Part III. Methods for automatically creating new FrameNets

(V9 30/3/09 00:14) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 207–208 1114 Boas_Ch08-P3 (p. 207)

8. Using FrameNet for the semantic analysis of German: annotation, representation, and automation

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal

1. Introduction

This chapter reports on the Saarbrücken Lexical Semantics Annotation and Analysis (SALSA) project, whose main goals are (1) the exhaustive semantic annotation of a large German corpus resource with FrameNet frames and frame elements[1] (Fillmore et al. 2003), including the generation of a frame-based lexicon from the annotated data, and (2) the induction of data-driven models for automatic frame semantic analysis as well as their application in practical Natural Language Processing (NLP) tasks.

A fundamental assumption of this project, which began in the summer of 2002, is that English FrameNet frames can be re-used for the semantic analysis of German. This assumption rests on the nature of frames as coarse-grained semantic classes which refer to "prototypical situations" (Fillmore 1985). To the extent that these situations agree across languages, frames should be applicable cross-linguistically (see also Boas 2005). While this is clearly a very attractive assumption, it must be empirically validated.

Unlike ontologies, FrameNet's structuring principles do not rely exclusively on conceptual considerations, but are linguistically grounded. A sense of a lemma can evoke a frame, and thus form a lexical unit (LU) for this frame, if this sense is syntactically able to realize the core frame

1. The FrameNet concept of "frame element" (FE) corresponds to the more general concept of "semantic role".

elements (FEs) "that instantiate a conceptually necessary component of a frame" (Ruppenhofer et al. 2006: 26). Consequently, frames may not be applicable to other languages if the subcategorization properties of lemmas in this language differ significantly from their English translations.

Among the issues that SALSA addressed is the extent to which cases of non-parallelism at the level of frames is correlated with typological differences across languages, in particular with respect to (syntactic) valency, and how to account for cross-linguistic divergences. In our work, we have found that the vast majority of frames can in fact be applied directly to the analysis of German – a language that is typologically close to English. The types of problems we encountered during our cross-linguistic work stem primarily from (1) general constructions in German that do not exist in English (such as particular uses of datives), and (2) lexicalization differences in particular semantic domains (such as movement).

The remainder of the paper is structured as follows. In Section 2, we describe the SALSA corpus annotation workflow, present our annotation scheme and process, and discuss various challenges that follow from particular choices of our approach, including (1) problems of coverage, (2) handling of special phenomena encountered in full text annotation (e.g., multiword expressions or metaphors), and (3) problems of vagueness and meaning distinctions. Section 3 discusses cross-lingual aspects of frame semantic annotation. We summarize our experience with frame semantic annotation for German on the basis of English FrameNet frames, as well as commonalities with and differences from related projects for other languages. The discussion also includes a description of our efforts in automated cross-lingual frame semantic resource creation. The final sections of the paper are devoted to the usage of the annotated corpus to induce automated analysis tools for NLP applications. In Section 4, we present Shalmaneser, a general shallow semantic parsing architecture for English and German. In Section 5, we discuss the SALSA RTE system, which utilizes frame semantic resources to investigate the usefulness of frame-semantic information for the NLP task of recognizing textual entailment (Dagan et al. 2005).

2. SALSA: Semantic Annotation and Lexicon Building for German

The main objective of the SALSA project is the creation of lexical semantic resources for German within the framework of Frame Semantics (Fillmore 1985). Similar to PropBank (Palmer et al. 2005), SALSA extends an

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 210)

existing German treebank, the TIGER treebank (Brants et al. 2002), with a layer of lexical semantic annotations, focusing on verbal predicates.

A first corpus was released in summer 2007 and consisted of about 500 German verbal predicates of all frequency bands plus some deverbal nouns, totaling about 20,000 annotated instances.

2.1. Corpus-driven resource creation

The SALSA project differs from FrameNet in that it is primarily concerned with providing an exhaustive annotation of the entire corpus as a basis for obtaining large-scale NLP resources with as complete coverage as feasible. Therefore, SALSA analyzes the entire TIGER corpus lemma by lemma, whereas FrameNet proceeds frame by frame, extracting relevant examples from different sections of the British National Corpus. Since we regard ourselves more as users of the existing FrameNet resource than as creators of a comparable German FrameNet, we are released from the requirement of systematically describing all possible frames and their realization patterns, as FrameNet aspires to. At the same time, our exhaustive annotation policy forces us to analyze all instances of a lemma in the corpus, which often requires the creation of proto-frames on the fly, as described in Section 2.3. Also, exhaustive annotation requires addressing frequently occurring phenomena with limited compositionality (such as idioms or support verb constructions), as well as cases of ambiguity and vagueness (see Section 2.4). In contrast, FrameNet primarily analyzes predicates with a clear syntax-semantics mapping that illustrate lexicographically relevant "core" meanings. Despite these differences, the two methods are converging in practice in that FrameNet is starting to pursue corpus-driven full-text annotation, while SALSA is extracting a general lexicon resource from corpus annotations and spends considerable efforts on proto-framing.

2.2. Annotation scheme and annotation practice

To annotate, we employ SALTO, a graphical annotation tool designed and implemented for SALSA (Burchardt et al. 2006a), which is shown in Figure 1. Freely available for research purposes (see Section 7), SALTO supports annotation in a simple drag-and-drop fashion and can also be used more generally for the graphical annotation of treebanks with any kind of relational information. SALTO uses SALSA/TIGER XML, a general XML format for input and output (see Section 4 for details), and additionally supports corpus management and quality control.

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 211)

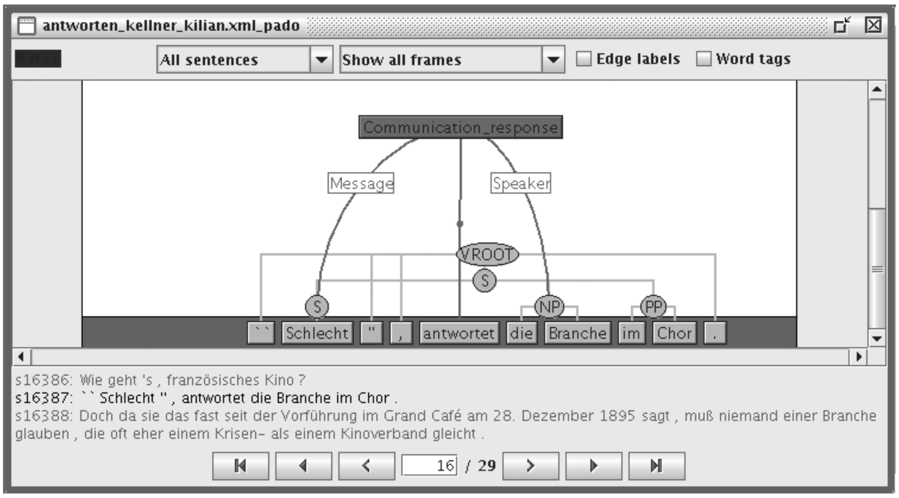

Figure 1. Annotation example: *"Schlecht", antwortet die Branche im Chor.* ("Badly", the industry sector answers in unison.)

We annotate frame-semantic information on top of the syntactic structure of the TIGER corpus, with a single flat tree for each frame: The root node is labeled with the name of a frame. The edges of the syntactic constituents are labeled with the names of FEs defined for the frame. Figure 1 shows a simple annotation instance: the verb *antwortet* ('answers') evokes the frame `Communication_response`. The NP subject *die Branche* ('the industry sector') is annotated with the FE SPEAKER and *schlecht* ('badly'), under a sentence (S) node, with the FE MESSAGE. In contrast to FrameNet, we annotate only *core* FEs (see Section 1). Moreover, we adjust the span of FEs to existing syntactic structures where possible.

Like PropBank, SALSA follows a corpus-based approach, aiming at full-text corpus annotation by covering all instances of a particular lemma in the corpus. To make this procedure feasible for annotators, annotation proceeds lemma-wise: for each lemma in the running text of the TIGER corpus, we extract all corpus sentences in which it occurs. The resulting subcorpora are given to pairs of annotators for parallel annotation, together with a list of candidate frames that seem appropriate. The annotators consult the frame definitions in FrameNet, and may also choose additional frames from FrameNet for novel uses they encounter in a given subcorpus. As a result of our corpus-based full-text annotation practice,

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 212)

we face two major challenges: one has to do with coverage, the other with the treatment of special linguistic phenomena.

2.3. Coverage and proto-frames

A major problem for exhaustive annotation is that FrameNet is still under development, and thus does not yet cover all senses of the lemmas that we annotate. Another, more subtle problem, are frequent usages whose meanings are clear in the context, but difficult to relate to lexicographical prototypes.

To assess FrameNet coverage for a given lemma and to spot missing senses, we thus extract a small sample of sentences containing instances of this lemma in the TIGER corpus prior to annotation. For each instance, we check whether there is a FrameNet frame that provides a felicitous analysis. The decision is based on the two criteria detailed in Ellsworth et al. (2004: 18–19): (1) Does the meaning of the instance meet the frame definition? (2) Can all important semantic arguments of the instance be described in terms of the FEs? In unclear cases, we also check annotated FrameNet example sentences for similar usages to get a better understanding for the full range of a frame.

The extraction process results in a list of instances for the current lemma which cannot be described in terms of existing frames. We group these into coarse-grained "sense groups" and construct a proto-frame for each group. The resulting proto-frames are lemma-specific, i.e., contain only a single lexical unit. Table 1 shows a proto-frame constructed for the "to be counted (among a group)" sense of *rechnen* ('to count as').

Table 1 illustrates that the SALSA proto-frames are similar to FrameNet frames – they have a textual definition, a set of FEs with

Table 1. Example of a proto-frame for one sense of *rechnen (zu)* ('count (as)')

Frame: `Rechnen.Unknown3`		
An Item is construed as an example or member of a specific Category. In contrast to Categorisation, no Cognizer is involved. In contrast to Membership, the Category does not have to be a social organisation.		
FEs	ITEM	**Die Philippinen und Chile** <u>rechnen</u> zu den armen Ländern der Region.
	CATEGORY	Die Philippinen und Chile <u>rechnen</u> **zu den armen Ländern der Region**.

FrameNet-style names, and annotated example sentences. They follow a simple naming convention, e.g. `Rechnen.Unknown3`, which marks the third proto-frame constructed for the lemma *rechnen*. The proto-frames are lemma-specific and not intended as final descriptions for the senses. They form a sense inventory for German that finds immediate application in our annotation process, allowing us to semantically annotate all corpus instances in the running text, even if not at the same level of generalization as provided by FrameNet frames.

We envisage that our proto-frames can form the input to a lexicographic generalization process for the further development of FrameNet. To support this integration, our proto-frames are defined at roughly the same level of granularity as FrameNet frames. In addition, we list frame-to-frame relations for proto-frames to indicate their relationship to both FrameNet frames and other proto-frames. For example, for `Rechnen.Unknown3` we record that it is identical to a proto-frame for *zählen* ('to count'): in the example sentence in Table 1, *rechnen* can thus be paraphrased by *zählen*.

To illustrate the quantitative relation between the coverage of FrameNet and of our proto-frames, we computed preliminary statistics on a dataset of 12,437 annotation instances and found that the average number of frames per lemma was 2.33, composed of 1.6 FrameNet frames and 0.73 SALSA proto-frames. In other words, less than one third of the lemma senses in our corpus was not covered by FrameNet. To gauge the degree of semantic granularity of our proto-frames, we compared the average number of lexical units (i.e., frames) of our lemmas to the average number of synsets (i.e., senses) for verbs in GermaNet. We found that our annotation was more fine-grained (2.33 frames per lemma) than the 2.2 synsets per verb in GermaNet (Hamp and Feldweg 1997). This is at least partly due to our treatment of idioms and metaphoric readings as additional, full-fledged senses of lemmas (see Burchardt et al. 2006b for more details).

2.4. Special phenomena

In standard annotation cases, there is a strong one-to-one mapping between syntactic and semantic structure: a frame is evoked by a single word, and its FEs link to syntactic (i.e., subcategorized) arguments of the word. An example is shown in Figure 1 above. However, due to our exhaustive annotation policy, we frequently encounter cases of limited

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 214)

Table 2. Phenomena with limited compositionality (LC)

	246 Lemmas		*nehmen*	
	Number	%	Number	%
Compositional	**10,820**	**87.0**	**42**	**17.4**
Metaphor	707	5.7	38	15.8
Support	597	4.8	132	45.8
Idiom	313	2.5	29	12.0
LC	**1,617**	**13.0**	**199**	**82.6**
Total	**12,437**	**100.0**	**241**	**100.0**

compositionality (“LC-phenomena”) in which frame choice, argument choice, or both, diverge from such a straightforward mapping between syntax and semantics. Three prominent cases of LC-phenomena which we encounter in our annotation are support verb constructions, idioms, and metaphors. As Table 2 illustrates, they occur quite frequently, constituting almost one seventh of the 12,000-instance corpus sample mentioned above. For high-frequency (and typically highly polysemous) verbs such as *nehmen* (‘to take’), they even make up the majority of instances. We now discuss our criteria for distinguishing the three LC-phenomena as well as our annotation schemes for each of them.

2.4.1. Support verb constructions

A support verb construction (SVC) is a combination of a verb with a “bleached” or abstract meaning (e.g. causation or perspectivization) with a predicative noun, which is typically its object. The noun constitutes the semantic head of the phrase and is usually treated as the frame-evoking element. An example is *Abschied nehmen* (‘to take leave’), where *Abschied* evokes the `Departure` frame. Often, the SVC can be paraphrased with a morphologically related verb (e.g., *sich verabschieden* (‘to say good-bye’)). Currently, SALSA annotates verbal parts of SVCs with a pseudo frame `Support`, whose only FE SUPPORTED points to the supported noun phrase. This annotation makes SVCs retrievable and thus available for a subsequent more elaborate analysis of the syntax-semantics interaction between the verbs and nouns involved.

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 215)

2.4.2. *Idioms*

We identify idioms by three criteria. They are multi-word expressions that are for the most part fixed, and which have to be understood as a whole while their figurative meaning is not recoverable synchronically from their literal meanings. An example is *(etwas) in Kauf nehmen* (literally 'to take (something) into purchase'), which means *to put up with (something)*. Figure 2 shows an instance of this idiom, *Die Gläubiger nehmen Nachteile in Kauf* ('the creditors put up with disadvantages'). As can be seen, we annotate the idiom as a whole as the frame-evoking element, which here evokes the frame `Agree_or_refuse_to_act`. The semantic arguments of the idiom are annotated as normal FEs – *die Gläubiger (the creditors)* fill the role SPEAKER, *Nachteile (disadvantages)* fill the role PROPOSED_ACTION.

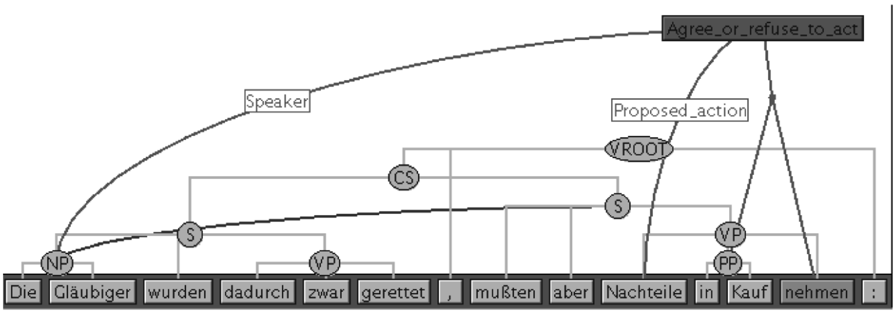

Figure 2. Multi-word target for idiom *in Kauf nehmen* ('to put up with s.th.')

2.4.3. *Metaphors*

Metaphors are distinguished from idioms through the existence of a figurative reading which is recoverable from their literal meaning. Following Lakoff's ideas on metaphorical transfer involving source and target domains (Lakoff and Johnson 1980), we annotate metaphorical expressions with two frames – a source frame representing the literal meaning, and a target frame representing the figurative meaning.

As an example, consider the metaphor *unter die Lupe nehmen* ('to put (literally: take) under a magnifying glass'). The source analysis is shown in Figure 3, where the verb *nehmen* ('take') is annotated as a frame-evok-

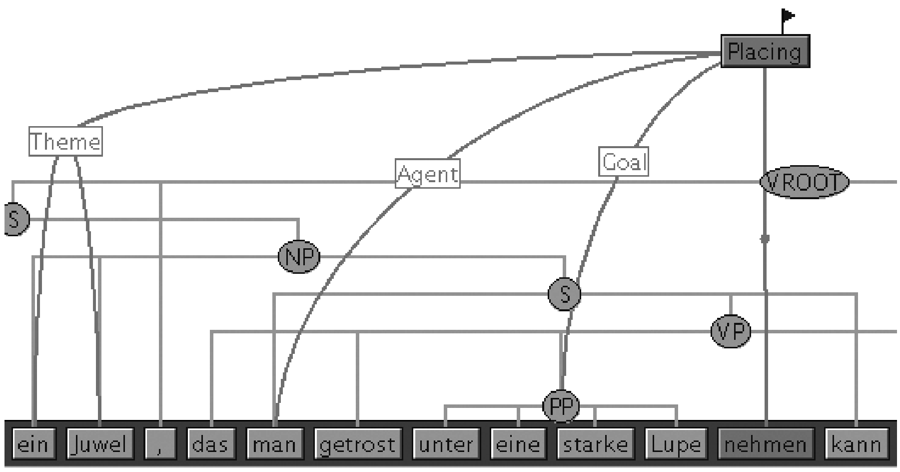

Figure 3. Analysis of the source (literal) reading of the metaphor *unter eine Lupe nehmen* (lit.: 'to take under a magnifying glass'). The frame `Placing` is introduced by the verb only

ing element, which introduces the frame `Placing`.[2] All arguments of *nehmen* are analyzed as ordinary FEs of `Placing`: *ein Juwel* ('a jewel') is the THEME that is taken, *man* ('one') is the AGENT who does the taking, and *unter die Lupe* ('under a magnifying glass') is the GOAL, the eventual position of the THEME.

The corresponding target reading is shown in Figure 4. Here, the frame `Scrutiny` is introduced by the fixed part of the metaphor, *unter die Lupe nehmen*.

We often found target (figurative) meanings difficult to describe in terms of (existing) FrameNet frames. In order to maintain our rate of annotation, we chose to restrict the annotation of difficult cases to source readings. During a later phase, these samples will then be retrieved for a more comprehensive analysis.

The double annotation using a source and a target frame facilitates modeling the construction of this metaphor as a transfer from a (concrete)

2. The most salient sense of the German verb *nehmen* is best analyzed with the frame `Taking`. However, *nehmen* can also be used with a directional argument expressing a GOAL, as in the example at hand. These cases are better analyzed using the frame `Placing`.

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 217)

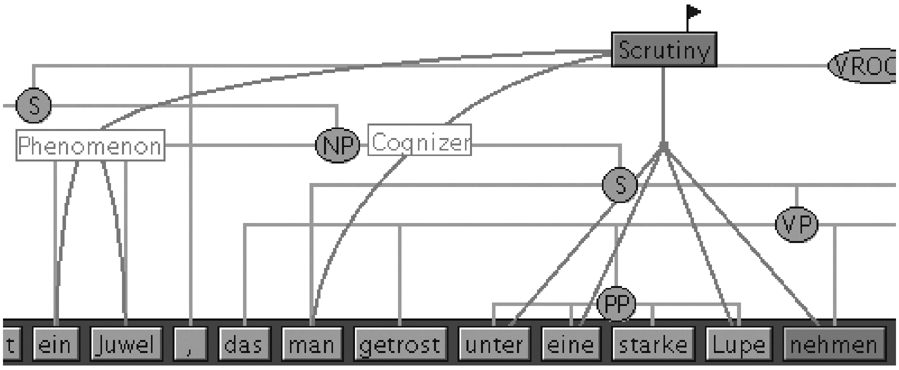

Figure 4. Analysis of the target (figurative) reading of the metaphor *unter eine Lupe nehmen* (lit.: 'to take under a magnifying glass'). The frame `Scrutiny` is introduced by the complete metaphor.

putting event to a (more abstract) investigation event. This illustrates that source and target frames describe complementary properties of metaphors: The source frame models the syntactic realization patterns of the arguments of the main predicate, while the target frame captures the figurative meaning.

Source/target frame pairs can be used to study *argument transfer* from source to target predicates. In simple cases, the transfer establishes a direct correspondence between source and target frames, including all arguments. In the example *Das Postfach explodiert* ('The mailbox explodes'), the source frame `Change_of_phase` with its role UNDERGOER directly maps onto the target frame `Expansion` with the role ITEM. As a more complex case, consider *unter eine* ***starke*** *Lupe nehmen* ('to put under a **strong** magnifying glass'). The corresponding transfer scheme in Figure 5 exemplifies a case of *argument incorporation*: the FE GOAL of the `Placing` frame is absorbed by the frame-evoking element of the SCRUTINY frame; in addition, the modifier *starke* ('strong'), which does not constitute a FE on the source side, constitutes the FE DEGREE of the target frame.

It is important to keep in mind that such transfer schemes do not answer the question about which factors trigger the metaphorical transfer for a specific utterance. However, they can model the interpretation process underlying metaphors to a certain degree. This, in turn, provides a description of the relation between source and target frames for specific metaphors, which facilitates expressing generalizations over patterns of FE shifts.

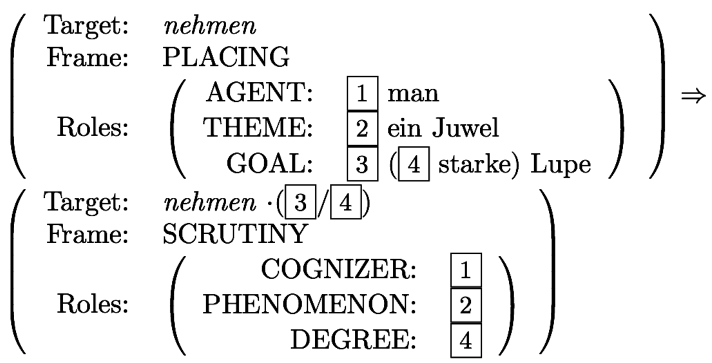

Figure 5. Transfer scheme for *Die Klangkultur ist ein Juwel, das man getrost unter eine starke Lupe nehmen kann*. ('Their sound is a jewel which stands up to any type of scrutiny.')

With this overview of how SALSA deals with LC-phenomena, we now discuss how underspecification and distinctions between FEs are handled in our workflow.

2.4.4. Underspecification

It is well-known that there are cases of vagueness in semantic annotation, where the assignment of only a single label (such as a frame, or an FE) would not be appropriate, and annotators should be able to assign more than one label (see Kilgarriff and Rosenzweig 2000). Allowing this type of annotation makes it possible to retrieve vague cases and avoids forcing the annotators to adopt ad-hoc choices for decisions which are impossible to make reliably.

SALSA annotation must address the vagueness problem both at the level of frames and FEs. As an illustration for frames, consider the verb *bemerken* ('to notice/comment') in (1). *Bemerken* typically introduces meaning components of two frames simultaneously, namely `Statement` (like *say*) and `Becoming_aware` (like *notice*). Neither frame alone conveys the complete meaning of *bemerken*, and forcing annotators to make an unambiguous decision would presumably result in inconsistent annotations.

(1) Kein Wunder, dass Gerhard Schäfer in seinem Buch derzeit eine "Renaissance der Verbindungen in den neuen Ländern" bemerkt. (TIGER s11777)

'(It is) not surprising that Gerhard Schäfer **notices/comments on** a "renaissance of fraternities in the new states".'

The metonymic sentence in (2) exemplifies a similar case at the FE level. Here, one frame is evoked, namely `Request`, but one of the FEs is vague. *Ein Antrag* ('a motion') describes the MEDIUM used to convey the demand, but it also refers metonymically to the SPEAKER. Again, no single annotation can capture the complete meaning.

(2) Die nachhaltigste Korrektur <u>fordert</u> [**ein Antrag** MEDIUM/SPEAKER]
'The most radical change is <u>demanded</u> by [**a motion** MEDIUM/SPEAKER].'

In such cases, SALSA annotators can assign more than one frame (or more than one FE of the same frame), connecting the multiple assignments by an underspecification link. Underspecification does not have an a priori disjunctive ("only one of the two labels fits, but it is impossible to decide which") or conjunctive ("both labels apply simultaneously to some extent") interpretation since it has been argued that this meta-level question is often as difficult to decide as the object-level question of which label to choose (see Kilgarriff and Rosenzweig 2000).

Underspecification is particularly useful to represent borderline instances of phenomena with limited compositionality. Notorious cases are the distinction between support constructions and metaphors, as well as between transparent metaphors and idioms that are no longer transparent.

2.4.5. Difficult role distinctions

FrameNet often uses ontological criteria to differentiate between closely related but mutually exclusive FEs. Such configurations arise, for example, between pairs of FEs that stand in a systematic metonymical relationship (as opposed to incidental cases of metonymy discussed in the last paragraph). Since these are difficult to distinguish with annotations, we defined, where necessary, higher-level FEs which generalize over the problematic FEs.

For example, in the FrameNet frame `Waiting`, a PROTAGONIST awaits the coming about of an EXPECTED_EVENT or a SALIENT_ENTITY in which it is involved. While the two crossed-out roles can be distinguished in examples (3) and (4), example (5) contains an argument that is neither a clear-cut EXPECTED_EVENT nor a SALIENT_ENTITY. We have therefore defined a new FE, called EXPECTED_EVENT_SALSA in the `Waiting` frame. This FE allows us to describe all three instances in (3)–(5) in the same manner, generalizing over EXPECTED_EVENT, SALIENT_ENTITY, and problematic borderline cases.

(3) Luise wartet [darauf, dass das Telefon klingelt ~~EXPECTED_EVENT~~ EXPECTED_EVENT_SALSA]
‘Luise waits [for the phone to ring ~~EXPECTED_EVENT~~ EXPECTED_EVENT_SALSA].’

(4) Luise wartet [auf ihren Mann ~~SALIENT_ENTITY~~ EXPECTED_EVENT_SALSA]
‘Luise is waiting for [her husband ~~SALIENT_ENTITY~~ EXPECTED_EVENT_SALSA].’

(5) Viele Wähler in Rußland haben immer [auf eine starke Sozialdemokratie EXPECTED_EVENT_SALSA] gewartet.
‘Many voters in Russia have always waited [for a powerful social democracy EXPECTED_EVENT_SALSA].’

2.5. Consistency control

Figure 6 shows the global structure of the annotation workflow in SALSA: Each dataset for a given lemma is annotated independently by two annotators (trained undergraduate students). Because of the double annotation process, a fair number of annotation mistakes can be detected automatically, and resolved in a double adjudication step: After annotation, the two annotated versions of a dataset are automatically merged into a single copy in which annotation differences are marked. The conflicts are resolved independently by two SALSA researchers. Almost all disagreements which remain after adjudication are truly difficult cases. Many are idiosyncratic problems, i.e., problems with particular instances. An example is that of referential ambiguities, which can lead to ambiguous FE assignments, or conceptual problems with respect to the FrameNet inventory. Examples of the latter are systematic problems in distinguishing FEs, or usages which meet frame descriptions only partially, or combine aspects of several frames. In cases where the adjudicators cannot reach an unanimous decision, underspecification is used as a last resort.

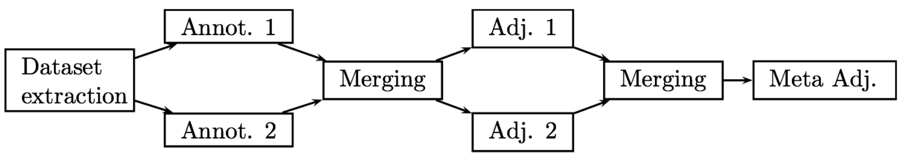

Figure 6. SALSA workflow: annotation and quality control

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 221)

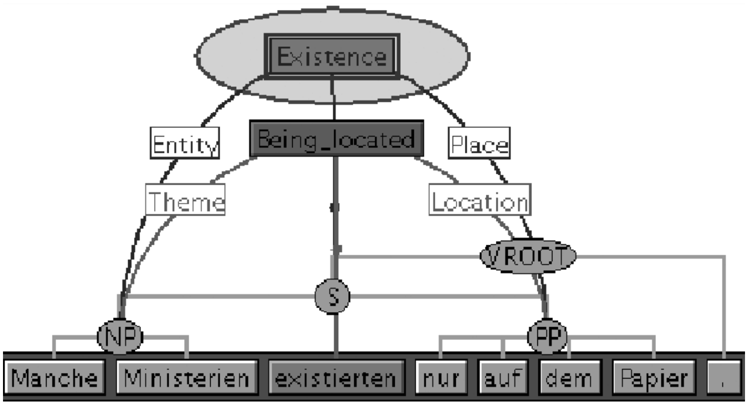

Figure 7. Inter-annotator difference: `Existence` vs. `Being_located`

The SALTO tool is used to manage the whole workflow, including dataset extraction and merging. In a special adjudication mode, SALTO guides the user specifically through those differences to allow for manual inspection and correction. Figure 7 shows an example of inter-annotator disagreement: One annotator tagged the word *existieren* ('exist') with the frame `Existence`, while the other annotator chose `Being_located`. The SALTO tool circled `Existence` to show that this is the next annotation choice to be either confirmed or denied by the adjudicator.

2.5.1. *Computing agreement*

It is typically best practice for annotation projects to report chance-corrected agreement, such as the kappa statistic (Siegel and Castellan 1988). However, as discussed in Burchardt et al. (2006b), kappa is only applicable to categorization tasks with fixed numbers of items and categories. Since these conditions do not apply to our setting, we do not report kappa; instead we report percentage agreements according to a strict evaluation metric (labeled exact match).

On the basis of two independently annotated and two adjudicated versions, we compute inter-annotator agreement and inter-adjudicator agreement. We consider frame selection and FEs assignment individually, due to their different characteristics. According to our method of computing agreement, inter-annotator agreement is 85% for frames and 86% for FEs for matching frames. Inter-adjudicator agreement is 97% for frames and 96% for FEs. Informally, annotators agree in more than 4/5 of all instances; adjudication creates consensus for another 4/5 of the disagreements. These numbers indicate substantial agreement, which demonstrates that the task is well-defined.

2.5.2. Limits of the four-eye principle

Quality control using inter-annotator agreement can only identify errors caused by individual annotation differences between annotators. If both annotators make the same error, it cannot be detected automatically. This limits the effectiveness of quality control by inter-annotator agreement with regard to systematic mistakes.

For this reason, we draw random samples from all completely annotated lemma-frame-pairs, which are then inspected for possible systematic annotation mistakes. We have also experimented with *intra*-annotator agreement, trying to automatically detect errors by finding "outliers" with non-uniform behavior. However, due to the LU-specific nature of semantic annotation, even correctly annotated datasets can show discrepancies.

2.6. From corpus to lexicon

One of the outcomes of the SALSA workflow illustrated in Figure 6 above is a frame-based lexicon model for German. This lexicon stores the information from the annotated corpus in a hierarchical model in description logics (Spohr et al. 2007). The model includes frame descriptions with their syntax-semantics linking patterns and frequency distributions.

Extracting a separate lexicon from the corpus offers a number of advantages. It allows the modular definition of generalizations over typically fine-grained annotation categories for individual instances as well as quantitative generalizations over these instances. The example in Table 3 shows that this kind of generalization is particularly crucial for information about the mapping between syntax and semantics. This information is extracted in way similar to the FrameNet lexical entry reports. Fine-grained categories like NN (normal noun), NE (named entity), and PPER (personal pronoun) lead to the fragmentation of the corpus-derived mapping information and makes it susceptible to noise in the data. We therefore introduce generalized categories to discover linguistically meaningful and more robust regularities.

A second advantage of the separate lexicon is that it allows practically arbitrary "views" of the data, e.g., grouping information by lemma, by frame, or by phenomenon. All lexicon entries provide links to the annotation instances, thus grounding the lexicon in the corpus.

Table 3. Generalizations over syntactic categories in the lexicon

Frame.Role	Annotated Category	Generalized Category
Placing.Theme	NN	NounP
Placing.Theme	NE	NounP
Placing.Theme	PPER	NounP
Statement.Message	S	VerbP
Statement.Message	VP	VerbP

A benefit of the use of description logics for lexicon modeling is that it is a very general representation format. It supports consistency control of the annotated data and can serve as a machine-readable repository of lexical data for NLP applications, as well as a data source for linguistic research. The latter point is supported by the query mechanism SeRQL which allows the flexible retrieval of data from description logics databases.

3. Cross-lingual aspects

3.1. The applicability of FrameNet frames for the annotation of German

The fact that our German corpus annotation is based on frames and FEs that were originally created for English raises the question of the applicability of frame semantic descriptions to other languages (see Boas 2005). In our experience, the vast majority of FrameNet frames can be re-used fortuitously to describe German predicate-argument structures. Nevertheless, some FrameNet frames require adaptation and modification. Below, we discuss two central types of problems, namely missing FEs and differences in the linguistic realization of frame structures.

3.1.1. Missing Frame Elements

We found a number of frames derived on the basis of English that were well suited for the semantic description of German lexical units, apart from the problem of German verbs realizing dative objects for which no appropriate FE is defined in the frame. Many of these cases are instances of the *external possessor construction*, in which a possessor of a verb's object is realized as an argument of the verb itself. While this construction is quite frequent in German, its use in English is known to be quite restricted; for example, it Hole (2005: 238) recently noted that "English beneficiary objects are heavily constrained [. . .]".

As an example, consider the frame `Taking`, in which an AGENT takes possession of a THEME by removing it from a SOURCE. In English, the SOURCE, usually realized as a *from*-PP, can be either a source location or a former possessor. It is not possible to realize both as separate, full-fledged arguments of a predicate, although the possessor may be incorporated in the source location ("from his hand"). Thus, FrameNet does not distinguish between the two. In contrast, the German verb *nehmen* ('to take') can realize location and possessor simultaneously as arguments, as the following example illustrates:

(6) Er <u>nahm</u> [ihm POSSESSOR] [das Bier THEME]
He took him the beer
[aus der Hand SOURCE]
out of the hand

To handle such cases, we add new FEs – here a FE POSSESSOR, thereby splitting the FrameNet FE SOURCE into a location-type SOURCE and a distinct POSSESSOR.

3.1.2. Differences in the lexicalization of frames

The meanings of German verbs sometimes cut across the frame distinctions designed on the basis of English data. An example is the German verb *fahren* ('to drive'), which encompasses both English *drive* (frame `Operate_vehicle`, with the FE DRIVER) and *ride* (frame `Ride_vehicle`, with the FE PASSENGER). In German, context often does not disambiguate between the two frames, which makes it difficult to make a decision between these alternative frames. Consider (7), where German *fahren* is fully underspecified as to whether the people referred to (*they*) were drivers or passengers of the 14 vehicles.

(7) In 14 Armeefahrzeugen <u>fuhren</u> sie von dem abgezäunten Gelände, das der Besatzungsmacht 28 Jahre lang als Hauptquartier gedient hatte.
'With 14 army vehicles they <u>departed</u> from the enclosed area that had served the occupying forces as headquarter for 28 years.'

In the case at hand, FrameNet has introduced the frame `Use_vehicle`, which subsumes both `Operate_vehicle` and `Ride_vehicle`.

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 225)

While this higher-level frame has no lexicalization in English, it is the right level to describe the meaning of German *fahren* in examples such as (7). In general, such cases need to be discussed from a multilingual perspective. In the ongoing annotation effort, we resort to underspecification (see Section 2.4.4). A possible area for future work is to find cross-lingually valid redefinitions for problematic frames, in cooperation with FrameNet and other partners.

3.2. SALSA and FrameNet projects for other languages

While SALSA frame annotation is done on a corpus with complete, deep syntactic annotation, Berkeley FrameNet (and FrameNet projects for other languages) annotate examples on the basis of unparsed corpus sentences, where syntactic information is added exclusively for annotated roles, either manually or semi-automatically. This is mirrored at the technical level in the choice of storage format: FrameNet's "lexical unit report" XML files represent annotations one frame at a time, and characterize role spans by way of character spans of the sentence string. In contrast, SALSA uses SALSA/TIGER XML (Erk and Padó 2004), an extension of TIGER XML, a description formalism originally used for syntax trees, and extended to semantic annotation. SALSA/TIGER XML can represent an arbitrary number of frames and roles (as shown in Figure 7, for example), defining their span in terms of (sets of) syntactic constituents. Several steps have been taken, however, to harmonize the different frame-semantic resources.

Our first goal was to allow the exchange of annotated data between projects. Mutually convertible data formats make it possible to develop common toolboxes, e.g., for modeling, consistency checking, or simply visualization using the SALTO tool (see Section 2.2). SALSA subcorpora and FrameNet lexical unit (LU) reports form the most appropriate level of granularity for data exchange: One SALSA subcorpus for a lemma corresponds to a set of LU reports, one for each reading of the lemma (i.e., frame). The direction SALSA → FrameNet is comparatively simple, since it only consists of removing most of the syntactic structures, retaining just the constituents labeled with FEs. The reverse direction (FrameNet → SALSA) is also fairly straightforward in that the span-based characterization of roles, in conjunction with categorial or functional information, can be used to define a partial syntactic and semantic structure in SALSA/TIGER XML. This is restricted to the annotated target word and FEs. In practice, the conversion direction was implemented in a different, pragmatically motivated way, in the context of developing a shallow semantic parser (see Section 4 for details): The conversion FrameNet → SALSA was implemented in the shape of an input filter that reads FrameNet LU reports, runs an automatic wide-coverage syntactic parser on the sentences, and converts the character-based annotation into a constituent-based annotation. Even though the accuracy of the automatic analysis cannot be guaranteed, this procedure makes it possible to train a shallow semantic parser directly on FrameNet data.

A further step, which builds directly on the ability to exchange annotated data, is to develop methods to compare and contrast data from more than one language in a flexible and comfortable manner. This goal has been realized in the lexicographical domain by FrameSQL, a database-oriented browser for the FrameNet database developed by Sato (2003). This tool has been extended to allow the contrastive display of FrameNet information for different languages, first for the language pair English–Spanish (Subirats and Sato 2004), and later also for English–German.

As Figure 8 shows, it is possible to compare the lexical units of two languages for the same frame, and their valencies. This represents a first step to facilitate the study of cross-lingual commonalities and divergences in the frame semantic paradigm.

An important area for future research is the development of a cross-lingual, declarative lexicon model that is modular and powerful enough to represent both SALSA-style and FrameNet-style representations, together with annotated examples and statistical generalizations.

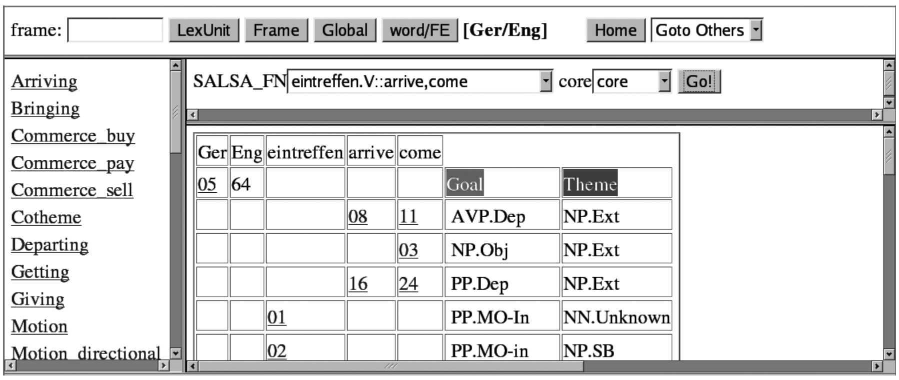

Figure 8. Sato Tool snapshot contrasting English *arrive* and *come* with German *eintreffen*

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 227)

Our current efforts in building a frame-based lexicon from German corpus annotations in Spohr et al. (2007) is a first step towards this goal.

3.3. Cross-lingual projection for resource creation

As already discussed, English FrameNet frames are well suited to describe predicate-argument structures of different languages. In this context, the question arises as to how the annotation effort can be kept minimal whenever a new language is analyzed. More specifically, we are interested in methods which can automate at least part of this process.

At SALSA, we approached this task by using annotation projection, a strategy that exploits translational information from large parallel corpora to transfer semantic annotation across languages (see Pitel (this volume) for an alternative approach). More specifically, we re-used the manual effort expended on the creation of the English FrameNet to create comparable frame-semantic resources for French and German. This task naturally divides into two subproblems: (1) the induction of frame-semantic lemma classifications (i.e., lists of admissible frame-evoking elements for frames); and (2) the creation of a corpus of sentences with annotation of FEs.

With regard to (1), we developed a general language-independent architecture to bootstrap frame-semantic lemma classifications. We found that high-quality classifications can be induced for new languages by concentrating on translation pairs of source and target language lemmas which are especially likely to be frame-preserving. This property can be established even on the basis of shallow linguistic knowledge by exploiting the distributional profile of translation pairs in a large parallel corpus. For example, in experiments on the EUROPARL corpus (Koehn 2005), we constructed FrameNet-sized lemma classifications for both German and French with a precision of 65% to 70%, comparable to the size of Berkeley FrameNet (Padó and Lapata 2005a).

As for the induction of semantic role annotation for German sentences, provided that the frames match, the main task is to establish a mapping between subsentential phrases of source and target sentences that constitute possible roles. This problem can be phrased as a graph optimization problem, using word alignments to describe the pairwise cross-lingual similarity of phrases. In an experimental evaluation (Padó and Lapata 2005b), we demonstrated that FEs can be projected with an accuracy of up to 69% f-score (75% precision) when English manual FE annotation is used. When an imperfect state-of-the-art automatic shallow semantic parser is used to analyze the English text, the performance degrades to 57% f-score. However, this is mostly a problem of recall: the precision remains very high at 74%, indicating that it is possible to produce high-quality semantic annotation for new languages even from noisy data.

While the fully automatic methods for both types of information still fall short of the quality of manually created resources, their use can speed up resource development for new languages considerably, or serve as a "rough-and-ready" resource if no manual effort can be expended at all.

4. Automation

In this section, we present our strategies for shallow semantic parsing. Shallow semantic parsing is important for all NLP applications that benefit from deeper text understanding, such as the applications that Manning (2006) calls "Information Retrieval++": question answering, information extraction, and customer response systems. The availability of robust and accurate systems that can produce shallow semantic parses for free text is a crucial step towards the usability of role-semantic information in applications, such as the recognition of textual entailment (cf. Section 5). Shallow semantic parsing can be divided into Word Sense Disambiguation (WSD) (in FrameNet: an assignment of frames to frame-evoking elements) and Semantic Role Labeling (SRL) – in FrameNet, the assignment of FEs. While WSD is one of the oldest NLP tasks (Ide and Véronis 1998), SRL has only recently become a task of considerable interest in the computational linguistics community, beginning with the seminal study by Gildea and Jurafsky (2002).

4.1. SHALMANESER: A system for shallow semantic parsing

Research on shallow semantic parsing is in its early stages, requiring further steps both on the level of the analysis and its application. For this reason, we have developed a system for shallow parsing in SALSA, called SHALMANESER (the Shallow semantic parser). SHALMANESER fills the need for a shallow semantic parser which is publicly available and which can be used as a "black box" to obtain semantic role analyses for text without the need to consider the intricacies of shallow semantic parsing (comparable to current syntactic parsers). While developed for English and German, the system is easily applicable to other languages as well.

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 229)

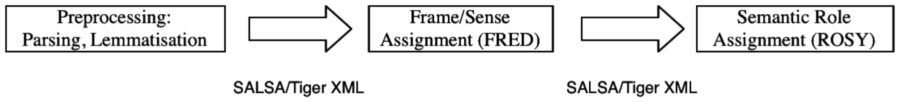

Figure 9. The SHALMANESER toolchain

The structure of SHALMANESER is illustrated in Figure 9. It takes plain text as input, which is first lemmatized, part-of-speech tagged, and syntactically analyzed. Semantic information is then added in two consecutive steps, WSD and SRL: First, the frame disambiguation system assigns semantic classes (senses) to lemmas. Then, the FE assignment system adds FEs to surrounding constituents. Both sense and FE assignments are modeled as supervised learning tasks. Sense assignment is decided on the basis of the lexical context and syntactic properties of lemmas (Erk 2005). For FE assignment, we rely both on syntactic features (e.g., path from FEE to constituent) and lexical features, which, although sparse, provide crucial information (see Erk and Padó 2005).

SHALMANESER uses the SALSA/TIGER XML format described in Section 3.2. Thus, the SALTO annotation tool can be used to inspect and manually modify the assigned frames and roles within a graphical interface. More generally, an open extensible architecture like the one offered by SHALMANESER allows for a modular view of semantic analysis. Semantic classes and roles are just one particular type among the many kinds of semantic information that are potentially helpful in NLP applications. The last years have seen impressive progress in the accurate computation of individual kinds of semantic information. These comprise lexical information (ontological status, lexical relations, polarity) and structural information (scope, modality, anaphoric and discourse structure).

4.1.1. Using SHALMANESER

SHALMANESER is designed with two application scenarios in mind. In an "end user scenario", pre-trained classifiers for English and German are available for exploring the use of role-semantic information in different NLP settings (see Section 7 for details). In a "research scenario", the modular architecture facilitates the integration of additional processing modules. Furthermore, we keep the processing components encapsulated to make them easily adaptable to new features, parsers, languages, or classification algorithms.

Researchers primarily interested in a robust system for shallow semantic analysis can use the pre-trained classifiers for English and German provided with SHALMANESER. A single command starts the analysis of plain text input, encompassing syntactic analysis, frame assignment and role assignment. More specifically, the training data for English is the FrameNet release 1.2 dataset, consisting of 133,846 annotated BNC examples for 5,706 lemmas. For German, the training data is a portion of the SALSA corpus (Erk et al., 2003), 17,743 annotated instances covering 485 lemmas.

The other aim of SHALMANESER is to allow research in semantic role assignment on a high level of abstraction and control. Studies in this area typically involve a comparative evaluation of different experimental conditions, e.g. the activation and deactivation of model features. In SHALMANESER, these parameters can be specified declaratively in experiment files.

4.2. Evaluation

The WSD and the SRL systems were evaluated against 10% held-out data from the FrameNet and SALSA datasets. The SHALMANESER WSD system obtained an accuracy of 93% (baseline: 89%) for English and 79% (baseline: 75%) for German. The high baseline for English is due to the fact that FrameNet, whose workflow progresses one frame at a time, provides an incomplete sense inventory for many words (but see below). The SHALMANESER SRL system was evaluated separately for the tasks of argument recognition (Is the constituent a role or not?) and argument labeling (If it is a FE, which FE is it?). The results are summarized in Table 4.

Table 4. SRL evaluation results

		argrec		*arglab*
Data	Prec.	Rec.	F	Acc.
English	0.855	0.669	0.751	0.784
German	0.761	0.496	0.600	0.673

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 231)

4.3. Handling incomplete coverage

Adequate coverage is a general problem of automatic semantic analysis, and frame-based shallow semantic parsing is not an exception. The main problem is that FrameNet is still under development, and frames have not been defined for all senses of all lemmas. The most difficult class in this respect is formed by lemmas for which there are no existing frames. Processing these cases requires more lexicographic (and presumably manual) effort. However, there are two classes of lemmas with incomplete coverage that can be treated (semi-)automatically, namely (a) lemmas which are not listed in FrameNet, but presumably fall under an existing frame, and (b) lemmas that are listed, but for which only a subset of the senses is covered by existing frames.

To provide an approximate semantic analysis for the lemmas in class (a) we developed the "Detour to FrameNet" system (Burchardt et al. 2005a). It exploits the larger coverage of WordNet (Fellbaum 1998) to (heuristically) assign existing FrameNet frames that approximate the lemma's meaning. The Detour system generates candidate frames on the basis of WordNet synonyms and hypernyms of the given lemma. It then selects the best fitting frame(s) with a weighting scheme. The Detour system can be used in combination with SHALMANESER, to assign analyses to otherwise unknown lemmas. Alternatively, it can be used on its own, e.g., to generate suggestions for manual annotation in order to speed up the annotation process.

Lemmas of class (b) pose a problem because when one of the senses of a target word is missing from the lexicon, standard WSD algorithms will always incorrectly assign one of the existing senses, wrongly assuming that all applicable sense labels for a target word are known. An example is shown in Figure 10, where a sentence from the Hound of the Baskervilles has been analyzed by SHALMANESER. FrameNet lacks a sense of "expectation" or "being mentally prepared" for the verb *prepare*, so *prepared* is assigned the sense COOKING_CREATION, a possible but improbable analysis.[3] Such erroneous labels can be fatal when further processing builds on the results of shallow semantic parsing, e.g. for drawing inferences.

To address this problem, we developed an approach to detect occurrences of unknown senses (Erk 2006) based on the method of "outlier

3. Unfortunately, the semantic roles have been mis-assigned by the system. The word *I* should fill the FOOD role while *for a hound* should be assigned the optional RECEIVER role.

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 232)

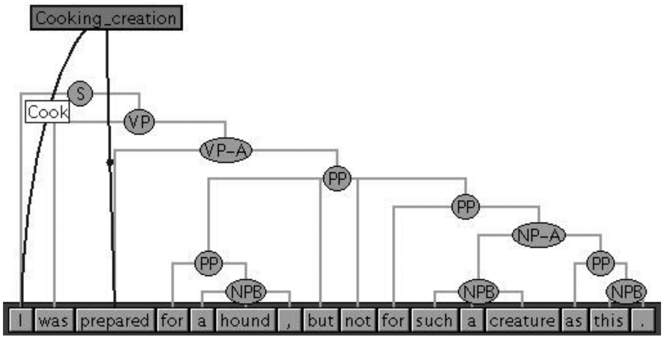

Figure 10. Wrong assignment due to missing sense: Example from "The Hound of the Baskervilles"

detection". An outlier detection model is trained on a set of positive examples only, deriving form it some model of "normality" to which new objects are compared. Its task is then to decide whether a new object belongs to the same set as the training data. For unknown sense detection, we constructed an outlier detection model based on the training occurrences of all senses of the target word. Whenever a new occurrence of the word is classified as an outlier, it is considered an occurrence of an unknown sense. In an evaluation of FrameNet 1.2 data, designating one sense of each lemma as an unknown sense, the best parameter set achieved a precision of 0.77 and a recall of 0.81 in detecting occurrences of unknown senses.

5. Applications

One of the aims of the SALSA project is to explore the usefulness of frame semantic descriptions in language technology. FrameNet descriptions differ from alternative lexical semantic descriptions, such as those found in PropBank, in that they combine different types of semantic information: (i) coarse-grained sense classification in terms of conceptual classes, i.e., frames, (ii) their predicate-argument structure, in terms of FEs, and (iii) semantic relations between frames, in terms of FrameNet's frame hierarchy (Fillmore et al. 2004). As a lexical-semantic framework, it crucially differs from truth-conditional semantic frameworks such as Montague Semantics or Discourse Representation Theory, in disregarding sentence-semantic phenomena such as tense, modality, quantification, or scope.

One application which has recently been successfully approached with frame-based processing is question answering (QA). In textual question answering (Fliedner (2006), Kaisser (2005)), frames present an attractive representation level for matching questions and potential answers. For question answering from structured knowledge bases Frank et al. (2007) applied a somewhat different strategy, which also highlighted the cross-lingual appropriateness of frames. They used frames as an intermediate layer which enabled the automatic translation of (multilingual) natural language questions to structured queries over (language-independent) domain ontologies.

5.1. Textual entailment

In this section, we focus on a problem related to questions answering, namely Recognizing Textual Entailment. Textual Entailment is a relation holding between a text (T) and a hypothesis (H). It holds "if the meaning of H can be inferred from the meaning of T, as would typically be interpreted by people." (Dagan et al. 2005: 1). An example where textual entailment holds is given in (8).

(8) T: In 1983, Aki Kaurismäki directed his first full-time feature.
H: Aki Kaurismäki directed a film.

Checking for textual entailment can be taken as a semantic verification step for many information access tasks. For example, a summarization system might generate (8H) as a summary of (8T); in this context, textual entailment can subsequently be used to ensure the consistency of the summary with the original information.

Modeling Textual Entailment has been institutionalized in the form of the yearly PASCAL Recognizing Textual Entailment (RTE) Challenge, where training data in terms of Text-Hypothesis pairs is provided together with human judgments about whether textual entailment holds or not. The task is then to model this relation and to predict whether entailment holds or not for unseen test data.

5.2. The SALSA contribution to the RTE challenge

Our hypothesis for approaching the RTE task is that FrameNet's coarse-grained conceptual classification and role-semantic analysis offers a useful abstraction layer with a significant degree of normalization across lexical predicates, parts of speech and syntactic argument realization, i.e., diathe-

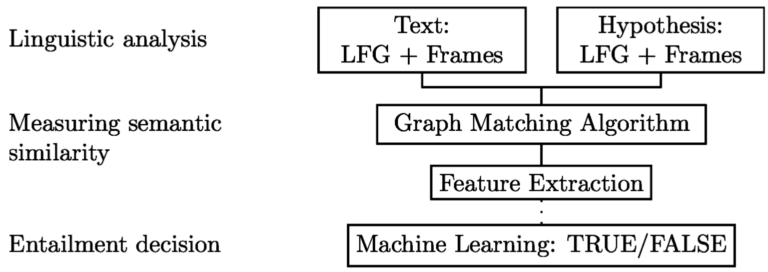

Figure 11. SALSA RTE Architecture

sis variations. Moreover, like WordNet, and based on its hierarchy of frames, FrameNet allows us to determine different types of semantic similarity measures (cf. Burchardt et al. 2005a).

Note, however, that frame semantic analysis on its own is not sufficient for the task. A theoretical issue that needs further consideration is that decisions about entailment often require additional types of information, such as fine-grained lexical information, (e.g., *rise* and *fall* are antonyms), sentence-level of information (e.g., negation or modality), or additional world knowledge. A more practical issue is coverage: At present, we cannot expect to always obtain complete analyses of free texts. We remedy this situation by combining different frame semantics with other resources in a layered approach that provides diverse kinds of information and supports a fall back in the case of missing or partial analyses.

The overall design of our system is shown in Figure 11. The linguistic analyses of H and T are graph structures (with their computation detailed below). They are taken as input to a module that computes *semantic similarity* by way of a graph matching algorithm. Different types of matches (e.g. functional-syntactic, frame-semantic) are recorded and marked as being safe or defeasible depending on the respective matching rules. Further measures of similarity are the size and connectedness of the resulting match graph. These similarities then serve as input to a statistically trained model which "decides" whether entailment holds or not.

The linguistic analysis part of the system is shown in Figure 12. It is centered around a frame-semantic projection on top of a symbolic LFG grammar (Frank and Erk 2004, Frank and Semecky 2004). We employ the English LFG grammar developed at PARC (Riezler et al. 2002), whose f-structure trees serve as an anchor for all information provided by the other resources. The frame-semantic annotations are produced by SHALMANESER and the Detour system (Burchardt et al. 2005a), and subse-

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 235)

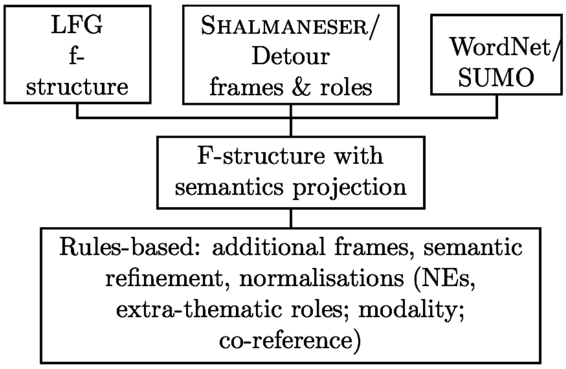

Figure 12. Linguistic analysis component of the SALSA RTE System

quently enriched with information from the WordNet and SUMO ontologies, using a WSD system (Banerjee and Pedersen 2003) and mappings from WordNet to SUMO (Niles and Pease 2003), respectively. Subsequently, the LFG F-structure is evaluated by a heuristic rule-based component to gather information about additional phenomena such as co-reference, modality, etc.

We can now present a complete example. Figure 13 illustrates the LFG and frame semantic analysis of T and H of (8) in the two boxes. The LFG information is displayed on the left of each box, the corresponding frame semantic projection on the right side. The frame `Behind_the_scenes` has been assigned to *direct* and *film* by the automatic frame and FE assignment systems. Based on the Named Entity Recognizer of the LFG, the `People` frame has been assigned in the rule-based refinement step. Because of a disambiguation problem, *feature* was not assigned a frame. However, in the graph matching process, both *feature* and *film* are recognized as a deep syntactic object (**dobj**) of the main predicate. At the same time, a defeasible match based on WordNet has been found to relate both predicates. This provides evidence that the semantic similarity between T and H is very high. H can thus be taken as "fully covered" by T and the statistical model successfully confirms entailment in this case.

The SALSA RTE system participated in the RTE-2 challenge (Burchardt and Frank 2006). With 59% accuracy it scored in the middle range of all participating systems. We take this as evidence that frame semantic analysis integrated with syntactic, lexical, and other types of knowledge resources is a promising basis for large-scale semantic processing.

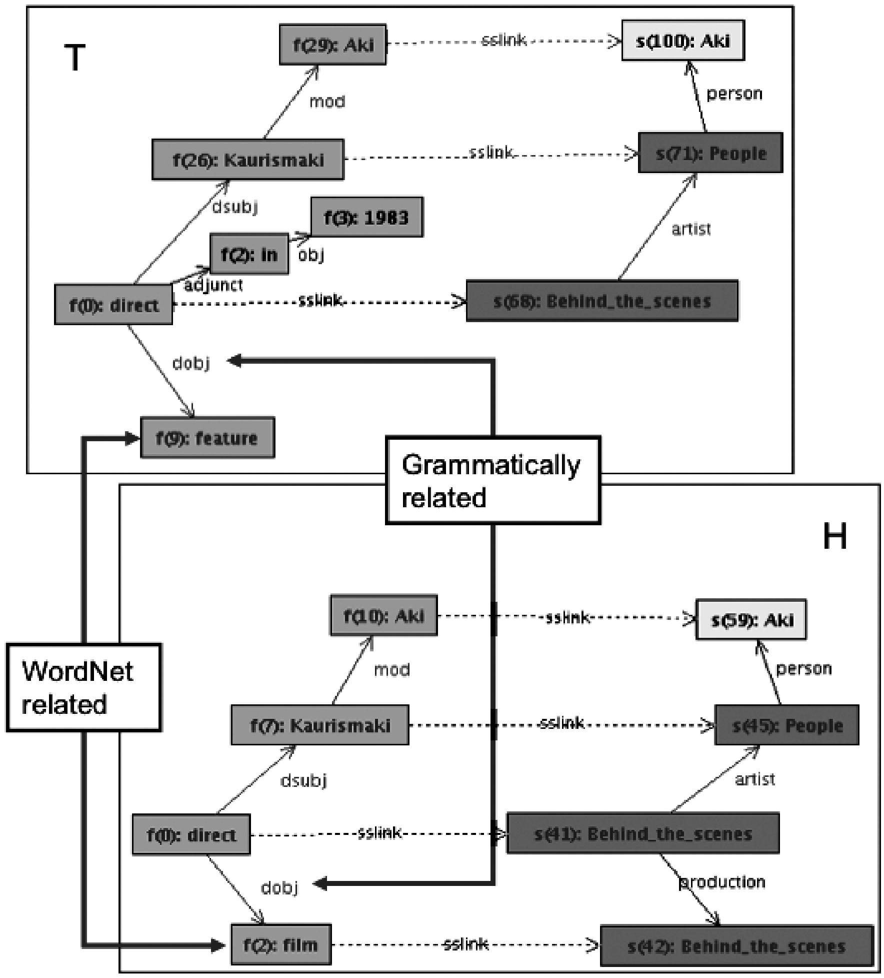

Figure 13. Analysis of example (8)

Ultimately, we envisage that frame-based analyses will be even more competitive in future years of the RTE Challenge, for which an extension to larger chunks of text is planned. We have already studied the interactions of frame semantic structures with discourse phenomena (Burchardt et al. 2005b), and found that frame semantic structures are tightly interrelated with discourse phenomena, and thus may serve as an informative component in models of discourse structure.

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 237)

6. Summary and outlook

In this paper we discussed various aspects in which the current phase of the SALSA project has investigated the annotation, representation and implementation of Frame Semantics, as realized in Berkeley FrameNet. Our results are both practical and theoretical. On the practical side, we have made the following software tools and resources available to the research community:

– The SALTO tool provides a convenient graphical interface for frame-semantic annotation and supports the frame annotation workflow from corpus extraction to quality control
– The SHALMANESER system is employed for shallow, statistical frame-semantic processing
– The DETOUR system offers approximate frame descriptions for missing entries in the FrameNet database
– The SALSA/TIGER corpus provides frame-semantic annotations for German newspaper texts, plus a queryable lexicon that stores the frame-semantic information extracted from the annotated corpus

On the theoretical side, we gained a number of significant insights. First, the initial hypothesis that Frame Semantics provides an appropriate and powerful framework for cross-lingual meaning descriptions has been impressively corroborated by the large-scale re-usability of Berkeley FrameNet frames for the description of German predicate-argument structures. Our successful approach to automatic cross-lingual projection of frame-semantic information from English to German and French bolsters the claim.

Second, we explored the feasibility of large-scale exhaustive frame-semantic annotation of text documents. We demonstrated that the annotation of all kinds of borderline cases and special phenomena of limited compositionality is indeed feasible. Moreover, we showed that frame-semantic annotation supports the systematic modeling of phenomena such as metaphors in an interesting way.

Third, we successfully employed frame-semantic resources for language technology tasks like RTE and Question Answering, confirming our conviction that frame-semantic resources constitute a valuable tool for all kinds of semantically informed natural-language applications.

From our experience, the most pressing issue restricting the extensive use of frame information in language-technology applications is the somewhat limited coverage of frame-semantic resources. Manual lexicon development or manual semantic annotation appears to be too time consuming to quickly arrive at a full coverage high-quality frame-semantic lexicon within the next three to five years. Therefore, we will concentrate on the further development of automated techniques of lexical semantic acquisition in the next phase of SALSA. We thus intend to speed up the development of frame-semantic resources with broader coverage by exploring the use of linguistically informed data expansion techniques and ways to access and integrate complementary knowledge provided by upper-model ontologies into a frame-semantic lexicon.

Acknowledgements

The research reported here was funded by the German Research Foundation (DFG) under Grant PI 154/9-2. We are grateful to the Berkeley FrameNet team and the Cross-lingual FrameNet Group for fruitful collaboration.

7. Appendix: SALSA Resources

The SALSA resources listed below are freely available for academic research.

SALTO

The SALTO tool was implemented at CLT Sprachtechnologie GmbH under the direction of Daniel Bobbert. It is implemented in Java and was tested successfully under Windows, Linux, SunOS and Mac OS X. SALTO can be downloaded from the SALSA project homepage at http://www.coli.uni-saarland.de/projects/salsa/page.php?id=software.

SHALMANESER

The SHALMANESER semantic analysis system is written in Ruby. It makes use of several third-party software systems, as described in the documentation. The system has been tested successfully under Linux. SHALMANESER can be downloaded from http://www.coli.uni-saarland.de/projects/salsa/page.php?id=software.

A WordNet Detour to FrameNet

The Detour system is written in Perl, and is available from the CPAN archive at http://search.cpan.org/~reiter/FrameNet-WordNet-Detour/. It requires FrameNet and WordNet as external resources.

SALSA Release 1.0

The first SALSA release in 2007 contains of a portion of the frame-annotated SALSA/TIGER corpus, together with FrameNet-style documentation of the FrameNet frames used in the annotation as well as the proto-frames developed by SALSA. This release includes a queryable lexicon model that stores the corpus-extracted lexicon data. The release is accessible from the SALSA homepage, at http://www.coli.uni-saarland.de/projects/salsa/page.php?id=release1.0.

8. References

Banerjee, Satanjeev and Ted Pedersen
2003 Extended gloss overlaps as a measure of semantic relatedness. In: *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, 805–810.

Boas, Hans C.
2005 Semantic frames as interlingual representations for multilingual lexical databases. In: *International Journal of Lexicography* 18.4: 445–478.

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*: 24–41.

Burchardt, Aljoscha, Katrin Erk, and Anette Frank
2005a A WordNet Detour to FrameNet. In: Bernhard Fisseni, Hans-Christian Schmitz, Bernhard Schröder, and Petra Wagner (eds.), *Sprachtechnologie, mobile Kommunikation und linguistische Resourcen (Computer Studies in Language and Speech 8.)*, 408–421. Frankfurt am Main: Peter Lang.

Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, and Sebastian Padó
2006a SALTO – a versatile multi-level annotation tool. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation.*

Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal
2006b The SALSA corpus: a German corpus resource for lexical semantics. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation.*

Burchardt, Aljoscha, and Anette Frank
2006 Approaching textual entailment with LFG and FrameNet frames. In: *Proceedings of the RTE-2 Workshop*, 92–97.

Burchardt, Aljoscha, Anette Frank, and Manfred Pinkal
2005b Building text meaning representations from contextually related frames – a case study. In: *Proceedings of the 6th International Workshop on Computational Semantics*, 66–77.

Dagan, Ido, Oren Glickman, and Bernardo Magnini
2005 The PASCAL recognizing textual entailment challenge. In: *Proceedings of the First Challenge Workshop, Recognizing Textual Entailment*, 1–8.

Ellsworth, Michael, Katrin Erk, Paul Kingsbury, and Sebastian Padó
2004 PropBank, SALSA and FrameNet: How design determines product. In: *Proceedings of the Workshop on Building Lexical Resources From Semantically Annotated Corpora at LREC 2004.*

Erk, Katrin
2005 Frame assignment as word sense disambiguation. In: *Proceedings of the 6th International Workshop on Computational Semantics.*

Erk, Katrin
2006 Unknown word sense detection as outlier detection. In: *Proceedings of the joint Human Language Technology Conference and Annual Meeting of the North American Chapter of the Association for Computational Linguistics*, 128–135.

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: *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 537–544.

Erk, Katrin and Sebastian Padó
2004 A powerful and versatile XML format for representing role-semantic annotation. In: *Proceedings of the 4th International Conference on Language Resources and Evaluation.*

Erk, Katrin and Sebastian Padó
2005 Analyzing models for semantic role assignment using confusability. In: *Proceedings of the joint Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 668–675.

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 241)

Fellbaum, Christiane (ed.)
1998 *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Fillmore, Charles J.
1985 Frames and the semantics of understanding. In: *Quaderni di Semantica* 4.2: 222–254.

Fillmore, Charles J., Collin F. Baker, and Hiroaki Sato
2004 FrameNet as a "Net". In: *Proceedings of the 4th International Conference on Language Resources and Evaluation*.

Fillmore, Charles J., Christopher R. Johnson, and Miriam R. L. Petruck
2003 Background to FrameNet. *International Journal of Lexicography* 16.3: 235–250.

Fliedner, Gerd
2006 Towards natural interactive question answering. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation*.

Frank, Anette and Katrin Erk
2004 Towards an LFG syntax-semantics interface for Frame Semantics annotation. In: Alexander Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing*, 1–12. Heidelberg: Springer Verlag.

Frank, Anette, Hans-Ulrich Krieger, Feiyu Xu, Hans Uszkoreit, Berthold Crysmann, Brigitte Jörg, and Ulrich Schäfer
2007 Question answering from structured knowledge sources. *Journal of Applied Logic, Special Issue on Questions and Answers: Theoretical and Applied Perspectives* 5.1: 20–48.

Frank, Anette, and Jiří Semecký
2004 Corpus-based induction of an LFG syntax-semantics interface for Frame Semantic processing. In: *Proceedings of the 5th International Workshop on Linguistically Interpreted Corpora*, 39–46.

Gildea, Daniel and Daniel Jurafsky
2002 Automatic labeling of semantic roles. *Computational Linguistics* 28.3: 245–288.

Hamp, Birgit and Helmut Feldweg
1997 GermaNet: a Lexical-Semantic Net for German. In: *Proceedings of the ACL/EACL97 workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, 9–15.

Hole, Daniel
2005 Towards a unified voice account of dative binding in German. In: Claudia Maienborn and Angelika Wöllstein (eds.), *Event Arguments: Foundations and Applications*, 213–242. Tübingen: Niemeyer.

Ide, Nancy and Jean Véronis
1998 Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics* 24.1: 1–40.

Kaisser, Michael
2005 QuALiM at TREC 2005: Web-Question-Answering with FrameNet. In: *Proceedings of the 2005 Edition of the Text Retrieval Conference, TREC 2005*.

Kilgarriff, Adam and Joseph Rosenzweig
2000 Framework and results for English Senseval. *Computers and the Humanities. Special Issue on SENSEVAL 34* 1–2, 15–48.

Koehn, Phillip
2005 Europarl: A parallel corpus for statistical machine translation. In: *Proceedings of the MT Summit X*.

Lakoff, George and Mark Johnson
1980 *Metaphors we live by*. Chicago: University of Chicago Press.

Manning, Christopher D.
2006 *Local textual inference: It's hard to circumscribe, but you know it when you see it – and NLP needs it*. Manuscript, Stanford University. http://nlp.stanford.edu/~manning/papers/LocalTextual-Inference.pdf.

Niles, Ian and Adam Pease
2003 Linking lexicons and ontologies: mapping WordNet to the suggested upper merged ontology. In: *Proceedings of the International Conference on Information and Knowledge Engineering*, 412–416.

Ohara, Kyoko Hirose, Seiko Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito, and Shun Ishizaki
2004 The Japanese FrameNet project: An introduction. In: *Proceedings of the Workshop on Building Lexical Resources from Semantically Annotated Corpora at LREC 2004*.

Padó, Sebastian and Mirella Lapata
2005a Cross-lingual bootstrapping for semantic lexicons. In: *Proceedings of the 22nd National Conference on Artificial Intelligence*, 1087–1092.

Padó, Sebastian and Mirella Lapata
2005b Cross-lingual projection of role-semantic information. In: *Proceedings of the joint Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing,* 859–866.

Palmer, Martha, Dan Gildea, and Paul Kingsbury
2005 The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics* 31.1: 71–106.

Riezler, Stefan, Tracy H. King, Ronald M. Kaplan, Richard Crouch, John T. Maxwell III, and Mark Johnson
2002 Parsing the Wall Street Journal using a Lexical-Functional Grammar and Discriminative Estimation Techniques. In: *Pro-*

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 243)

ceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 271–278.

Ruppenhofer, Josef, Michael Ellsworth, Miriam R.L. Petruck, and Jan Scheffczyk
2006 FrameNet II: Extended Theory and Practice. http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126.

Sato, Hiroaki
2003 FrameSQL: A software tool for the FrameNet database. In: *Proceedings of the 3rd Conference of the Asian Association for Lexicography* 251–258.

Siegel, Sidney and N. John Castellan
1988 *Nonparametric statistics for the Behavioral Sciences*, 2nd edition. London: McGraw-Hill.

Spohr, Dennis, Aljoscha Burchardt, Sebastian Padó, Anette Frank, and Ulrich Heid
2007 Inducing a Computational Lexicon from a Corpus with Syntactic and Semantic Information. In: *Proceedings of the 7th International Workshop on Computational Semantics*, 210–221.

Subirats, Carlos and Miriam R.L. Petruck
2003 Surprise: Spanish FrameNet! In: *Proceedings of the Workshop on Frame Semantics, XVII. International Congress of Linguists*.

Subirats, Carlos and Hiroaki Sato
2004 Spanish FrameNet and FrameSQL. In: *Proceedings of the 4th International Conference on Language Resources and Evaluation*.

(V9 30/3/09 00:12) WDG (155mm×230mm) TimesNRMT 1114 Boas pp. 209–244 1114 Boas_06_Ch08 (p. 244)