" constituents, where we give the extracted entity the label it would have in its non-extracted context (but see below).

In two important ways FrameNet annotations differ from the practices of other projects represented in this workshop. First, limiting the scope over which the annotator seeks FEs to the clause or phrase headed by the target LU is entirely appropriate in a lexicographic project: the basic semantic and syntactic combinatorial needs of a verb or adjectives can be shown in structurally simple sentences and with short phrases. There appear to be no cases in which the valence properties of particular nouns, verbs and adjectives include patterns that occur only in complex structures, and the user of a dictionary (whether native speaker or language learner) would usually prefer shorter, clearer examples. For any project devoted to the analysis of full sentences, or complete texts, by contrast, one must deal with every sentence that comes along, simple or not.

Secondly, there is no reason in a purely lexicographic approach to annotate many instances of a particular valence pattern, and that means that FrameNet does not attempt to provide information about relative frequency of valence patterns or lexical collocations for given LUs. A valence description can be complete with a small number of distinctive examples; if one pattern occurs 80% of the time and another only 5%, we need just enough examples of each to document the existence of the pattern. The same would not be true of a project primarily concerned with providing fodder for machine learning of valence patterns; in such a case, the researcher would want as many annotated instances as possible for each LU, ideally with example types having a frequency distribution representative of running text in the domain of the intended application.

A project annotating text for purposes of information extraction would have still different needs, including searching in preceding or following sentences for information about frame role participants. (We will discuss some recent FrameNet moves in this direction in Sect. 5.2.).

2.3. Null Instantiation

As we created valence descriptions for LUs, we became aware that such an account of the syntactic realizations of FEs for a given LU would not be complete without some description of the conditions that license the omission of an FE. Some of these arose from the expansion of annotated sentence types; clearly, for example, imperatives and passives both license the omission of agents; we created a dummy "word" in such sentences, and labeled it with the appropriate FE and called the grammatical function "Constructionally null instantiated" (CNI). More important from a lexicographic point of view, are those cases in which FEs are omitted with no clear licensing by a syntactic construction. We distinguished two such cases, those where the content of the omitted FE is recoverable from context (linguistic or extralinguistic), as in Ex. (3-a) and those where it is not, as in (3-b). In the former, the sentence is not felicitous unless the addressee can be supposed to know exactly what contest the speaker is referring to; this kind of "zero anaphora" we refer to as Definite Null Instantiation (DNI). In the latter, although the beverage is clearly some form of alcohol, the situation need not provide any further information as to what Jan indulged in for the sentence to be appropriate; we mark such cases as Indefinite Null Instantiation (INI).[11]

(3) a. We won!

b. Jan drank at the party.

3. The Transition from FrameNet I to II

3.1. Layered Annotation

Over the course of the FrameNet I period, we had accumulated a long list of things that we would like to do, but were unable to do because of the limitations of the data structures and the software itself. When we received funding for FrameNet II, we fundamentally rethought and rewrote the entire software system, preserving some of the look-and-feel of the FN1 system, but radically altering the basic data structures and the software itself.

The most crucial problem was that we were representing our annotations as SGML tags embedded in the actual text of the sentences. Aside from technical problems arising from interspersing the annotation with the text, there are a number of fundamental problems resulting from such a representation. The fact that the FE, GF, and PT were all attributes on a single SGML element (and that SGML elements of the same type cannot be nested) meant that they had to be coterminous; we were unable to properly represent situations where one FE was contained within another. For example, in Ex. (4-a), with *hit* as the target in the Cause_harm frame, we say that *him* is the FE VICTIM, and *on the arm* expresses the FE BODY_PART. Then in Ex. (4-b) we would like to say that *his arm* expresses BODY_PART, and *his* **also** expresses the VICTIM. Inserting tags to annotate this would result in malformed SGML.

(4) a. The bullet hit [him] [on the arm]

b. The bullet hit [[his] arm]

[10] But we do not choose the maximally "simple" sentences in which all arguments are represented by pronouns; we prefer sentences with lexical material whose semantics shows a good example of the frame. Thus, *The bullet hit him on the arm* is a better example for *hit.v* in the Cause_harm frame than *It hit him*

[11] INI includes more than what is traditionally referred to as "omitted objects": in the case of the INI element in a sentence like *That depends*, the omitted element would be expressed as a PP rather than a direct object, e.g., *on the situation*.

The solution we arrived at was to convert all the data from text with SGML markup to entries in a relational database, implemented in MySQL; the structure of the database was designed to mirror the conceptual structure of frame semantics, so far as practical. Thus there is now a table of frames, a table of frame elements, each associated with a frame, and a table of lemmas. Each record in the table of lexical units contains a pointer to a frame and a lemma, directly embodying the concept of the pairing of a form (or a set of forms) with a meaning (partially represented by the choice of frame, and supplemented by an ordinary dictionary-style definition contained in the record for the LU).[12] The use of standoff annotation also allows us to mark more than one target word per sentence, each with its own set of annotated FEs., etc. While this is rarely needed for purely lexicographic purposes, it is essential when we begin full annotation of running text, discussed in Section 5.2.

4. The FrameNet II period

4.1. Frame-to-Frame Relations

The theory of Frame Semantics has long asserted the existence of a rich collection of frames at various levels of abstractness, from the most abstract, in which the frame elements would simply be the thematic roles of Case Grammar[13], to quite specific frames, which would contain the majority of lexical units and (partially) encode the differences among them. The more specific frames were supposed to be subtypes of the more abstract meaning, among other things, that their roles (FEs) would be subtypes of the more abstract roles.

During FrameNet I, we had merely speculated on such relations. With the move to a relational database, we suddenly found ourselves able to represent such frame-to-frame relations, and hence, to make lots of decisions about just what sort of frame hierarchy we wanted, and how it could best be represented. Our general principle was that, since the project was intended to be basically linguistic and lexicographic, we would build whatever hierarchy we needed from the bottom up, creating first the frames needed to represent the commonalities and differences among lexical items. Where generalizations above that level needed to be made, we would create higher-level frames, but we did not seek to build a "complete" ontology. We did work with a consciousness of some sort of very high-level abstract frames like Event which would contain FEs like THEME, PLACE and TIME, and we did create these frames and started defining the links between them and more specific frames.

Questions immediately arose. One was whether a hierarchical relation meant that **all** the FEs of the parent frame were inherited by the child; we found instances in which what seemed to be a more specific, child frame never (or rarely) expressed an FE that was clearly part of the parent. A related question was whether we actually needed to create separate FEs for each child frame, or could rely on some sort of FE inheritance to supply many of the FEs needed in annotation of the lower-level frames. Once again, we resolved to proceed bottom-up; all the FEs needed for annotation would be created in the lower-level frames, and whatever inheritance relations they had to higher-level frames would be created explicitly. This meant more work for us, but it enabled us to continue to define FEs locally and quite specifically. To deal with the question of incomplete FE inheritance, we defined two types of frame-to-frame relations: **Inheritance** proper, in which all of the types and structure of the parent frame were inherited, and the **Using** relation, in which only a subset of the parent FEs were inherited. We also defined a third type of frame relation, **Subframe**, referring to subevents of complex events; both the complex event and its subevents are simply frames, although we sometimes use suggestive names like "scenario" for frames with subframe structure.[14]

4.2. Multiple inheritance

Frame semantics suggested monotonic, possibly multiple inheritance relations, and that was what we implemented. We found a number of cases in which different parent frames seemed to contribute different aspects of meaning to the child. Consider, for example, the frames related to employment, whose frame-to-frame relations are shown in Fig. 1. The highest level is a **background** frame Employment_scenario, which has FEs EMPLOYEE, EMPLOYER, and TASK, but does not itself contain any LUs. It does, however, have three subframes, Employment_start, Employment_continue, and Employment_end, representing the salient phases of the process.

On the next level down are Employee's_scenario and Employer's_scenario, both of which have a Using relation to Employment_scenario, including binding to its three FEs, and each of which has three subframes for the three phases of the process. The Employee's_scenario (on the left) is a **perspectival** frame, and its three subframes frame the situation from the EMPLOYEE's perspective: Get_a_job, Being_employed, and Quitting. The subframes of the Employer's_scenario are from the converse point of view: Hiring, Employing and Firing. Finally, two other perspectival frames, Intentionally_act and Intentionally_affect are inherited by the transitions at the beginning and end of the Employee's_scenario and Employer's_scenario respectively. Getting a job or quitting a job are intentional actions, but they are not framed in terms of their effects on others (although they usually do have effects on others such as employers, spouses, etc.). Hiring and Firing, on the other hand, are intended to affect the employee.

4.3. Reframing

From time to time, we have reanalyzed a general semantic area, adding frames, redrawing frame boundaries, and moving LUs from frame to frame. We do not undertake this lightly, but we are always trying to make our frame definitions and annotations more consistent with each other. For example, in the course of FrameNet II, we tried to consistently separate causative events, inchoative events, and

[12]For more detail on the database structure, see (Baker et al., 2003; Fillmore et al., 2001)

[13]For a recent statement of the situation, see Fillmore (2003).

[14]Of course, nearly any event can be considered to have some sort of subevents; the question here is where that breakdown has **linguistic** relevance.

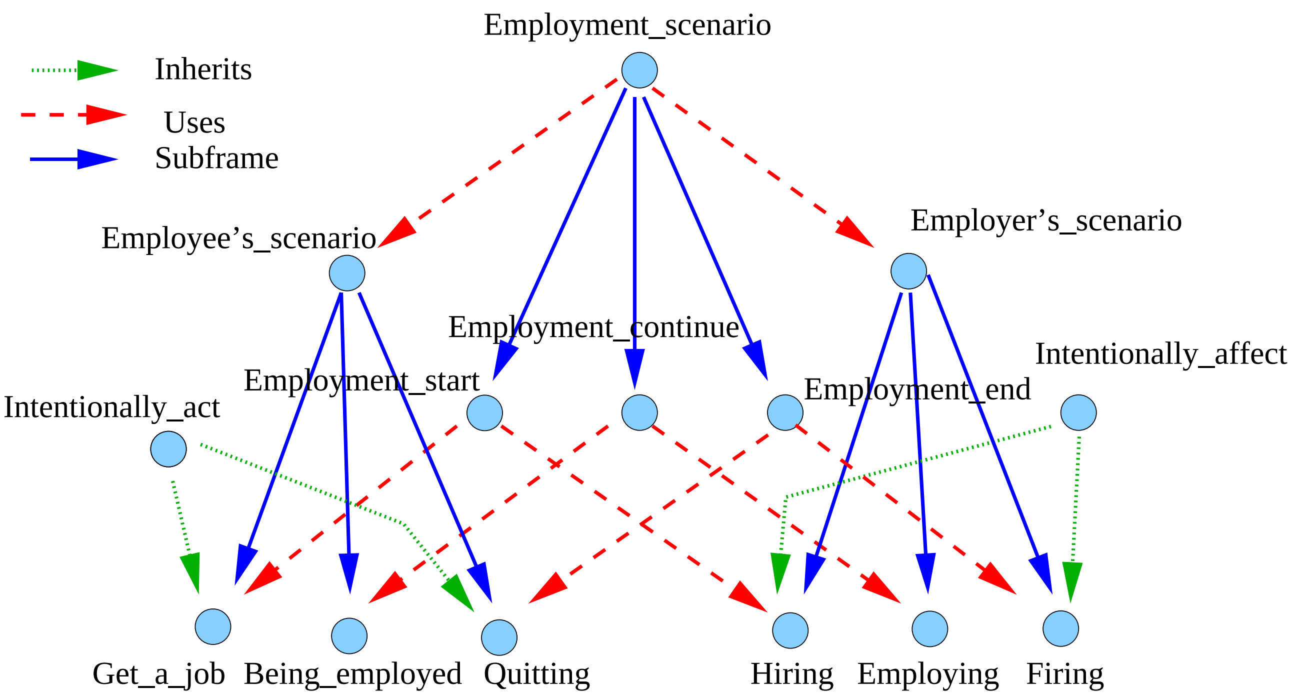


Figure 1: The Employment Scenario and its Subframes

statives into separate frames, even when they are referring to the same sort of situation. Often the same words are used in two or more of these frames. E.g., the verb *attach* can be either causative as in Ex. (5-a), or inchoative (5-b), while the adjective *attached* is stative (5-c), so these sentences would be in three separate frames.

(5) a. He attached the trailer to the back of the truck.

b. The trailer attaches to the back of the truck with a bolt.

c. The trailer is attached to the back of the truck with a bolt.

During FrameNet II, we have developed software to assist in this task of "reframing", including moving whole LUs or a group of annotated sentences from one frame to the other, while changing the FEs to point to the new frame, according to a mapping developed by a staff member. Without a relational database and the software tools, such a task would be extremely time-consuming.

5. Current Projects and Future Directions

5.1. Coreness and FE-to-FE relations within a Frame

It has been clear from the beginning that not all FEs are of equal importance in a frame. At one point, we were hopeful that we could assign semantic role ranks to each FE in a frame, and that their syntactic behavior would largely be predictable from their rank. Recently we have adopted what we hope will be a simpler system in which all FEs are classified as either **core**, **peripheral**, or **extra-thematic**. Core FEs are those which are essential to the definition of the frame itself, such as the SPEAKER and ADDRESSEE in the Statement frame or the COOK and the PRODUCED_FOOD in the Cooking_creation frame. Peripheral FEs are those which are inherently part of the situation, but not central to the definition of the frame. For example, it is ontologically necessary for all events to take place at some time and in some place, but the PLACE and TIME FEs are rarely central to event frames, such as Statement or Cooking_creation. Finally, there are participants in some events that are not really part of the frame, but are introduced by some other construction, such as the RECIPIENT in Ex. (6). Though they are not strictly part of the frame, we sometimes include them in the list of FEs so that they will be available for annotation.

(6) [$_{\text{COOK}}$ Loretta] COOKED [$_{\text{PRODUCED\_FOOD}}$ couscous] [$_{\text{RECIPIENT}}$ for her friend].

We have been doing a lot of checking of our annotation data, trying to ensure that all the core FEs are either marked in the sentence or marked as null instantiated. But this exercise has made us more conscious of another level of complexity; we now need to take into account relations between the FEs within a frame.

(7) a. [$_{\text{INTERLOCUTORS}}$ They] 'd come back CHATTING [$_{\text{MANNER}}$ merrily] . [$_{\text{TOPIC}}$ INI]

b. [$_{\text{INTERLOCUTOR\_1}}$ Malcolm Anderson] was in the room , CHATTING [$_{\text{INTERLOCUTOR\_2}}$ with the police photographer] . . .

c. [$_{\text{INTERLOCUTOR\_1}}$ Malcolm Anderson] was in the room , CHATTING. [$_{\text{INTERLOCUTOR\_2}}$ INI]

The most clearcut case has to do with FEs such as INTERLOCUTOR_1, INTERLOCUTOR_2, and INTERLOCUTORS in the Chatting frame, exemplified in Ex. (7). In sentences such as (7-a), the two sides of the conversation are expressed by one NP; in those like (7-b), they are separate constituents, with INTERLOCUTOR_2 usually expressed as a PP_{WITH} or PP_{TO}. Clearly one pattern or the other must occur; (7-c) must be considered a case of the latter pattern, with INTERLOCUTOR_2 omitted (INI). Thus we could describe the relations among these three FEs by saying that INTERLOCUTOR_1 **requires** INTERLOCUTOR_2 (and vice-versa) and that they both **exclude** INTERLOCUTORS (and vice-versa). Patterns like this are familiar to most linguists from morphology and phonology, and strongly suggest that we should make some sort of generalization about a "proto-FE" that can be expressed either jointly or separately. It is this "proto-FE" that is really core, rather than any of its realizations.

Another sort of relation that we are beginning to recognize is a little fuzzier. Consider the most frequent motion frame, Self-motion, which has FEs SOURCE, PATH and GOAL. Clearly any motion proceeds from some point, along some path to some other point, and these are part of the definition of motion itself, so we would like to consider all three as core FEs. But we find that relatively few sentences express all three; it seems to be a fact about the structuring of information in English that a single phrase headed by a noun or verb in this frame rarely overtly expresses more than one or two of these FEs. In order to facilitate the process of checking annotation, we have created a relation called '**coreset**, to indicate that these core FEs form a set, and that a sentence in which only one of them is marked should not be considered to be missing annotation. This also suggests some sort of underlying "proto-FE", but the boundaries here are much fuzzier, as another FE, AREA also seems to be part of the set. This is an active area of investigation, but we have included the coreset, requires, and excludes relations in our released data for others' consideration and suggestions.

5.2. Full Text Annotation

We have long had in mind that ultimately, the strengths of the frame semantic approach would be most apparent if entire stretches of text were fully annotated in the FrameNet style. This means that every frame evoking element would be marked as a target, and that most (or all) of the rest of the text would be labeled as frame elements; an operation which would compose the meanings of these labelings would produce at least a good start on a deep representation of the meaning of the text (Fillmore and Baker, 2001).

Accordingly, early in FN2 we added several tables to the database that represent the corpus, document, and paragraph from which a sentence comes. This table was initially empty, as all the sentences came from the BNC and the document, paragraph and sentence number information was not available. But when we added the North American newswire corpora from the LDC, we processed those files so as to embed the information on document structure. (This was easier as the news articles all had XML markers for document start and end and paragraph divisions, unlike our (early) version of the BNC.)

Now we have begun collaboration with the PropBank project at U Penn to annotate with FrameNet labels a portion of the Wall Street Journal texts that they have marked in PropBank. The objective is to annotate enough text in both styles so that another team, led by Dan Jurafsky, can work on learning how to translate between the two styles of annotation and use the combined annotated corpus as a resource for training semantic parsers.[15]

The experience of trying to annotate **all** the frame-evoking words in running text has been enlightening in several ways. First, contrary to our usual practice of selecting relatively short, clear sentences, we are having to annotate sentences that are longer and have more complex structure; although we have not run into any unsurmountable problems, we are having to consider some syntactic structures that we would otherwise avoid.

Second, we are annotating some rather common LUs by virtue of their appearance in the text that we happen not to have covered before. In the long run, this will no doubt improve FrameNet's coverage of the general vocabulary, but, since we are keeping to our principle of fully defining the frames and FEs as we create them, we are having to make a major effort to define new frames. We estimate that we need roughly 250 new frames to cover the first 125 sentences of text, which amounts to 50% increase in our total frame inventory, and we are devising the technical means to speed up this process.

And we are finally doing what we have long envisioned with regard to higher-level rhetorical relations, such as sentence-initial *but* and *now*, and pronouns which refer to the state of affairs described in the preceding sentence. In these cases, we are forced to mark FEs across phrase, clause, and even sentence boundaries; we also need to annotate parts of speech such as conjunctions, pronouns and prepositions which we have not dealt with before.

Most of the changes of annotation practice do not involve any changes to the software or basic workflow. In order to move ahead with the business of creating new frames, we have omitted the usual practice of annotating all the LUs in each frame before moving on to the next, although we have found in the past that it is best to do so, if we want to be sure that we have the definition of the frame itself and the FEs "right". We have even omitted to annotate the usual number of examples for each LU, necessary to produce the full lexical entry, annotating only the sentences in the PropBank texts where they appear. This means that we will have to revisit these LUs and go through the standard process of extracting a few hundred examples and annotating 20 or so before we can consider them finished. We will also need a way of marking the often complex, confusing sentences from the newswire so that they will be distinguished from the shorter, clearer ones suitable for a human-readable dictionary. They may, however, be valuable for machine learning, as they exemplify more difficult, boundary cases for classification.

5.3. Multiword Expressions and Representing Constructions

We have sought to avoid partisanship in our choice of syntactic theories in our annotation, but it is impossible to

[15]This collaboration is funded by an NSF ITR grant, on which Jurafsky is the PI.

avoid such questions completely. Not too surprisingly, if forced to choose, we would generally favor a Construction Grammar approach (Kay, 2004; Kay, 2002; Kay, 1998)

According to Construction Grammar, **all** linguistic forms are constructions; some of them, such as the Subject-Predicate construction are purely syntactic, while others are purely lexical, such as the the word *decide*. A finite clause whose main verb is *decide* represents a unification of the lexical construction *decide* with the Subject-Predicate construction. Given that lexical units and syntactic patterns are varieties of the same sort of object, it is not surprising that there should be intermediate varieties that are partially lexical and partially syntactic. Many MWEs are mainly lexical, with just a minimum of syntax attached to them, such as English verb-particle lemmas that either license or prohibit intervening material between their two parts.[16] Somewhere in the middle of the continuum, we have constructions such as *What's X doing Y*, which combines the words *what BE* and *doing* with a unique set of implicatures and pragmatic constraints. (Kay and Fillmore, 1999)

We have long discussed how FrameNet might represent constructions which cannot simply be regarded as sequences of lexemes. As we attempt to fully annotate running text, we increasingly feel the need for such a representation. For example, many ways of expressing people's ages are used in news stories: *58-year-old Horace Philpot, Horace Philpot is 58, Horace Philpot is 58 years old*, and even *Horace Philpot, 58, was found sleeping…* Information retrieval systems need to be able to recognize all of these as different ways of expressing the same information. We would like to treat them all as different constructions, and connect them all with the same frame.

Part of the difficulty is that Construction Grammar has thus far not been complete and precise enough to be implemented as a parser.[17] Thanks to recent work by Anette Frank and her colleagues (in this volume), we see a possibility that the F-structures of LFG might be used for this purpose, but this is still contingent on a number of factors, not least finding a good way to represent a set of conditions on an F-structure in our database. Some in our group have also suggested that HPSG might be conceptually closer to Construction Grammar, but there are no immediate plans for representing constructions in an HPSG formalism. Since we would want to annotate enough examples of each construction to produce a description of it comparable to our current lexical entries, we would need to search for a large number of examples from the corpora. The availability of reliable, broad-coverage parsers for many languages for both LFG and HPSG is another argument in favor of at least using one of them for such searches, if not directly for our internal representation.

[16]Of course, this is an oversimplification. Some typically allow a theme NP between the verb and the particle, but some manner adverbials can also intervene (*clean it almost completely out*). The idea of lexical items with some sort of grammar richer than subcategorization attached to them is, of course, common to all lexicalized theories of grammar, including HPSG, LFG, XTAG, etc.

[17]Although such an effort is underway in the Neural Theory of Language group at ICSI, `http://www.icsi.berkeley.edu/NTL`.

6. Conclusions

We hope that we have shown in this paper that annotation policies are likely to evolve during the course of a project both for "internal" reasons, having to do with the increasing understanding of the ramifications of any policy decision, however straightforward it may seem, and "external" reasons, relating to the varieties of uses to which a collection of annotated text can be put and the varieties of text that are to be annotated.

In recent years, there has been increasing recognition in the NLP community of the importance of lexical resources in general and, in particular, of the value of cross-training to develop lexical resources from annotated corpora and semantic parsers based on such lexical resources that can then be used to annotate more text. Different projects will continue to have their particular methods, starting point, and goals, but we would like to believe that the time has come for increased cooperation among among builders of grammars, lexicons, research corpora and ontologies. We hope that FrameNet's past lexicographic accomplishments and ongoing work on full-text analysis will make it possible for the project to contribute to increasing collaboration in this field, across annotation projects and across languages, with a common goal of building better lexical resources, and thus, better systems for natural language understanding and generation.

7. References

Atkins, B. T. S., 1994. Analysing the verbs of seeing: a frame semantics approach to corpus lexicography. In Susanne Gahl, Christopher Johnson, and Andrew Dolbey (eds.), *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*. Berkeley Linguistics Society.

Baker, Collin F., Charles J. Fillmore, and Beau Cronin, 2003. The structure of the FrameNet database. *International Journal of Lexicography*, 16(3).

Braasch, Anna, 1994. There's no accounting for taste - except in dictionaries. In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, and P. Vossen (eds.), *Proceedings*. Amsterdam: EURALEX.

Burnard, Lou and Guy Aston, 1998. *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh University Press. `http://www.natcorp.ox.ac.uk/`.

Emele, Martin and Ulrich Heid, 1994. DELIS: Tools for corpus-based lexicon building. In *Proceedings of Konvens-94*. Heidelberg: Springer.

Fellbaum, Christiane (ed.), 1998. *WordNet:An Electronic Lexical Database*. The MIT Press.

Fillmore, Charles J., 1993. Frame semantics and perception verbs. Contribution to the Dagstuhl Seminar, forhtcoming in Hans Kamp, James Pustejovsky (eds.): Universals in the Lexicon: At the Intersection of Lexical Semantic Theories.

Fillmore, Charles J., 1995. The hard road from verbs to nouns. In *In Honor of William S.-Y. Wang*. Pyramid Press.

Fillmore, Charles J., 2003. Valency and semantic roles: the concept of deep structure case. In Vilmos Agel, Ludwig M. Eichinger, Hans-WernerHellwig, Hans Jürgen Heringer, and Henning Lobin (eds.), *Dependency and Valency: An International Handbook of Contemporary Research*, volume 1. Walter de Gruyter, pages 457–75.

Fillmore, Charles J. and Collin F. Baker, 2001. Frame semantics for text understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop*. Pittsburgh: NAACL.

Fillmore, Charles J., Charles Wooters, and Collin F. Baker, 2001. Building a large lexical databank which provides deep semantics. In Benjamin Tsou and Olivia Kwong (eds.), *Proceedings of the 15th Pacific Asia Conference on Language, Information and Computation*. Hong Kong.

Heid, Ulrich, 1994. Relating lexicon and corpus: Computational support for corpus-based lexicon building in DELIS. In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, and P. Vossen (eds.), *EURALEX 1994 Proceedings*. Amsterdam: Vrije Universiteit.

Heid, Ulrich and Katja Krüger, 1996. A multilingual lexicon based on Frame Semantics. In Lynn Cahill and Roger Evans (eds.), *Proceedings of AISB96 Workshop on Multilinguality in the Lexicon*. University of Sussex, UK.

Kay, Paul, 1998. An informal sketch of a formal architecture for Construction Grammar. In *proceedings of the Conference on Formal Grammar, HPSG and Categorial Grammar*. Saarbrücken. To appear.

Kay, Paul, 2002. English subjectless tagged sentences. *Language*, 78:453–81.

Kay, Paul, 2004. Pragmatic aspects of grammatical constructions. In Laurence Horn and Gregory Ward (eds.), *In Handbook of Pragmatics*. Blackwell.

Kay, Paul and Charles J. Fillmore, 1999. Grammatical constructions and linguistic generalizations: The *What's X Doing Y?* construction. *Language*, 75:1–33.

Kingsbury, Paul and Martha Palmer, 2002. From TreeBank to PropBank. In Manuel González Rodríguez and Carmen Paz Suárez Araujo (eds.), *Third International Conference on Language Resources and Evaluation, LREC-02, Las Palmas, Canary Islands, Spain, May 28- June 3, 2002*. Gran Canaria: European Language Resources Association.

Kipper, Karin, Hoa Trang Dang, and Martha Palmer, 2000. Class-based construction of a verb lexicon. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*. Austin, Texas: AAAI.

Kohl, Karen T., Douglas James, Robert C Berwick, Franklin Cho, Zeeshan Khan, and Naoyuki Nomura, to appear. Representing verb alternations in WordNet. In *WordNet: An electronic lexical database and some of its applications*. MIT Press.

Levin, Beth, 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Miller, George A., R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–44. Guest Editor: George A. Miller.

Ostler, Nicholas, 1995. Perception vocabulary in five languages–towards an analysis using frame elements. In S. Johanssen and S. Oksefjell (eds.), *Machine Translation and the Lexicon. Third International EAMT Workshop*. Berlin: Springer Verlag.

Palmer, Martha, Dan Gildea, and Paul Kingsbury, to appear. The proposition bank: An annotated corpus of semantic roles. Submitted to Computational Linguistics.

The Japanese FrameNet Project: An Introduction

Kyoko Hirose Ohara, Seiko Fujii*, Toshio Ohori*, Ryoko Suzuki,
Hiroaki Saito, Shun Ishizaki*****

Keio University
4-1-1 Hiyoshi, Kohoku-Ku, Yokohama City, 223-8521, Japan
ohara@hc.cc.keio.ac.jp

Abstract

This paper presents an overview of the Japanese FrameNet (JFN) research project, which began in July 2002. The goal of JFN is to create a corpus-based lexicon of Japanese described in terms of frame semantics. An important question being asked by JFN is whether Japanese words can be described in FrameNet style, i.e., along the same lines as English words. This point is illustrated in this paper with an example of preliminary analysis of Japanese motion verbs. The Japanese FrameNet can be described as a lexicographic project with an eye to finding out similarities and differences between Japanese and English pertaining to their lexical and grammatical structures.

1. Introduction

This paper gives an overview of the Japanese FrameNet (hereafter JFN) research project, which was launched in July 2002. JFN tries to create a corpus-based lexicon of Japanese described in terms of frame semantics (Fillmore, 1982).

JFN is headquartered on Hiyoshi Campus of Keio University and includes researchers from Keio University and University of Tokyo. So far, a corpus has been chosen and is being expanded and a tool for extracting data from the corpus has been implemented. A pilot study is being undertaken to analyze motion and communication verbs in Japanese.

The rest of the paper is structured as follows. Section 2 describes the goals of JFN. Section 3 illustrates the corpus and computational tools used in the project. Section 4 gives an example of preliminary analysis of Japanese motion verbs.

2. Project Goals

The ultimate goal of JFN is to produce a FrameNet-style database of Japanese words (cf. Fillmore et al., 2003). The resulting database will thus contain valence descriptions of Japanese words and a collection of annotated corpus attestations. In producing this database we will explore whether Japanese words can be described along the same lines as English words, employing the same frame semantic approach.

In the first phase of the project, which lasts until March 2005, we are trying to build a prototype of such a lexicon, focusing on analyzing and annotating Japanese motion and communication verbs. In the second phase, however, we intend to analyze words in other semantic domains as well.

3. The Corpus and the Tools for Analyzing Japanese

The JFN corpus currently contains 8 million sentences taken from the Mainichi Newspaper (CD-Mainichi Newspaper) and texts taken from novels and essays.

The search tool has been developed in JFN (cf. Ohara et al., 2003). The tool searches for both the root form and conjugated forms of a verb, adjective, or adjectival noun at the same time. Another key feature of the JFN search tool is the fact that it can be used with a dependency structure analyzer called CaboCha. CaboCha was developed at Nara Institute of Science and Technology and it performs morphological analysis as well as syntactic parsing of any Japanese sentence. Although CaboCha sometimes parses colloquial sentences incorrectly, using our search tool together with CaboCha enables us to add any text to our corpus.

Currently there are three display modes in the JFN search tool: the parse tree mode; the morphological analysis mode; and context display mode. Figure 1 shows a snapshot of the parse tree mode:

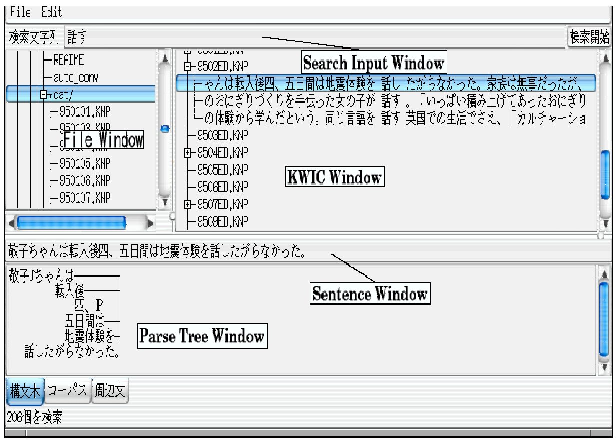


FIGURE 1. A Screenshot of the parse tree mode of the JFN search tool

The entire screen consists of five windows: the search input window to input a keyword to be searched; the file window to specify file(s) in which a keyword is searched; the KWIC window displaying all the sentences containing the keyword and allowing the user to highlight any sentence by clicking on it; the sentence window showing

---

* University of Tokyo, 3-8-1 Komaba, Meguro-Ku, Tokyo 153-0041, Japan.
** Keio University, 3-14-1 Hiyoshi, Kohoku-Ku, Yokohama City, Kanagawa Pref. 223-8522, Japan.
*** Keio University, 5322 Endo, Fujisawa City, Kanagawa Pref. 252-8520, Japan.

the highlighted sentence; and the parse tree window, which displays the tree of a highlighted sentence.

The JFN search tool is written in Ruby script language, and runs on Linux and Solaris operating systems as well as on various Windows platforms. We plan to make it publicly available.

4. Applicability of English-based Frames to Japanese

An important question being asked by JFN is whether Japanese words can be described in FrameNet style, i.e., along the same line with English words. Although our analysis and annotation of corpus sentences are based on relevant frames and frame elements established for English in the English FrameNet, we constantly ask whether it is necessary to establish frames and frame elements separately for Japanese. What follows is a preliminary analysis of Japanese motion verbs. As we have done elsewhere for Japanese communication verbs (ibid.), the analysis below attempts to deal with recognized differences in senses among Japanese motion verbs related in meaning, by refining frame elements already established in the English FrameNet.

Japanese has been characterized as a "path-type verb-framed language," since in order to encode a Path of motion, Japanese employs motion verbs unlike English, which employs prepositions and verb particles for the same purpose (cf. Talmy, 1985; 1991; 2000; Matsumoto, 1997; Kageyama, 2001). Japanese is thus abundant in motion verbs which at the same time describe various paths. *Wataru* 'go across' and *koeru* 'go beyond, go over,' are examples of such motion+path verbs. Both of the verbs can be analyzed as evoking the Path_Shape frame, since the two verbs describe motion in terms of the shape of the Path traversed by the Theme that moves. Thus, *wataru* 'go across' can be used with an accusative-marked direct object NP denoting a Path as in (1a) and (1b) (cf. Kunihiro et al., 1982). In (1a) *kawa* 'river' denotes an area that lies between two points in space, while in (1b) *hasi* 'bridge' refers to a medium or a passage that is constructed between the two points:

(1) a. *nanmin ga* <u>*kawa o*</u>
refugees NOM river ACC
watatta
went.across
'The refugees went across (crossed, traversed) the river.'

b. *nanmin ga* <u>*hasi o*</u>
refugees NOM bridge ACC
watatta
went.across
'The refugees crossed the bridge.'

On the other hand, although the verb *koeru* 'go beyond' takes an accusative-marked direct object NP such as *kawa* 'river' in (2a) just like *wataru* 'go across' does, *hasi* 'bridge' typically cannot be its direct object as shown in (2b):

(2) a. *nanmin ga* <u>*kawa o*</u>
refugees NOM river ACC
koeta
went.beyond
'The refugees went beyond (passed) the river.'

b. **nanmin ga* <u>*hasi o*</u>
refugees NOM bridge ACC
koeta
went.beyond
(Intended meaning) 'The refugees passed the bridge.'

Furthermore, *koosaten* 'intersection' can be the direct object of *wataru* as in (3a), but not of *koeru* as shown in (3b). However, when (3b) is used to depict a situation in which the child not only crossed the intersection but also went beyond it, then the sentence becomes acceptable:

(3) a. *kodomo ga zitensya de*
child NOM bike INSTR
<u>*koosaten o*</u> ***watatta***
intersection ACC went.across
'The child crossed the intersection by bike.'

b. **kodomo ga zitensya de*
child NOM bike INSTR
<u>*koosaten o*</u> ***koeta***
intersection ACC went.beyond
(Unacceptable with the reading) 'The child passed the intersection by bike (and stopped there).'
(Acceptable with the reading) 'The child went by bike past the intersection.'

It thus seems necessary to identify sub-categories of the frame element Path such as Route and Boundary, in order to describe the different kinds of Path that the two verbs above and others take. That is, *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. Therefore, the annotations of (1a), (2a), and their equivalents in English would be as follows:

(4) a. Japanese

(cf. 1a)

Theme	Path.Route
nanmin ga	*kawa o*
refugees NOM	river ACC
NP	NP

watatta
went.across

b. English

Theme		Path
The refugees	**went**	across the river.
NP		PP

(5) a. Japanese

(cf. 2a)

Theme	Path.Boundary
nanmin ga	*kawa o*
refugees NOM	river ACC
NP	NP

koeta
went.beyond

b. English

Theme		Path
The refugees	**went**	**beyond** the river.
NP		PP

With these contrastive analyses of Japanese and English, we hope to eventually build a bilingual lexicon, to be used by Japanese learners of English as well as by machine translation (cf. Boas, 2002). Such a lexicon will especially be effective as an encoding dictionary for Japanese learners of English.

5. Conclusion

This paper has outlined the goal, computational environments, and a preliminary analysis of JFN. In conclusion, our current effort can be described as a lexicographic project with an eye to finding out similarities and differences between Japanese and English pertaining to their lexical and grammatical structures.

Aknowledgements

We wish to thank Takuya Kosugi, Hideaki Shirakawa, and Kiyoko Uchiyama, for their valuable discussions and comments regarding this paper. The Japanese FrameNet is funded by a grant from Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Boas, H.C. (2002). Bilingual FrameNet Dictionaries for Machine Translation. In Proceedings of the Third International Conference on Language Resources and Evaluation (LREC2002). Vol. IV. (pp. 1364-1371).

CaboCha
http://cl.aist-nara.ac.jp/~taku-ku/software/cobocha

CD-Mainichi Newspaper 1995-2002.

Fillmore, C.J. (1982). Frame semantics. In Linguistics in the Morning Calm. (pp. 111-137). Seoul: Hanshin Publishing Co.

Fillmore, C.J., Johnson, C.R., and Petruck, M.R.L. (2003). Background to Framenet, International Journal of Lexicography, Vol. 16.3: 235-250.

Kageyama, T. ed. (2001). Nichi Ei Taishoo Dooshi no Imi to Koobun [Contrastive Analysis of Japanese and English: Semantics of Verbs and Constructions]. Tokyo: Taishukan Shoten.

Kunihiro, T., Shibata, T., Nagashima, Y., Yamada, S., and Asano, Y. (1982). Kotoba no Imi. [Meanings of Words]. Vol.3. Tokyo: Heibonsha.

Matsumoto, Y. (1997). Kuukan Idoo no Gengo Hyoogen to Sono Kakuchoo [Linguistic Expressions of Motion in Space and Their Extensions]. In Tanaka, S. and Matsumoto, Y. Nichi Eigo Hikaku Sensho. Roku: Kuukan to Idoo no Hyoogen [A Series on Contrastive Analysis of Japanese and English, 6: Expressions of Space and Motion]. (pp. 125-230). Tokyo: Kenkyusha.

Ohara, K.H., Fujii, S., Saito, H., Ishizaki, S., Ohori, T., and Suzuki, R. (2003). The Japanese FrameNet Project: A Preliminary Report. In Proceedings of Pacific Association for Computational Linguistics (PACLING'03) (pp. 249-254).

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In Shopen, T. (Ed.), Language typology and syntactic description, Vol. 3, (pp. 57-149). Cambridge: Cambridge University Press.

Talmy, L. (1991). Path to realization: A typology of event conflation. Berkeley Linguistics Society 17 (pp. 480-519).

Talmy, L. (2000). Toward Cognitive Semantics. Cambridge, MIT Press.

Spanish FrameNet and FrameSQL[1]

Carlos Subirats and Hiroaki Sato

Autonomous University of Barcelona (Spain) and Senshu University (Japan)
Carlos.Subirats@uab.es, hiroaki@isc.senshu-u.ac.jp

Abstract

In this paper, we introduce the Spanish FrameNet Project which is creating an online lexical resource for Spanish, based on Frame Semantics and supported by corpus evidence. Spanish and English lexical units in the Emotion and Motion frames are compared and differences between the lexicalisation patters and constructions in both languages are described. The paper also introduces FrameSQL which is a web-based application to search and view the Spanish FrameNet data on the web browser as well as the Berkeley's FrameNet data. The application handles both of the data seamlessly, showing Spanish and English lexical units belonging to the same frame on the same window. It makes it easier to compare semantic structures of the two lexicons.

Background to Spanish FrameNet

The Spanish FrameNet Project (http://gemini.uab.es/SFN) is creating an online lexical resource for Spanish, based on Frame Semantics (Fillmore, 1982, 1985) and supported by corpus evidence. The "starter lexicon" will be available to the public by January 2006, and will contain at least 1000 lexical items –predicative verbs, nouns and adjectives– representative of a wide range of semantic domains. The aim is to document the range of semantic and syntactic combinatory possibilities (valences) of predicates in specific senses, through:

- human approved and automatic annotated example sentences and
- automatic capture and organization of the annotation results.

The Spanish FrameNet (SFN) database will be in a platform-independent format, and it will be able to be displayed and queried via the web and other interfaces. The SFN database will act both as a dictionary and a thesaurus. The dictionary features include:

- definitions, tables showing how frame elements are syntactically expressed in sentences containing each word,
- annotated examples from the corpus: human approved and automatically annotated, and an alphabetical index.

Like a thesaurus, words are linked to the semantic frames in which they participate, and frames, in turn, are linked to wordlists and to related frames. The basic assumption of Frame Semantics is that each word evokes a particular frame and possibly profiles some element or aspect of that frame. Semantic frames are schematic representations of situations involving various participants, props, and other conceptual roles, each of which is called a frame element (FE). The semantic arguments of a predicating word correspond to the frame elements of the frame (or frames) associated to that word. A frame semantic description of a lexical unit identifies the frames which underlie a given meaning and specifies the ways in which frame elements are realized in structures headed by the word (See Johnson et al., 2002). For example, consider the Judgement_communication frame which deals with communicating a positive or negative judgment of an Evaluee to an Addressee, e.g. *alabar* (praise) or *criticar* (criticize). This frame minimally includes the FEs Communicator, Evaluee and Addressee. Sentence (1) below is a canonical example of a verb in the Judgement_communication frame.

(1) Max **elogió** a Eva ante los directivos de la empresa.
Max praised to Eva in-front-of the directors of the company
Max praised Eva before the company directors.

Here, *Max* fills the role of Communicator; *Eva* is the Evaluee; and *los directivos de la empresa* is the Addressee. Note that the Addressee is expressed in (1) above, but it may not be realized in other sentences, as shown in sentence (2) below, where *la actuación de la empresa* is the Evaluee, but the Addressee is not instantiated.

(2) Sara **denunció** la actuación de la empresa.
Sara reported the performance of the company
Sara reported the company performance.

Each frame element tag is part of a set of three tags, consisting of the frame element, like Communicator, Evaluee, etc., the grammatical function and the phrase type of the annotated constituent. The mappings between the semantic and syntactic information given in the triples of annotation for the set of sentence types in which a

[1] Spanish FrameNet is a research project which is sponsored by the Department of Science and Technology of Spain (Grant No. TIC2002-01338), and is developed at the Autonomous University of Barcelona in cooperation with the FrameNet Project, administered at the International Computer Science Institute in Berkeley, CA. The authors would like to thank their colleagues, Rocío Donés, Jordi Duran, Mercedes García, Lídia Moya, and Marc Ortega for their assistance.

given lexical unit occurs constitutes its valence. The goal of Spanish FrameNet is to annotate corpus citations and to discover the valence patterns for a large number of words showing how those valence patterns are instantiated in actual sentences. Each Spanish FrameNet entry will provide links to other lexical resources, including Spanish EuroWordNet synsets and syntactic subcategorization frames. The project's deliverables will consist of the SFN database itself: lexical entries for individual word senses, frame descriptions, and annotated subcorpora.

The SFN project is based on the evidence offered by a 330 million-word corpus which includes both New World (60%) and European Spanish (40%). The corpus is POS tagged and lemmatized with a tool that uses an electronic dictionary of Spanish of 600,000 forms, both single (92%), and multi-word lexical units (8%), basically multi-word nouns (85%), like *bomba atómica* (atomic bomb), *carga de profundidad* (depth charge), and multi-word adverbs (9%) like *a ciegas* (unknowingly), *por ahora* (so far), etc. Multi-word verbs like *tener en cuenta* (to take into account) and lexicalized prepositional phrases with support verbs like *estar de moda* (to be in fashion) are tagged and lemmatized with transducers. SFN uses the Corpus Workbench software from the Institut für Maschinelle Sprachverarbeitung of the University of Stuttgart[2] for searching the corpus. The semantic and syntactic annotation is carried out by using the FNDesktop, the system developed by the Berkeley FrameNet Project. The input of the FNDesktop are sentences that have been automatically extracted from the corpus, and then POS tagged and lemmatized (Subirats and Ortega, 2000). The extraction of subcorpora where predicates appear in all their relevant constructions provide annotators with examples of each possible syntactic configuration in which a given lexical item can occur. Annotators then select sentences for annotation that best illustrate the ways in which frame elements are realizad syntactically. Figure 1 shows an actual sentence from the database annotated with the FNDesktop.

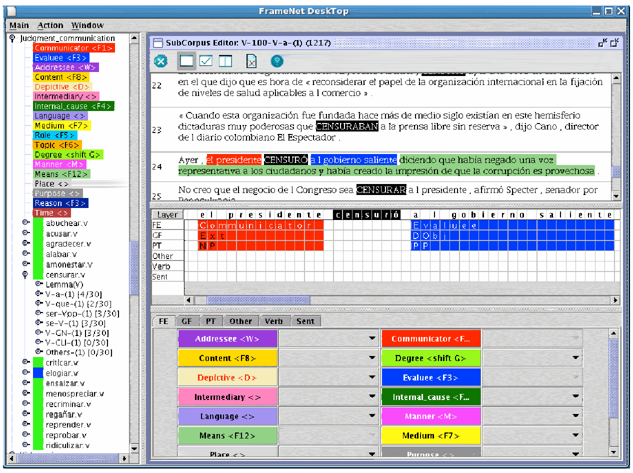


Figure 1: Annotation of a sentence in the Judgement_communication frame

[2] http://www.ims.uni-stuttgart.de/

Different lexicalization patterns in English and Spanish emotion predicates

SFN is studying areas of the lexicon that parallel existing English FrameNet desriptions. Most of the frames defined so far are valid cross-linguistically, because frames are meant to charachterize conceptual structures at a basic level of description.

Valence descriptions provided by SFN and FN can be used to study different lexicalization patterns in English and Spanish. Thus, for instance, *sorprender* (to surprise) in (3) is a Cause_emotion verb characterizing an event, in which an agent seeks to cause an emotion on an Experiencer.

(3) Juan **sorprendió** a María al contarle la verdad.
Juan surprised to Maria on explaining-her the truth
Juan surprised María by telling her the truth.

The reflexive verb *sorprenderse* (to get surprised) in (4) and the adjectival past participle *sorprendido* (surprised) with the support verb *estar* (to be) in (5) are two Experiencer_objet predicates in which the Experiencer is the subject and the Stimulus is the object.

(4) María se **sorprendió** de que Juan cantase.
María REFL surprised of that Juan Sang
María got surprised when Juan sang.

(5) María está **sorprendida** de que Juan cante.
María is surprised of that Juan sang
María is surprised that Juan sang

Both *sorprenderse* in (4) and *sorprendido* in (5) express parts of the complex event characterized by *sorprender* in (3): *sorprenderse* is an inchoative verb which characterizes the beginning of an event and *sorprendido* expresses the ongoing state which occurs after the above mentioned beginning. Therefore, *sorprenderse* and *sorprendido* are simpler parts of the complex event *sorprender* (cf. Subirats and Petruck, 2003).

This analysis allows us to study the lexicalization pattern differences among English and Spanish emotion predicates. Both Spanish and English lexicalize the causative meaning with two verbs, namely, *surprise* and *sorprender*. On the contrary, there exists a difference in the lexicalization of the inchoative meaning: Spanish uses the reflexive verb *sorprenderse*, whereas English uses the construction made by *get* and the adjectival past participle *surprised*. As a result, English only has the LU *surprised* in the Experiencer_subject frame and Spanish has two LUs, that is, the reflexive verb *sorprenderse*, and the adjective *sorprendido* (cf. Figure 2). These differences can be verified thanks to FrameSQL, an application which allows to compare predicates or predicate-related constructions in the frames which share the same name, and therefore the same characteristics in English and Spanish.

	Stative being in a state	**Inchoative** entering into a state	**Causative** putting into a state
Spanish	**Experiencer_subject**		**Cause_emotion**
	estar V-PP	**V REFL**	**V**
	estar sorprendido	sorprenderse	sorprender
English	**Experiencer_subject**		**Cause_to_experience**
	be V-PP	**get V-PP**	**V**
	be surprised	get surprised	surprise

Figure 2: Different lexicalization patterns in Spanish and English emotion predicates (cf. Subirats and Petruck, 2003).

Different constructions in English and Spanish motion predicates

Comparative valence descriptions between SFN and FN have still shown other differences. For instance, motion predicates in Spanish, like the majority of predicates from other frames, accept Purpose FEs, such as *para pedirle dinero a un amigo* in (6) below.

(6) **Voy** a San Francisco para pedirle dinero.
Go to San Francisco To ask-him money
a un amigo.
to a friend
I go to San Francisco to ask a friend for money.

However, many motion predicates in Spanish accept an Intentional FE, such as *a ver un amigo* in (7), which expresses the intention of the motion event, which is semantically different from the purpose.

(7) **Voy** a San Francisco a ver a un amigo.
Go to San Francisco to see to a friend
I go to San Francisco to see a friend.

Intentional FEs not only have a different meaning from Purpose FEs, but they are also syntactically different. In this way, the Intentional FE *a ver un amigo* in (7) is a prepositional object and, therefore, it is not only a conceptual argument of the target, but also a syntactic argument. On the contary, the extrathematic Purpose FE *para pedirle dinero a un amigo* in (6) is an adjunct which is not a syntactic argument of the target. The semantic difference between Intention and Purpose allows both FEs to be present in the same sentence, such as in (8), acting as different conceptual arguments of the same target.

(8) Juan **fue** a San Francisco a visitar a
John went to San Francisco to visit a
un amigo para pedirle dinero.
a friend to ask-him money
John went to san Francisco to visit a friend and ask him for money.

There is a clear difference between Spanish and English. While in Spanish, there are two conceptual and syntactic arguments attached to the same target, in English there are two coordinated sentences with two different targets. Thus English uses another construction to express the same meaning.

FrameSQL

FrameSQL is a web-based application to search and view the Berkeley FN data on the web browser (Sato, 2003). Since its data structure is basically the same as that of SFN, FrameSQL can handle the SFN data with a little modification. The application stores the FrameNet data in an MySQL database, and executes various searches in the SQL language, when users select search parameters on the web browser. The application handles both of the FN data seamlessly, showing Spanish and English lexical units belonging to the same frame on the same window. It makes it easier to compare semantic structures of the two lexicons.

FrameSQL has several search modes. Figure 3 shows a basic menu for searching the Spanish lexical unit *elogiar* of the Judgement_communication frame. The search menu consists of four panes: the upper one for selecting search modes, the middle-left for specifying frames and lexical units to search and view, the middle-right for setting search parameters, and the bottom for showing help files and search results.

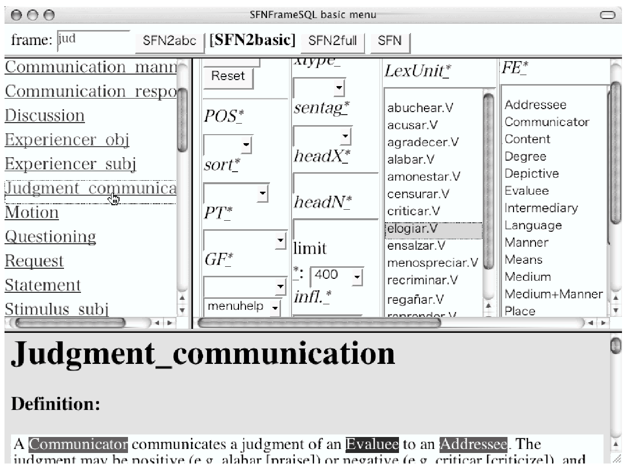


Figure 3: Basic search menu of FrameSQL

Figure 4 shows search results of the lexical unit *elogiar*. The bottom pane summarizes how each of FEs are used in annotated examples. Each line in the bottom pane consists of the number of annotated examples (Num), two hyperlinks to English FrameNet (Sloppy, Exact), and a set of FEs and LU used in annotated examples (FE/LUset). The left numbers are hyperlinked to annotated examples. For example, when a user clicks on the hyperlink *01* of the last line which have the FE/LU set *Communicator+elogiar.V+Evaluee+Role*, annotated examples with this set appear on the middle-left pane of Figure 4.

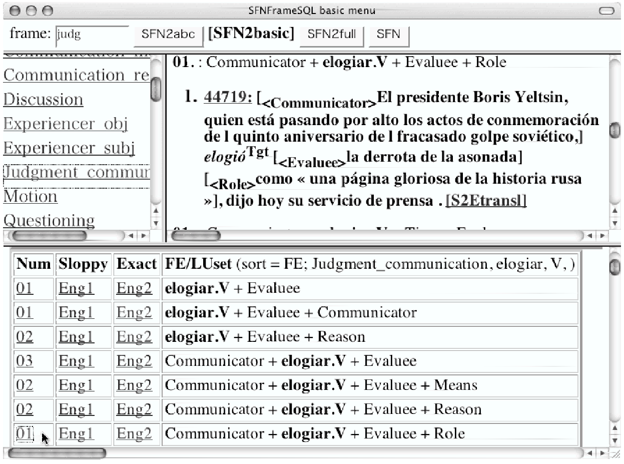


Figure 4: Search results of *elogiar*

The hyperlinks to English FrameNet lead to English annotated examples which have similar FE sets to Spanish ones. For example, when a user click on the Engl link of the FE/LUset *Communicator+elogiar.V+Evaluee+Role*, English annotated examples of the Judgement_communication frame appear in the middle-left pane that contain this FE set, as in Figure 5.

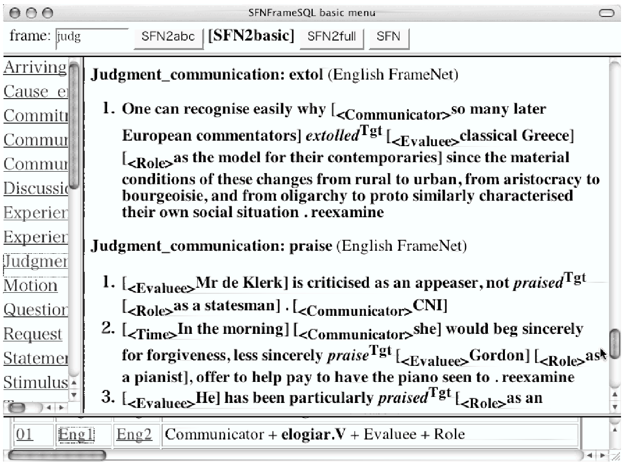


Figure 5: Accessing English FrameNet

References

Baker, C.F., C.J. Fillmore and B. Cronin. (2003). The Structure of the FrameNet Database, International Journal of Lexicography, Volume 16(3), 281--296.

Baker, C.F. and J. Ruppenhofer (2002) FrameNet's Frames vs. Levin's Verb Classes. In J. Larson and M. Paster (Eds.) In Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society. (pp.27–38).

Fillmore, C.J. (1985). Frames and the semantics of understanding. In Quaderni di Semantica 6(2), 222-254.

Fillmore, C.J. (1982). Frame Semantics. In Linguistics in the Morning Calm (pp.111–137). Seoul: Hanshin Publishing Co.

Fillmore, C.J., C.R. Johnson and M.R.L. Petruck. (2003). Background to FrameNet. International Journal of Lexicography 16(3), 235–250.

Fillmore, C.J. and H. Sato. (2002). Transparency and Building Lexical Dependency Graphs. In J. Larson and M. Paster (eds.) Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society (87—99).

Johnson, C.R., M.R.L. Petruck, C.F. Baker, M. Ellsworth, J. Ruppenhofer, and C.J. Fillmore. (2002). FrameNet: Theory and Practice. Berkeley, CA: International Computer Science Institute, Technical Report-02009.

Sato, H. (2003). FrameSQL: A Software Tool for FrameNet. In ASIALEX '03 Tokyo Proceedings (pp. 251–258), Asian Association of Lexicography.

Subirats, C. and M. Ortega. (2000). Tratamiento automático de la información textual en español mediante bases de información lingüística y transductores. Estudios de Lingüística del Español 10: http://elies.rediris.es/elies10/

Subirats, C. and M. R. L. Petruck. (2003). Surprise: Spanish FrameNet! Presentation at Workshop on Frame Semantics. Prague (Czech Republic): International Congress of Linguists. July 29, 2003.

PropBank, SALSA, and FrameNet: How Design Determines Product

Michael Ellsworth[*], **Katrin Erk**[†], **Paul Kingsbury**[‡], **Sebastian Padó**[†]

[*]International Computer Science Institute
1947 Center St., Berkeley, CA, USA
infinity@icsi.berkeley.edu

[†]Computational Linguistics, Saarland University
Im Stadtwald 17, 66123 Saarbrücken, Germany
{erk,pado}@coli.uni-sb.de

[‡]Institute for Research in Cognitive Science
3401 Walnut St., Philadelphia, PA, USA
kingsbur@unagi.cis.upenn.edu

Abstract

We compare three projects that annotate semantic roles: PropBank, FrameNet, and SALSA. The first part of our analysis is a comparison of the different word sense distinction criteria underlying the annotation. Then, we study the effects of these criteria at the level of actual phenomena that require annotation. In particular, we discuss metaphor, support constructions, words with multiple meaning aspects, phrases realizing more than one semantic role, and nonlocal semantic roles.

1. Introduction

During the last few years, corpora with semantic role annotation have received much attention, since they offer rich data both for empirical investigations in lexical semantics and large-scale lexical acquisition for NLP applications.

However, semantic role annotation of text is a complicated endeavor, whose product is deeply influenced by the initial design philosophies and policy choices of a project. We examine key differences between three annotation projects, FrameNet (Johnson et al., 2002), PropBank (Kingsbury and Palmer, 2002), and SALSA (Erk et al., 2003), and the consequences of these differences. After introducing the goals of the projects, we compare the criteria for determining the words senses underlying the annotation. Then, we discuss the consequences of these choices at the level of actual annotation.

2. PropBank, FrameNet and SALSA

FrameNet is primarily a lexicographical project. Its starting point is the observation that words can be grouped into semantic classes, the so-called 'frames', representations for prototypical situations or states. Each frame provides its set of semantic roles. The Berkeley FrameNet project is building a dictionary which links frames to the words and expressions that can introduce them in text. Examples from the BNC (Burnard, 1995) serve to illustrate typical usages.

The more practical aim of PropBank, on the other hand, was to obtain a complete semantic role annotation of the Penn Treebank (Marcus et al., 1994). The PropBank lexicon was added first to facilitate annotation, and later evolved into a resource on its own. No higher-order organization was established at first, so for each unique verb sense, a 'frameset' was constructed that consists of the set of semantic roles at its accompanying syntactic realizations.

SALSA uses the FrameNet dictionary as the basis for its annotation but, like PropBank, pursues an exhaustive annotation of its corpus, the TIGER corpus (Brants et al., 2002), a German newspaper corpus. Different from FrameNet, however, SALSA is not committed to always assigning a single sense (frame) to a target expression, or a single semantic role to a constituent. In cases of systematic as well as idiosyncratic ambiguity and vagueness, annotators may assign more than one frame or semantic role and mark the occurrence as being 'underspecified'.

3. Criteria for frameset and frame creation

In this section, we describe the criteria used for grouping instances of role-introducing expressions (targets) into senses, i.e. frames (in FrameNet) and framesets (in PropBank), respectively. SALSA uses FrameNet's criteria.

3.1. PropBank

Since the purpose of the PropBank lexicon was primarily to provide a description of every verb in the Penn Treebank II corpus in all their attested usages, it was kept as agnostic as possible with respect to higher-level generalizations. Recall from above that framesets are verb-specific, and even though polysemous verbs could possess multiple framesets, in general senses were merged into single framesets whenever possible. Distinction of senses, and therefore creation of distinct framesets, was triggered by both syntactic and semantic properties.

One important criterion is the number of possible semantic roles. For example, the verb *afford* is given two framesets, on the basis of pairs of sentences such as:

(1) These days Nissan can **afford** that strategy, even though profits aren't exactly robust. (wsj_0286)

(2) Last year the public was **afforded** a preview of Ms. Bartlett's creation in a tablemodel version, at a BPC exhibition. (wsj_0984)

Although each sentence has two realized semantic roles, the passive morphology in (2) indicates that a third role is possible. The same is not true for (1), which leads to the creation of two framesets:

afford.01	'be able to sustain the cost of something'
arg0:	entity sustaining cost
arg1:	costly thing
afford.02	'provide, make available'
arg0:	provider
arg1:	thing provided
arg2:	recipient

This pair of sentences also serves to illustrate how it is not the number of roles appearing in any sentence which determines the framing, but the number of possible roles across a variety of syntactic alternations such as active/passive or causative/inchoative.

Even if the number of roles is the same, framesets are also distinguished when the meanings of the usages are sufficiently different, that is if a relatively proficient speaker of English will be able to distinguish between these senses. For example, the verb *stem* also takes two framesets[1], each with two roles, given pairs of sentences such as:

(3) Travelers Corp.'s third-quarter net income rose 11%, even though claims **stemming** from Hurricane Hugo reduced results $40 million. (wsj_0144)

(4) If the company can start to ship during this quarter, it could **stem** some, if not all of the red ink, he said. (wsj_1973)

PropBank therefore assumes the following two framesets:

stem.01	'arise'
arg1:	entity arising, coming about
arg2:	arising from what?
stem.02	'stanch, cause to stop flowing'
arg0:	causer of non-flowing
arg1:	thing no longer flowing

Because roles are defined per verb, the classification of individual verbs into higher-level classes is not trivial. Most framesets make reference to VerbNet (Kipper et al., 2002) classes, a refinement of Levin's (1993) scheme, and efforts are underway to discover natural classes of verbs based on patterns of usage (Kingsbury and Kipper, 2003).

3.2. FrameNet

FrameNet practice differs fundamentally from the process described for PropBank in not considering syntactic differences (except inasmuch as these correlate with semantics). This means that FrameNet can consider verbs, adjectives, nouns, prepositions, adverbs, and even multiword expressions (such as *pull the wool over X's eyes* in the Prevarication frame) on the same footing, despite any structural differences between them, since it is only their meaning which matters.

FrameNet's semantic criteria for creating frames also differ from those of PropBank in taking the senses as less predefined. FrameNet first collects and analyzes the corpus attestations of target words (or idiomatic phrases) judged to have semantic overlap (as determined by consulting thesauri, dictionaries, and native intuitions). The attestations are divided into semantic groups, noting especially the semantic roles (frame element) of each (but ignoring pragmatic and general constructional differences, as PropBank does), and then combining these small groups into frames. Note that the resulting groupings need not correspond to the initial groupings given by thesauri, etc. The factors which may serve to differentiate or relate the groups of attestations include the following:

1. For the target:
   (a) The basic denotation of the targets may differ, such as in the case of *blue* and *broken* which refer to completely different kinds of states. Obviously this is a diagnostic which is easy in some cases and hard in others. A more difficult case is the basic meanings of *take* vs. *receive*, which share lots of implications about a Theme changing hands. It is simply unclear whether these are exactly the same kind of thing. The difficulty of forming an intuitive type-hierarchy for events is why other criteria are needed.
   (b) The presuppositions, expectations, and concomitants of the targets may differ. For example, *cross-examine* evokes a courtroom session, a preceding event of questioning by an opposing legal party, etc., differentiating it from the simpler *examine*. By this feature, *receive* and *take* would be differentiated, since *receive* presupposes another willing agent participating as the Donor and *take* does not.
2. For semantic roles:
   (a) Their number and type, (e.g. *kill* has a role not present for *die*)
   (b) Interrelations (e.g. Purpose refers to a particular kind of mental state of an Agent, as opposed to Means which refers to an action of an Agent)
   (c) Profiling (e.g., the difference of *buy* and *sell*, in which both refer to a Buyer and a Seller, but in the case of *buy* the Buyer is portrayed as more saliently controlling the action, vs. *sell*, in which the Seller is portrayed as more salient), and
   (d) The semantic preferences and restrictions the target imposes (e.g. *tie* requiring the Connector be a long, flexible object).

Grouping usages according to close matches of such features allows FrameNet to form "minimal" frames; the more-inclusive final frames are then formed by loosening some of these conditions such as 2d., so that *tie* and *staple* can be

[1] This neglects two other senses, unseen in the Wall Street Journal: 'remove the stems from something which inherently has a stem' as in *John stemmed the cherries* and 'reduce something down to just a stem' as in a morphological stemmer/lemmatizer.

grouped despite the constraints on what kind of Connector they specify.

Conversely, these semantic considerations (especially 2b.) led FrameNet to draw a distinction between causative and inchoative cases that PropBank does not make. Lexical membership in a FrameNet frame entails that for each use of a target, all of the core frame elements must be semantically present. Inchoatives do not entail the existence of a Cause or Agent, as can be seen by comparing *the rain ended* to the infelicitous $^{?}$*(someone/something) ended the rain.* The inchoative and causative uses of *end* thus belong to the frames Process_end and Cause_to_end respectively.

4. Consequences in the Annotation

The different aims of PropBank, FrameNet and SALSA are reflected in the practice of annotation. PropBank limits itself to annotating the literal meaning of the target, preferring small, incremental, easily-attained goals. FrameNet and SALSA follow Fillmore (1985), which states that *'Frame Semantics does not seek to draw an a priori distinction between semantics proper and (an idealized notion of) text understanding'* and consequently try to annotate what is actually understood. This makes the task more complex but should finally yield a more informative annotation.

Semantic annotation has to deal with large classes of phenomena for which the meaning is either hard to pin down or subject to debate. We now show the consequences of different annotation choices of the three frameworks for such phenomena for both tasks of frame(set) assignment and semantic role assignment. For the first task, we discuss metaphors, support cases, and instances with multiple meaning aspects, while issues for the second task are phrases realizing multiple semantic roles and nonlocal semantic roles.

4.1. Metaphor

Metaphors are abundant even in newspaper texts. A recent study of a 100k word corpus found that roughly 54% of all motion terms were used metaphorically (Tewari, 2003). (5) is a case in point.

(5) Viele meinen, dass Perot mit seinem Befehlston auf dem Capitol gegen eine Wand **laufen** würde. (Tiger s13)
(Many think that Perot would **walk** into a brick wall on the Capitol with his commanding tone.)

In such cases, annotation projects have to decide between annotation the 'source' (literal) or 'target' (metaphorical) meaning (following Lakoff and Johnson's (1980) terminology). However, the border between metaphor proper, and cases that are lexicalized so far as to be indiscernible as a metaphor, is often not clear-cut, as in (6). *Get through [a difficult time]* could be characterized as a metaphor with a Motion source, but can also be seen as lexicalized so far to have become a separate sense of *get*.

(6) Der "Pluralismus von Erklärungen" aus der CDU/CSU-FDP-Koalition zeige, dass die Einigkeit über die Pflegeversicherung nur "vorgetäuscht" worden sei, "um über die Sommerpause zu **kommen**", sagte Klose.
(The "multiplicity of explanations" given by the CDU/CSU coalition showed that they only "pretended" to agree on nursing care insurance "in order to **get through** the summer break", Klose said.)

PropBank. PropBank, for the most part, takes a consistently literal analysis of such constructions. A later pass of annotation is planned, in which instances will be flagged as being metaphorical. Nevertheless, there are cases when metaphor is unavoidable. These tend to occur with the most frequent verbs, those with the most leached-out underlying semantics. When these are common enough, they can be thought of as being just another sense of the verb and thus acquire a new frameset. The division between a true metaphor and a different sense is not clear, however: how often is often enough?

FrameNet. FrameNet decides between conventionalized metaphors, like (5), and nonce metaphors, such as in (7), whose unique meaning is determined by its special context. Conventionalized metaphors are annotated with the target frame, while nonce metaphors are ignored, or in rare cases they are annotated and tagged with the sentence-level tag "Metaphor".

(7) A small gust of laughter **blew** through him , and left him smiling . (BNC)

SALSA. In the finished SALSA corpus, both the source and the target frame will be assigned. To speed up annotation, however, the tagging of metaphoric instances is split up into two passes. In the current first pass, the instance in question is marked as metaphoric, and either the source or the target frame is tagged (with a flag indicating which of the two it is). The annotators mark whichever of the two frames is easier to determine; the default is the source, since the target meaning is sometimes hard to pin down in terms of frames. (8) shows such a case.

(8) Den einen **geht** der Schritt **zu weit**, den anderen nicht **weit genug**. (TIGER s10471)
(For some this **goes too far**, for some, not **far enough**.)

(8) talks about some cognitive scale, maybe one of acceptability. But the target sense can only be described on a very abstract level, much more abstract than is usual in frame descriptions.

4.2. Support constructions

Support constructions are non-compositional multiword expressions[2] in which a governing verb and/or preposition are optional for lending the phrase, semantically headed by a noun or adjective, its essential meaning. Putting it slightly more formally, a support construction involves (1) an adjective or noun that denotes a state or event and is the source

[2]Non-compositionality is tested by substitutability, replacing the words of the phrase with likely synonyms. If the synonyms do not allow the phrase to retain a similar meaning, then it is non-compositional and should be annotated as a unit.

of virtually all the meaning of the phrase and (2) syntactically governing verbs or prepositions with only simple, grammatical meaning which do not have the same meaning independently of the target.

The simplest cases are phrases like *take a bath*, which evokes the Grooming frame, in which *bath* (as in *his bath lasted three hours*) all by itself evokes the exact same frame. *Be in possession (of)* provides another clear, but slightly more complicated example. Here, *be* and *in* are supports, because when we compare *John is in possession of the estate.* and *John's possession of the estate*, the differences in meaning are not framal differences.

One obvious problem that supports present for any semantic annotation project is how to recognize and record the cases, and how to record the differences between cases. A further basic problem is what types of 'minor' meaning change are allowed for the supports themselves, such as causativity, aspectual change, etc, and how to record the differences between them.

For support, as for the other phenomena we have discussed, there are borderline cases that could be characterized as support as well as something else. This problem occurs particularly often with high-frequency verbs that can denote situatedness, like *put*, *lie*, *stand*. The trouble with cases like (9) is that they could be analyzed either as a simple case of support, or as a metaphor with the frame Being_situated as a source.

(9) Zwar **liege** die Verantwortung allein bei der Bundesregierung , doch angesichts der nicht unerheblichen Gefährdung der eingesetzten Soldaten habe man eine breite Zustimmung gesucht, sagte ein Sprecher. (TIGER s1307)
(While responsibility **lies** solely with the federal government, broad agreement had been sought in view of the considerable danger for the soldiers, a spokesman said.)

FrameNet. The types of meaning change allowed by FrameNet for supports include:

Vanilla: the support adds virtually nothing to the target (like the *take a bath* example above).

Aspectual: the support changes the temporal focus of the event portrayed by the target, e.g. *get/go/fall into a (foul) mood* vs. (the vanilla) *be in a (foul) mood*.

Point-of-view: the support changes the profiled point-of-view of the target, e.g. *undergo* in *undergo a physical exam* vs. *give a physical exam*, with the patient's and doctor's points of view respectively.

Causative: the support adds another participant and the idea of causation of the basic scene. These generally occur paired with a non-causative support, as in *put in a (foul) mood* vs. *be in a (foul) mood*; *bring into play* vs. *come into play*; *give a headache* vs. *have a headache*, or the idiosyncratic *show a good time* vs. *have a good time*.

Currently, FrameNet annotates supports with a special tag, and only when they occur in the context of a noun or adjective target that is already being examined. There is no annotation of supports as targets themselves, and no systematic way of recognizing instances of the separate types given above.

PropBank. PropBank dodges the entire issue by lumping all support constructions for each verb into a single frameset, described as 'support'. These framesets usually take two or three roles, of which one is the noun which is the real predicate and the others are the roles of the nominal. For example,

(10) [$_{\text{Arg0}}$ John] **made** [$_{\text{Arg1}}$ a shrewd guess about Mary's intentions].

For those cases where the predicate nominal is deverbal, the Nombank project at New York University is annotating the semantic role structure using the PropBank lexical frames, so a sentence such as (10) will receive a second, overlapping structure:

(11) [$_{\text{Arg0}}$ John] made a [$_{\text{ArgM-MNR}}$ shrewd] **guess** [$_{\text{Arg1}}$ about Mary's intentions].

SALSA. In the first pass, SALSA tags all the cases recognized by FrameNet above just as the Pseudo-frame Support. This is somewhat similar to the PropBank treatment. A second pass over the support cases is planned, giving them a deeper, FrameNet-style analysis.

4.3. Words with several simultaneous meaning components

There are words with several simultaneous meaning components, which are unlike polysemy in that the different meanings are not a question of context, but rather refer to two simultaneous situations at once. This can be either restricted to a single instance or systematic. For example, many verbs can be systematically used to describe both the content of a communication and its manner:

(12) And don't expect many complete games by pitchers – perhaps three out of 288, **laughs** *t* Mr. Fingers, the former Oakland reliever. (wsj_0214)

The following idiosyncratic case demonstrates that such cases can show a blend of the syntactic patterns of the two single usages. (13) has both the direct speech of a "communication", and the direct object of the "impede" meaning:

(13) "Sorry, you cannot enter", he **blocked** the way.

The question of how to annotate these cases has an obvious impact on the usefulness of the annotation, since in order to be aware of the full meaning potential of the expression, one would need to indicate both (or all) components.

PropBank. To annotate an instance such as (12), PropBank creates a new frameset for *laugh*; while the main frameset includes only a single role for the laugher, this new frameset must also include a role for the utterance spoken while laughing. Since this behavior is systematic for 'manner of speaking' verbs (including *laugh, cry, wheeze* and many others), this policy can lead to a proliferation of framesets. The same is true for idiosyncratic cases such as (13).

FrameNet. In FrameNet, blended frames are constructed for targets that systematically exhibit several meaning aspects, like (12), while idiosyncratic cases such as (13) are not treated[3].

SALSA. In SALSA, instances with multiple meaning aspects can be marked with more than one frame, in accordance with the general underspecification principle used in SALSA annotation. For (12), the applicable frames would be Statement and something like Physical_obstruction.

A particularly difficult case arises when a sentence might be seen to evoke multiple senses or not, depending on the view of the reader. Unlike the cases above, where the multiple senses are clearly present, the senses available in the sentences below are much more subtle and optional:

(14) Such a thought would never **cross my mind**.

(15) I must admit I feel a tad embarrassed though, as the idea of focusing on the local market first didn't even **cross my mind**… (www.webhostingtalk.com/archive/thread/232858-1.html, Feb 24, 2004)

Literally this construction means that the speaker would not think of X, that X would never occur to him or her. But (14) also has the connotation of not wanting to do (the thing which was referred to). (15), seemingly identical to (14), seems to not invoke this secondary sense, apparently related to the lack of the modal *would*. Both examples, (14) and (15), share the idea of Invention, and the first example also includes the idea of Desiring. How many of these senses should annotators mark for each of these sentences? Should annotators simply tag (14) as a case of Invention, or should the secondary sense also be indicated?

The annotation of these examples is unproblematic in PropBank since there are no syntactic peculiarities. FrameNet would differentiate the two examples with one target *would cross mind* and another simply *cross mind*. The first would be in a frame which inherits from Desiring, the second would not. SALSA treats such borderline examples on a case-by-base basis, letting annotators decide between single-frame annotation and underspecification on the basis of the prevailing overall meaning of the sentence.

4.4. Phrases realizing multiple semantic roles

We now turn to phenomena that concern semantic role assignment rather than frame(set) assignment. The first phenomenon parallels target words with multiple meaning aspects: lexical material that simultaneously bears multiple semantic roles. This situation often arises with a plural constituent within which two separate semantic roles have been merged, as the contrast between (16) and (17) shows:

(16) Argentine negotiator Carlos Carballo will **meet** with banks this week. (wsj_0021)

(17) The economic and foreign ministers of 12 Asian and Pacific nations will **meet** in Australia next week to discuss global trade as well as regional matters such as transportation and telecommunications. (wsj_0043)

Sentence (18) is much more complex. The expression *under the hand dryer* is certainly the Place role of the drying, but it also indicates the Instrument of the drying, the hand dryer.

(18) We immediately rushed to the ladies, washed Jessica carefully in the sink and **dried** her under the hand dryer. (BNC)

Note that in this example a certain amount of knowledge about hand dryers (namely that they usually blows hot air downward) is required of the annotator, as well as some degree of inference, in determining that the Place doubles as the Instrument of the drying. Note also that the assignment of Instrument to *under the hand dryer* is defeasible and can be overwritten, for example by continuing the sentence by *… using lots of paper towels*.

PropBank. Under PropBank there is no provision for a single constituent to bear multiple labels, so the annotators are forced to choose. For these and similar cases a hierarchical notion of semantic roles was developed, preferring lower-numbered to higher-numbered labels and numbered labels to ArgMs, which are felt to be universal and adjunct-like. In (17), with a choice between Arg0 and Arg1, the lower numbered label, Arg0, is used, and in (18), the instrumental Arg2 is used in preference to the location ArgM-LOC. While annotators have experienced little difficulty with this policy, it might pose interesting challenges to systems attempting to interpret PropBank annotations.

FrameNet/SALSA. In FrameNet, *meet* in the sense of (17) is in the Discussion frame, which has one role for the subject participant when it refers to the collective Interlocutors, and two other roles (Interlocutor_1 and Interlocutor_2) for the subject and complement respectively when these denote participants separately. The relationships between the roles (Interlocutor_1 requires Interlocutor_2 and Interlocutors excludes Interlocutor_1 and Interlocutor_2) are specifically encoded in the database. Sentence (18) is annotated for both semantic roles, Place and Instrument, in the FrameNet corpus.

4.5. Nonlocal semantic roles

Since relatively few sentences, especially in more formal or journalistic registers, contain only one clause, the question of the scope of annotation often arises. How far away from a verbal head does one look for roles of that verb? Interestingly, roles that are realized nonlocally show the same characteristics as the Instrument role in (18): World knowledge, as well as some inference, is required to a much larger degree than usual to assign these semantic roles, and the assignment is defeasible, i.e. it is possible to change the way semantic roles are assigned by setting the expressions in

[3] One possibility of analyzing (13) would be to annotate block.v just for Physical_obstruction, and to introduce an extrathematic Message role to that frame's definition. Like all extrathematic roles, it would be introduced by some kind of construction with its own separate semantics and form constraints – i.e. it is not introduced by the target, but the target can unify with it.

question in a different context. For example:

(19) Besitzer von Zweifamilienhäusern, die vor 1987 gebaut oder **gekauft** haben. . . (TIGER s975)
(Owners of two-family homes who have built or **purchased** before 1987). . .

Gekauft (purchased) evokes the frame Commerce_sell. *Zweifamilienhäusern* (two-family homes), which is not in the maximal projection of *gekauft*, may be inferred to be the Goods role of this frame. This inference is defeasible, though. Suppose this sentence occurs within a text about buying stocks. Then the Goods may be the stocks instead of the houses.

Noun targets are especially problematic in that most of their roles are usually realized nonlocally and are therefore defeasible, such as in (20):

(20) Vor Jahren, als Helmut Kohl erstmals ganz unten war [. . .], machte [. . .] Günter Oettinger bundesweit mit einer Rücktritts**forderung** von sich reden. (TIGER s1862)
(Years ago, when Helmut Kohl was on the rocks for the first time, Günter Oettinger brought himself into public awareness with a **demand** for resignation.)

Forderung (demand) evokes the frame Request. Neither the Speaker nor the Addressee of the request are realized locally. The Speaker is probably Günter Oettinger, but he might also be just a medium. The Addressee is probably Helmut Kohl; nevertheless, all these inferences can be overridden by context.

FrameNet FrameNet allows annotators to annotate nonlocal arguments when they participate in any of a number of recognized nonlocal constructions such as questions and fronting, or in general when we can recognize that the assignment of the semantic roles is not defeasible by context.

SALSA. In SALSA nonlocal semantic roles are included in the annotation, for three reasons: First. annotators usually have strong intuitions about these nonlocal semantic roles and tend to annotate them when this is not explicitly prohibited. Second, these nonlocal, defeasibly inferred semantic roles constitute interesting data on inferences people make when understanding sentences. Third, local and nonlocal semantic roles can be clearly distinguished through the syntactic structure, which makes it possible to sort out nonlocal roles whenever that is required.

PropBank. PropBank, being built upon the existing syntactic parse in the Penn English Treebank II, makes use of the 'traces' (overt markers on empty nodes, coindexed with their lexical antecedents) present in the treebank to find nonlocal arguments.

There are cases, however, when there is a genuine ambiguity as to the antecedent of a trace, such as in (21):

(21) Commonwealth Edison is **seeking** about $245 million in rate increases [*T*-1] to pay for Braidwood 2. (wsj_0015)

In this example, the trace [*T*-1] could point to *Commonwealth Edison*, who will be doing the paying after all, or to *$245 million in rate increases* which is the instrument of paying. There are many cases of this agent/instrument ambiguity in trace chains, leading PropBank to choose the agent in all cases.

5. Conclusion

Whether in handling metaphors, identifying support cases, or assigning a single sense to a role-bearing expression, all three projects have to deal with corpus instances that lie on the borderline between different categories and defy clear classification. Our comparison has shown that the theoretical differences between the three projects lessen in view of actual annotation, however the mechanisms the three projects use in dealing with borderline cases differ. PropBank tends to formulate general policies (e.g. preferring arguments with lower numbers when more than one role label applies to a phrase), FrameNet includes systematic cases, but excludes idiosyncratic borderline cases from its consideration, and SALSA allows for more than one tag through underspecification, allowing for a later analysis based on annotated underspecified instances.

One deeper question that our comparison has highlighted is: How much context information should annotators use in determining the tag to be assigned, and how much inference are they allowed to perform to divine the meaning of an expression? This question is probably most prominent in the *cross one's mind* example (14) for frame(set) assignment, and for semantic role assignment the *hand dryer* example (18) and the nonlocal cases. While all three projects have to allow some use of context in determining the meaning of a phrase, FrameNet and SALSA have to take it into account to a larger degree since they are assigning the meaning that is actually understood, while PropBank mostly focuses on the literal meaning.

6. References

Brants, S., S. Dipper, S. Hansen, W. Lezius, and G. Smith, 2002. The TIGER treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*. Sozopol.

Burnard, L., 1995. *User's guide for the British National Corpus*. British National Corpus Consortium, Oxford University Computing Services.

Erk, K., A. Kowalski, S. Pado, and M. Pinkal, 2003. Towards a resource for lexical semantics: A large German corpus with extensive semantic annotation. In *Proceedings of ACL-03*. Sapporo, Japan.

Fillmore, C.J., 1985. Frames and the semantics of understanding. *Quaderni di Semantica*, IV(2).

Johnson, C. R., C. J. Fillmore, M. R. L. Petruck, C. F. Baker, M. J. Ellsworth, J. Ruppenhofer, and E. J. Wood, 2002. FrameNet: Theory and Practice. `http://www.icsi.berkeley.edu/~framenet/book/book.html`.

Kingsbury, P. and K. Kipper, 2003. Deriving verb-meaning clusters from syntactic structure. In *Workshop on Text Meaning, HLT/NAACL 2003*. Edmonton, Canada.

Kingsbury, P. and M. Palmer, 2002. From TreeBank to PropBank. In *Proceedings of LREC-2002*. Las Palmas, Spain.

Kipper, K., M. Palmer, and O. Rambow, 2002. Extending PropBank with VerbNet semantic predicates. In *Workshop on Applied Interlinguas, held in conjunction with AMTA-2002*. Tiburon, CA.

Lakoff, G. and M. Johnson, 1980. *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Levin, B., 1993. *English Verb Classes and Alternations*. Chicago, IL: University of Chicago Press.

Marcus, M., G. Kim, M.A. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger, 1994. The Penn Treebank: Annotating predicate argument structure. In *ARPA Human Language Technology Workshop*. Plainsboro, NJ.

Tewari, A., 2003. Detecting metaphors involving motion verbs in online business news sources. `www.eecs.berkeley.edu/~ambuj/research/reports/cs182report.pdf`.

Validating and Improving the Czech WordNet via Lexico-Semantic Annotation of the Prague Dependency Treebank

Jan Hajič,[†] Martin Holub,[*] Marie Hučínová,[*] Martin Pavlík,[*] Pavel Pecina,[*] Pavel Straňák,[*] Pavel Martin Šidák[*]

[†]Institute of Formal and Applied Linguistics, [*]Center for Computational Linguistics
Faculty of Mathematics and Physics, Charles University
Malostranské náměstí 25, Prague, Czech Republic
{hajic, holub, hucinova, pavlik, pecina, p.stranak, sidak}@ufal.mff.cuni.cz

Abstract

We give a brief report on our experience with lexico-semantic annotation of a Czech linguistic corpus. We use the Czech WordNet (CWN) as a repository of lexical meanings and we annotate each word which is included in the CWN. The statistics of the annotated data is used as a feedback for validating and improving the coverage and quality of the CWN. We also discuss some methodological questions.

1. Introduction

Generally, the annotation of linguistic corpora usually consists of a sequence of processes corresponding to several levels of annotation. In the Prague Dependency Treebank (PDT; see (Hajič et al., 2001b), (Hajič et al., 2001a)), the annotation can be viewed as a gradual enrichment of text by several types of labels in the following sequence: raw text — tokenized text — morphologically analyzed and lemmatized text — syntactically annotated text — lexico-semantically annotated text.

Lexico-semantic annotation (if the process is manual, done by humans) or *tagging* (if it is automatic, performed by a machine) means assigning a semantic tag from an a priori given set to *each* relevant lexical unit in a text. Lexical units which we deal with during this process are lemmas of words;[1] the relevant ones are those of the autosemantic parts of speech, namely all nouns, adjectives, verbs, and adverbs.

In this paper, symbol $\mathrm{T}_p(l)$ denotes a set of *possible semantic tags* which can be assigned to lemma l. Note that the members of $\mathrm{T}_p(l)$ always make a list of options from which a human annotator selects a correct tag for the lemma l in a given context.

The purpose of lexico-semantic annotation or tagging is to distinguish between different meanings of semantically ambiguous lemmas that can emerge when a lemma is used in different contexts. Undoubtedly, the lexico-semantic information given by correctly assigned semantic tags may be very important for many NLP tasks.

This paper concentrates on our practical experience with lexico-semantic annotation and empirical observations rather than on theoretical questions. At the very beginning, to start the lexico-semantic annotation of the PDT, we had to make two crucial decisions:

1. What system of semantic tags should we use for the lexico-semantic annotation?

   One possibility is to use a well known type of lexical database called WordNet (Fellbaum, 1998). Then, the basic semantic elements are *synsets*, sets of synonyms. As we annotate Czech texts, we decided to employ the Czech WordNet (CWN, (Smrž, 2003)) as a semantico-lexical basis for the annotation even though this choice is not a matter of course.

2. Moreover, it is also problematic how to employ the system of the synsets. In other words, how should we establish $\mathrm{T}_p(l)$ for each relevant lemma using the WordNet?

   For $\mathrm{T}_p(l)$ one can simply take the set of synsets which contains exactly the given lemma, while more complicated solutions permit even various sets of synsets to serve as semantic tags.

   Our current approach described in section 2. is very close to the first option, yet in section 6. we also discuss the latter one as in our opinion it is a way how to eliminate or at least reduce the undesirable impact of high granularity of the WordNet.

The rest of the paper is organized as follows: in section 2. we describe the process of manual annotation, our annotation tool, and how we deal with the CWN. Section 3. first introduces some information about texts we have annotated, and then the statistics of the performed annotation. Two applications are shown in sections 4. and 5. We validate the famous Yarowsky's hypothesis "one sense per collocation" and use the annotated data for validating and improving the CWN. Finally, we discuss the relation between the granularity of semantic tags and the inter-annotator agreement. Section 7. briefly summarizes the main contributions.

[1]The lemmas at the syntactical level of the PDT form a set of *tectogrammatical lemmas*, which is different from the set of lemmas at the morphological level (Hajič and Honetschläger, 2003). However (despite lexico-semantic analysis being placed only after the syntactical level), we currently use the lemmas produced by morphological analyzer for various practical or technological reasons.

Incorrect Reflexivity	l is reflexive but CWN knows only its non-reflexive form or vice versa.
Missing Positive Sense	l is positive, but CWN includes only its negative form.
Missing Negative Sense	l is negative, but CWN includes only its positive form.
Incorrect Lemma	The lemma l assigned to the word is incorrect (therefore the synsets proposed are incorrect too).
Figurative Use	The word is used in a metaphorical or other figurative way.
Proper Name	Assigned to proper names not included in the CWN.
Unclear Word Meaning in the Text	The meaning of l is unclear (therefore no synset can be assigned).
Unclear CWN Sense	The meaning of a synset is unclear and no other proposed synset can be used.
Missing More General Sense	At least one of the proposed synsets corresponds to the meaning of l, but is too specific and so expressing only part of l.
Missing Sense	None of the synsets proposed expresses the meaning of l and more specific exceptions can not be used.
Other Problem	Assigned if no other category can be used.

Table 1: List of the exceptions ordered by their preference.

2. Annotation using the Czech WordNet

The CWN was originally developed as a part of the EuroWordNet project (see (EuroWordNet, 2004), (Vossen, 1998)). Since then it was extended and is still being developed as a part of the BalkaNet Project (BalkaNet, 2004); currently, it consists of 28,392 synsets (including nouns, adjectives, verbs, and adverbs) (Smrž, 2003).

We use the CWN to obtain the set of possible semantic tags $\mathrm{T}_p(l)$ for each relevant lemma. In the process of annotation, each annotated lemma is assigned the best tag from this set.

2.1. Semantic tags based on the CWN synsets

In this paper, basic lexical units of the CWN (i.e. elements of synsets) are called *literals*. Literals which consist of exactly one lemma are called *uniliterals*, the other are called *multiliterals*.

Given a lemma l, the members of the set $\mathrm{T}_p(l)$ are

1. all synsets with a uniliteral consisting of l, and
2. some synsets with multiliterals (especially with those containing l) selected by a special procedure based on the CWN hypernymy/hyponymy relation (Pavlík, 2002).

2.2. Annotation environment

We use a graphical annotation tool.[2] The input file is a morphologically annotated text from the PDT with the corresponding $\mathrm{T}_p(l)$ sets encoded. The window of the application is split into four parts (see Fig. 1). When the annotator loads the input file, the text is displayed in the area marked A. In column B the annotator can see the list of lemmas of the words to be annotated. When the annotator chooses a lemma in column B, it is highlighted in the area A and he can see a list of possible tags $\mathrm{T}_p(l)$ in area C. To decide which synset from the offered list best represents the meaning of the word, the annotator can browse the synsets displayed in area C and review their English glosses (if present in the CWN), their hypernym synsets and the glosses of these hypernyms in area D. This way the annotator can see at the same time the annotated word in its full context and all the necessary information about its $\mathrm{T}_p(l)$ to select the best tag.

2.3. Instructions for annotators

The annotators must always assign exactly one synset or exception[3] to each relevant word and they are instructed to try to assign a uniliteral synset first. Only if no uniliteral synset is usable, they examine the multiliteral synsets (if present). If and only if no synset from $\mathrm{T}_p(l)$ can be assigned, the annotators choose one of the exceptions given in Table 1. First eight exceptions should be chosen preferably. Only if none of them is used, exception 'Missing Sense' can be assigned. Only if neither of the mentioned options is applicable, the annotator assigns the last exception 'Other'.

3. Annotation statistics

The long-term goal of our project is the complete annotation of the PDT 1.0 (Hajič et al., 2001a). After one year of annotation we have processed 11,014 sentences containing 125,129 words, mostly from the domain of economics. This is about 15 % of the PDT.

The entire annotation was performed independently by two human subjects (postgradual students with linguistic education) having identical instructions described in section 2. The average time needed for processing a typical document containing about 50 sentences by one annotator was 1 hour. Such a document contains approximately 100 to 280 words to be annotated.

Now we present a summary of the annotated data and some statistics.

3.1. Summary of the data distribution

In terms of lexical semantics, only *autosemantic* words (nouns, adjectives, verbs, and adverbs)[4] can be the subject of semantic tagging. There were 69 % such words in the annotated text. However, only words present in the CWN

[2]The program called DA was designed and implemented by Jiří Hana.

[3]In contrast to SemCor (Landes et al., 1998).

[4]Numerals are sometimes considered autosemantic words too, but usually they are not the subject of semantic annotation.

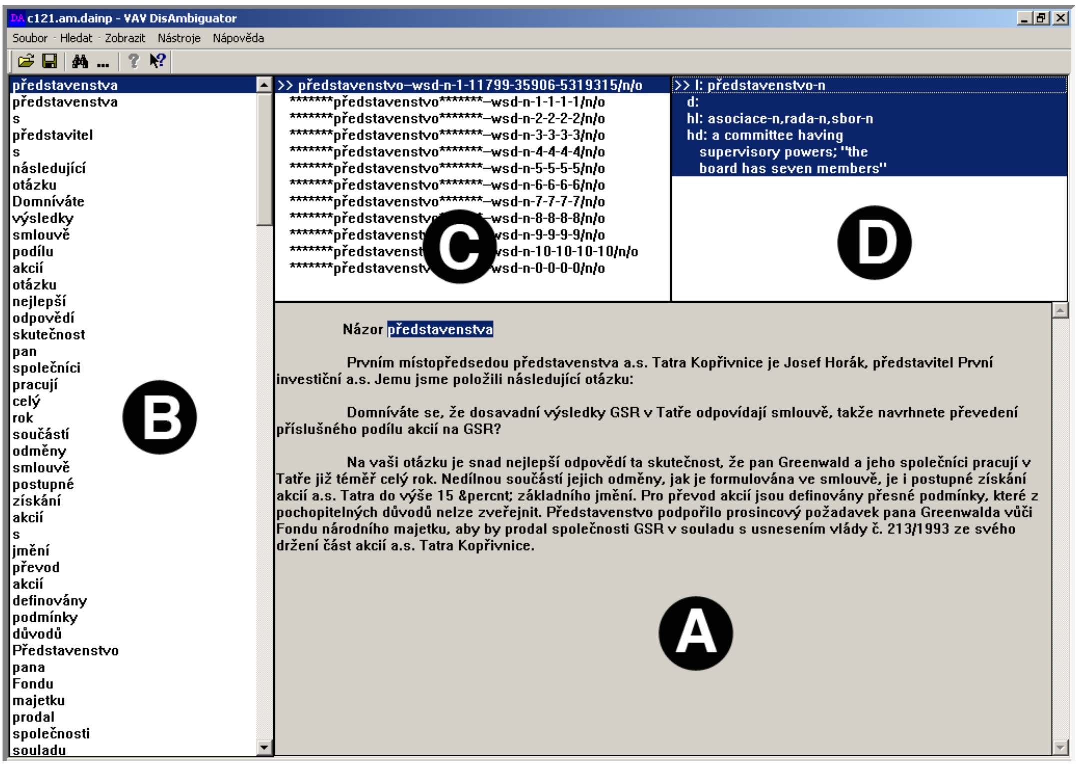


Figure 1: A screenshot of the annotation tool DA.

were annotated because they have at least one possible tag to be assigned. 34 % of all words fullfiled this condition but only 24 % were ambiguous (i.e. had more than one possible tag). This implies that only about 1/2 of all autosemantic words in a given text can be subject of automatic word sense disambiguation and only 1/3 are really ambiguous (according to the CWN). Detailed counts are given in the following table.

All words	125 129	100.0 %	
Autosemantic words	85 965	68.7 %	100.0 %
Annotated words	42 900	34.3 %	49.9 %
Ambiguous words	30 091	24.0 %	35.0 %

Table 2: Word counts in annotated text.

70 % of annotated words were nouns, 20 % were verbs, and 10 % were adjectives. Since the CWN version we worked with does not contain any adverbial synsets, no adverbs were annotated.

Detailed summary of part-of-speech (POS) distribution is given in Table 3. The relative counts are with respect to counts of autosemantic words. These numbers refer to “coverage” of annotated texts with words from the CWN. Generally, the coverage is poor, but varies strongly depending on POS.

Only 70 % of nouns, 26 % of adjectives, and 46 % of verbs occur at least in one synset and thus could be processed by annotators. Now let us see how difficult this work was.

As described in section 2., there are three types of semantic tags used for annotation: uniliteral synsets, multiliteral synsets, and exceptions. The average numbers of tags of different types which could be selected for one word are in Table 4. A typical annotated word had 3 possible uniliteral and 7 multiliteral synsets in the set of possible tags $\mathrm{T}_p(l)$. Considering only those words with more than one possible tag, they have 3.8 uniliteral synsets and 9 multiliteral ones. Multiliteral synsets appeared almost exclusively in the tag sets of nouns.

POS	Autosemantic		Annotated		Ambiguous	
N	43 315	100 %	30 184	70 %	22 294	51 %
A	16 519	100 %	4 272	26 %	3 107	19 %
V	18 421	100 %	8 444	46 %	4 690	25 %
D	7 710	100 %	0	0 %	0	0 %

Table 3: Absolute and relative word counts per POS.

POS	Annotated words			Ambiguous words		
	U	M	E	U	M	E
N	2.8	9.8	11	3.5	12.1	11
A	3.0	0.1	11	4.7	0.1	11
V	3.8	0.0	11	4.9	0.0	11
All	2.9	6.9	11	3.81	9.0	11

Table 4: Average numbers of possible tags of all types for annotated and ambiguous words with respect to their POS, and in total. (U stands for uniliterals, M for multiliterals, and E for exceptions.)

Although multiliteral synsets appeared in sets $\mathrm{T}_p(l)$ very often, annotators used them rather rarely (0.6 % of words), which is in accordance with their instructions (see section 2.3.). Uniliteral synsets were assigned to 82 % of all annotated words. 17.4 % of words were tagged by an exception. See details for relevant POS in Table 5.

POS	U	M	E
N	85.8	1.2	13.0
V	62.9	0.0	37.1
A	90.9	0.0	9.1
All	82.0	0.6	17.4

Table 5: Average usage (in %) of uniliteral synsets (U), multiliteral synset (M), and exceptions (E) per POS and in total.

3.2. Inter-annotator agreement

All kinds of linguistic annotation are usually performed by more than one annotator. The reason is to obtain more reliable and consistent data. In order to learn this reliability we can measure inter-annotator agreement, a relative number of cases when selections of the annotators were identical. This number gives also evidence of how difficult the annotation is. Manually annotated data is often used to train systems for automatic assigning relevant tags (tagging). Inter-annotator agreement gives an upper bound of accuracy of such systems.

POS	U	UM	UME
N	64.7	65.1	70.9
V	44.5	44.5	63.8
A	71.0	71.0	74.6
All	61.4	61.6	69.9

Table 6: Inter-annotator agreement (in %) on selection of the same: uniliteral synset (U); uniliteral or multiliteral synset (UM); uniliteral or multiliteral synset or exception (UME).

Table 6 shows the inter-annotator agreement measured from various points of view. Basic agreement on selection of uniliteral synsets was 61.4 %. If we consider both uniliteral and multiliteral synsets the inter-annotator agreement increases only by 0.2 %. Overall inter-annotator agreement on all possible types of tags is 69.9 % – almost 1/3 of all processed words are not annotated reliably. This number varies depending on POS: verbs were significantly more difficult to assign a correct uniliteral synset.

Generally speaking, the inter-annotator agreement is relatively low but it does not necessarily imply that annotators had problems to distinguish word meanings. They rather had problems to select the most suitable options that would correspond to their opinion.

According to the CWN, some words occurring in the annotated texts had up to 18 senses (see Table 7). Surprisingly, the inter-annotator agreement does not depend on the degree of ambiguity. It ranged from 15 % to 80 % regardless of the number of possible tags. We can conclude that the size of word tag sets is probably not what causes the low inter-annotator agreement.

Ambiguity	Words	Agreement (%)
1	12809	79
2	11154	75
3	7071	70
4	5466	54
5	2270	56
6	1034	51
7	819	39
8	547	53
9	329	63
10	162	72
11	612	80
12	69	52
13	68	38
14	90	41
15	13	15
16	369	60
17	18	0
18	72	50

Table 7: Overall inter-annotator agreement in relation to degree of word sense ambiguity in the CWN.

3.3. Sense Distribution

In Table 4 we show the average word sense ambiguity in our text according to the CWN. Although this number is relatively high (3 uniliteral plus 9 multiliteral synsets), the real average sense ambiguity of words according annotators is only 1.47. Put differently, all annotated words were assigned only 1.47 different tags in average.

Omitting the cases of disagreement, 62.4 % of all annotated words were always assigned only one synset.

Some more details are given in Table 8.

Amb	N	V	A	Total
1	61.2	56.4	73.2	62.4
2	28.7	28.4	19.5	27.3
3	7.9	10.7	0.7	7.2
4	0.7	4.1	2.6	1.4
5	1.0	0.3	4.0	1.4
6	0.5	0.0	0.0	0.3

Table 8: Word sense distributions in relation to degree of ambiguity.

4. Related experiments

Manual semantic annotation (and also other types of manual annotation) is a time-consuming and therefore expensive process. One way to make this work easier is to use a user-friendly application providing a comfortable environment for annotator's decision making and tag assignment.

Another (but disputable) method is to preprocess unannotated text and automatically tag unambiguous phenomena or prepare the most likely tags for each word occurrence. This approach has two problematic aspects: usually, automatic annotation is not perfect and annotator should review computer's results; but then the annotator can excessively incline to computer's preferred selections.

An example of the latter method is an application of Yarowsky's hypothesis "One sense per collocation" (Yarowsky, 1995) saying that all occurrences of a word in the same collocation have the same meaning. Thus, annotators could process only the first occurrence of each collocation and then this choice would be automatically assigned to all the other occurrences of this collocation.

We obtained a list of significant collocations occurring in the PDT more than 5 times (for the method see (Pecina and Holub, 2002)) and extracted those collocations that appear in our semantically annotated text. There were 3,741 such collocations, 964 unique.

First we have separately validated this hypothesis on the texts annotated by each annotator, and then only on words that were assigned the same tag by both annotators.

Semantic annotation	a)	b)
Annotator A	86.22	77.25
Annotator B	86.42	71.03
Annotator A+B agreement	97.88	96.24

Table 9: Validity (in %) of Yarowsky's hypothesis "One sense per collocation" for words in collocation occuring a) at least once and b) at least twice in the annotated text.

Results of this experiment can be found in the column a) of Table 9. Considering only the reliable annotation from both annotators, the hypothesis is valid for 97.88 % of words and this fully corresponds to Yarowsky's observation on English.

We obtained worse results on all annotated words – taking separately from both annotators – only about 86 %, which however coresponds to the low inter-annotator agreement. The annotators had difficulties to select appropriate tags, consequently they sometimes annotated words with the same meaning with different synsets (low consistency of annotation).

Results in column b) of Table 9 are from experiments using words occurring in the text more than once. They are unsurprisingly lower.

5. Validating and improving the Czech WordNet

Based on our experience with semantic annotation we point out some issues concerning the coverage and quality of the CWN:

- Less than 50 % of nouns, adjectives and verbs in annotated texts occur in the CWN.
- Only 30 % of all nouns, adjectives and verbs were succesfully annotated with a CWN synset.
- Some of very common meanings of frequent words are not covered by the CWN.
- Only 12 % of all CWN synsets were assigned to a word.

These facts give us evidence of (i) uneven distribution of the CWN synsets and (ii) insufficient word coverage.

One of the important outcomes of our work is valuable information which can lead to quality improvement of the CWN and that cannot be obtained in other way. We can provide the authors of the CWN with

- distribution of synset elements for individual synsets;
- distribution of synsets for individual words;
- more or less specific information about missing synsets, percentage and specification of their types (which correspond to the kinds of the exceptions, see Table 1.).

5.1. Comparing two CWN versions

The CWN version 1.2a, which we have been using, has 24,855 synsets, whereas the newly developed version 1.8d has 28,392 synsets. 3537 synsets were added in total, but more importantly many synsets were verified and changed, some wrong synsets were deleted and new once added, some of them based on our feedback.

Valency frames were also added to many verb synsets, which should simplify annotator's decisions and improve consistency of annotations. Most importantly, CWN 1.2a did not include any adverbial synsets. Consequently none of the 7710 adverbs in our texts has been annotated. The version we have been using does not include Czech glosses and not all synsets have an English gloss. Some English glosses also do not fit the Czech synsets. In contrast, CWN 1.8d includes many Czech glosses that fit the synsets and also includes example sentences.

We expect that using the new CWN version will lead to an improvement of the inter-annotator agreement by eliminating some sources of common errors. However, the high granularity of the WordNet senses, which also often causes inter-annotator disagreement, is a problem sui generis.

6. Discussion on semantic tags and the inter-annotator agreement

We have mentioned two main issues related to our work: insufficient quality of the CWN and poor inter-annotator agreement. The latter one can be tackled by changing our annotation methodology.

As mentioned in the introduction, one of the fundamental questions is what system of semantic tags (i.e. $\bigcup \mathrm{T}_p(l_i)$) should be used for the lexico-semantic annotation. This is closely related to the problem of granularity.

High granularity of the WordNet senses, i.e. the fact that words in the WordNet often have too many senses with only fine distinctions, is probably the most usual argument against the WordNet.

To reduce the impact of this undesirable granularity we can allow the annotators:

(i) assign more than one proposed synset or

(ii) assign a hypernym of a proposed synset.

The option (i) would probably worsen the inter-annotator agreement on synsets and exceptions (currently 69.9 %), yet it would also reduce the number of words annotated with exceptions (24.6 %), so the impact on agreement on synset selection is unclear. The option (ii) states the question how general hypernyms we should allow to be used as semantic tags (since the more general the tag the less information provided).

7. Summary

Our semantic annotation of the PDT has two major applications:

1. Lexico-semantic tags are a new kind of labels in the PDT and will become a substantial part of a complete resource of training data, which can be exploited in many fields of NLP.

2. The process of annotation provides a substantial feedback to the authors of the CWN and significantly helps to validate and improve its quality.

   To our best knowledge, the only comparable annotated corpus that can be used for WordNet validation is English SemCor (Landes et al., 1998), cf. also (Stevenson, 2003); as for the other languages, our project seems to be unique.

8. Acknowledgments

This work has been supported by the Ministry of Education, project Center of Computational Linguistics (project LN00A063). Here we would also like to express our grateful thanks to Karel Pala, the head of the CWN development team, for his kind cooperation, helpful commentaries, and fruitful suggestions.

9. References

BalkaNet, 2004. Project website. http://www.ceid.upatras.gr/Balkanet/.

EuroWordNet, 2004. Project website. http://www.illc.uva.nl/EuroWordNet/.

Fellbaum, Christiane (ed.), 1998. *WordNet, An Electronic Lexical Database*. Cambridge: MIT Press, 1st edition.

Hajič, Jan, Eva Hajičová, Petr Pajas, Jarmila Panevová, Petr Sgall, and Barbora Vidová-Hladká, 2001a. Prague dependency treebank 1.0 (Final Production Label). Published by Linguistic Data Consortium, University of Pennsylvania.

Hajič, Jan and Václav Honetschläger, 2003. Annotation lexicons: Using the valency lexicon for tectogrammatical annotation. *Prague Bulletin of Mathematical Linguistics*, (79–80):61–86.

Hajič, Jan, Barbora Vidová-Hladká, Eva Hajičová, Petr Sgall, Petr Pajas, Veronika Řezníčková, and Martin Holub, 2001b. The current status of the prague dependency treebank. In R. Mouček K. Taušer V. Matoušek, P. Mautner (ed.), *TSD2001 Proceedings*, *LNAI 2166*. Berlin Heidelberg New York: Springer-Verlag.

Landes, Shari, Claudia Leacock, and Randee I. Tengi, 1998. Building semantic concordances. In Christiane Fellbaum (ed.), *WordNet, An Electronic Lexical Database*, chapter 8. Cambridge: MIT Press, 1st edition, pages 199–216.

Pavlík, Martin, 2002. *Semantic Disambiguation of Terms in DIS (in Czech)*. Master's thesis, Faculty of Mathematics and Physics, Charles University, Prague.

Pecina, Pavel and Martin Holub, 2002. Semantically significant collocations (in czech). Technical Report TR-2002-13, UFAL/CKL.

Smrž, Pavel, 2003. Quality Control for Wordnet Development. In Petr Sojka, Karel Pala, Pavel Smrž, Christiane Fellbaum, and Piek Vossen (eds.), *Proceedings of the Second International WordNet Conference—GWC 2004*. Brno, Czech Republic: Masaryk University.

Stevenson, Mark, 2003. *Word Sense Disambiguation: The Case for Combinations of Knowledge Sources*. CSLI Studies in Computational Linguistics. Stanford, California: CSLI Publications.

Vossen, Piek (ed.), 1998. *Introduction to EuroWordNet*. Computers and the Humanities. Kluwer Academic Publishers, 1st edition.

Yarowsky, David, 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Meeting of the Association for Computational Linguistics*. Cambridge, MA.

Generalisations over Corpus-induced Frame Assignment Rules

Anette Frank

German Research Center for Artificial Intelligence
DFKI GmbH
Saarbrücken, Germany
frank@dfki.de
and
Computational Linguistics Department
Saarland University
Saarbrücken, Germany
frank@coli.uni-sb.de

Abstract

In this paper we discuss motivations and strategies for generalising over instance-based frame assignment rules that we extract from frame-annotated corpora. Corpus-induced syntax-semantics mapping rules for frame assignment can be used for automatic semantic role labelling of unparsed text, but further, to extract linguistic knowledge for a lexical semantic resource with a general syntax-semantics interface. We provide a data analysis of a comprehensive rule set of corpus-induced frame assignment rules, and discuss the potential of applying different types of generalisations and filters, to obtain a uniform extended data set for the extraction of linguistic knowledge.

1. Introduction

Various research groups are currently concerned with the creation of large-scale lexical semantic resources that provide information about predicate-argument structure. The Berkeley FrameNet project (Baker et al., 1998), following Fillmore's theory of frame semantics (Fillmore, 1976), is building a large semantic lexicon, including the definition of frames and semantic roles, and a corpus of manually annotated sentences. A strictly corpus-based approach is carried out with 'PropBank' (Kingsbury et al., 2002) – a manual semantic role annotation on top of the PennII Treebank.

There are first approaches for learning stochastic models for semantic role assignment from annotated corpora; e.g. (Gildea and Jurafsky, 2002; Fleischman et al., 2003). Probabilistic models for semantic role assignment systems will eventually be used for automated semantic annotation in NLP applications, but they can also be used, in a bootstrapping architecture, to learn increasingly refined probabilistic models from extended training sets, by application of meta-learning strategies, such as active learning.

The current models for stochastic role assignment models are essentially corpus-based. Yet, besides the development of systems for automated role labelling, there is also interest in a general lexical semantics resource that can be formalised and integrated into alternative NLP systems.

In our work we investigate techniques for automated induction of rules for automatic semantic role assignment from semantically annotated corpora.[1] In this paper we discuss strategies for generalising over corpus-induced frame assignment rules. We provide a data analysis of a comprehensive rule set, and discuss the potential of applying different types of generalisations and filters, to obtain a uniform extended data set – for semi-automatic acquisition of new training data, and the extraction of linguistic knowledge.

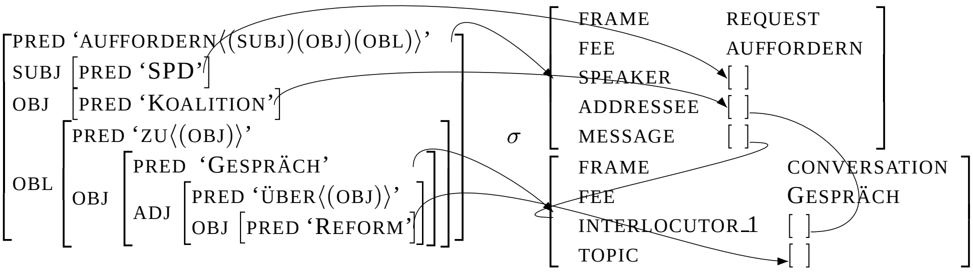
[1] The work is conducted in the context of the SALSA project; see (Erk et al., 2003) and http://www.coli.uni-sb.de/lexicon.

2. Deep syntactic analysis for semantic role labelling

Since semantic role assignment is based on a syntactic annotation layer, automated processing for semantic role assignment on unparsed text requires an interface between a syntactic analyser and the targeted semantic annotation. Current competitions explore the potential of shallow parsing as a basis for semantic role labeling. However, (Gildea and Palmer, 2002) have emphasised the role of deeper syntactic analysis for semantic role assignment. We follow this line, and explore the potential of deep syntactic analysis for semantic role labelling, choosing Lexical Functional Grammar (Bresnan, 2001) as underlying syntactic framework.

In a first study, (Frank and Erk, 2004) discuss advantages of semantic role assignment on the basis of functional syntactic analyses as provided by LFG parsing, and present an LFG projection architecture for frame semantics. In this architecture, frames are projected from f-structure representations, as displayed in Figure 1. The semantic projection is defined by lexical entries of frame evoking predicates, which map f-structure nodes for grammatical functions to frame semantic roles in a frame semantics projection. The projection of frames in context can yield partially connected frame structures. In Figure 1, *Gespräch* projects to the MESSAGE role of REQUEST, but it also introduces a frame of its own, CONVERSATION. Thus the CONVERSATION frame, by coindexation, is an instantiation, in context, of the MESSAGE of REQUEST. Figure 2 displays how these mappings can defined in a classical LFG co-description projection architecture, by use of functional descriptions; see (Frank and Erk, 2004) for details.

As an alternative to the co-description approach, we implemented frame projection in a description-by-analysis (DBA) architecture. In co-description, semantics projection is tightly intervowen with grammar definitions and the parsing process. The DBA approach, by contrast, is more



SPD fordert Koalition zu Gesprächen über Reform auf.
’SPD requests coalition to talk about Reform’

Figure 1: LFG projection architecture for Frame Annotation

auffordern V,
$(\uparrow$PRED$)$=‘AUFFORDERN$\langle(\uparrow$SUBJ$)(\uparrow$OBJ$)(\uparrow$OBL$)\rangle$’
...
$(\sigma(\uparrow)$ FRAME$)$ = REQUEST
$(\sigma(\uparrow)$ FEE$)$ = $(\uparrow$ PRED FN$)$
$(\sigma(\uparrow)$ SPEAKER$)$ = $\sigma(\uparrow$ SUBJ$)$
$(\sigma(\uparrow)$ ADDRESSEE$)$ = $\sigma(\uparrow$ OBJ$)$
$(\sigma(\uparrow)$ MESSAGE$)$ = $\sigma(\uparrow$ OBL OBJ$)$

Figure 2: Frame projection in lexical entry (co-description)

```
pred(X,auffordern),
subj(X,A), obj(X,B), obl(X,C), obj(C,D)
==>
+’s::’(X,SemX), +frame(SemX,request), +fee(X,auffordern)
+’s::’(A,SemA), +speaker(SemX,SemA),
+’s::’(B,SemB), +addressee(SemX,SemB),
+’s::’(D,SemD), +message(SemX,SemD).
```

Figure 3: Frame projection rule (as a transfer rewrite rule)

modular. Here, frame projection rules apply to completed f-structure representations produced by the LFG parser.

The DBA approach is realised by use of a transfer rewrite system.[2] The system allows the definition of rewrite rules that apply to an f-structure context and introduce, on their right-hand side, a semantic projection for frames: the specific FRAME evoked by the frame evoking element (FEE), i.e., the triggering predicate in the f-structure. The rules further define the projection of frame-specific semantic roles from particular local (or sometimes non-local) functional paths (such as SUBJ, OBJ, OBL OBJ), starting from the f-structure node of the frame evoking predicate. The example of Figure 3 is equivalent to the co-description variant in Figure 2, and thus yields the same frame projection, displayed in Figure 1.

3. Corpus-based induction of an LFG–frame semantics interface

(Frank and Semecky, 2004) present a method for the automatic induction of LFG-based frame assignment rules from semantically annotated corpora. This method was first applied to the SALSA corpus (Erk et al., 2003), a German newspaper corpus enriched with frame semantic annotations. The SALSA annotations are built on, and extend the syntactically annotated TIGER corpus (Brants et al., 2002). In (Frank and Semecky, 2004) the frame semantic annotations of the SALSA/TIGER corpus were ported to a ’parallel’ TIGER corpus of corresponding LFG f-structure analyses (Forst, 2003). Figure 3 displays an example of a frame assignment rule that was extracted from the resulting frame-extended LFG SALSA/TIGER corpus. (Frank and Semecky, 2004) further present first experiments to apply the resulting computational syntax-semantics interface for frame semantics in an LFG parsing architecture, using a wide-coverage LFG grammar of German.[3]

A similar architecture for corpus-based induction of a frame semantics interface was recently developed in the context of the Senseval-3 task on semantic role labeling for English.[4] Here, the basis was a subset of the English frame annotated sentences of the FrameNet project (Baker et al., 1998), and the wide-coverage stochastic English LFG grammar developed at Parc (Riezler et al., 2002). The grammar provided a ’parallel’ LFG corpus with most-probable analyses for the annotated sentences. Similar to the methods applied for SALSA/TIGER, we port the frame annotations to the LFG parsed sentences, and extract frame assignment rules that can be applied to new sentences in an LFG parsing-transfer architecture.

In both scenarios, the next steps towards an automated system for LFG-based frame assignment involve the design of probabilistic models to select the most probable frame assignments from the choice of possible assignments that are generated by application of the corpus-induced frame assignment rules proper – as well as generalisations of these rules, which account for unseen configurations.

Besides the development of a probabilistic semantic role labelling system, the aim of the SALSA project is to acquire generalised linguistic knowledge, i.e. a frame semantic lexicon with a well-defined syntax-semantics interface, from a large frame-annotated German corpus. It is also in view of this more ambitious aim that we are concerned with a closer inspection of the corpus-induced syntax-semantic mapping rules for frame assignment.

[2] The system comes as a module of the grammar development platform XLE (http://www2.parc.com/istl/groups/nltt/). It was designed and implemented by Martin Kay (Xerox Parc) for a Machine Translation prototype; see (Frank, 1999). Recent enhancements to the system were realised by Richard Crouch.

[3] The German LFG grammar is being developped at the IMS, University of Stuttgart.

[4] This was done in joint work with Katrin Erk and Ulrike Baldewein.

4. Generalisations over corpus-induced frame assignment rules

In this section we discuss motivations and strategies for generalising over sets of instance-based frame assignment rules that we extract from frame annotated corpora. In Section 5 we provide a quantitative evaluation of the rule set we extracted from the English FrameNet corpus sentences that were provided as training data in the Senseval-3 semantic role labeling task.

On the basis of this evaluation, Section 6 reviews the potential of the proposed generalisations over corpus-induced frame assignment rules: for abstraction of a general linguistic knowledge base, and for the targeted acquisition of training material in an active learning scenario, to develop increasingly refined stochastic models for frame assignment on the basis of continuously extended training corpora.

4.1. Motivations

Corpus-based extraction of frame assignment rules is confronted with two problematic issues: quality and coverage.

Quality It is well-known from treebank-based grammar induction that corpus-based acquisition and formalisation of linguistic knowledge is confronted with the problem of noise in the data. In our case, noise can be imported from various sources: (i) mistakes and inconsistencies in the manual syntactic or semantic annotations; (ii) problems in the automated mapping from corpus specific syntactic annotation schemes to the LFG f-structure encoding; (iii) problems in the extraction of frame assignment rules from the frame-enriched LFG corpora, and finally (iv) parsing errors or missing coverage of the underlying LFG grammars.

Coverage The problem of coverage is specific to the nature of lexical semantic corpus annotation. Lexical semantic annotation is confronted with a severe sparse data problem, since we may not encounter a large-enough variety of predicates in specific senses and constructions within manageable sizes of manually annotated corpora. E.g., while the SALSA corpus is comparable, in size, to the Penn Treebank, of the 4185 verbs (types), 1457 (34.81%) occur only once, and 3307 (79.02%) occur with frequency 1-10.

This sparse data problem is even more serious if we consider, as we do in SALSA, semi-automatic annotation of new corpus instances and learning of a principled syntax-semantics interface from corpus annotations: since there are multiple sources of noise in the data (see above), we may miss out a number of (already rare) corpus instances.

'Filling Gaps' In order to address these problems, we investigate the potential of various generalisations or 'filters' over instance-based rule sets, which can be used to identify and 'fill gaps' in the base of corpus samples.

Targeted acquisition of new corpus data to fill these gaps will enable the extraction of more homogeneous syntax-semantics mapping constraints for the final semantic lexcion resource. Most importantly, though, this way of acquiring new corpus material can be used to support active learning techniques, by providing a selection of 'informative' novel annotation instances, i.e. novel training instances that are promising candidates for improving stochastic models for automated frame assigment.

In the following we present different aspects of generalisations over corpus-based frame annotation instances. These range from linguistically motivated generalisations to distributional criteria regarding the densitiy of annotation samples for different classes of annotation events.

4.2. Linguistic generalisations

LFG f-structures provide a level of representation that abstracts away from surface-syntactic variations that are irrelevant for frame assignment (such as word order, long-distance phenomena or coordination). On the other hand, f-structures are genuine *syntactic representations* that differ from semantic predicate argument structures in that they do represent *functional syntactic* variants that are not distinguished in the semantic representation.

Diatheses A prominent example is the active-passive diathesis. Due to the sparseness of data we encounter with current sizes of annotated corpora, we may or may not encounter both active and passive constructions for a given frame evoking predicate and its specific semantic role configuration. This 'gap' in the training data may be compensated by the use of a greater variety of features in stochastic modelling for role assignment, but the lack of generalisation will be problematic for automated methods in building a final lexicon resource from the corpus-induced rule sets.

In order to fill such gaps in the training corpus we can generate missing active or passive variants of frame projection rules, and apply them to candidate sentences extracted from unparsed corpora. Sentences that receive the targeted annotation can be presented to annotators for acknowledgement, and – on approval – can be added to the set of training samples. On the basis of the extended corpus, we can extract more general frame assignment rules, with disjunctive constraints to account for active and passive constructions (see Figure 4). This will lead to a more homogeneous frame semantic lexicon resource, and will increase the coverage of automated frame assignment models when applied to unseen text.

```
pred(X,auffordern),
{ passive(X,−), subj(X,A), obj(X,B)
| passive(X,+), subj(X,B), obl_ag(X,A) },
obl(X,C), obj(C,D)    ==>
+'s::'(X,SemX), +frame(SemX,request), +fee(X,auffordern)
+'s::'(A,SemA), +speaker(SemX,SemA),
+'s::'(B,SemB), +addressee(SemX,SemB),
+'s::'(D,SemD), +message(SemX,SemD).
```

Figure 4: Generalisation over active-passive diathesis

Non-local frame element assignments Another source of gaps in the annotation samples are frames that occur in non-local syntactic contexts. In case the evoking predicate is not, alternatively, found in a local syntactic context, the extracted rules will not be able to annotate the same frame in a more general, local context.

The LFG formalism provides a significant capacity for argument localisation (in long-distance, coordination, raising and control constructions). However, there are constructions where arguments cannot be localised on syntactic grounds. A classical example are constructions involving anaphoric control, such as gerunds.

In example (1), from the FrameNet data, the THEME role of the frame evoking predicate, *disappear*, was annotated as the passive SUBJ of the main clause, while the FEE is contained in the clausal ADJUNCT phrase (cf. Figure 5), while the local subject of the adjunct clause is a non-overt pronominal SUBJ. The functional path from the f-structure node of the frame evoking predicate to the f-structure of the THEME role is inside-out and non-local: ((ADJUNCT $ ↑) SUBJ).[5] Starting out from the local f-structure ↑ of the frame evoking element *disappear* the path leads inside-out via the set-valued ADJUNCT function to the dominating node (ADJUNCT $ ↑). From this node, the path leads outside-in via the function SUBJ to the f-structure of *sword*.

(1) *The Solland Sword was lost for many years, having disappeared during the destruction of Solland by Gorbad Ironclaw's Orcs .*

Similar to the active-passive distinction, in cases were our rule set does not comprise the corresponding local variant of the identified non-local frame assignment rule, we can generate an alternative local assignment rule, here looking for a local SUBJ of the frame evoking predicate in active voice. We can use such rules to automatically annotate sentences from unparsed corpora, again presenting the targeted instances to annotators for acknowledgement. With this method, we systematically extend the set of general, local frame assignment rules.

The identified patterns of typical non-local path descriptions can, moreover, serve as a 'functional bridge' in non-local annotation contexts. That is, we can state generic frame assigning rules that account for such 'bridging' non-local functional paths for frame element assignment. These can be triggered as fallback rules, to identify novel annotation instances in non-local configurations.

4.3. Abstractions from frame assignment rules

Finally, we can apply similar methods for acquiring novel annotation instances, by analysing the distribution of role assignments for a given frame, abstracting over the specific frame evoking elements that were found to invoke the frame. That is, from the FEE-specific annotations in the corpus we abstract classes of 'non-lexicalised frames' with syntactic mapping constraints. We can apply these generic frame assignment rules to novel corpus instances, where we condition the application to the set of FEEs that can trigger the given frame. We will further experiment with frame assignment rules that define clusters (instead of specific instances) of role-preposition correspondences.

5. Investigating corpus-induced samples of frame assignment rules

In this section we provide a data analysis of LFG frame assignment rules that we acquired from frame-annotated corpora. For this analysis, we concentrate on the rule set we induced from the FrameNet corpus data (Section 3).[6]

[5] ADJUNCTs are represented as set-valued f-structures. In functional path descriptions, reference to an element of a set is made by the path symbol '$' for 'in_set'.

5.1. Coverage

Due to the lexicographic approach of the FrameNet project, the English FrameNet data can be assumed to be rather homogeneous and balanced as to the quantitative distribution of frame evoking predicates and their constructional variants. By contrast, the mapping from the FrameNet annotations to LFG representations is currently based on the most probable analysis of the English LFG grammar, which may still feature wrong selections. Moreover, a number of frame element bracketings in the FrameNet annotations did not map to a unique f-structure node in the corrsponding LFG analysis, and hence did not yield frame assignments in the LFG-based frame-enriched corpus.[7]

These (interrelated) challenges are reflected in the coverage figures of Table 1, with 90.19% of sentences that receive frame element annotations, yet only 67.41% coverage at the level of overall frame element assignments, measured against the target annotations in the original FrameNet corpus.[8] We obtain 1.77 frame element assignments per sentence in average, against 2.33 in the FrameNet data.

	abs no	in %	avg/s
s(entences)	24274	100	-
s with extracted fpaths	21893	90.19	-
target fes	57325	100	2.33 fe/s
extracted fpaths for fes	38643	67.41	1.77 fe/s

Table 1: Coverage: extracted fe-assignment paths

Table 2 gives an overview of the distribution of different *types* of functional path equations (fpaths) that lead from (the f-structure of) the frame evoking element (FEE) to (the f-structure of) its frame element (or semantic role) – for distinct FEEs, or abstracting over the FEE of a given frame. As expected, taking the assigned semantic roles into account (in fpath-role) leads to a greater variety of distinct fpath-role assignments, both for FEE-specific and – proportionally higher – for frame-specific assignment paths.

per FEE	all	min	max	avg.
fpath	11465	1	67	8.10
fpath-role	13477	1	79	9.52

per Frame	all	min	max	avg.
fpath	4211	22	292	105.28
fpath-role	5497	24	385	137.43

Table 2: Distribution of fpath types (per FEE, per Frame)

5.2. Active-passive diathesis

The above figures are not really informative as to how complete the distribution of the acquired frame assignment

[6] As the SALSA corpus is still under construction, our rule set is considerably smaller, and relatively unbalanced over frames. A data analysis on the basis of the more balanced and sufficiently varied FrameNet data therefore seemed to prove more indicative.

[7] We will further improve the mapping procedures from corpus annotations to LFG parses, so we expect the figures to improve.

[8] We lost 284 sentences of the original corpus that we could not map to f-structures for technical reasons. These sentences have not been subtracted from the FrameNet data counts in Table 1.

rules is for specific syntactic variants (i.e. fpath-role assignments) over the different classes – whether FEEs or frames.

A closer look is provided by Table 3, for the distributional patterns of fpath-role assignments in active-passive alternations. Almost half of the verb types do only appear in either active or passive constructions - and it is not clear from the counts whether there are missed-out alternations, or whether there are genuinely non-alternating verbal predicates.[9] Moreover, as is seen on the right-hand side, the proportion of local (subj,obj, obl_ag) fpaths found in active and passive constructions is very low (11.89–15.09% for active, and 12.09-20.48% for passive constructions).

Table 4 views the active-passive alternation from a different angle, by looking at passive-invariant semantic roles, i.e. the roles whose functional path assignment is (for given a frame, or a given FEE) never affected by the active-passive alternation. The frequency of such invariant fpath-role pairs (i.e. identical fpath-role assignments in a passive and active constructions) is very low.

verbs (types)		all vs. local fpaths			
		active		passive	
nonfragmented	590	all fp	7118	all fp	3028
active/passive	321	subj	1072	subj	620
passive only	24	obj	846	obl_ag	366
active only	245	obl_ag	2	obj	4

Table 3: Active-passive diathesis: distribution and fpaths

	all	passive-invariant	
FEE-fpath-role	4827	224	4.64%
Frame-fpath-role	2210	206	9.32%

Table 4: Passive-invariant fpath-role assignments

Closer inspection of the data underlying Table 4 shows that many fpath-role pairs are wrongly classified as passive-invariant due to a rare active or passive occurance that is produced by noise in the data (e.g. a wrong parse). Typical examples of such misclassifications are cases like *mumble*, occurring with SUBJ-SPEAKER assignment in both active and passive, yet with a distribution of 28 vs. 3. While these are rather clear weighted distributions, there are cases where the distribution is more unmarked (e.g. *murder* with a SUBJ-VICTIM distribution of 1 vs. 3 active vs. passive occurrences), and thus become difficult to distinguish from correct, but still infrequent distributions of correct instances of passive-invariant fpath-role pairs, in particular adjuncts.

This kind of noise in the data does clearly not only affect the identification of passive-invariant fpath-role assignments, but also the identification of active-passive alternating verbs in Table 3. That is, we observe a high number of instances that are identified as active-passive alternating, but on the basis of erroneous active or passive occurrences.

Filtering noise In order to filter such misclassifications, we computed a confidence weight for fpath-role assignments on the basis of their proportional distribution in passive vs. active assignments. The weight for a given fpath-role assignment in an active or passive construction, respectively, is computed by its relative frequency wrt. the overall number of fpath-role assignments in the respective voice, for a given FEE (or frame). This value we then used to experiment with different thresholds for computing counts on the active-passive distribution of fpath-role assignments.

As seen in Table 5, this filter reduces the number of active-passive alternating verb (type)s, by filtering erroneous instances from the base of counts. While the number of instances drastically reduces, only a small number of verb types are eliminated from consideration. On the other hand, the proportion of correct *local* functional subcategorisation paths in the retained set of fpath-role assignments increases with the threshold. For active verbs, the culmination point for positive filtering effects seems to be around .6. For passive verbs, we obtain the best filtering effect for subj with threshold .6, and for obl_ag with .7. Thus, the filters eliminate erroneous or otherwise rare occurrences.

verbs (types)		all vs. local fpaths					
		active			passive		
nonfrag	590	all fp	7118	in %	all fp	3028	in %
act/pass	321	subj	1072	15.06	subj	620	20.48
pass only	24	obj	846	11.89	obl_ag	366	12.09
act only	245	obl_ag	2		obj	4	
thresh .6	581	all fp	1470		all fp	741	
act/pass	309	subj	386	26.26	subj	211	28.48
pass only	25	obj	332	22.59	obl_ag	167	22.58
act only	247	obl_ag	1		obj	1	
thresh .7	580	all fp	1470		all fp	677	
act/pass	307	subj	386	26.26	subj	166	24.52
pass only	24	obj	332	22.59	obl_ag	160	23.63
act only	249	obl_ag	1		obj	1	

Table 5: Filters on active-passive diathesis

As a filter of noise in the computation of passive-invariant fpath-role assignments, we compute a weight for each fpath-role pair based on the relative frequency of passive as opposed to active occurrences (per FEE or frame). As seen in Table 6, this results in a radical reduction of passive-invariant fpath-role assignments, since many fpath-role occurrences do not show a sufficiently unbalanced distribution over active and passive, and thus do not exceed the threshold. This holds in particular for adjuncts and obliques which are clearly non-alternating functions. Selected application of the filter to functions that participate in the active-passive alternation, such as SUBJ and OBJ, shows moderate filtering effects that produce satisfactory results.[10]

threshold (.6)	filter on all fpaths		filter on subj/obj	
FEE-fpath-role	141/224	62.95	54/71	76.06%
frame-fpath-role	157/206	76.21	69/82	84.15%
threshold (.7)				
FEE-fpath-role	86/224	38.39	40/71	56.34%
frame-fpath-role	110/206	53.40	52/82	63.41%

Table 6: Filters on passive-invariant fpath-roles

[9]We only consider verbs whose functional context is not affected by fragmentary parses (nonfragmented).

[10]We will further experiment with weights that are parameterised for specific functional roles and patterns of argument structure variation, along the lines of (Merlo and Stevenson, 2001).

	all (w/o fragmented)				outside-in				inside-out (and outside-in)			
	abs	in %	types	in %	abs	in %	types	in %	abs	in %	types	in &
all lengths	38034	100/100	1582	100/100	31568	83/100	431	27/100	6466	17/100	1151	73/100
length 1	27567	72.48	97	6.13	27567	87.33	97	22.51	0	0.00	0	0.00
length 2	5967	15.69	218	13.78	3577	11.33	75	17.40	2390	36.96	143	12.42
length 3	3460	9.10	610	38.56	314	0.99	158	36.66	3146	48.65	452	39.27
length 4	820	2.16	456	28.82	63	0.20	57	13.23	757	11.71	399	34.67
length 5	187	0.49	169	10.69	47	0.15	44	10.21	140	2.17	125	10.86
length 6	29	0.08	28	1.77	0	0.00	0	0.00	29	0.45	28	2.43
length 7	4	0.00	4	0.25	0	0.00	0	0.00	4	0.00	4	0.35

Table 7: Path types

outside-in		inside-out (and outside-in)	
path	frequency	path	frequency
↑	9213	((OBJ ADJUNCT $ ↑) OBJ)	548
(↑ SUBJ)	5030	((OBJ $ ↑) SUBJ)	497
(↑ SPEC POSS)	3228	((ADJUNCT $ ↑) SUBJ)	240
(↑ OBJ)	3176	((SUBJ ADJUNCT $ ↑) SUBJ)	228
(↑ ADJUNCT)	2835	(($ ↑) $)	195
(↑ MOD)	2556	(($ ADJUNCT $ ↑) $)	160
(↑ ADJUNCT_OF)	1001	(($ OBJ ↑) $)	135
(↑ OBL_AG)	499	((ADJUNCT $ ↑) $)	133
(↑ ADJUNCT_IN)	314	(($ OBJ ↑) SUBJ)	123
(↑ OBL_WITH)	297	((XCOMP ADJUNCT $ ↑) XCOMP)	121

Table 8: Top ten frequent path types

5.3. Local and non-local frame assignment paths

Another issue that affects the homogeneity of the corpus-induced syntax-semantics interface for frame semantics is the nature and variety of functional paths that are extracted from frame-annotated sentences. As seen in Table 5, only a small proportion of fpaths involved in active-passive alternations is found to be local, i.e. involve a locally subcategorised SUBJ, OBJ, or OBL_AG grammatical function.

Path types Table 7 gives an overview of the distribution of path lengths in the fpath assignments we extracted from the FrameNet data. With increasing path length, the frequency of occurrences decreases, while the variety of fpath types increases. We further differentiate between *outside-in* paths (the path leads from the f-structure of the FEE downwards to an embedded f-structure node) and *inside-out* paths (leading from the FEE inside-out and outside-in to an f-structure node that is not dominated by the FEE).

Infrequent path occurrences are susceptible of noise in the data or are not expected to contribute valuable information in stochastic training. So, both for the extraction of linguistic knowledge and for stochastic training, we could set a frequency-based threshold on the length of paths to consider. A general cut-off for all paths to length $\leq$3 retains 97.27% of the coverage, and yields a reduction of path types to 58.47%. However, the frequency distributions for inside-out and outside-in path types are quite different. Also, the variety of fpaths is significantly higher for inside-out paths (73%) as opposed to outside-in paths (27%). A selective cut-off, restricting path length to $\leq$2 for outside-in, and $\leq$3 for inside-out paths leaves 96.44% coverage and 48.48% of path types; including path length 3 for inside-out yields 98.43% coverage with 73.70% of the path types.

As seen in Figure 8, inside-out fpaths of length 3 occur most frequently among inside-out fpaths, and two of them range among the top ten frequent fpaths overall.[11]

Thus, as an alternative to a cut of data based on path length, a cut-off on the basis of frequencies for individual fpaths could be more adequate for cautious filtering.

Generalising over non-local assignment paths Among the top ten inside-out fpaths we also find the non-local fpath described in Section 4.2. This fpath occurs with 135 verb types (210 tokens). For 4 verb types we do not find a corresponding local fpath in the extracted rule set. However, there are 501 verbs (4385 tokens) with local subject fpaths, while we have seen the non-local configuration only for 135 types. These remaining 370 types can be caught by generalised fall-back rules for the non-local variant, if in new corpus data they occur in the identified non-local context.

On the other hand, there are less frequent non-local paths that account for general syntactic configurations that we may encounter in new data, such as the coordination construction in (2). Here the FEE *occupants*, which triggers the RESIDENCE frame, takes as its LOCATION role the coordinated adjunct PP *of .. flats*. The coordinated adjunct is attached high to the coordinated noun heads *owners and occupants*. This high attachment is reflected in the f-structure, which differs from non-coordination.[12] The fpath we obtain is (($ ↑) ADJUNCT), crossing coordination inside-out.

(2) *give greater protection to the [[owners and occupants] [of shops, commercial premises, houses and flats]]*

We identified 97 instances of this pattern, for 38 predicates in 13 frames and for 16 roles. The corresponding local

[11] The element relation of set-valued ADJUNCTs does not contribute to the path length, but it does for coordination: (($ ↑) $).

[12] The grammar does not distribute ADJUNCTs in coordination.

fpath (adjunct_of) occurs in 1013 instances of 340 predicates in 33 frames and for 62 roles. Again, we can provide alternative local/non-local annotation rules, to account for non-local configurations that are not in the data set.

6. Implications

There are several conclusions that can be drawn from the data analysis in Section 5.

Filtering noise In order to be able to extract a lexical semantic resource with a general syntax-semantics interface from corpus annotations, we must acquire sufficiently large and varied corpus samples. We have seen for various examples that reliable generalisations can only be obtained if noise in the data can be eliminated by various kinds of frequency-based filters. Where appropriate, these should be combined to yield reliable confidence measures.

Targeted data acquisition On the basis of quantititive evaluations and an automated frame-assignment architecture, we can identify candidate sentences in unparsed text to 'fill gaps' in the pruned set of annotations, or to provide additional 'evidence' in cases of indiscriminative data counts. Thus, we can pursue a process of targeted data acquistion in an effective, and semi-automated way.

Rule generalisations As seen in Table 7, and in the analysis of the active-passive diathesis, there is a great variety of fpaths in the mapping to semantic roles, due to constructional varieties in the underlying corpus sentences. We identified related local and non-local fpath assignments, and more of these need to be established by data inspection. For such regular alternations, we can identify gaps for local variants, which we can fill with newly acquired data, for the extraction of a frame semantic lexicon with well-defined syn-sem mappings.

For the purpose of active learning techniques in stochastic model building, regular alternations and constructional variants in frame projection can be modeled by generalising frame assignment rules to account for the respective variants. This extends the coverage of automated frame assignment, and the stochastic models that are built on top of it.

Corpus-driven vs. lexicographic The SALSA project – a primarily corpus-driven annotation effort – will be confronted with additional challenges. In contrast to FrameNet data, assembled in a lexicographic effort, the TIGER corpus is less balanced and features novel annotation problems (idioms, support constructions, or metaphors). The need to acquire additional data by generalisations over existing annotations will be even more important in this scenario, to extend the base of annotations in a targeted way.

However, the TIGER annotations will provide a significant boost, for construction of an initial set of frame assignment rules and models for probabilistic selection. Acquisition of novel informative training data can be steered by data analysis and generalisations over existing annotations.

Interplay of statistical and symbolic techniques In sum, we propose to combine statistical techniques with a symbolic syntax-semantics interface for frame assignment, to support both the targeted acquisition of 'informative' training data and the extraction of a semantic lexicon with a well-defined syntax-semantics interface.

Acknowledgements The research reported here was conducted in the project LFG4SALSA, a cooperation project of the German Research Center for Artificial Intelligence, DFKI Saarbrücken and the Computational Linguistics Department of the University of the Saarland at Saarbrücken.

7. References

Baker, C. F., C. J. Fillmore, and J. B. Lowe, 1998. The Berkeley FrameNet project. In *Proceedings of COLING-ACL 1998*. Montréal, Canada.

Brants, S., S. Dipper, S. Hansen, W. Lezius, and G. Smith, 2002. The TIGER Treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*. Sozopol, Bulgaria.

Bresnan, J., 2001. *Lexical-Functional Syntax*. Oxford: Blackwell Publishers.

Erk, K., A. Kowalski, S. Padó, and M. Pinkal, 2003. Towards a Resource for Lexical Semantics: A Large German Corpus with Extensive Semantic Annotation. In *Proceedings of the ACL 2003*. Sapporo, Japan.

Fillmore, Charles J., 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, volume 280.

Fleischman, M., N. Kwon, and E. Hovy, 2003. Maximum entropy models for FrameNet classification. In *Proceedings of EMNLP'03*. Sapporo, Japan.

Forst, M., 2003. Treebank Conversion – Establishing a testsuite for a broad-coverage LFG from the TIGER treebank. In A. Abeillé, S. Hansen, and H. Uszkoreit (eds.), *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC '03)*. Budapest.

Frank, A., 1999. From Parallel Grammar Development towards Machine Translation. A Project Overview. In *Proceedings of Machine Translation Summit VII "MT in the Great Translation Era"*. Singapore.

Frank, A. and K. Erk, 2004. Towards an LFG Syntax–Semantics Interface for Frame Semantics Annotation. In A. Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing*, Lecture Notes in Computer Science, Vol. 2945. Springer Verlag, Heidelberg.

Frank, A. and J. Semecky, 2004. Corpus-based Induction of an LFG Syntax-Semantics Interface for Frame Semantic Processing. *to appear*.

Gildea, D. and D. Jurafsky, 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3).

Gildea, D. and M. Palmer, 2002. The Necessity of Parsing for Predicate Argument Recognition. In *Proceedings of ACL'02*. Philadelphia, PA.

Kingsbury, P., M. Palmer, and M. Marcus, 2002. Adding semantic annotation to the Penn TreeBank. In *Proceedings of the HLT Conference*. San Diego.

Merlo, P. and S. Stevenson, 2001. Automatic Verb Classification Based on Statistical Distributions of Argument Structure. *Computational Linguistics*, 27(3):373–406.

Riezler, S., T. H. King, R. M. Kaplan, R. Crouch, J. T. III Maxwell, and M. Johnson, 2002. Parsing the Wall Street Journal using a Lexical-Functional Grammar and Discriminative Estimation Techniques. In *Proceedings of the ACL'02*. Philadelphia, PA.

Appendix

(1) *The Solland Sword was lost for many years , having disappeared during the destruction of Solland by Gorbad Iron-claw's Orcs .*

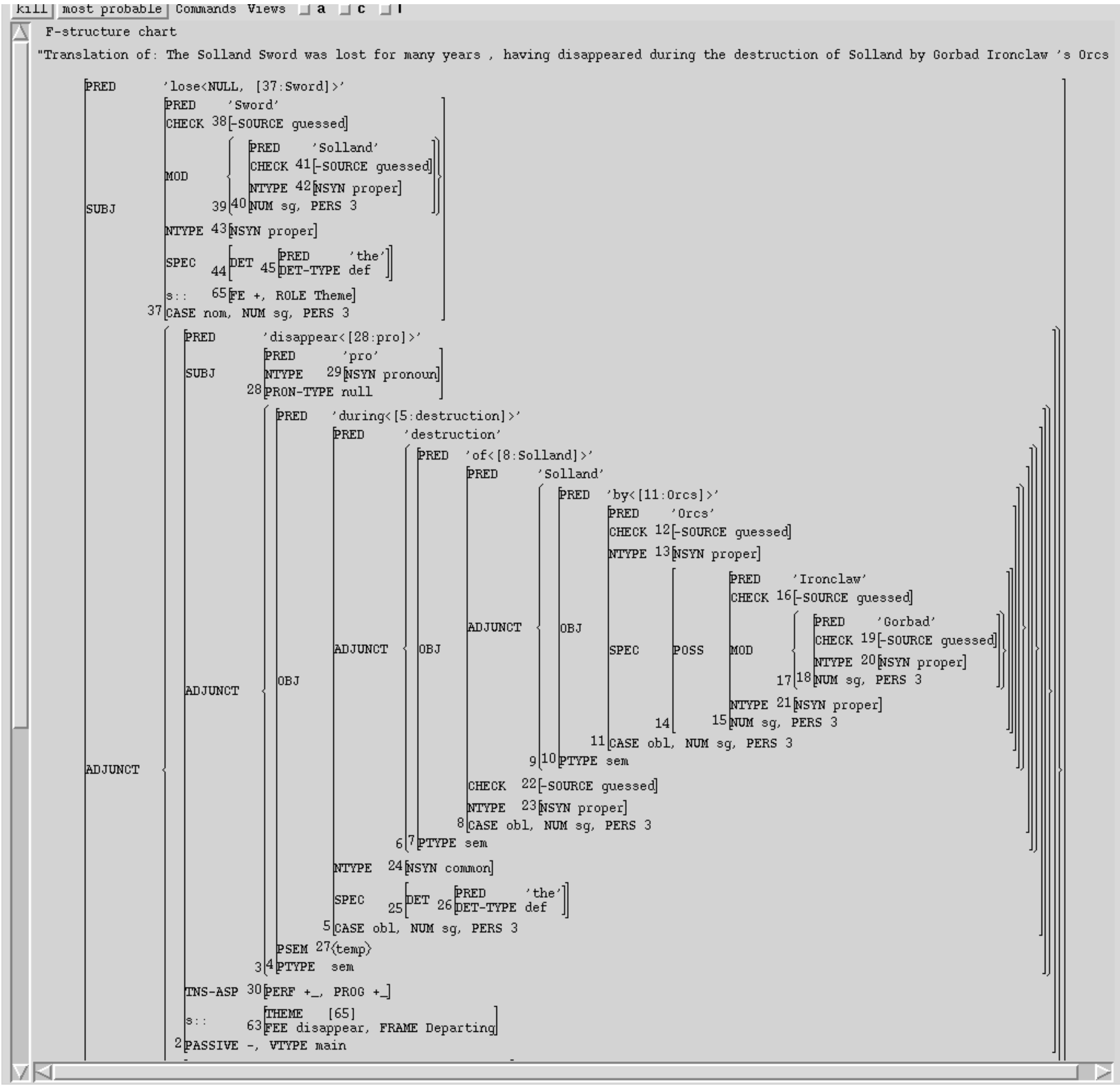


Figure 5: F-structure for example (1), with partial s-projection for frames

Annotating Predicate-Argument Structure for a Parallel Treebank

Lea Cyrus, Hendrik Feddes, Frank Schumacher

Arbeitsbereich Linguistik, University of Münster
Hüfferstraße 27, 48149 Münster, Germany
{lea,feddes,frank}@marley.uni-muenster.de

Abstract

We report on a recently initiated project which aims at building a multi-layered parallel treebank of English and German. Particular attention is devoted to a dedicated predicate-argument layer which is used for aligning translationally equivalent sentences of the two languages. We describe both our conceptual decisions and aspects of their technical realisation. We discuss some selected problems and conclude with a few remarks on how this project relates to similar projects in the field.

1. Introduction

Parallel corpora are widely accepted as a valuable data source for machine translation and other research. So far, however, the amount of linguistic annotation in these corpora is limited, and particularly multilingual corpora annotated with syntactic information are rare. Our goal is to build a treebank of aligned parallel[1] texts in English and German with the following linguistic levels: POS tags, constituent structure, functional relations and predicate-argument structure for each monolingual subcorpus, plus an alignment layer to "fuse" the two – hence our working title for the treebank, FuSe, which additionally stands for *fu*nctional *se*mantic annotation (Cyrus et al., 2003).

We use the Europarl Corpus (Koehn, 2002), which contains sentence-aligned proceedings of the European parliament in eleven languages and thus offers ample opportunity for extending the treebank at a later stage.[2] For syntactic and functional annotation we basically adapt the TIGER annotation scheme (Albert et al., 2003), making adjustments where we deem appropriate and changes which become necessary when adapting to English an annotation scheme which was originally developed for German.

The fusion of the language pair will take place on an alignment layer which connects the predicate-argument layers of both monolingual subcorpora. Only the alignment layer is explicitly defined for a language pair rather than for a single language. Apart from this layer, the subcorpora are monolingual resources in their own right.

Although, eventually, the treebank will prove useful for several fields of application, the most obvious one being machine translation, our main motivation is to contribute to linguistic research. The treebank will serve as a resource for both monolingual and contrastive analyses.

[1] In accordance with the terminology suggested in (Sinclair, 1994), we understand "parallel" to mean that the texts are translations of each other.

[2] There are a few drawbacks to Europarl, such as its limited register and the fact that it is not easily discernible which language is the source language. However, we believe that at this stage the easy accessibility, the amount of preprocessing and particularly the lack of copyright restrictions make up for these disadvantages.

2. Reasons for Predicate-Argument Structure

In a parallel treebank, it is necessary to capture the translational equivalence between two sentences. Our basic assumption is that this equivalence can best be represented by means of a predicate-argument structure. It is sometimes assumed that predicate-argument structure can be derived or recovered from constituent structure or functional tags such as subject and object.[3] While it is true that these annotations provide important heuristic clues for the identification of predicates and arguments, predicate-argument structure goes beyond the assignment of phrasal categories and grammatical functions, because the grammatical category of predicates and consequently the grammatical functions of their arguments can vary.

For instance, it is very common for an English verbal predicate to be expressed by a nominalisation in German, as is the case in the NPs in (1) and (2), where the English verb *nominate* is translated as the German noun *Nominierung*.

(1) their automatic right to nominate a member of the European Commission[4]

(2)
ihr automatisches Recht auf Nominierung eines
their automatic right on nomination of_a
Mitglieds der Europäischen Kommission
member of_the European Commission

The annotations of these noun phrases are shown in Figure 1.[5] It can be seen that the correspondence between NP_{508} and NP_{505} cannot be inferred from the constituent structure, since NP_{508} is an immediate constituent of an IE ("extended infinitive") while NP_{505} is deeply embedded in a PP. Neither can the correspondence of NP_{508} and NP_{505} be inferred from their respective functional categories, since NP_{508} is a direct object (OD) while NP_{505} is a modifier (AG: "genitive attribute"). However, the resemblance between these constituents becomes apparent when they are marked for their argument status, because they both fulfill a similar role.

[3] See e. g. (Marcus et al., 1994).

[4] Europarl:de-en/ep-00-02-15.al, 326. Note that throughout this paper, sentences are sometimes cited with irrelevant parts omitted.

[5] All figures are at the end of the paper.

We have therefore chosen to represent predicate-argument structure on a dedicated layer in our treebank in order to be able to capture the parallelism between translations and to use it as the basis for alignment.

3. Details of the Predicate-Argument Annotation

The predicate-argument structures used here consist solely of predicates and their arguments. Although there is usually more than one predicate in a sentence, no attempt is made to nest structures or to join the predications logically in any way.[6] The idea is to make the predicate-argument structure as rich as is necessary to be able to align a sentence pair while keeping it as simple as possible so as not to make it too difficult to annotate. In the same vein, quantification, negation, and other operators are not annotated. In short, the predicate-argument structures are not supposed to capture the semantics of a sentence exhaustively in an interlingua-like fashion.

3.1. Predicates and Arguments

In determining what a predicate is and how many there are in a sentence we rely on a few assumptions that are of a heuristic nature. One of these assumptions is that predicates are more likely to be expressed by tokens belonging to some word classes than by tokens belonging to others. Potential predicate expressions in FuSe are verbs, deverbal adjectives and nouns[7] or other adjectives and nouns which show a syntactic subcategorisation pattern. The predicates are represented by the capitalised citation form of the lexical item (e. g. NOMINATE). Homonymous or polysemous predicates are differentiated by means of a disambiguator, predicates are assigned a class based on their syntactic form, and derivationally related predicates form a predicate group.

Arguments are given short intuitive role names (e. g. ENT_NOMINATED) in order to facilitate the annotation process. These role names have to be used consistently only within a predicate group. If, for example, an argument of the predicate NOMINATE has been assigned the role ENT_NOMINATED and the annotator encounters a comparable role as argument to the predicate NOMINATION, the same role name for this argument has to be used.

Keeping the argument names consistent for all predicates within a group while differentiating the predicates on the basis of syntactic form are complementary principles, both of which are supposed to facilitate querying the corpus. The consistency of argument names within a group, for example, enables the researcher to analyse paradigmatically all realisations of an argument irrespective of the syntactic form of the predicate. At the same time, the differentiation of predicates makes possible a syntagmatic analysis of the differences of argument structures depending on the syntactic form of the predicate.

3.2. Binding Layer

All elements of the predicate-argument structure must be bound to elements of the phrasal structure (terminal or non-terminal nodes). These bindings are stored in a dedicated binding layer between the constituent layer and the predicate-argument layer.

When an expected argument is absent on the phrasal level due to specific syntactic constructions, the binding of the predicate is tagged accordingly, thus accounting for the missing argument. For example, in passive constructions like in Table 1, the predicate binding is tagged as `pv`. Other common examples are imperative constructions. Although information of this kind may possibly be derived from the constituent structure, it is explicitly recorded in the binding layer as it has a direct impact on the predicate-argument structure.

Sentence	wenn	korrekt	gedolmetscht	wurde
Gloss	if	correctly	interpreted	was
			↑	
Binding			`pv`	
			|	
Pred/Arg			DOLMETSCHEN	

Table 1: Example of a tagged predicate binding (Europarl:de-en/ep-00-01-18.al, 2532)

Bindings of arguments may be tagged as well, an example for this being object-control (cf. Table 2). To account for the deviant case of the subject of the embedded clause in an object-control construction, the binding of this argument is tagged (`oc-case`). With this information, a researcher or a machine learner will be able to ignore a specific argument which might distort statistics on the phrasal realisations of arguments.

The predicate binding is tagged as well to mark the entire object-control construction (`oc`). This tagging enables the researcher to filter out this specific predicate-argument structure, so as to ignore these constructions completely.

Section 4.1. will show that linking predicates or arguments to constituents cannot always be achieved by binding them to a single node in the constituent structure. In order to be flexible in this respect, the binding layer allows for complex bindings, with more than one node of the constituent structure to be included in and sub-nodes to be explicitly excluded from a binding to a predicate or argument.[8]

3.3. Alignment Layer

On the alignment layer, the elements of a pair of predicate-argument structures are aligned with each other. Arguments are aligned on the basis of corresponding roles within the predications. Comparable to the tags used in the binding layer that account for specific constructions (see

[6]Since the predicate-argument structure is always bound to the constituent structure (see Section 3.2.), it might well be possible to derive this information, e. g. through coordination structures and the hierarchical ordering of constituents.

[7]For all non-verbal predicate expressions for which a derivationally related verbal expression exists it is assumed that they are deverbal derivations, etymological counter-evidence notwithstanding.

[8]See the database documentation (Feddes, 2004) for a more detailed description of this mechanism.

Sentence	It was this which inspired	us	to propose	the same thing	with regard to state aid .
		↑	↑	↑	
Binding		oc-case	oc	[]	
		|	|	|	
Pred/Arg		PROPOSER	PROPOSE	PROPOSAL	

Table 2: Example of tagged predicate and argument bindings (Europarl:de-en/ep-00-01-18.al, 237)

Section 3.2.), the alignments may also be tagged with further information. This becomes necessary when the predications are incompatible in some way. Section 4.3. will give examples.

If there is no corresponding predicate-argument structure in the other language or if an argument within a structure does not have a counterpart in the other language, there will simply be no alignment. Section 4.2. provides an example where a predication is left dangling.

Table 3 gives an overview of the annotation layers as described in this section.

Layer	*Function*
Phrasal	constituent structure of language A
Binding	binding ↓ predicates/arguments to ↑ nodes
PA	predicate-argument structures
Alignment	aligning ↕ predicates and arguments
PA	predicate-argument structures
Binding	binding ↑ predicates/arguments to ↓ nodes
Phrasal	constituent structure of language B

Table 3: The layers of the predicate-argument annotation

4. Problematic Cases

In this section we will elaborate on some problematic cases of predicate-argument annotation which we have encountered so far, some of them particular to the annotation and alignment of predicate-argument structures for a language pair.

4.1. Binding Predicate-Argument Structure to Constituent Structure

It was mentioned in Section 3. that all predicates and arguments must be bound to either terminal or non-terminal nodes in the constituent structure. However, this is not always possible since in some cases there is no direct correspondence between argument roles and constituents. For instance, this problem occurs whenever a noun is postmodified by a participle clause: in Figure 2, the argument role ENT_RAISED of the predicate RAISE is realised by NP_{525}, but the participle clause (IPA_{517}) containing the predicate (*raised*$_6$) needs to be excluded, because not excluding it would lead to recursion. Consequently, there is no simple way to link the argument role to its realisation in the tree.

In these cases we link the argument role to the appropriate phrase (here: NP_{525}) and prune out the constituent that contains the predicate (IPA_{517}; see Section 3.2. for this mechanism), which results in a discontinuous argument realisation.

4.2. Coping with Modality

Generally, modal verbs are not considered to be predicates and are consequently not included in our predicate-argument database. This can cause a problem when a verbal predicate that is modified by a modal auxiliary in L1 (3) is represented by a deverbal noun in the corresponding sentence in L2 (4).

(3) The laws against racism must be harmonised.[9]

(4) Die Harmonisierung der Rechtsvorschriften
The harmonisation of_the laws
gegen den Rassismus ist dringend erforderlich.
against the racism is urgently necessary.

This can be illustrated by Figure 3: the realisation of the verbal predicate HARMONISE (*harmonised*$_6$) is modified by the modal auxiliary *must*$_4$. In the German sentence, the nominal predicate HARMONISIERUNG (*Harmonisierung*$_1$) is used. Here, the modality is expressed by a predicate of its own, namely ERFORDERLICH (*erforderlich*$_9$, 'necessary'). This second predicate does not correspond to any predicate in the English sentence.

It would be an easy way out to resort to annotating modal auxiliaries as if they were full verbs and consequently predicates, but we have opted against this makeshift solution. One has to keep in mind that the predicate-argument annotation is done monolingually and only later serves as the basis for alignment. It should not be assumed that the corresponding equivalent is known to the annotator during the annotation process. Even though the way a sentence is expressed in another language can give valuable insights into its structure and meaning, this should not go so far as to change the way the original language is annotated. This is particularly true since the idea behind the FuSe treebank is that it is in principle extendable and may well include languages other than English and German in the future. As it cannot be foretold what phenomena will be encountered once further languages are added, the decisions as to what is annotated and what is not should not be guided by cross linguistic considerations.

Thus, the simple fact alone that a predication in one language does not correspond to a predication in another should not induce one to alter the annotation praxis so as to make the two versions more compatible with each other. Modality, in particular, can be expressed in a variety of ways, and just because one of them is the realisation as a predicative adjective does not make, say, a modal adverbial like *certainly* a predicate. The same argumentation holds for modal auxiliaries.

[9] Europarl:de-en/ep-00-01-19.al, 489.

4.3. Incompatible Predications

Sometimes, the predications in two corresponding sentences express approximately the same idea but are otherwise incompatible with each other. This can be demonstrated with sentences (5) and (6), the annotation, argument structure and alignment of which are illustrated in Figure 4.

(5) Our motion will give you a great deal of food for thought, Commissioner[10]

(6) Eine Reihe von Anregungen werden wir Ihnen,
A row of suggestions will we you,
Herr Kommissar, mit unserer Entschließung
Mr. Commissioner, with our resolution
mitgeben
give

The incompatibility results from the fact that, while the predicates GIVE and MITGEBEN are roughly equivalent in meaning, the two sentences are organised differently with regard to their information structure. This has caused the two corresponding argument roles of GIVER and MITGEBER to be realised by two incompatible expressions representing different referents (NP_{500} vs. wir_5). The English version is somewhat metaphorical in that, unlike in the German sentence, there is no animate entity in this agent-like argument position. The actual agent is not realised as such and can only be identified by a process of inference based on the presence of the possessive pronoun our_0. To complicate matters even further, the translational equivalent of NP_{500} (i. e. the constituent realising the English GIVER), is not even an argument in the German sentence (PP_{508}).

Consequently, it seems impossible to reach a satisfactory alignment in this case: either two arguments with the same role but different meanings would have to be aligned, or else the alignment would rely solely on translational equivalence, which would reduce to absurdity our reasons for including predicate-argument structure.

We solve the problem as follows: since cases like this are at the same time potentially interesting for contrastive analyses and a hazard for applications using the treebank for automatic learning, we keep up the alignment on the basis of argument roles but tag the alignment (see Section 3.3.) between the arguments in question and thus mark them as being incompatible (`incomp`) with each other. This enables the interested researcher to formulate explicit searches for this alignment type while making it possible for applications to skip these cases if this is preferred.

Sentences (7) and (8) are a second case where we make use of the possiblilty to tag the alignment. Here, the adjectival predicate INAPPLICABLE in (7) is represented by the negated predicate ANWENDBAR ('applicable') in the German counterpart (8).

(7) the Directive is inapplicable in Denmark [11]

(8) die Richtlinie ist in Dänemark nicht anwendbar
the Directive is in Denmark not applicable

Since whether or not a predicate is negated does not alter its argument structure we do not annotate negation (see Section 3.). As this leads to an alignment of predicates with opposite meanings, we tag the alignment between the two predicates as `abs-opp` ("absolute opposites"). In theory, this method could also be applied to cases where a predicate is translated by its relational opposite (e. g. *buy* vs. *sell*). So far, however, we have not yet come across this type of translation in our data. It will be interesting to discover what types of incompatibility will come to light as the annotation proceeds.

5. Database Structure and Tools

We use ANNOTATE (Plaehn, 1998a) for the semi-automatic assignment (Brants, 1999) of POS tags, hierarchical structure, phrasal and functional tags. ANNOTATE stores all annotations in a relational database.[12] To stay consistent with this approach we have developed an extension to the ANNOTATE database structure to model the predicate-argument layer and the binding layer.

Due to the monolingual nature of the ANNOTATE database structure, the alignment layer (Section 3.3.) cannot be incorporated into it. Hence, additonal types of databases are needed. For each language pair (currently, English and German), an alignment database is defined which represents the alignment layer, thus fusing two extended ANNOTATE databases. Additionally, an administrative database is needed to define sets of two ANNOTATE databases and one alignment database. The final parallel treebank will be represented by the union of these sets (Feddes, 2004).

While annotators use ANNOTATE to enter phrasal and functional structure comfortably, the predicate-argument structures and alignments are currently entered into a structured text file which is then imported into the database. A graphical annotation tool for these layers is under development. It will make binding the predicate-argument structure to the constituent structure easier for the annotators and suggest argument roles based on previous decisions.

6. Relation to Other Projects and Outlook

This section will show briefly how our approach relates to other projects annotating some kind of predicate-argument structure, such as PropBank (Palmer et al., 2003) and FrameNet (Johnson et al., 2003), and how the alignment structures of the parallel treebank make up for certain drawbacks of our annotation scheme.

Since our annotation of predicates and their arguments is not a means in itself but to the end of aligning constituents of a parallel treebank, it is kept deliberately simple. It resembles the mnemonic descriptors clarifying the numbered arguments in the PropBank framesets. We do not, however, attempt any generalisation whatsoever: neither do we organise our predicates in frames, as is done by FrameNet and adopted by SALSA (Erk et al., 2003), nor do we follow the Levin classes (Levin, 1993), as is done in the PropBank project.

[10] Europarl:de-en/ep-00-01-18.al, 53.
[11] Europarl:de-en/ep-00-01-18.al, 2522.

[12] For details about the ANNOTATE database structure see (Plaehn, 1998b).

Some problems we encounter with our simple scheme could be avoided with a deeper predicate-argument structure. As the first example in Section 4.3. shows, predications which are incompatible in our scheme need not be incompatible in a FrameNet-like scheme: if the argument roles were deeper than our intuitive role names, i. e., if *our motion* in example (5) were not a GIVER but, e. g., a CAUSE, the incompatibility with the corresponding structure in (6) would not arise.

There are several reasons for us to stick to our simple approach. For one thing, a more complex scheme would make the annotation more susceptible to inconsistencies. Secondly, transferring the approaches mentioned above to other languages than English is not a straightforward matter. While this seems to be working quite well for the FrameNet frames (Erk et al., 2003), Levin's verb classes are inherently English and cannot be directly applied to German. In a later stage of the project, it might be possible to work through the predicate-argument database and map our very specific scheme to a more general one, e. g. by assigning each predicate to a frame and each argument to a frame element. However, other studies show that mapping one scheme onto another is far from trivial (Hajičová and Kučerová, 2002), and quite a lot of manual work will presumably be necessary.

Finally, we believe it is possible to exploit the corpus as a parallel lexical resource to see how different predicates can be clustered automatically by analysing their mappings in the other language. Figure 5 sketches the general idea. Suppose that in the English sub-corpus, two predicate-argument structures have different predicates (BUY and PURCHASE) which subcategorise for comparable arguments and express the same concept. In a FrameNet-like annotation, these predicates would be instantiations of the same frame (e. g. COMMERCIAL_TRANSACTION). In our scheme, neither are these predicates grouped in any way, nor do the comparable arguments get the same role names.

However, it is well conceivable that both predicates are translated identically in the corresponding German structures (e. g. by KAUFEN 'buy'). Since predicates and arguments are aligned to each other, the comparability of the predicates (BUY – PURCHASE) and their arguments (BUYER – PURCHASER and ENT_BOUGHT – ENT_PURCHASED) can be derived (cf. the dashed lines). It will then be instructive to investigate how these clusters compare to FrameNet frames and to explore to what extent such a data-driven approach to frame semantics is feasible.

7. References

Albert, Stefanie et al., 2003. TIGER Annotationsschema. Technical report, Universität des Saarlandes, Universität Stuttgart, Universität Potsdam. Unpublished Draft – 24 July 2003.

Brants, Thorsten, 1999. *Tagging and Parsing with Cascaded Markov Models: Automation of Corpus Annotation*. Saarbrücken: Saarland University.

Cyrus, Lea, Hendrik Feddes, and Frank Schumacher, 2003. FuSe – a multi-layered parallel treebank. Poster presented at the Second Workshop on Treebanks and Linguistic Theories, 14–15 November 2003, Växjö, Sweden (TLT 2003). `http://fuse.uni-muenster.de/Publications/0311_tltPoster.pdf`.

Erk, Katrin, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal, 2003. Towards a resource for lexical semantics: A large German corpus with extensive semantic annotation. In *Proc. ACL-03*.

Feddes, Hendrik, 2004. FuSe database structure. Technical report, Arbeitsbereich Linguistik, Münster University. `http://fuse.uni-muenster.de/Publications/dbStruktur.pdf`.

Hajičová, Eva and Ivona Kučerová, 2002. Argument/valency structure in PropBank, LCS Database and Prague Dependency Treebank: A comparative pilot study. In *Proc. LREC-2002*.

Johnson, Christopher R., Miriam R. L. Petruck, Collin F. Baker, Michael Ellsworth, Josef Ruppenhofer, and Charles J. Fillmore, 2003. FrameNet: Theory and practice. `http://www.icsi.berkeley.edu/~framenet/book/book.html`.

Koehn, Philipp, 2002. Europarl: A multilingual corpus for evaluation of machine translation. `http://www.isi.edu/~koehn/publications/europarl/`.

Levin, Beth, 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: Chicago University Press.

Marcus, Mitch, G. Kim, M. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger, 1994. The Penn Treebank: Annotating predicate argument structure. In *Proc. ARPA HLT Workshop*.

Palmer, Martha, Daniel Gildea, and Paul Kingsbury, 2003. The Proposition Bank: An annotated corpus of semantic roles. Submitted to *Computational Linguistics*, `http://www.cis.upenn.edu/~mpalmer/papers/prop.pdf`.

Plaehn, Oliver, 1998a. ANNOTATE Bedienungsanleitung. Technical report, Saarland University. `http://www.coli.uni-sb.de/sfb378/negra-corpus/annotate-manual.ps.gz`.

Plaehn, Oliver, 1998b. ANNOTATE Datenbank-Dokumentation. Technical report, Saarland University. `http://www.coli.uni-sb.de/sfb378/negra-corpus/datenbank.ps.gz`.

Sinclair, John, 1994. Corpus typology. Preliminary recommendations. Technical Report EAG-CSG/IR-T1.1 (Work in Progress), EAGLES. `http://www.ilc.cnr.it/EAGLES96/typology/typology.html`.

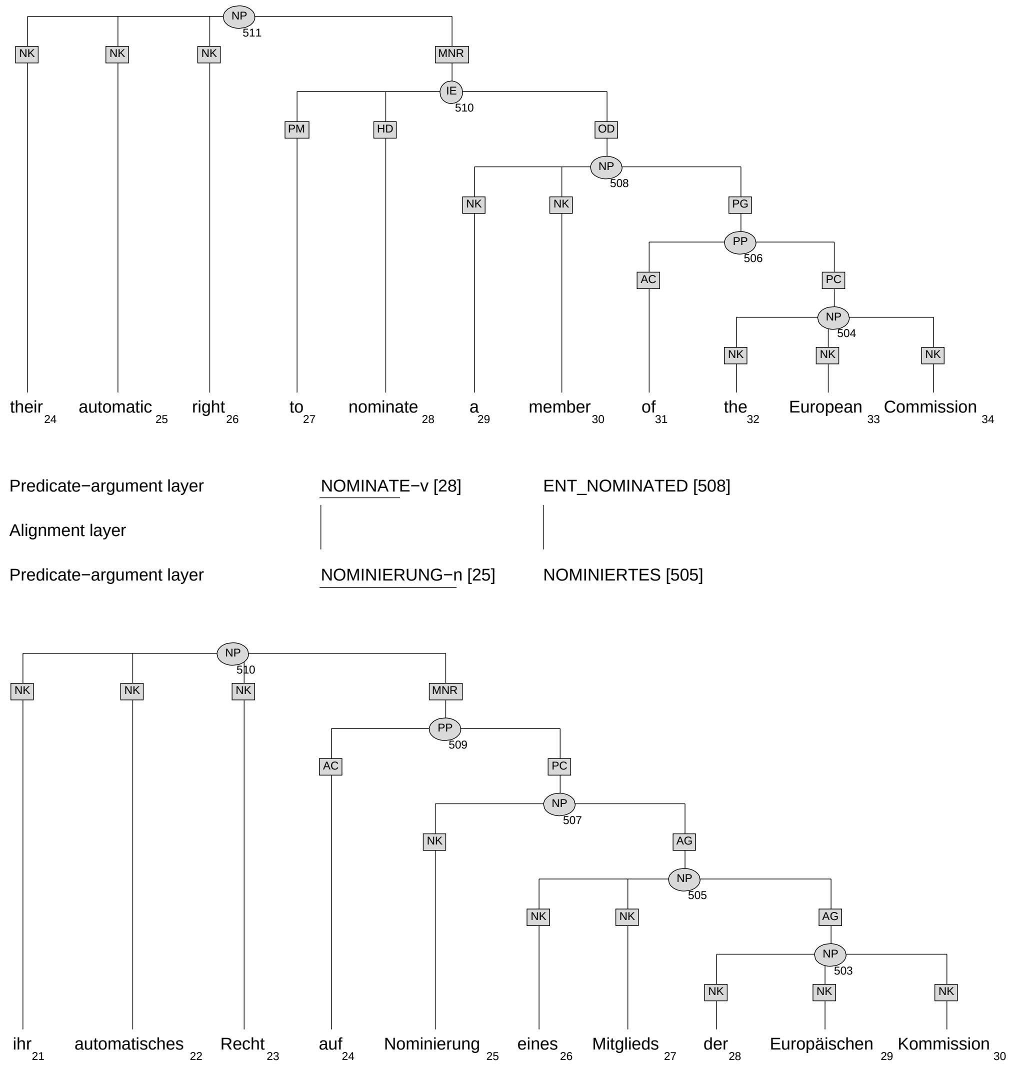


Figure 1: Alignment of a verb/direct-object construction with a noun/modifier construction

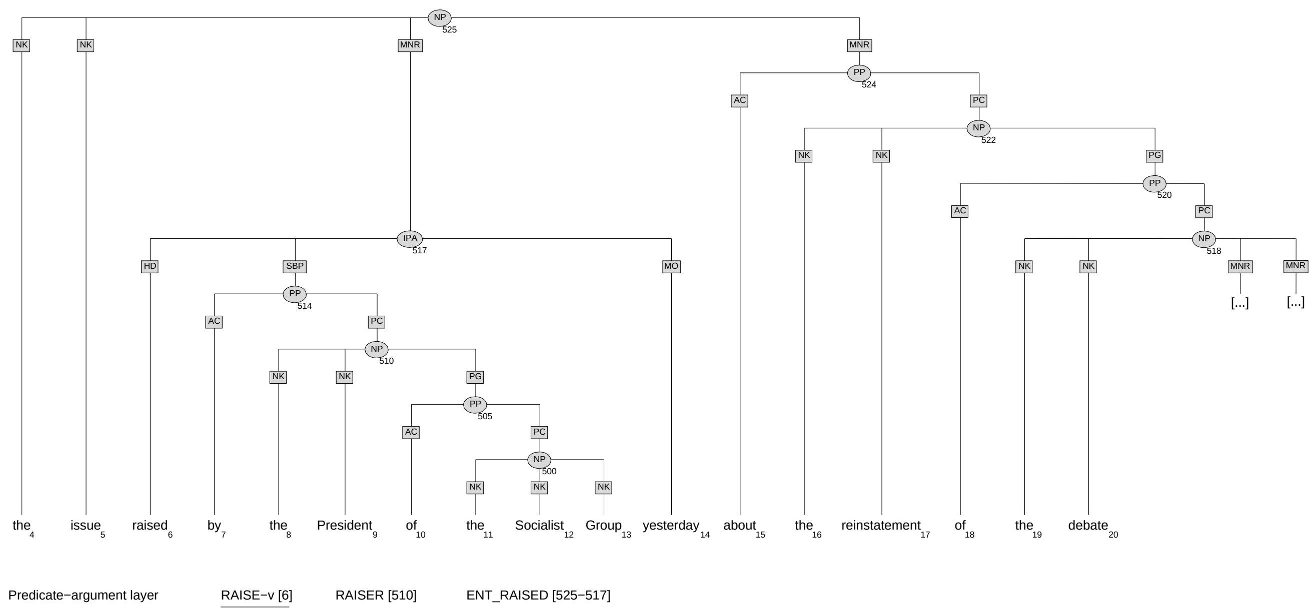


Figure 2: Complex constituent binding of an argument

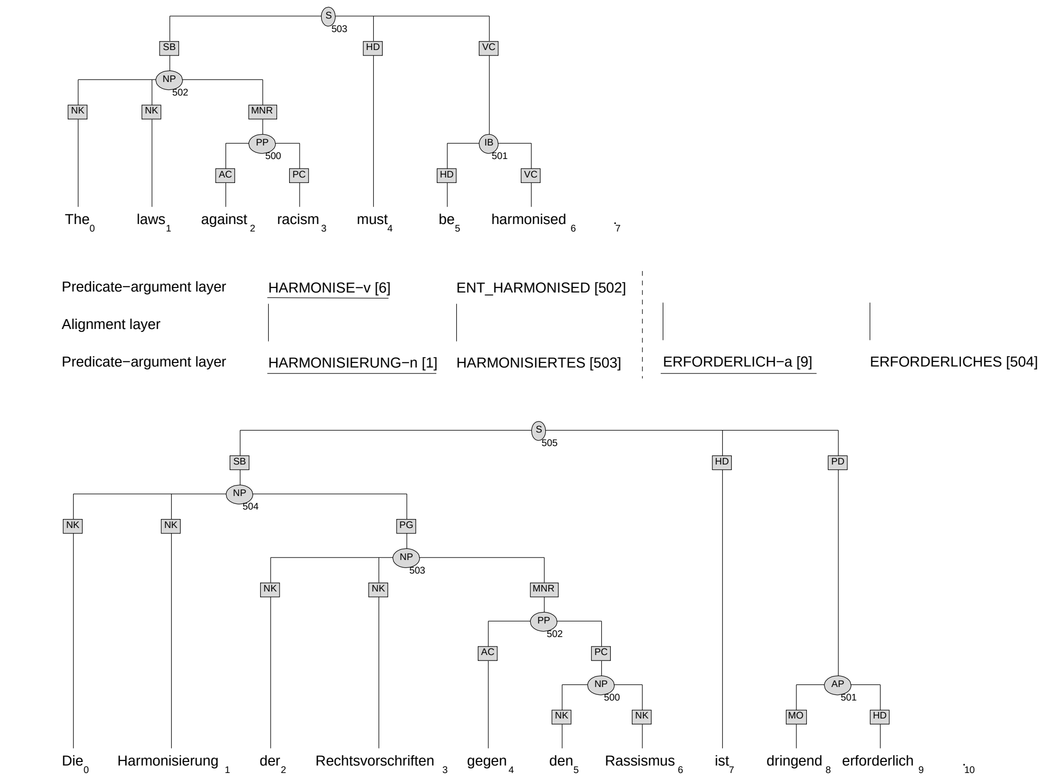


Figure 3: Modality

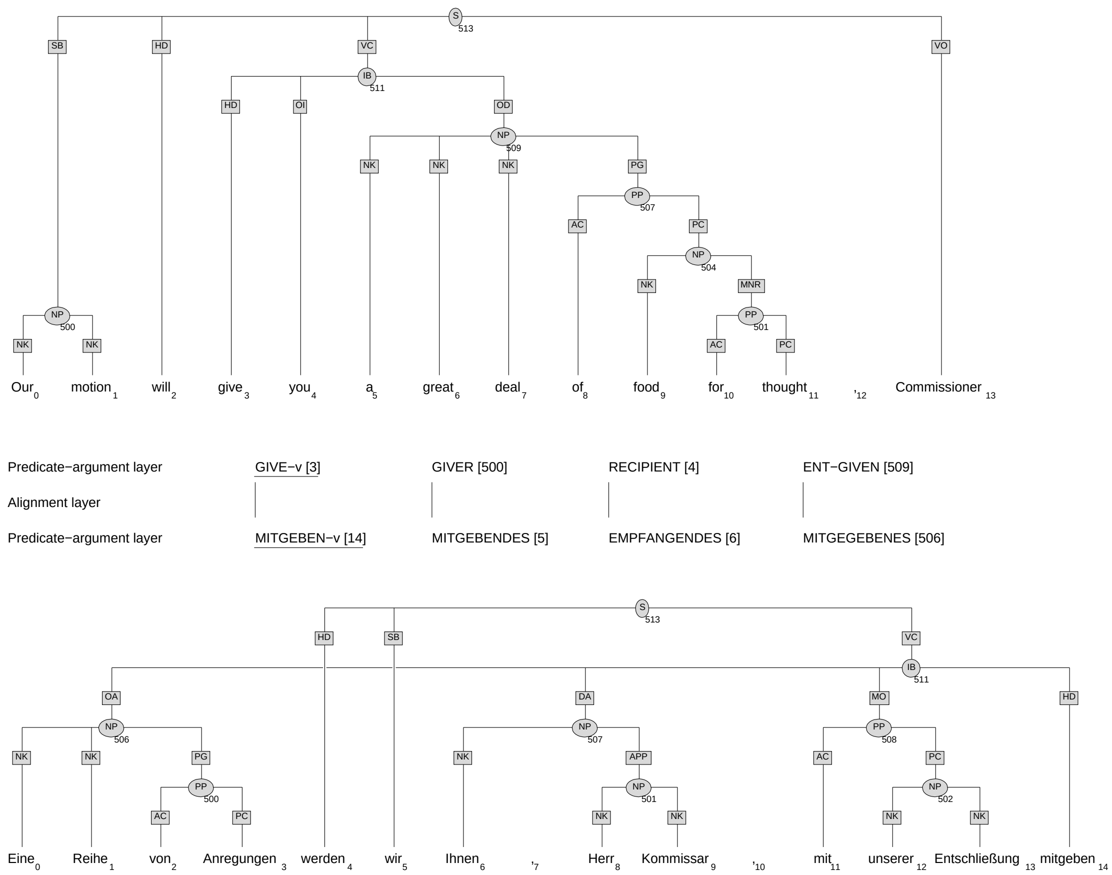


Figure 4: Incompatible predications

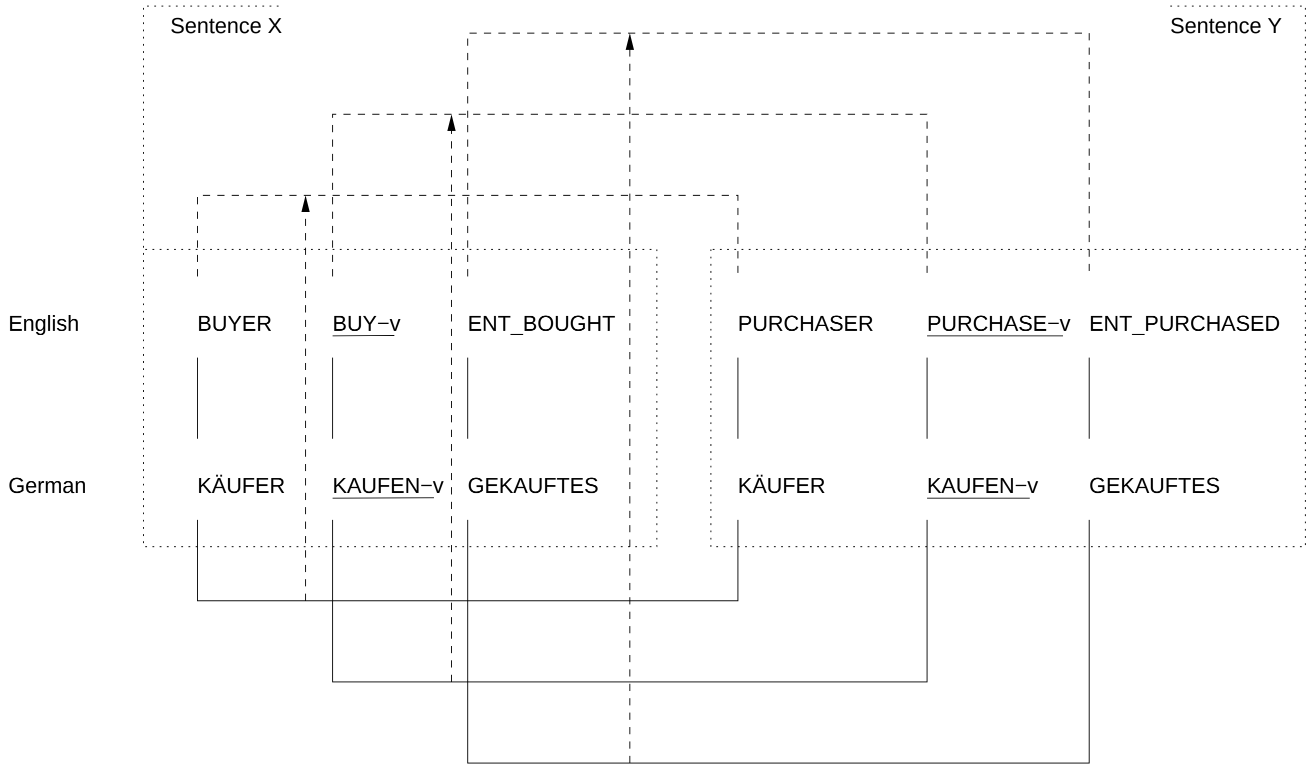


Figure 5: Deriving predicate clusters by exploiting alignment structures

Semantic Annotation and Lexico-Syntactic Paraphrase

Bonnie J. Dorr†, Rebecca Green†, Lori Levin‡, Owen Rambow∗, David Farwell§, Nizar Habash†∗, Stephen Helmreich§, Eduard Hovy¶, Keith J. Millerˆ, Teruko Mitamura‡, Florence Reederˆ, Advaith Siddharthan∗

†Institute for Advanced Computer Studies
University of Maryland
{bonnie, rgreen, habash}@umiacs.umd.edu

‡Language Technologies Institute
Carnegie Mellon University
{lsl, teruko}@cs.cmu.edu

∗Department of Computer Science
Columbia University
{rambow, as372}@cs.columbia.edu

§Computing Research Laboratory
New Mexico State University
{david, shelmrei}@crl.nmsu.edu

¶Information Sciences Institute
University of Southern California
hovy@isi.edu

ˆMITRE Corporation
{freeder, keith}@mitre.org

Abstract

The IAMTC project (Interlingual Annotation of Multilingual Translation Corpora) is developing an interlingual representation framework for annotation of parallel corpora (English paired with Arabic, French, Hindi, Japanese, Korean, and Spanish) with deep-semantic representations. In particular, we are investigating meaning equivalent paraphrases involving conversives and non-literal language use, as well as extended paraphrases involving syntax, lexicon, and grammatical features. The interlingua representation has three levels of depth. Each level is characterized by the types of meaning equivalent paraphrases that receive identical representations at that level.

1. Introduction

An important issue for computational linguists and lexicographers is the question of meaning-equivalent paraphrases, including lexical synonymy, conversives (*buy/sell*), idioms (*kick the bucket/die*), and more extended paraphrases, such as *Its network of eighteen organizations has lent a billion dollars to microenterprises* and *The network comprises eighteen organizations which have disbursed a billion dollars to microenterprises.*

Semantic annotation projects such as PropBank (Kingsbury and Palmer, 2002) and FrameNet (Baker et al., 1998) do not cover extended paraphrases. (Barzilay and McKeown, 2001) have proposed automatic methods for extraction of extended paraphrases. However such methods are subject to the usual pitfalls of machine learning systems, requiring large amounts of training data and having imperfect precision and recall.

Our approach in the IAMTC project (Interlingual Annotation of Multilingual Translation Corpora) is complementary to other semantic annotation projects and to projects that automatically label semantic paraphrases. Firstly, we are annotating texts in seven languages (Arabic, English, French, Hindi, Japanese, Korean, and Spanish). Secondly, we are investigating meaning-equivalent paraphrases by annotating multiple versions of the same text, usually one non-English source language and two independently produced English translations. The annotation scheme includes three levels of depth. Each level is characterized by the types of paraphrase that are resolved at that level.

2. Project Overview

The IAMTC project has four goals — development of an interlingual representation framework, annotation of bilingual corpora, development of semantic annotation tools, and design of evaluation metrics for assessing interlingual representations.

The methodology for developing the interlingual representation involves careful study of text corpora in seven languages. Each corpus is bilingual and multi-parallel. Each text in a corpus has at least three versions, a non-English original and at least two English translations. The comparison of annotated multi-parallel corpora sets us apart from other semantic annotation projects. The multiple parallel texts allow us to document naturally occurring paraphrases of the same meaning. The interlingua framework will include a formal definition of the three levels of representation, characterization of paraphrases that are resolved at each level, and coding manuals for each level.

The corpus annotation effort is aimed at making a publically available semantically annotated corpus that can be useful for any natural language processing application that seeks to include more semantic depth. We expect the corpus to be useful for improvement of machine translation, summarization, and information extraction.

Our efforts in tool development have led us not only to create tools for annotation, but also to build interfaces for comparing and reconciling annotations within and across annotators. Because our annotation tools facilitate the building of trees with feature structures at each node, we expect that the tools will be useful for other types of linguistic annotation beyond the IAMTC project.

Finally, an important part of the IAMTC project is to design and test a variety of new evaluation metrics for assessing the interlingual representations and choosing an appropriate granularity of meaning representation. Evaluation is generally important in language technologies, but is particularly important in an area like interlingua design that deals with meanings that are deep, multi-faceted, and not well-defined. Our evaluation metrics include inter-annotator agreement, intra-annotator consistency, and success in NLP applications.

3. The Interlingua Representation

Recognizing the complexity of interlinguas, we adopt an incrementally deepening approach, which allows us to produce relatively stable annotations while exploring alternatives at the next level down. We currently identify three levels of representation, referred to as IL0, IL1, and IL2. Each level of representation incorporates additional semantic features and removes existing syntactic ones.

IL0 is a deep syntactic dependency representation, constructed by hand-correcting the output of a dependency parser based on Connexor (`www.connexor.com`). Though this representation is purely syntactic, it abstracts as much as possible from surface-syntactic phenomena. For example, auxiliary verbs and other function words are removed from IL0. In addition, corresponding active and passive voice sentences receive the same representation in IL0. Thus it is more abstract than the Praguian Analytical level, but more syntactic than the Tectogrammatical level (Bohmova et al., 2003). IL0 is a useful starting point for IL1 in that syntactic dependencies are often indicative of semantic dependencies. Figure 1, which appears on the last page of the paper, shows the IL0 representation for the sentence *Sheikh Mohamed, who is also the Defense Minister of the United Arab Emirates, announced at the inauguration ceremony that "we want to make Dubai a new trading center."*

IL1 is an intermediate semantic representation. Open class lexical items (nouns, verbs, adjectives, and adverbs) are associated with concepts drawn from the Omega ontology (Hovy et al., 2003). Also at this stage, syntactic relations are replaced by semantic roles such as AGENT, THEME, and GOAL. However, IL1 is not an interlingua; it does not normalize over all linguistic realizations of the same semantics. Figure 2, which appears on the last page of the paper, shows the IL1 corresponding to the IL0 in Figure 1. Concept names and thematic role names added by the annotators are in upper case; some nodes are associated with more than one concept.

IL0 and IL1 have been documented with coding manuals and have been used by annotators to tag several texts. (See Section 4.) However, IL2, the deepest meaning representation, is still under development. The methodology for designing IL2 involves comparison of IL1's in the multi-parallel corpus in order to see how meaning equivalent IL1's can be reconciled or merged. IL2 is expected to normalize over:

- Conversives (e.g., *X bought a book from Y* vs. *Y sold a book to X*), as does FrameNet (Baker et al., 1998) at the more general level of Commercial_transaction.
- Non-literal language usage (e.g., *X started its business* vs. *X opened its doors to customers*).
- Extended paraphrases involving syntax, lexicon, and grammatical features (see example in introduction).

Figures 3-6 illustrate the relationship between IL1 and IL2. The examples are tentative at this point, since IL2 has not yet been formalized. Figure 3 shows the expected representation of *Mary bought a book from John* and *John sold a book to Mary* at IL1 and IL2. The IL1's for the two sentences are different because the verbs *buy* and *sell* use different participants as agents. However, the IL2 representation captures the common meaning of the buying and selling events, as has been suggested by many theories of meaning representation.

```
IL2:
  TRANSFER-POSSESSION
    [JOHN, source]
    [MARY, goal]
    [BOOK, theme]
    [PURCHASE, manner]

IL1 candidate #1:
   BUY
     [MARY, agent]
     [JOHN, source]
     [BOOK, theme]

IL1 candidate #2:
   SELL
      [JOHN, agent]
      [MARY, goal]
      [BOOK, theme]
```

Figure 3: **IL1 and IL2 for Conversives**

Figures 4-6 show an extended paraphrase in French and English. The English and French sentences are from parallel texts in the January 1997 edition of the UNESCO Courier, which is available in 29 languages and Braille. Figure 4 shows an English sentence and its IL1. The head of the English IL1 is the concept LEND. Figure 5 shows a French sentence and its IL1. The head of the French IL1 is the concept COMPRISE.

Figure 6 sketches some proposed mappings from IL1 to IL2, which would be needed in order to reconcile the

English:

Its network of eighteen independent organizations in Latin America has lent one billion dollars to microenterprises.

English IL1:

```
LEND
  [NETWORK, agent]
     [COMPRISE, mod]
         [ORGANIZATIONS, part]
  [DOLLARS, theme]
  [MICROENTERPRISE, goal]
```

Figure 4: **English Sentence and IL1**

French:

Le réseau regroupe dix-huit organisations indépendantes qui ont déboursé un milliard de dollars.

'The network comprises eighteen independent organizations which have disbursed a billion dollars'

French IL1:

```
COMPRISE
  [NETWORK, whole]
  [ORGANIZATIONS, part]
    [DISBURSE, mod]
       [NETWORK, agent]
       [DOLLARS, theme].
```

Figure 5: **French Sentence and IL1**

IL1-IL2 Mappings:

```
of/regroupe    <-> COMPRISE
lend/debourse <-> TRANSFER-MONEY
```

IL2:

```
COMPRISE:
  [NETWORK, whole]
  [ORGANIZATIONS, part]

TRANSFER-MONEY
  [NETWORK, agent]
  [DOLLARS, theme]
  [MIRCROENTERPRISE, goal]
```

Figure 6: **Conversion from IL1 to IL2**

IL1's from Figures 4 and 5. The words *of* and *regrouper* are found to express the concept COMPRISE. The argument, ORGANIZATION of both words help to confirm that *of* and *regrouper* describe the same relation. Similarly, the concept TRANSFER-MONEY is identified as a common concept for *lend* and *debourser*, which share two arguments, NETWORK and DOLLARS.

The range of paraphrase phenomena being addressed by the different representation levels is summarized in Table 1, which is based on examples from (Hirst, 2003), (Kozlowski et al., 2003), and (Rinaldi et al., 2003). The table indicates for which types we expect to produce normalized representations reflecting the similarity in meaning between paraphrases of that type and at which level the normalization will take place.

4. Work to Date

The IAMTC project has trained approximately ten annotators, each of whom has annotated twelve texts. The twelve texts consist of two English translations of each of six foreign language articles. All annotators have worked on the same texts. This section describes the annotation procedures.

In order to prepare IL0, the texts are first passed through the Connexor dependency parser. Project experts then edit the dependency trees using TrEd (Pajas, 1998). During the editing process, parsing errors are corrected and the project's conventions for dependency trees are enforced. The project's conventions concern the treatment of closed-class function words and copular sentences.

The IL0's are then passed to the annotators. The IL1 for a sentence is the result of assigning concepts and semantic roles to the nodes of the IL0. The IAMTC has developed an interface called TIAMAT for producing IL1's. Through TIAMAT, annotators can access the Omega ontology (Hovy et al., 2003), which contains concept names from WordNet (Fellbaum, 1998) and Mikrokosmos (O'Hara et al., 1998), and thematic role names from the Lexical Conceptual Structure (LCS) verb database (Dorr et al., 2001). Annotators are currently instructed to choose at least one WordNet concept and one Mikrokosmos concept for each content word in the text, or to choose a dummy concept if no suitable concept is found. Concepts that are covered in the LCS database are accompanied by a list of semantic roles. The annotators can assign a role to each of the verb's dependents, choosing either from the LCS frame or from a list of about fifteen roles that are defined in the coding manual. Annotators can also consult the IL0 in TrEd while they are using TIAMAT.

5. Evaluation

In our initial experiments, we have measured inter-annotator agreement of our semantic annotations. Many metrics have been proposed for measuring intercoder agreement in a coding task, including Kappa (Carletta, 1996) and a "Wood Standard" based on comparison of peers (Habash and Dorr, 2002). Since each text might be tagged by a different number of annotators and annotators may pick more than one sense per word, we are currently experimenting with metrics that take into account for each word the number of annotators and the number of senses that the ontology makes available for that word.

The evaluation process also involves consistency checking and reconciliation. Consistency checking involves comparison of meaning equivalent (synonymous or nearly syn-

Relationship Type	Example	Where Normalized
Syntactic variation	The gangster killed at least 3 innocent bystanders. *vs.* At least 3 innocent bystanders were killed by the gangster.	IL0
Lexical synonymy	The toddler sobbed, and he attempted to console her. *vs.* The baby wailed, and he tried to comfort her.	IL1
Morphological derivation	I was surprised that he destroyed the old house. *vs.* I was surprised by his destruction of the old house.	IL2
Clause subordination vs. anaphorically linked sentences	This is Joe's new car, which he bought in New York. *vs.* This is Joe's new car. He bought it in New York.	IL2
Different argument realizations	Bob enjoys playing with his kids. *vs.* Playing with his kids pleases Bob.	IL2
Noun-noun phrases	She loves velvet dresses. *vs.* She loves dresses made of velvet.	IL2
Head switching	Mike Mussina excels at pitching. *vs.* Mike Mussina pitches well. *vs.* Mike Mussina is a good pitcher.	IL2
Overlapping meanings	Lindbergh flew across the Atlantic Ocean. *vs.* Lindbergh crossed the Atlantic Ocean by plane.	IL2
Comparatives vs. superlatives	He's smarter than everybody else. *vs.* He's the smartest one.	Not normalized
Different sentence types	Who composed the Brandenburg Concertos? *vs.* Tell me who composed the Brandenburg Concertos.	Not normalized
Inverse relationship	Only 20% of the participants arrived on time. *vs.* Most of the participants arrived late.	Not normalized
Inference	The tight end caught the ball in the end zone. *vs.* The tight end scored a touchdown.	Not normalized
Viewpoint variation	The U.S.-led invasion/liberation/occupation of Iraq . . . You're getting in the way. *vs.* I'm only trying to help.	Not normalized

Table 1: **Relationship Types Underlying Paraphrase**

onymous) words in parallel texts. This a preliminary step toward identifying meaning equivalent sentences whose IL1's can and cannot be merged into a single IL2. It also helps us to evaluate the ontology and TIAMAT. In order to evaluate the ontology, we are interested in the extent to which there exist nodes that can express the common meaning of near synonyms. In assessing TIAMAT we would like to know how easy it is to navigate through the ontology in order to find the nodes that express the common meaning of near synonyms.

Reconciliation is the process of comparing two or more annotations produced by different people. There is a tool available for displaying multiple annotations with color coding for agreement or disagreement. The reconciliation process is conducted partly by each annotator separately and partly by interaction between the annotators. We are interested in finding out whether agreement on subsequent annotations increases as a result of reconciliation.

Another criterion to evaluate is the usefulness of the interlingual representations in NLP tasks. Since the ultimate goal is to generate a representation that is useful for MT (among other NLP tasks), we plan to measure the ability to generate accurate surface texts from the representation. We plan to use an available generator, Halogen (Langkilde and Knight, 1998). Sentences will be generated from interlinguas and then compared with the originals through a variety of standard MT metrics (ISLE, 2003). This will serve to determine whether the elements of the representation language are sufficiently well defined and whether they can serve as a basis for inferring interpretations from semantic representations or (target) semantic representations from interpretations.

6. Identification of Paraphrases

The ability to discern paraphrases is beneficial to virtually all linguistic applications, including information retrieval, information extraction, question-answering, text summarization, and machine translation. In the IAMTC framework, two sentences with the same IL2 are considered paraphrases even if they have different IL1's. The IL2 annotation on the corpus will allow us to easily study the different surface realizations of a given meaning pattern. Our intention is that these corpora will be used to improve the accuracy and robustness of semantic analysis in many NLP applications.

Acknowledgments

This work was supported by the National Science Foundation under grant number IIS-0325695. We would also like to thank our hard working annotators and graduate students.

7. References

Baker, Collin F., Charles J. Fillmore, and John B. Lowe, 1998. The Berkeley FrameNet Project. In Christian

Boitet and Pete Whitelock (eds.), *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*. San Francisco, California: Morgan Kaufmann Publishers.

Barzilay, R. and K.R. McKeown, 2001. Extracting paraphrases from a parallel corpus. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*. Toulouse, France.

Bohmova, A., J. Hajic, E. Hajicova, and B. Hladka, 2003. *The Prague Dependency Treebank: Three-Level Annotation Scenario*. Anne Abeillé (ed.) Treebanks: Building and Using Syntactically Annotated Corpora, Kluwer Academic Publishers.

Carletta, J.C., 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

Dorr, Bonnie J., Mari Olsen, Nizar Habash, and Scott Thomas, 2001. LCS Verb Database. Technical Report Online Software Database, University of Maryland, College Park, MD. Http://www.umiacs.umd.edu/ ˜bonnie/ LCS_Database_Docmentation.html.

Fellbaum, Christiane, 1998. *WordNet: An Electronic Lexical Database*. MIT Press. Http://www.cogsci.princeton.edu/˜wn [2000, September 7].

Habash, Nizar and Bonnie J. Dorr, 2002. Handling Translation Divergences: Combining Statistical and Symbolic Techniques in Generation-Heavy Machine Translation. In *Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002*. Tiburon, CA.

Hirst, Graeme, 2003. Paraphrasing paraphrased. In *Keynote address for The Second International Workshop on Paraphrasing: Paraphrase acquisition and Applications*. Association for Computational Linguistics.

Hovy, Eduard, Andrew Philpot, Judith L. Klavans, Ulrich Germann, and Peter T. Davis, 2003. Extending metadata definitions by automatically extracting and organizing glossary definitions. In *Proceedings of the National Conference on Digital Government Research*. Boston, MA.

ISLE, 2003. Towards Systematizing MT Evaluation. In *Proceedings of MT Summit IX Evaluation Workshop*. New Orleans, Louisiana.

Kingsbury, P. and M. Palmer, 2002. From Treebank to PropBank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*. Las Palmas, Spain.

Kozlowski, Raymond, Kathleen F. McCoy, and K. Vijay-Shanker, 2003. Generation of single-sentence paraphrases from predicate/argument structure using lexico-grammatical resources. In *Proceedings of the Second International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP2003)*. ACL 2003.

Langkilde, I. and K. Knight, 1998. Generation that exploits corpus-based statistical knowledge. In *Proceedings of the 36th Annual Meeting of the Association for computational Linguistics joint with the 17th International Conference on Computational Linguistics, ACL/COLING*. Montreal, Canada.

O'Hara, Tom, Kavi Mahesh, and Sergei Nirenburg, 1998. Lexical acquisition with wordnet and the mikrokosmos ontology. In *Proceedings of the ACL Workshop on the Use of WordNet in NLP*. Montreal, Canada.

Pajas, Petr, 1998. Tree editor manual. Technical report, CLSP Summer Sorkshop, Johns Hopkins University, Baltimore, MD.

Rinaldi, Fabio, James Dowdall, Kaarel Kaljurand, Michael Hess, and Diego Moll, 2003. Exploiting paraphrases in a question-answering system. In *Proceedings of the Second International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP2003)*. ACL 2003.

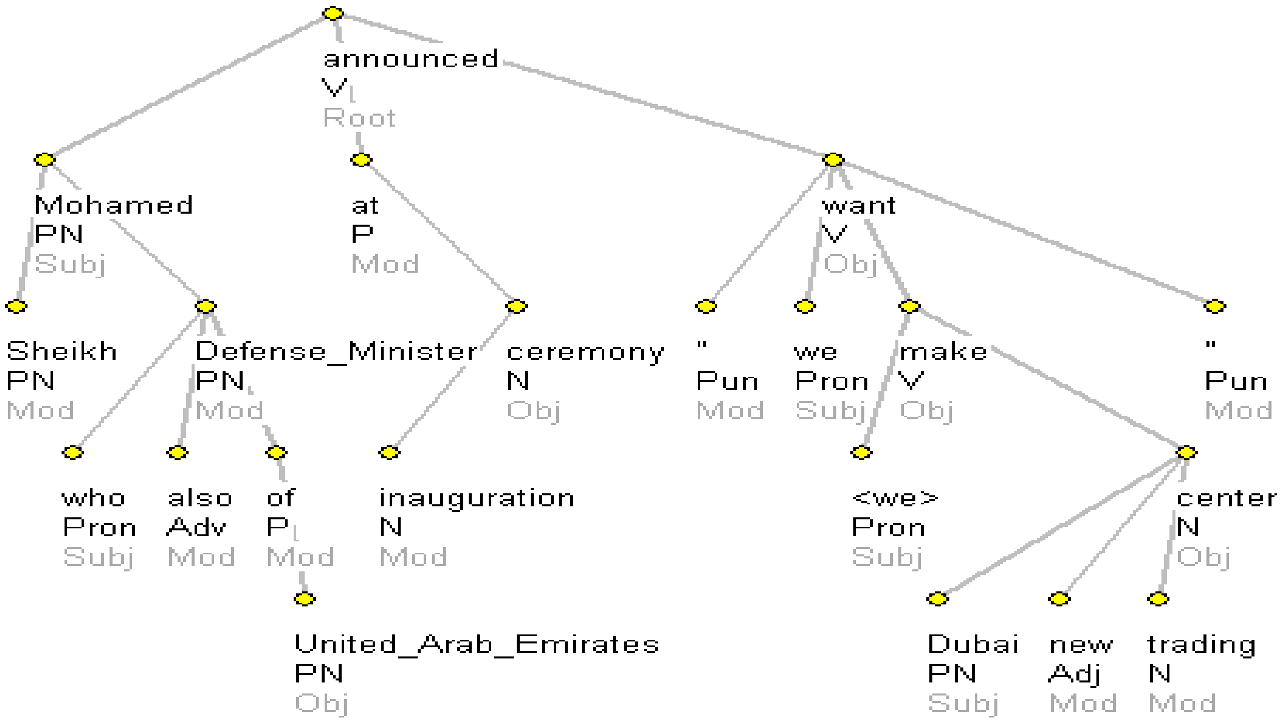


Figure 1: **An IL0 Dependency Tree**

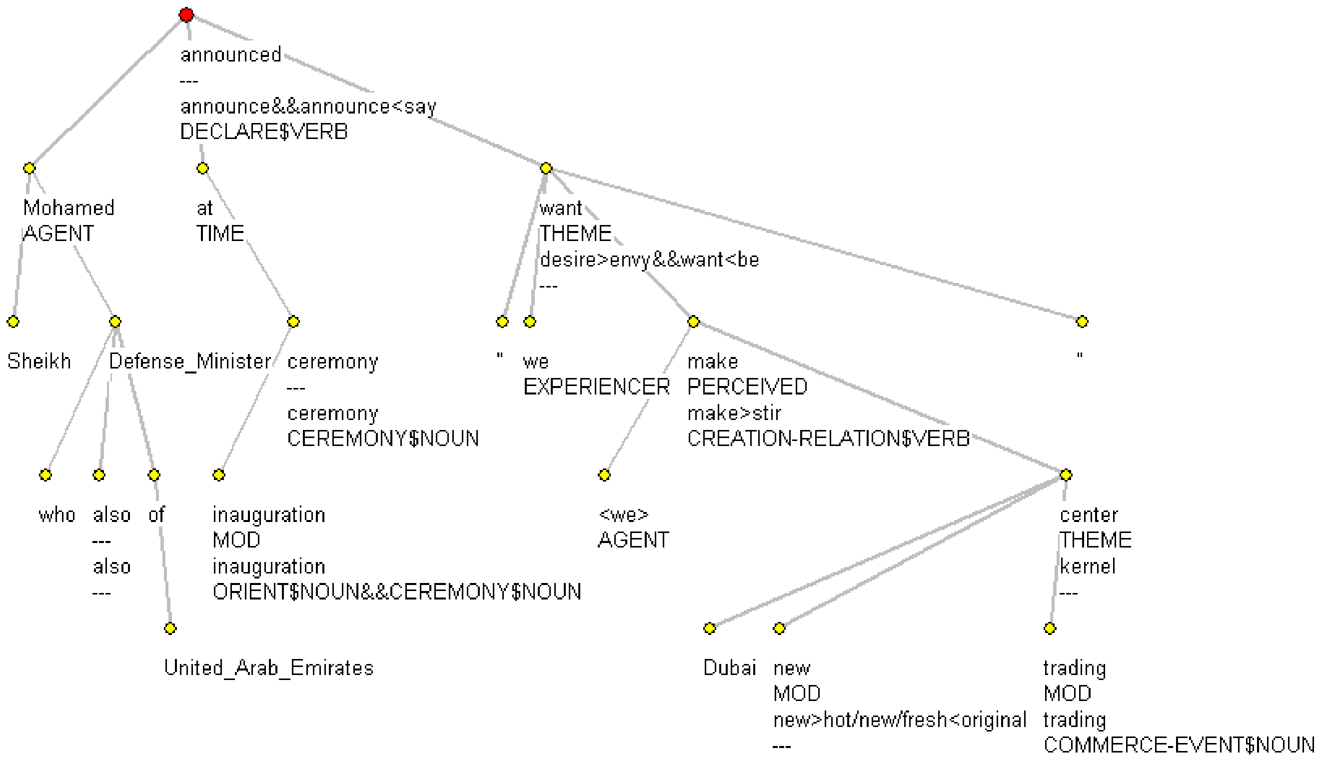


Figure 2: **An IL1 Representation**

Building a computational lexicon and ontology with FrameNet

Myroslava O. Dzikovska*, **Mary D. Swift**†, **James F. Allen**†

* Human Communication Research Centre
University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, United Kingdom
m.dzikovska@ed.ac.uk
† Department of Computer Science
University of Rochester, Rochester, NY 14627-0226
{james,swift}@cs.rochester.edu

Abstract

This paper explores FrameNet as a resource for building a lexicon for deep syntactic and semantic parsing with a practical multiple-domain parser. The TRIPS parser is a wide-coverage parser which uses a domain-independent ontology to produce semantic interpretations in 5 different application domains. We show how semantic information from FrameNet can be useful for developing a domain-independent ontology. While we used FrameNet as a starting point for our ontology development, we were unable to use FrameNet directly because it does not have links between syntax and semantics, and is not designed to include selectional restrictions. We discuss changes that needed to be made to the FrameNet frame structure to convert it to our domain-independent LF Ontology, the additions we made to FrameNet lexicon, and the resulting differences between the systems.

1. Introduction

This paper explores FrameNet(Johnson and Fillmore, 2000) as a resource for building a lexicon for deep syntactic and semantic parsing with a practical multiple-domain parser. Semantic corpus annotation such as FrameNet is an important way to ensure reliability and ease of use of semantic representations. Achieving inter-annotator agreement results in semantic classes that can be reliably distinguished by humans, unlike, for example, WordNet synsets (Miller, 1995), which are often difficult to differentiate for human annotators. An open question, however, is whether the FrameNet classes and frame elements can be obtained and used automatically. There has been some work in this area, in particular, on learning FrameNet frame elements from corpora (Gildea and Jurafsky, 2002) and on using them in the SMARTKOM project (Chang et al., 2002). However, the extent to which FrameNet annotations will be usable in practical applications is still an open issue.

In this paper, we describe our experience in using FrameNet in the process of building a multi-domain conversational dialogue system. The TRIPS system is a dialogue assistant which has been applied to 5 different application domains. Our lexicon uses frame structures as a domain-independent semantic representation, and therefore FrameNet is an attractive source of semantic information. We used the FrameNet classes as a starting point for our ontology development.

We made our top-level ontology for parsing consistent with the FrameNet ontology, and this helped us by identifying the verb classes that can be reliably distinguished by human lexicon developers when defining entries in a computational lexicon. FrameNet also provides semantic roles, but it does not provide links between lexical entries and the frames, and it does not contain selectional restrictions. In creating those links, we changed the representation in order to simplify lexicon maintenance, making it easier to define syntax-semantics mappings and selectional restrictions in the lexicon and ontology.

We describe the needs of a wide coverage parser and grammar using the TRIPS parser as a realistic example in Section 3; we then discuss the changes that needed to be made in our domain-independent ontology from the FrameNet formalism (Section 4), and compare the resulting lexicons (Section 5). Our experience can be useful for the designers of other NLP systems, as well as guidance for further development of semantic annotation schemes which can be used in natural language understanding systems.

2. Background

Typically, a parsing and semantic interpretation system requires an ontology as a source of semantic types and a lexicon with the following information for every word:

- Syntactic features;
- Subcategorization frames;
- Semantic representation;
- For every subcategorization frame, the correspondence between syntactic and semantic structures.

A number of lexicon and ontology projects provide parts of the necessary information. Among the resources frequently used for natural language processing tasks are syntactic features and subcategorization frames in COMLEX (Macleod et al., 1994), word senses in WordNet (Miller, 1995) and EuroWordNet (Vossen, 1997), and semantic representations of world knowledge in CyC (Lenat, 1995). Of particular interest to our project is FrameNet, which provides semantic frame representations based on the analysis of corpus examples, and VerbNet (Kipper et al., 2000), which provides subcategorization frames and correspondence between those and verb semantics.

Though each of these lexicons and ontologies provides some of the requirements we listed above, there is no single resource which integrates all the information necessary for parsing. We found that FrameNet and VerbNet entries were the most useful for our purposes, as we discuss in more detail in the following sections. Integration of all the required information presents significant challenges, primarily in making sure that during parsing the correct semantic type can be chosen for the word, and correct semantic argument labels are assigned to all its arguments. We found that in a practical system simplifications may be necessary to achieve efficiency and accommodate the fact that the system cannot rely on the world knowledge available to humans annotating corpus examples.

3. The TRIPS parser

Before describing the use of FrameNet in the TRIPS ontology, we discuss in more detail the TRIPS parser and its representational requirements. The TRIPS parser is a chart parser which utilizes 3 main knowledge sources: a wide-coverage domain-independent grammar, a domain-independent lexicon, and a domain-independent ontology, as elaborated below.

Our wide-coverage domain-independent grammar has been developed and tested in 5 different spoken dialogue domains. It has been tested on human-human speech corpora (Swift et al., 2004), and provides good coverage of complex structures including gaps, relative clauses, complex noun phrases *etc*. The grammar rules build up a domain-independent logical form used for discourse processing, discussed below.

Our domain-independent lexicon provides word definitions for the grammar. Each word definition has to include the syntactic features, subcategorization frames and the linking between syntax and semantics to allow the parser to build the logical form. While our lexicon is not yet as large as the projects like WordNet, it offers wide coverage in our domains, which results in many ambiguous lexical entries. On average, there are 1.26 syntax-semantics patterns per word, and for verbs this figure is 1.60. The ambiguity in lexical entries necessitates the development of mechanisms for semantic disambiguation. In our project, we use domain-independent selectional restrictions expressed as feature sets as our primary disambiguation mechanism.[1]

Finally, our domain-independent ontology, which we call the *LF Ontology*, is the source of semantic types that provides the semantics for entries in the domain-independent lexicon.It includes the repository of all semantic types defined in the system, as well as selectional restrictions to help disambiguation. The relationship between the LF Ontology and FrameNet is discussed in the rest of the paper.

Using the domain-independent grammar and lexicon linked to the LF ontology, the TRIPS parser produces a domain-independent logical form. This is a flat unscoped neo-Davidsonian representation, using event arguments and semantic roles. It is similar to QLF (Alshawi et al., 1991) and Minimal Recursion Semantics (Copestake et al., 1995) in that it uses identifiers to link the (non-recursive) terms together. An example representation for *load the truck with oranges* is shown in Figure 1.

(SPEECHACT sa1 SA_REQUEST :content e123)
(F e123 LF::Filling*load :Agent pro1 :Theme v1 :Goal v2)
(IMPRO pro1 LF::Person :context-rel *YOU*)
(THE v1 (SET-OF LF::FOOD*orange))
(THE v2 LF::Vehicle*truck)

Figure 1: The LF representation of the sentence *load the oranges into the truck.*

The representation identifies the sense of the main verb *load* as an instance of concept LF::Filling, corresponding to the FrameNet frame *filling*. Moreover, it identifies *oranges* as a :Theme of the filling action, that is, the object being moved, and *truck* as a :Goal of the filling action. Since it is an imperative, the parser also infers an implicit pronoun as a subject of the sentence, corresponding to the :Agent role.

Unlike traditional QLF representations, which typically use n-place predicates, we use named arguments (which we call **semantic roles**) in our representations, as it is done in neo-Davidsonian representations and description logic. It makes it easier to provide uniform representations connected to different syntactic alternations (*e.g.*, the only difference between *the window broke* and *the hammer broke the window* is that the former does not have an instrument role filled in), and we hope to be able to use the role-based representations for some syntactic generalizations, as discussed in Section 6.

In this example, the role names defined for LF::Filling are exactly the same as those for the filling frame in FrameNet. This is not always the case, and the need to change the role structure for the LF ontology is discussed in Section 4.

In the rest of the section, we discuss the specific requirements the parsing system places on its lexicon and domain-independent ontology. These are the motivations for choosing FrameNet as an appropriate domain-independent ontology, but also for the changes needed for its use in a computational system.

3.1. Ontology design considerations

When providing the semantic information for parsing described in the previous section, the development of our system is influenced by two main goals: support for efficient wide-coverage parsing, and also fast lexicon acquisition. The first requirement means that the information provided in the lexicon should be sufficient to parse sentences encountered in the domains quickly. Therefore, we need to reduce the parser search ambiguity whenever possible while maintaining the wide coverage of the system. The second requirement means that new word definitions should be possible to define automatically, or, if defined by hand (as we are currently doing), the information necessary to define a lexical entry should be easy to obtain. Either the lexicon developer should be able to define a word from the examples of other similar words already defined in the lexicon, or, if no similar words were defined previously, the relevant information should be easy to obtain from online

[1]Another option would be statistical disambiguation, but it proves difficult for spoken dialogue domains, where corpus data are difficult and costly to collect. We have demonstrated that domain-independent selectional restrictions improve parsing speed and accuracy in our lexicon (Dzikovska, 2004).

resources. In particular, we would like to be able to obtain the semantic class of the word from FrameNet, and then find a way to link the syntactic structure with the frame elements.

Specifically, our decisions about the ontology were influenced by the following considerations:

- **The level of abstraction**. The semantic predicates used during interpretation must be specific enough to allow the system to draw reasonable inferences about the world. For example, using the same predicate MOVE to denote verbs such as *run*, *walk* and *drive* loses important distinctions between the meanings, such as speed and whether a vehicle is involved. At the same time, we want the semantic predicates to be such that the system has a reasonable chance of selecting the correct sense during the interpretation process. For example, WordNet lists 16 senses for the verb move, including "change location", "move as so to change position", "cause to move" and "change residence". Disambiguation between those senses is difficult even for human annotators, and extensive reasoning about context is necessary to select the correct sense is not feasible given the current state of the art for dialogue systems. FrameNet offers the appropriate level of abstraction for word senses, as discussed below.

- **The compositionality of meaning representations** In a domain-independent ontology, we would like the meanings of the complex phrases to be compositional, built from the meanings of their components. For example, consider a sentence *Submit a purchase order*. In a system that only knows about submitting purchase orders, this is an atomic action. Therefore, it can potentially be represented as a single concept in the system ontology, SUBMITPURCHASEORDER(P), where *p* is parameter which corresponds to the purchase order to submit. This representation may be the most efficient for domain reasoning, but if there are other things that can be submitted, such as proposals or application, this leads to a proliferation of concepts: SUBMITPROPOSAL(P), SUBMITAPPLICATION(A). This is not a desirable situation for parsing, because it results in additional ambiguity in constructions like *submit it*, which then become multiply ambiguous between interpretations with different possible meanings of *submit*.

- **Efficiency**. For a dialogue system, the speed of interpretation is crucial for effective operation, and we would like to use as much semantic information as possible during parsing to speed up and improve disambiguation.

- **Syntax-semantics mappings**. In order to use an ontology in a parsing system, we need to be able to link the syntactic structures to corresponding ontological representations. This needs to be specified in our lexicon; ideally, it should be available directly from a lexicon developed together with the ontology, otherwise, it needs to be acquired later, during construction of our parsing lexicon. The properties of the ontology, including the level of abstraction and compositionality, and also the arguments associated with each type, should facilitate syntax-semantics mapping. For example, if an ontology requires collecting phrases like *from Pittsford* and *to Avon* into a single PATH frame, then special handling for path adverbials has to be implemented in the grammar, adding to the complexity of the system. FrameNet has simple frame elements, which are easy to obtain during parsing. However, there are issues with disambiguating them, as discussed in Section 4.

In our analysis, the FrameNet frames offer the right level of abstraction for a computational system. The guideline we use in our lexicon is to consider two senses of a word different only if we can distinguish them automatically (i.e. based on subcategorization patterns and domain-independent selectional restrictions) in most circumstances. Because FrameNet was developed based on corpus examples, with frames which can be reliably distinguished by human annotators, the frame structures offer the right level of abstraction as word senses in a computational system. In addition, because the frames are expected to cover a large number of examples, they offer a good level of compositionality, representing generic situations with parameters to be filled in the roles.

3.2. Syntax-semantics templates and the LF Ontology

FrameNet is missing a crucial piece of information - syntax-semantics mappings, which are necessary to obtain our logical form representations. An example lexical entry in our lexicon is shown in Figure 2. It defines the verb *load* and 2 syntactic patterns. The pattern defined by AGENT-THEME-GOAL-TEMPL encodes the information that in a sentence *Load the oranges into the truck* the (implicit) subject will fill the :Agent role, the direct object is a noun phrase which will fill the :Theme role, and the prepositional complement is a prepositional phrase using the preposition *into*, and filling the :Goal role.

The syntax-semantic mappings have to be defined for all lexical entries. In defining them, we encounter issues with semantic role names similar to those we encountered when defining appropriate word senses. When a mapping between syntactic and semantic arguments is defined, the semantic arguments in the given frame must be defined on a level of abstraction appropriate to draw inferences about the world, but possible to disambiguate based on syntactic structure and selectional restrictions. We found that some FrameNet frame elements did not satisfy those criteria, which necessitated changes to the ontology structure discussed in Section 4.

The syntax-semantics templates are combined with selectional restrictions in our ontology to provide semantic disambiguation. Selectional restrictions are not part of the FrameNet database, we added them to our LF representation to provide the parser with the information necessary for disambiguation. For example, the LF ontology entry for LF::Filling is shown in Figure 3. It is a subtype of a more general LF::Motion frame (the addition of hierarchical structure to the LF Ontology is discussed in the next

```
(a) (load
      (wordfeats (morph (:forms (-vb))))
      (senses
        ((LF-Parent LF::Filling)
         (TEMPL AGENT-THEME-GOAL-TEMPL)
         (Example "Load the oranges into the truck"))
        ((LF-parent LF::Filling)
         (TEMPL AGENT-GOAL-THEME-TEMPL)
         (Example "Load the truck with oranges"))
        ) ))

(b) (AGENT-THEME-GOAL-TEMPL
      (SUBJ (NP) Agent)
      (DOBJ (NP) Theme)
      (COMP (PP (ptype into)) Goal optional)

(c) (AGENT-GOAL-THEME-TEMPL
      (SUBJ (NP) Agent)
      (DOBJ (NP) Goal)
      (COMP (PP (ptype with)) Theme)
```

Figure 2: Defining words in the lexicon (a) Lexicon definitions for the verb *load* in the LF::Filling sense; (b) The template used to define the syntactic pattern for *load the oranges into the truck* (c) The template used to define the syntactic pattern for *load the truck with oranges*

section). As such, it inherits a basic set of arguments, which are :Theme, :Source and :Goal.

```
(define-type LF::Motion
 :sem (Situation (Aspect Dynamic))
 :arguments
   (Theme (Phys-obj (Mobility Movable)))
   (Source (Phys-obj))
   (Goal (Phys-obj))

(define-type LF::Filling
 :parent LF::Motion
 :sem (Situation (Cause Agentive))
 :arguments
   (Agent (Phys-obj (Intentional +)))
   (Goal (Phys-obj (Container +))))
```

Figure 3: LF type definitions for LF::Motion and LF::Filling. In the lexicon, feature vectors from LF arguments are used to generate selectional restrictions based on mappings between subcategorization frames and LF arguments.

The LF definitions contain selectional restrictions on the arguments expressed in terms of semantic feature sets. Features encode basic meaning components used in semantic restrictions, such as form, origin and mobility for physical objects. For example, the :Theme argument is defined as *Phys-obj (Mobility Movable)* to reflect the fact that it has to be a mobile object, as opposed to generally fixed objects such as cities and mountains. LF::Filling places an additional restriction on its :Goal, requiring that it has to be a container.

The semantic feature set we utilize is a domain-independent feature set developed using EuroWordNet (Vossen, 1997) as a starting point, and extended by incorporating lexico-syntactic generalizations from other linguistic theories (Dzikovska et al., to appear). The set of features is limited to 3-10 per word. The small size of the feature set provides the lexicon developers with an easy to use framework in which to express semantic properties of words for selectional restrictions, because each word only needs to be classified along a small set of dimensions. However, the small feature set size limits the expressivity of the selectional restrictions, so not every possible restriction can be captured in it (see Section 4 for an example).

In our work on domain-independent lexicon development we found this approach a useful compromise. While it is small enough to keep lexicon development simple, it covers enough of the basic properties of words to significantly improve parsing speed and accuracy in two evaluation domains (Dzikovska, 2004). Selectional restrictions as feature sets offer further advantages in terms of efficient implementation and domain customization (Dzikovska et al., 2003). Therefore, in our lexicon we distinguish the word senses and semantic arguments which can be disambiguated based on syntactic structure and selectional restrictions expressible in terms of our feature set. This has a direct impact on the decision to simplify frame role structures discussed in the next section.

4. Adapting FrameNet for the TRIPS LF Ontology

We made two major changes to our ontology that diverged from FrameNet representation: we added a hierarchical structure and reduced the number of distinct frame elements (which we call roles). The FrameNet ontology is mostly flat, even though it contains many frames subsuming verbs that have identical argument structures. While FrameNet is designed to represent the hierarchies of frames, currently only about one-third of the frames in FrameNet inherit from other frames (Gildea, personal communication). In cases where frames included similar words but reflected finer meaning distinctions, we collected them under a common parent. For example, *Suasion1*, *Suasion2* and *Suasion3* include a group of verbs such as *encourage, convince, induce,* which have the same set of roles, but the difference in meaning comes from whether the addressee forms an intention to act. From the point of view of argument structure and selectional restrictions these frames are identical, so we collect them under a general parent and use the same set of selectional restrictions.

Table 1 shows the statistics about the number of LF types at different levels of our hierarchy. Level 0 types are types that do not inherit from anything, level 1 are types with 1 parent, and so on. The first 2 levels in our ontology were created artificially, because we needed special types for parsing: a unique root in the ontology, a type which unifies with nothing else ("-"), and another type which unifies with anything but "-". Thus, the contentful entries start at level 2, and we have 7 root entries that do not inherit from anything, 103 entries at depth 1. The majority of the types

Level	Frame Count
0	1
1	2
2	7
3	103
4	170
5	207
6	103
7	44
8	10
9	9

Table 1: The number of LF types at different levels of our LF hierarchy

we use are at depth 2 or 3 (170 and 207 respectively), but the hierarchy goes up to 6 levels deep, mostly in the parts of ontology where objects are classified.

In the process of developing our ontology, we had to add types to support problem solving and planning actions, which were absent in the version of FrameNet we utilised: FRAMENET II Release 1.0. For example, it did not have a classification for the word *need*, which occurs frequently in our dialogues, so we defined a new LF::Necessity frame in our lexicon.[2] Other words common in our task-oriented domains but not currently found in FrameNet are *cancel*, *revise*, *schedule*. Sometimes words were defined within FrameNet, but we needed to define additional senses because the FrameNet frame did not cover the common usage in our domain. For example, the word *change* is defined only as an instance of frame *Transformation*, where an entry is transformed into something else, like in *change the rabbit into a hat*. In one of our domains, a frequent usage is *Change the dial to VDC* (i.e., change the setting, but not the dial itself). So we created a new LF::Change-state frame to account for this sense. Similarly, the adjective *open* is defined as *Candidness* in FrameNet, corresponding to usages like *She was open with us about the party*, with synonyms such as *candid*, *forthright*, *etc*. In our domains, *open* has to do with physical accessibility, *The route is open*, or *there is an open door*. These senses are not suitable for the words grouped in the *Candidness* frame, thus we established the LF::Openness frame to account for them.

The hierarchical structure provides a level of generalisation in the ontology that makes it easier to include and maintain selectional restrictions. For that purpose, we also simplified the frame elements in our ontology. FrameNet utilises situation roles, so a *driving* situation involves a *driver* role, whereas the *communication* situation has a *communicator*. However, these roles may be seen as instances of a general *agent* role, which is an intentional being doing the action. A limited number of role names simplifies the inheritance in the LF Ontology by allowing us to define a general restriction (e.g., agents are intentional beings) high in the hierarchy tree.

For purposes of mapping between syntax and semantics, a smaller number of role names facilitates the definition of these mappings, because it creates opportunities for generalisation. For example, many motion verbs will use exactly the same set of syntax-semantics mappings, and not having the distinctions between "driver" and "self-mover" makes it easier to add new verbs by example.

More importantly, we found some frame elements too specific or too dependent on pragmatic information to be distinguishable during parsing. For example, the frame *closure* defines 2 separate frame elements: "Container-portal", for example *flap* in *Close the tent flap*, and "Containing-object", *coat* in *buttoned her coat*. Both can occur as direct objects of relevant verbs. Human annotators are able to distinguish those based on common sense knowledge. For parsing, however, selectional restrictions expressed with a limited set of semantic features are not specific enough to make this determination. Moreover, to our knowledge there is no reasoner able to make this distinction in a domain-independent manner. Therefore, we made the decision to define a more general :Theme role for our LF type LF::Closure, which covers both those semantic arguments. The relevant distinctions, if necessary, can be made by the domain specific reasoners using our customization mechanisms (Dzikovska et al., 2003).

The decision to use a reduced, more general set of roles has an advantage for fast acquisition of lexical entries. Many linguistic theories make syntactic generalisations based on semantic classes (see for example (Levin, 1993), (Jackendoff, 1990)). While we do not use such generalisations yet, we designed our ontology to facilitate those in the future, as discussed in Section 7. For example, the VerbNet lexicon defines the verb *close* with *agent*, *patient*[3] and *instrument* roles, and defines the corresponding subcategorization frames and syntactic variations. This generalization is only possible with more general role names, and we hope to use it in the future to speed up the development of syntax-semantics mappings.

5. Evaluation

In this section, we present statistics about our current lexicon, and how it compares with the FrameNet ontology.[4] Currently, our LF Ontology contains 656 LF types, corresponding to different concepts. The complete statistics for our lexicon is shown in Table 2. We have 2446 words total in our lexicon, 1999 of which are open class words - adjectives, nouns, verbs and adverbs, with 2248 different word senses. The system uses 37 semantic roles, considerably fewer than FrameNet, which has 554 frame elements.

We compared our lexicon with the FrameNet version 1.0. Table 3 shows the number of lexical items for each part of speech which were defined in both lexicons, in TRIPS lexicon only, and in FrameNet lexicon only.

It is interesting to note that while FrameNet is much larger in size than the TRIPS lexicon, there's a considerable number of lexical items, in all categories, which do

[2] *need*, and other words we cite in our examples, are also missing from the latest web version of FrameNet, FrameNet II release 1.1.

[3] which corresponds to our :Theme.

[4] The FrameNet statistics in this section are from FrameNet II Release 1.0 unless otherwise noted.

POS	Count	Senses	Synt. var.	Comment
ADJ	422	1.07	1.12	Adjectives
N	875	1.06	1.09	Nouns
ADV disc	36	1.08	1.11	Discourse adverbials
ADV	221	1.32	1.55	Adverbs (including adverbial prepositions)
V	490	1.29	1.71	Verbs
NAME	22	1.00	1.00	Names
PUNC	10	1.00	1.00	Punctuation signs
UTTWORD	121	1.01	1.01	Discourse words like OK, yes, yeah, etc.
OTHER	249	1.02	1.04	Other parts of speech for functional words, including ART, PREP, QUAN, CONJ, PRO, NUMBER
Total	2446	1.12	1.26	

Table 2: Lexicon statistics in our system

POS	Common	Trips only	FrameNet only
Adj	114	308	1072
N	285	582	2479
V	225	232	1774

Table 3: Lexicon statistics

not overlap between those lexicons. Part of the problem is that the comparison is with an older version of FrameNet II (release 1.0) and the current release (1.1) is much richer. However, manual inspection of the data and comparison with the release 1.1 data available on the Web still shows significant non-overlapping areas. For verbs, these include

- Verbs dealing with plans and goals: *achieve, accomplish, complete etc.*
- Verbs dealing with intentions and permissions: *need, authorize, assume, trust etc.*
- Verbs dealing with mutual understanding in a conversation: *recap, reformulate, misunderstand*
- Verbs with particles common in spoken language: *look for, back up, dig out etc.*

Verbs with particles do not appear to be consistently annotated in FrameNet, so the number of verbs listed as in TRIPS but not FrameNet may include some of those that in FrameNet are annotated as senses belonging to a verb ignoring the particle. When we excluded verbs with particles from the counting, the number of verbs defined in TRIPS but not FrameNet was 164, still a substantial difference. Moreover, when a particle is not included with the verb annotation, it poses a significant problem for a parser, because particles provide important syntactic clues during parsing and disambiguation, and loss of this information adds ambiguity to the process.

We did not analyze in detail the differences between nouns and adjectives, but based on several spot-checks, it appears that this is an area that has been developed in FrameNet II Release 1.1, which now defines many common adjectives and nouns such as colour names and common foods. The biggest differences appear to be in words that are essential for coverage in our domains, which are transportation and computer purchasing. Therefore, TRIPS defines the names for many physical objects such as *bus, dvd, cd-drive* which are not part of the FrameNet lexicon. This points to the issue we need to deal with in our future work. Our data suggest that the text corpora that are the basis of FrameNet are quite different from the task-oriented spoken dialogue corpora, and that's why there are a number of words important in our domains which are currently not included in the FrameNet database. If the LF types for those are added to our ontology, we need to address synchronization issues with further FrameNet updates.

5.1. Role structure evaluation

As discussed above, the names of semantic roles, much as the names of the frames themselves, have to be at the right level of abstraction in order to facilitate a connection with syntax. Therefore, during the development of the LF Ontology we needed to simplify the FrameNet role structure. The FrameNet version we evaluated contained 554 frame elements. We discussed in Section 4 the problems that this caused in efficiently acquiring lexical entries and in frame element disambiguation. In contrast, TRIPS has 37 roles used in subcategorization frames. This number is considerably easier to manage in defining syntax-semantics mappings, and for disambiguation.

The TRIPS role set, though developed independently, is similar in size and structure to the role set in another semantic lexicon, VerbNet (Kipper et al., 2000), which also aims to link syntactic and semantic structure. A detailed comparison can be found in (Dzikovska, 2004). In brief, VerbNet has 28 roles, 8 of which are the same as those used in the TRIPS LF lexicon. We did not conduct the formal evaluation of the consistency of the rest of the role set, but, generally speaking, the rest of the role sets intersect, but VerbNet makes finer distinctions in some cases (splitting :Theme into *theme* and *patient*). In addition, TRIPS contains semantic roles for classifying adjective, adverb and noun arguments, not covered by the VerbNet lexicon. We plan to resolve the differences and use VerbNet selectional restrictions and syntactic patterns to extend coverage of our verb lexicon as part of our future work.

In comparing the role sets it is also important to note that FrameNet intends to cover all parts of the sentence relevant to the event, be they verb arguments or adjuncts expressed by adverbs or even clauses. This results in some highly specific frame element names, such as “Abundant-entities”, “arguer” or “manifestation-of-bias”, each of which occurs only in a single frame. In our evaluation, 313 of frame elements appeared in one frame only. At the same time, the 6 most common frame elements, “Manner”, “Time”, “Degree”, “Place”, “Means” and “Purpose”, are handled as adverbial senses in the TRIPS lexicon, with the exception of a small number of verbs which subcategorize for them. For example, usually :Time-duration role is realized by an adverbial, as *for 5 minutes* in *She completed the task in 5 minutes*. But for 2 frames, LF::Take-time and LF::Leave-time explicitly subcategorize for it as a direct object, *e.g.,* It takes 5 minutes to complete. In the TRIPS lexicon there are 4 roles which appear with only 1 frame, and 2 of those are realized as adverbials in other constructs, so they are not unique labels for a given frame, but just exceptional cases of arguments typically handled by adverbials.

The large number of role names difficult to disambiguate for the parser is the main reason why we were unable to use FrameNet directly in our lexicon. The distinction between subcategorized arguments and adjuncts (generally coming from adverbials) is very important in parsing and semantic disambiguation, and that FrameNet does not mark it in its frame element structure makes it difficult to use directly in a practical NLP parser.

6. Future Work

Our work highlights both the usefulness of FrameNet as a basis for building a computational ontology and lexicon, and its limitations as a source representation for parsing. FrameNet provides word meanings which can be reliably distinguished by humans, which makes lexicon development easier, and frame representations are convenient for natural language processing because they are easy to obtain from linguistic structure and allow us to encode optional arguments. However, to facilitate connections to syntax and allow for possible syntactic generalisations, we needed to modify the information available in FrameNet by adding hierarchy and using a smaller set of role names.

In the future we plan to include syntactic generalizations based on syntactic alternations as done in VerbNet (Kipper et al., 2000). Currently there is no direct mapping between TRIPS and VerbNet classes. Our ontologies were developed independently, because the VerbNet database was unavailable at the time; additionally, VerbNet is not designed to cover other word classes, such as nouns and adjectives, and we developed our lexicon to provide semantic roles representations for all open-class words. As mentioned above, our analysis shows a significant overlap between our semantic roles and VerbNet roles. In our evaluation, we also noted a “core” set of roles, including “Agent”, “Cause”, “Source”, “Goal”, “Theme”, which, after the general frame elements typically implemented by adverbials we mentioned before, are the most frequent frame elements used in FrameNet. This raises issues of standardisation and developing a general set of roles suitable both for semantic analysis and for syntactic generalisations, and we are working on mappings between the TRIPS and VerbNet roles, and possibly between the TRIPS and FrameNet roles.

We also need to address the coordination between FrameNet and TRIPS ontologies. Our ontology is based on FrameNet, but it is not synchronised with the current FrameNet version, because of the changes and additional information necessary in our representations. Currently, instead of trying to synchronize our ontologies directly, we are working on a learning module which uses FrameNet and other resources to propose meanings of novel words as an aid to human lexicon developers.

7. Conclusions

In conclusion, this paper discusses FrameNet as a source of semantic information for a deep syntactic parser. Our wide coverage parser needs an ontology as a source of domain-independent word senses, and FrameNet provides a well-documented source of reliably distinguishable semantic classes. For use in our practical dialog system, however, we needed to streamline aspects of the FrameNet data for efficiency. There remain open questions, especially the extent to which such streamlining can be handled automatically as both systems develop in parallel, which need to be addressed in future work.

8. Acknowledgments

We thank Mark Core and Dan Gildea for useful comments and discussion for this paper. This material is based upon work supported by a grant from The Office of Naval Research under grants number N00014-01-1-10115 and N00014-03-1-0780 under subcontract to University of West Florida, from DARPA, under research grant number F30602-98-20133, DARPA under grant number NBCHD030010 via subcontract to SRI number 03-000223 and The National Science Foundation under grant number E1A-0080124. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of above named organizations.

9. References

Alshawi, Hiyan, David M. Carter, Björn Gambäck, Stephen G. Pulman, and Manny Rayner, 1991. Transfer through quasi logical form: A new approach to machine translation. Technical Report T91020, SICS, Stockholm, Sweden.

Chang, N., J. Feldman, R. Porzel, and K. Sanders, 2002. Scaling cognitive linguistics: Formalisms for language understanding. In *Proceedings of the First International Workshop on Scalable Natural Language Understanding*. Http://smartkom.dfki.de/Vortraege/icsi-ai2.pdf.

Copestake, Ann, Dan Flickinger, Rob Malouf, Susanne Riehemann, and Ivan Sag, 1995. Translation using Minimal Recursion Semantics. In *Proceedings of the 6th. International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95)*. Leuven, Belgium.

Dzikovska, Myroslava O., 2004. *A Practical Semantic Representation For Natural Language Parsing*. Ph.D. thesis, University of Rochester.

Dzikovska, Myroslava O., Mary D. Swift, and James F. Allen, 2003. Integrating linguistic and domain knowledge for spoken dialogue systems in multiple domains. In *Proceedings of IJCAI-03 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*. Acapulco.

Dzikovska, Myroslava O., Mary D. Swift, and James F. Allen, to appear. Customizing meaning: building domain-specific semantic representations from a generic lexicon. In Harry Bunt (ed.), *Computing Meaning, Volume 3*, Studies in Linguistics and Philosophy. Kluwer Academic Publishers.

Gildea, Daniel and Daniel Jurafsky, 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.

Jackendoff, R., 1990. *Semantic Structures*. The MIT Press.

Johnson, Christopher and Charles J Fillmore, 2000. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings ANLP-NAACL 2000*. Seattle, WA.

Kipper, Karin, Hoa Trang Dang, and Martha Palmer, 2000. Class-based construction of a verb lexicon. In *Proceedings of the 7th Conference on Artificial Intelligence (AAAI-00) and of the 12th Conference on Innovative Applications of Artificial Intelligence (IAAI-00)*. Menlo Park, CA: AAAI Press.

Lenat, Douglas B., 1995. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38.

Levin, Beth C., 1993. *English Verb Classes and Alternations: a Preliminary Investigation*. Chicago, IL: University of Chicago Press.

Macleod, Catherine, Ralph Grishman, and Adam Meyers, 1994. Creating a common syntactic dictionary of English. In *SNLR: International Workshop on Sharable Natural Language Resources*. Nara.

Miller, G., 1995. Wordnet: A lexical database for English. *Communications of the ACM*, 38(5).

Swift, Mary D., Myroslava O. Dzikovska, Joel R. Tetreault, and James F. Allen, 2004. Semi-automatic syntactic and semantic corpus annotation with a deep parser. In *Proceedings of LREC-2004*.

Vossen, Piek, 1997. Eurowordnet: a multilingual database for information retrieval. In *Proceedings of the Delos workshop on Cross-language Information Retrieval*.

Towards Using FrameNet for Question Answering

Gerhard Fliedner

DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken
and
Saarland University
Computational Linguistics
D-66123 Saarbrücken
Gerhard.Fliedner@dfki.de

Abstract

We describe an approach to Question Answering (QA) that is centred around the idea of translating both the text collection and the user's question to the system into a lexical semantic representation based on FrameNet. As this representation abstracts away from a number of issues of text surface, this leads to a meaning-oriented search. A system for automatically deriving FrameNet representations from German texts has been implemented and is described. We then turn to implementation issues in the envisaged implementation of a QA system based on this translation.

1. Introduction

Most Question Answering systems today use the following approach (Hirschman and Gaizauskas, 2001): After processing the user's question, a document search is done using keywords from the question. The actual search is thus done based on surface words, mostly using indexing techniques. The retrieved candidate documents or passages are then processed either statistically (Zhang and Lee, 2003) or with 'deeper' linguistic methods (Moldovan et al., 2002a; Moldovan et al., 2002b; Hovy et al., 2001; Elworthy, 2000). The answer that is returned to the user is mostly a text snippet from the original text; more recent approaches also employ generation techniques. These systems are in general very efficient, but have a number of potential disadvantages, such as imperfect precision and high reliance on answer redundancy (Light et al., 2001).

We propose a meaning-oriented approach to Question Answering for German that is centred around the idea of automatically annotating the text collection to be searched with lexical semantic structures based on FrameNet. The user's questions are also automatically translated into meaning-oriented FrameNet representations and matching is done directly on these structures.

We want to make use of the idea of pre-processing the text to enrich it with more structure. As basis for the annotation format, we have chosen FrameNet, using especially its concept of semantic valency to abstract away over linguistic issues on the text surface. By doing the FrameNet translation process 'off-line', i. e. at document indexing time, the actual search can efficiently be done at retrieval time over structured data.

We have implemented a system for annotating German text using a cascade of 'flat' parsers. Its FrameNet coverage is yet small, but will grow with the increasing coverage of FrameNet for German. This system is eventually to form the core of a meaning-oriented QA system, which is still in its design phase.

This paper is organised as follows: We first give a short overview of FrameNet and describe how it can be useful in QA. We give an overview of the system we have implemented for translating text into FrameNet. We then turn to a description of the planned QA system together with the open implementation issues, giving examples for the issues and the solutions we currently envisage.

2. FrameNet for Semantic Annotation

FrameNet is a lexical database resource containing valency information and 'abstract' predicates. English FrameNet is developed at the ICSI, Berkeley, CA[1] (Baker et al., 1998; Johnson et al., 2002). Development of a German FrameNet is currently underway within the context of the SALSA project[2] (Erk et al., 2003).

FrameNet is used to describe semantic roles and thus the relations between participants and objects in a situation as predicate-argument-structures: Frames describe prototypical situations with all the objects, participants and actions belonging to them (e. g., a purchase with buyer, seller, goods etc.). Therefore, words that are semantically similar will receive comparable descriptions with identical role labels. This comparability not only holds for synonyms, but also for antonyms, for converse relations (such as *buy* and *sell*) and also across parts of speech (e. g., verb *vs.* noun).

This representation is therefore especially suited as an abstraction level for applications where the surface wording is less important than the contents. This is the case for Information Management systems: In IE and IR, especially in QA, it is more important to extract or find the right contents; differences in wording are more often a hindrance in this process than not.

Matching against FrameNet structures instead of just words in an index would, for example, allow to find an answer to the question *"Who bought Mannesmann?"*, no matter if the relevant text passage originally reads *"Voda-*

[1] framenet.icsi.berkeley.edu/~framenet

[2] *The Saarbrücken Lexical Semantics Annotation and Analysis Project*, www.coli.uni-sb.de/lexicon/

fone took over Mannesmann in 2000.", "*Mannesmann was sold to Vodafone.*", or "*Vodafone's purchase of Mannesmann. . .* ".

When using approaches like Tectogrammatical Structures (Böhmová et al., 2003) or PropBank (Palmer et al., 2004), an additional inferencing step would be needed here to relate, for example, the Actor of buying and the Addressee of selling.

We plan to build a Question Answering system that makes use of this annotation process: The text collection to be searched is first annotated with FrameNet structures. This representation is stored in a way that can efficiently be accessed. The user's questions to the system are then translated into FrameNet structures using the same translation process. These representations are then matched to find the answer.

The annotation of the texts with FrameNet structures can roughly be compared with the Information Extraction task of template filling. However, FrameNet frames are linguistically motivated and grounded. They do not in general describe scenarios, as e. g., the Scenario Templates in the Message Understanding Conferences (Marsh and Perzanowski, 1998). Thus, frames are less domain dependent than IE templates and do not share the disadvantage of IE systems, namely that they have to be adapted for new domains in a rather labour-intensive process (Appelt et al., 1995).

3. Deriving FrameNet Structures from Text

Our system for annotating German texts with FrameNet structures uses a cascaded parsing process of flat parsers. This approach was introduced under the name of easy-first-parsing (Abney, 1996). It generally leads to a more robust overall system. Only an exemplary FrameNet coverage is implemented so far. Using the evolving German FrameNet lexicon, we plan to extend the coverage. In the following, we will shortly describe the modules employed in the derivation of FrameNet structures.

Texts are first tokenized and the words are morphologically analysed. We employ the Gertwol two-level morphology system that is available from Lingsoft Oy, Helsinki. It covers the German morphology with inflection, derivation and compounding and has a lexicon of approximately 350,000 stems.

The next step is the analysis of the sentence structure based on the German sentence topology. German sentences have a relatively rigid structure of topological fields that helps to determine the sentence structure. It is used to identify subordinate clauses and other clausal constructions as well as verb clusters using a context free grammar (Braun, 2003).

We then identify named entities in the text. Our method for named entity recognition is based on hand-crafted regular expressions, supported by a gazetteer with several thousand entries. At the moment, we recognize company names and currency expressions, as well as some person names. Our grammars are derived from the German NE grammars developed in our project (Bering et al., 2003).

Named entity recognition is followed by a chunker for noun phrases (NPs) and prepositional phrases (PPs) (Fliedner, 2002). The NP chunker uses extended finite state automata that allow the recognition of NPs embedded in other NPs – a common phenomenon in German. The chunker can integrate NEs recognised in the previous step into larger NPs, allowing for complex coordination and modification phenomena.

The results of the previous steps are put together into one overall structure that we have called PReDS (Partially Resolved Dependency Structure, (Braun, 2003)). PReDS is a syntacto-semantic dependency structure that retains a number of syntactic features (like prepositions of PPs) while abstracting away over others (like active/passive). It is therefore somewhat similar to Tectogrammatical Structures (Böhmová et al., 2003) or Logical Forms (Elworthy, 2000).

In a last step, the resulting PReDS structure is translated into FrameNet structures (Fliedner, 2004a). This translation uses weighted rules matching sub-trees in the PReDS. The rules can be automatically derived from a preliminary version of a FrameNet database containing valency information on an abstract syntactic level (using for example notions like deep subject to avoid different descriptions for active and passive on the one hand, but retaining prepositions as heads of PPs on the other hand).

Our system has not yet been systematically evaluated. We plan to conduct an evaluation in two steps: Firstly, to evaluate all modules contributing to the result separately as far as possible. We are currently investigating the possibility of using the TIGER corpus for German (Brants et al., 2002) as a 'gold standard', using an evaluation technique based on grammatical relations (Carroll et al., 2003). Secondly, for an end-to-end evaluation of the FrameNet annotation process, we plan to eventually use the FrameNet corpus for German under development in the SALSA project (Erk et al., 2003) as the gold standard.

First walk-through analyses for a limited number of sentences suggest that currently around 75 % of the sentences in our test corpus of business news texts receive a PReDS representation. For the target words that receive a FrameNet representation, in the majority of cases the 'core' frame elements (i. e. the central arguments such as BUYER and SELLER for COMMERCE) are correctly assigned, whereas 'non-core' elements (such as TIME and LOCATION) do significantly worse. We hope to improve this by the planned introduction of sortal information: On the one hand, we plan to extend the types of Named Entities that are recognised (introducing, e. g., date expressions). On the other hand, we intend to add sortal information from GermaNet, the German version of WordNet (Kunze and Lemnitzer, 2002).

4. Implementation Issues of the QA system

We now turn to the question how these structures can actually be used in Question Answering. We describe some of the modules and techniques needed and preliminary ideas on how to handle some of the issues connected with them. We will illustrate the discussion with some examples.

As a basis for the examples, we use one sentence from our newspaper corpus of business news texts (simplified from *Süddeutsche Zeitung*, 2 January 1995). The glossed sentence is shown in (1), two central frames automatically derived from it in (2) and (3).

(1)	*Lockheed*	*hat*	*von*	*Großbritannien*
	Lockheed	has	from	Great Britain
	den	*Auftrag*	*für*	*25 Transportflugzeuge*
	the	order	for	25 transport planes
	erhalten.			
	received.			

‘Lockheed has received an order for 25 transport planes from Great Britain.’

(2)

GETTING
TARGET: *erhalten*
DONOR: *Großbritannien*
RECIPIENT: *Lockheed* [ORGANISATION]
THEME: [1] *Auftrag für 25 Transportflugzeuge*

(3) [1]

REQUEST
TARGET: *Auftrag*
MESSAGE: *25 Transportflugzeuge*
SPEAKER:
ADDRESSEE:

So far, no full sortal information on the frame elements is present in this representation: Only for those identified by the Named Entity Recogniser described above do we have the relevant information (as ORGANISATION for *Lockheed* in the example). As mentioned above, we plan to use both an enhanced NE recognition and to add sortal information from GermaNet. This information should be carefully used: Firstly, such sortal information will be limited to words that are within the coverage of GermaNet. Secondly, there is the well-known problem of metonymies and other non-literal meanings.

Frame merging. The frame in (3) is only partly filled: The REQUEST frame does not yet contain any information on the SPEAKER and the ADDRESSEE of the request, i. e. who is requesting the 25 planes from whom. This information is not present in the text in the form of grammatical relations on the surface and needs to be transferred from the GETTING frame. This is similar to template fusion in IE, where typically more than one template is instantiated and the overall information can only be gathered by merging the templates. In IE systems, this is often done by a set of rules describing (in-) equality constraints over template slots (Appelt et al., 1995).

From another point of view, this can also be seen as an inference over the known information using additional inference rules. We are investigating ways to best achieve this merging of frames, either based on hand-crafted rules or on machine learning techniques. We currently believe that a combination of the two will probably provide the most robust results.

Question Typology. For the question answering process, we need to translate the user’s question into the corresponding FrameNet structure. This translation must be accompanied by a question type recognition. Processing the question to find the focus and expected answer type has proven to be an important issue in QA systems (Hermjakob, 2001; Harabagiu et al., 2002). We plan to use patterns over FrameNet structures to match a number of question types in a similar way. Thus, a question like *„Von wem hat Lockheed einen Auftrag erhalten?“* (‘From whom has Lockheed received an order?’) should receive a representation like (4) that can be matched against (2) to produce the result *Großbritannien* (Great Britain). Note that the expected answer type PERSON_OR_ORGANISATION can, in this case, be derived from the question word itself, as *von wem?* asks for a person or an organisation.

(4)

GETTING
TARGET: *erhalten*
DONOR: ? [PERSON_OR_ORGANISATION]
RECIPIENT: *Lockheed*
THEME: *Auftrag*

Matching. We currently assume that the representation of the text would take the form of a network of frame instantiations. That is, each frame as shown in (2), (3) would represent one situation described in the text. They would be linked, for example would the THEME element of the GETTING frame contain a link to the REQUEST frame, not a textual representation, as indicated by the co-reference symbols. This idea, however, has to be enhanced in several ways. We will describe some of the issues in the following.

In order to do the actual matching, the FrameNet annotation of the text collection needs to be stored efficiently. This can in principle be done in a standard relational database. We have described above the advantage of not having to do an exact matching with regard to the surface structure. This is ensured as long as the source text and the question receive the same frame representation. However, this will have to be enhanced at least by a match with semantic hyponyms and hypernyms. FrameNet provides information on sub-frames and super-frames. Additionally, GermaNet also provides information on hyponyms and hypernyms. We plan to make use of this information during the matching process, allowing, for example, for a question that contains a *buy* not only to match the COMMERCE_BUY frame, but also the more general GETTING frame that contains words like *obtain* or *acquire* and *vice versa*. In addition, *Flugzeug* (plane) should find *Transportflugzeug* (transport plane), etc. A database search using an ontological information for searching could, for example, be defined as an XML database search with an ontology extension (Schenkel et al., 2003).

Missing Frames. We need to introduce ‘pseudo-frames’ to make up for missing FrameNet data. The FrameNet coverage, especially for German, is not yet perfect. That means that not all words and concepts will receive a FrameNet representation. In such cases, a pseudo-frame will have to be introduced. As such a pseudo-frame would, of course, contain no information on the semantic roles of the target word’s arguments, these relations would have to remain underspecified. One could, for example, use the grammatical relations to label them.

If a user would, for example., enter the question *„Von wem hat Lockheed einen Auftrag gekriegt?“* (‘From whom has Lockheed received an order?'), the colloquial verb *kriegen* would not receive a FrameNet representation but rather a pseudo-frame like that shown in (5).

(5)

PSEUDOFRAME_KRIEGEN.V
TARGET: *kriegen*
?DEEP_SUBJ: *Lockheed*
?DEEP_OBJ: *Auftrag*
?PP_VON: ? [PERSON_OR_ORGANISATION]

When more specific information cannot be found in the text representation, the search process would find the GETTING frame above by leaving the frame element names (i. e. the roles) underspecified and could tentatively output the slot matching the wild-card in the search, guided by the sortal information. This should, however, be accompanied by a warning to the user, since the result is uncertain.

Sortal Information. In the above example, we have concentrated on providing some underspecified information on the semantic relations of an unknown word. Quite often, however, this will not be necessary: Many nouns, for example, do not introduce any obvious roles. Therefore, it will be sufficient in many cases to add sortal information from GermaNet (if present) and relate it to the FrameNet hierarchy – these relations between FrameNet and WordNet concepts has been introduced with the latest versions of FrameNet. This might, for example, help by identifying the transport planes in the example as a sort of plane, and thus as a VEHICLE in FrameNet.

It is, to a large extent, an open question how to deal with cases of sortal mismatches. These are traditionally handled by type coercion in many theoretical frameworks. On closer inspection, our example above contains such a case: The MESSAGE of the REQUEST frame in (3) contains a representation of *25 Transportflugzeuge* (25 transport planes). However, we would expect a MESSAGE to contain an event. This event remains underspecified here. If, therefore, a user asks a question that specifies this event like *„Wen hat Großbritannien mit dem Bau von Transportflugzeugen beauftragt?"* ('To whom has Great Britain given an order for the construction of transport planes?'), namely as a construction event, the matching is no longer straightforward: Either an underspecified event would have to be introduced in the textual representation, triggered by the type mismatch, or the matching phase needs to allow for matching such different representations.

Matching Interlinked Frames. Another question is how the matching of questions that produce two or more interlinked frame representations is done. As an example, consider the question *„Welches Volumen hat der Auftrag für Transportflugzeuge, den Lockheed von Großbritannien bekommen hat?"* ('What is the size of the order for transport planes that Lockheed has received from Great Britain?'). Here, on the one hand the order must be identified with the representation in (2) and (3) above, then, the size of the order must be found. (In the text collection, it is actually specified in the next sentence.) In 'classical' database searches this could efficiently be done by a join over tables. It is still not clear if all questions containing more than one relation can be translated directly in this fashion. We believe that in the remaining cases directed inferences could help.

Inferencing. We are currently looking into the question if adding a directed inferencing step in this process would help. Some recent QA systems have such an inference module to improve the search (Moldovan et al., 2002a). One example discussed there is finding out that committing suicide is a form of dying. We are currently investigating the possibility of automatic inferencing over the FrameNet structures similar to the approach taken there. The important point here is to direct and constrain the inference: Full, undirected inference tends to be very time-consuming. Inferences could, as mentioned above, also help in cases were the granularity of the question representation does not match that of the text collection.

5. Conclusions

We have presented a system for automatically annotating German texts with FrameNet structures. This annotation process is eventually to be used as the core of a QA system that uses direct matching of meaning-oriented FrameNet representations of both the text collection and of the user's question. We think that this approach can help to abstract away from questions of surface wording of texts and questions in a principled way.

However, this approach also comes with a number of questions concerning the implementation of the QA system. We have presented some of them, together with ideas on how they may be solved. We think that the implementation of a QA system along the lines of the approach sketched here is practicable.

An important part of the development of the overall QA system will be an evaluation to see if a FrameNet-based system can really help to improve the performance of a QA system measurably. This is notoriously difficult to achieve (Hirschman and Gaizauskas, 2001). We currently plan to ensure this by a combination of Wizard of Oz experiments to find out more about user needs towards QA systems with prototype evaluations (Fliedner, 2004b).

An interesting additional question for the future is whether using FrameNet as a core representation will help to facilitate building a cross-lingual QA system. The structures themselves do not depend on the language that is represented. One could imagine annotating, for example, an English document base with FrameNet representations and use this database to match the FrameNet representations of German questions. This would open up interesting perspectives, as cross-lingual QA is expected to gain in importance in the future (Magnini et al., 2003).

6. Acknowledgements

The work described in this paper is carried out within the Collate projects (*Computational Linguistics and Language Technology for Real Life Applications*), conducted by DFKI GmbH (partly jointly with Saarland University), funded by the German Ministry for Education and Research, Grant numbers 01 IN A01 B and 01 IN C02.

7. References

Abeillé, Anne (ed.), 2003. *Building and Using Parsed Corpora*. Dordrecht: Kluwer Academic Publisher.

Abney, Steven, 1996. Partial parsing via finite-state cascades. In *Proceedings of the ESSLLI Workshop on Robust Parsing*. Prague.

Appelt, Douglas E., Jerry R. Hobbs, John Bear, David Israel, Megumi Kameyama, and Mabry Tyson, 1995. SRI International FASTUS system MUC-6 results and analysis. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*. Columbia, MD.

Baker, Colin F., Charles J. Fillmore, and John B. Lowe, 1998. The Berkeley FrameNet project. In *Proceedings of the COLING-ACL*. Montreal, Canada.

Bering, Christian, Witold Drożdżyński, Gregor Erbach, Clara Guasch, Petr Homola, Sabine Lehmann, Hong Li, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, Atsuko Shimada, Melanie Siegel, Feiyu Xu, and Dorothee Ziegler-Eisele, 2003. Corpora and evaluation tools for multilingual named entity grammar development. In *Proceedings of Multilingual Corpora Workshop at Corpus Linguistics 2003*. Lancaster.

Böhmová, Alena, Jan Hajič, Eva Hajičová, and Barbora Hladká, 2003. The Prague Dependency Treebank. A three-level annotation scenario. In (Abeillé, 2003), chapter 7, pages 103–127.

Brants, Sabine, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, and George Smith, 2002. The TIGER Treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*. Sozopol, Bulgaria.

Braun, Christian, 2003. Parsing German text for syntacto-semantic structures. In *Proceedings of the Lorraine-Saarland Workshop Series "Prospects and Advances in the Syntax/Semantics Interface"*. Nancy, France.

Carroll, John, Guido Minnen, and Ted Briscoe, 2003. Parser evaluation. using a grammatical relation annotation scheme. In (Abeillé, 2003), chapter 17, pages 299–316.

Elworthy, David, 2000. Question answering using a large NLP system. In *Proceedings of the 9th Text REtrieval Conference (TREC 9)*. Gaithersburg, MD.

Erk, Katrin, Andrea Kowalski, and Manfred Pinkal, 2003. A corpus resource for lexical semantics. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS)*. Tilburg, The Netherlands.

Fliedner, Gerhard, 2002. A system for checking NP agreement in German texts. In *Proceedings of the ACL Student Workshop*. Philadelphia, PA.

Fliedner, Gerhard, 2004a. Deriving FrameNet representations: Towards meaning-oriented question answering. Accepted for *International Conference on Applications of Natural Language to Information Systems (NLDB)*.

Fliedner, Gerhard, 2004b. Issues in evaluating a question answering system. To Appear in *Proceedings of LREC Workshop "User-Oriented Evaluation of Knowledge Discovery Systems"*.

Harabagiu, Sanda, Dan Moldovan, Marius Paşca, Rada Mihalcea, Mihai Surdeanu, Răzvan Bunescu, Roxana Gîrju, Vasile Rus, and Paul Morărescu, 2002. Falcon: Boosting knowledge for answer engines. In *Proceedings of the 9th Text REtrieval Conference (TREC 9)*. Gaithersburg, MD.

Hermjakob, Ulf, 2001. Parsing and question classification for question answering. In *Proceedings of the ACL 2001 Workshop on Open-Domain Question Answering*. Toulouse, France.

Hirschman, L. and R. Gaizauskas, 2001. Natural language question answering: the view from here. *Natural Language Engineering*, 7(4):275–300.

Hovy, Eduard, Ulf Hermjakob, and Chin-Yew Lin, 2001. The use of external knowledge in factoid QA. In *Proceedings of the 10th Text REtrieval Conference (TREC 2001)*. Gaithersburg, MD.

Johnson, Christopher R., Charles J. Fillmore, Miriam R. L. Petruck, Colin F. Baker, Michael Ellsworth, Josef Ruppenhofer, and Esther J. Wood, 2002. FrameNet: Theory and practice. Internet: `http://www.icsi.berkeley.edu/~framenet/book/book.html`.

Kunze, Claudia and Lothar Lemnitzer, 2002. Germanet – representation, visualization, application. In *Proceedings of the LREC 2002*, volume V. Las Palmas.

Light, Marc, Gideon S. Mann, Ellen Riloff, and Eric Breck, 2001. Analyses for elucidating current question answering technology. *Natural Language Engineering*, 7(4):325–342.

Magnini, Bernardo, Simone Romagnoli, Alessandro Vallin, Jesús Herrera, Anselmo Peñas, Víctor Peinado, Felisa Verdejo, and Maarten de Rijke, 2003. The Multiple Language Question Answering Track at CLEF 2003. In *Working Notes for the CLEF 2003 Workshop*. Trondheim, Norway.

Marsh, Elaine and Dennis Perzanowski, 1998. MUC-7 evaluation of IE technology: Overview of results. In *Message Understanding Conference Proceedings MUC-7*. Washington, DC.

Moldovan, Dan, Sanda Harabagiu, Roxana Girju, Paul Morarescu, Finley Lacatusu, Adrian Novischi, Adriana Badulescu, and Orest Bolohan, 2002a. LCC tools for question answering. In *Proceedings of the Eleventh Text REtrieval Conference (TREC 2002)*. Gaithersburg, MD.

Moldovan, Dan, Marius Paşca, Sanda Harabagiu, and Mihai Surdeanu, 2002b. Performance issues and error analysis in an open-domain question answering system. In *Proceedings of the 40th Annual Meeting of the ACL*. Philadelphia, PA.

Palmer, Martha, Dan Gildea, and Paul Kingsbury, 2004. The Proposition Bank: An annotated corpus of semantic roles. Submitted to *Computational Linguistics*.

Schenkel, Ralf, Anja Theobald, and Gerhard Weikum, 2003. Ontology-enabled XML search. In Henk Blanken, Torsten Grabs, Hans-Jörg Schek, Ralf Schenkel, and Gerhard Weikum (eds.), *Intelligent Search on XML Data. Applications, Languages, Models, Implementations, and Benchmarks*, number 2818 in Lecture Notes in Computer Science (LNCS). Berlin: Springer, pages 119–131.

Zhang, Dell and Wee Sun Lee, 2003. A language modeling approach to passage question answering. In *Proceedings of 12th Text REtrieval Conference (TREC 2003)*. Gaithersburg, MD.