FrameNet for the Semantic Analysis of German: Annotation, Representation and Automation

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

November 26, 2008

Contents

1. Introduction

This Chapter provides an overview of the activities in the Saarbrücken Lexical Semantics Annotation and Analysis (SALSA) project, which runs since summer 2002. The chief aims of the project are (i), the *exhaustive semantic annotation* of a large German corpus resource with FrameNet frames and semantic roles (Baker et al., 1998), including the *induction of a frame-based lexicon* from the annotated data, and (ii), the induction of *data-driven models for automatic frame semantic analysis* as well as their application in practical NLP tasks.

A fundamental assumption of this project is that frames developed in the Berkeley FrameNet project for the description of English can be used as well for the semantic analysis of German. In other words, we assume that frames form a largely *language independent* inventory of semantic classes. While this is clearly a very attractive assumption, it requires empirical investigation. In the area of syntax, for example, it has been found that although contemporary grammar theories may offer frameworks that can be used to describe the syntactic structure of all languages, a major effort is required to devise cross-linguistically consistent grammar models for a multitude of languages (see e.g., Butt et al. (2002)). While in the area of semantics cross-lingual parallelism is a much more difficult notion, in the case of Frame Semantics we see good chances for cross-lingual parallelism of its descriptions. This is a consequence of the way in which frames and their roles are defined in Frame Semantics (Fillmore, 1985): Frames are defined primarily on the conceptual level as “prototypical situations”, and their roles correspond to participants of this situation, typically characterised by reference to the properties they exhibit and the inferences they allow. Moreover, frames are devised as rather coarse-grained conceptual classes. To the extent that these agree across languages, frames can be said to be universally applicable.

However, unlike ontologies, FrameNet’s semantic descriptions do not rely exclusively on conceptual considerations; membership of a predicate in a frame has to be *grounded* linguistically by the predicate’s syntactic ability to realise the *core* frame elements. The *core* frame elements are those “that instantiate a conceptually necessary component of a frame”. For example, the SPEAKER, MESSAGE and ADDRESSEE roles of the COMMITTMENT frame are all core roles, while TIME, PLACE and REASON are not (see Ruppenhofer et al. (2005) for a discussion). As a consequence, non-parallelism on the frame level can occur in case the subcategorisation properties of predicates in a new language differ vastly from conceptually similar English predicates.

It is an open research issue to what extent cases of non-parallelism at the level of frames are correlated with typological differences across languages, in particular with respect to (syntactic) valency, and how to account for cross-linguistic divergences. In the case of SALSA, it has turned out that the vast majority of frames can in fact be applied for the analysis of German – a language that is is comparatively close to English. A number of problems we observed for cross-lingual parallelism between English and German relate to (a) general constructions in German which do not exist in English (such as datives), and (b) idiosyncratic differences in particular semantic domains.

1.1. Plan of the Paper.

In Section 2, we describe the SALSA corpus annotation efforts, presenting the annotation scheme and process, and discussing various challenges that follow from particular choices of our approach. Section 3 discusses cross-lingual aspects of frame semantic annotation. We summarise our experience with frame semantic annotation for German on the basis of English FrameNet frames, as well as commonalities with and differences to related projects for other languages, including 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. We present SHALMANESER, a general shallow semantic parsing architecture for English and German (Section 4) and a system building on frame semantic resources, the SALSA RTE system, that was built to investigates the usefulness of frame-semantic information in practical Natural Language Processing tasks, in particular, the Recognising Textual Entailment (RTE) Challenge (Section 5).

2. SALSA: Semantic Annotation and Lexicon Building for German

The main objective of the SALSA project is the creation of lexical semantics resources for German within the framework of Frame Semantics. Similar to PropBank (Palmer et al., 2005), SALSA has chosen a corpus-based approach, extending an existing German treebank, the TIGER treebank (Brants et al., 2002), with a semantic layer of lexical semantic annotations. We re-

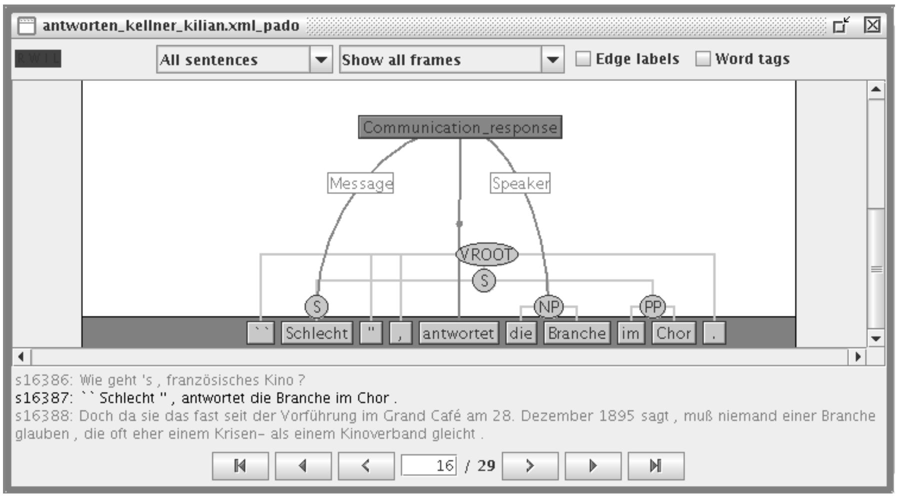

Figure 1. Annotation example: *"Badly", the industry sector answers in unison.*

strict our attention to predicates with a proper predicate-argument structure, currently focussing on verbal predicates, parallel to PropBank practice. Annotation proceeds one predicate at a time and is *exhaustive* in that all corpus instances of the predicate are annotated.

A first release, which is scheduled for early 2007, will consist of about 500 German verbal predicates of all frequency bands plus some deverbal nouns, with a total size of around 20,000 annotated instances.

2.1. Annotation Scheme and Annotation Practice

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 labelled with a *frame*, which can be interpreted as a semantic class, or sense. The edges are labelled with the names of the semantic roles or *frame elements (FEs)* that are defined for the frame, and point to syntactic constituents. Figure 1 shows a simple annotation instance: the verb *antwortet* ("answers") introduces the frame COMMUNICATION_RESPONSE. The NP subject *die Branche* is annotated as realising the frame element SPEAKER and *schlecht,* under an S node, as MESSAGE.

In contrast to FrameNet, we annotate only *core* frame elements (see Section 1). We also use the existing syntactic structure as a guidance for the span of frame elements.

The picture in Figure 1 is a screenshot of SALTO, a graphical annotation tool designed and implemented for the specific needs in SALSA (Burchardt et al., 2006a). However, SALTO can be used more generally for the graphical annotation of treebanks with any kind of relational information in a simple drag-and-drop fashion. SALTO uses a general XML format for input and output, SALSA/TIGER XML (Erk and Padó (2004), see Section 4 for details), and additionally supports corpus management and quality control. SALTO is freely available for research purposes (cf. Section 7).

Much alike PropBank, SALSA follows a corpus-based approach, aiming at *exhaustive* corpus annotation. That is, we aim at annotating all instances of a particular predicate in the corpus. To make this feasible for annotators, annotation proceeds *lemma-wise*: for each lemma we consider, we extract all TIGER sentences that contain the corresponding predicate. The resultig subcorpora are given to (pairs of) annotators for double annotation, together with a list of candidate frames that seem appropriate. The annotators consult the frame definitions in FrameNet, and can also choose additional frames from FrameNet for novel uses they encounter in a given subcorpus.

As a result of our corpus-based, exhaustive annotation practice, we are confronted with two major challenges: one has to do with coverage, the other with the treatment of special linguistic phenomena.

2.2. Coverage Issues

Coverage problems arise from two main sources. The first is a general problem: Even though FrameNet is continually being extended, it does not yet cover the complete "word sense space". The second, more subtle, problem is a result of a our exhaustive annotation strategy: Since we have to analyse each and every instance of a predicate, we also face productive usages whose meaning is clear in the context, but difficult to relate to lexicographical prototypes.

These problems require us to ascertain for each new predicate that all of its senses are covered by FrameNet frames. To do so, we draw a small sample of TIGER instances prior to annotation.

For each instance, we check whether there is a FrameNet frame that provides a felicitous analysis for it. The decision is based on the criteria detailed

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** rechnen zu den armen Ländern der Region.
	CATEGORY	Die Philippinen und Chile rechnen **zu den armen Ländern der Region.**

Table 1. Example of a Proto-frame for *rechnen (zu)* ("*count (as)*") .

in Ellsworth et al. (2004): Does the meaning of the instance meet the frame definition? Can all important semantic arguments of the instance be described in terms of the frame elements? In cases of doubt, we also check annotated FrameNet example sentences for similar usages.

We found that a sample size of twenty is a reasonable compromise between keeping the effort practicable and encountering the most important senses.

2.2.1. Proto-frames

For the majority of German predicates, the process described above results in a list of instances with non-covered readings. We group these into coarse-grained "sense groups" and construct a *predicate-specific proto-frame* for each group. Table 1 shows a proto-frame we constructed for the *to be counted (among a group)* sense of *rechnen*.

Similar to FrameNet frames, the SALSA proto-frames have a textual definition, a set of roles with FrameNet-style names, and annotated example sentences. The proto-frames follow a simple naming convention, e.g. RECHNEN.UNKNOWN3, which marks the third proto-frame constructed for the predicate *rechnen*.

Since SALSA is not a lexicographic project, the SALSA proto-frames are not intended as finalized descriptions of these senses. Nevertheless, our predicate-specific proto-frames can provide input for the further development of FrameNet: We attempt to keep proto-frames 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

	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**

Table 2. Phenomena with limited compositionality (LC)

other proto-frames. For example, for RECHNEN.UNKNOWN3 we record that it is identical to a proto-frame for *zählen*; in the example sentence, *rechnen* can be paraphrased by *zählen*.

We computed prelimiary statistics on a dataset of 12,437 annotated instances. We found that the average number of frames per predicate was 2.33, composed of 1.6 FrameNet frames and 0.73 proto-frames. In other words, somewhat less than one third of the predicate senses in our corpus was not covered by FrameNet. The average polysemy in SALSA (2.33) is higher than the current average WordNet verb polysemy (2.2); this is at least partly due to our treatment of idioms and metaphoric readings as additional senses of predicates, Also, these numbers do not yet reflect the grouping of lexical proto-frames into "larger" frames. More details can be found in Burchardt et al. (2006b).

2.3. Special Phenomena

In standard annotation cases, there is a strong parallelism between syntactic and semantic structure: a single-word predicate lexically introduces a frame, whose frame elements link to syntactic (i.e. subcategorised) arguments, as in the example in Figure 1. However, due to our exhaustive annotation policy, we frequently encounter cases of *limited compositionality* in which frame choice, argument choice, or both, diverge from this simple picture. The main phenomena are support verb constructions, idioms, and metaphors. Their frequencies, computed on the same corpus sample used above, are shown in Table 2.

Almost one seventh of this sample constituted instances of these phenomena. For high-frequency, and therefore highly polysemous, verbs such as *nehmen (take)*, these phenomena even constitute the majority of instances.

For each of these special phenomena, we have developed criteria for their distinction, as well as special annotation schemes. These are briefly outlined below.

2.3.0.1. Support Verb Constructions.

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

2.3.0.2. Idioms.

We use three criteria for identifying idioms: Idioms are multi-word expressions which are (a) (for the most part) fixed, (b) introduce the meaning as a whole, and (c) whose understood meaning is not synchronically recoverable from their literal meaning. An example is *Nachteile in Kauf nehmen*, literally *to take disadvantages into purchase*, meaning *to put up with disadvantages*. Our annotation scheme for idioms is to annotate the complete multi-word expression as the frame-evoking element; arguments do not require special treatment.

2.3.0.3. Metaphors.

Metaphors are distinguished from idioms by 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), in case of metaphors we annotate two frames: a *source* frame

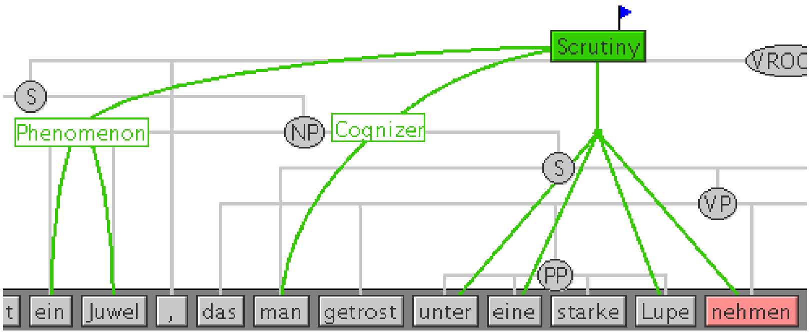

Figure 2. Multiword target for frame SCRUTINY: *Unter eine Lupe nehmen* (lit: *take under a magnifying glass*).

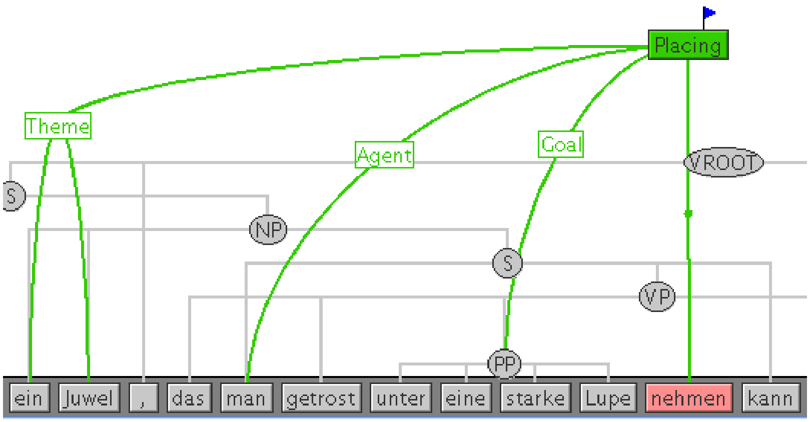

Figure 3. Metaphor source frame PLACING: *Unter eine Lupe nehmen* (lit: *take under a magnifying glass*).

to represent the literal meaning, and a *target* frame to represent the figurative meaning. As an example, consider *unter die Lupe nehmen (to put* (literally: *take*) *under a magnifying glass).* The source frame is TAKING, and the target frame is SCRUTINY, which models the construction of this metaphor as a transfer from a (concrete) putting event to a (more abstract) investigation event.

We attempt to annotate both frames for all metaphorical instances, and mark their status as *Source* and *Target*. Being the result of a complex interpretation process, the target meaning is often difficult to describe. We annotate these cases with the source frame only in order to sustain annotation speed.

In a later stage, these samples can be retrieved for a more comprehensive analysis.

Transfer Schemes for Metaphors.

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 understood meaning. Those instances which have received source *and* target frames can be used to study *transfer schemes*, including information about *argument change*. The SALSA annotation seems well-suited for this task, since frames as sense classes provide an empirically founded, fine-grained vocabulary to describe transfer processes. In addition, roles can be used to describe argument mappings that occur in transfer schemes.

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 Fig. 4 shows a case of *argument incorporation*: the GOAL role of PLACING is absorbed in the frame-evoking element of SCRUTINY; in addition, the modifier *starke (strong)*, which does not fill a role on the source side, fills the DEGREE role in the target frame.

Transfer schemes such as the one shown here do not answer the question as to which factors trigger the metaphorical transfer for a specific utterance. However, they can model the interpretation process of metaphors to a certain degree, and provide a description of the relation between source and target for specific metaphors, which makes it possible to express generalisations over patterns of role shift.

2.3.0.4. Vagueness.

It is a well-known fact that in semantic annotation there are cases of vagueness in which the assignment of only a single label to a markable would not be appropriate (Kilgarriff and Rosenzweig, 2000). For such cases, annotators

$$
\left(\begin{array}{ll}
\text{Target:} & \textit{nehmen} \\
\text{Frame:} & \text{PLACING} \\
\text{Roles:} & \left(\begin{array}{rl}
\text{AGENT:} & \boxed{1}\ \text{man} \\
\text{THEME:} & \boxed{2}\ \text{ein Juwel} \\
\text{GOAL:} & \boxed{3}\ (\boxed{4}\ \text{starke) Lupe}
\end{array}\right)
\end{array}\right) \Rightarrow
\left(\begin{array}{ll}
\text{Target:} & \textit{nehmen} \cdot (\boxed{3}/\boxed{4}) \\
\text{Frame:} & \text{SCRUTINY} \\
\text{Roles:} & \left(\begin{array}{rl}
\text{COGNIZER:} & \boxed{1} \\
\text{PHENOMENON:} & \boxed{2} \\
\text{DEGREE:} & \boxed{4}
\end{array}\right)
\end{array}\right)
$$

Figure 4. 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 scrutiny.”)

should be able to assign more than one label. This makes it possible to retrieve vague cases, and it avoids forcing the annotators to make impossible choices.

SALSA annotation faces the problem of vagueness both at the level of frames and frame elements. As an example for frames, occurrences of the verb *feststellen* (*remark*) often introduce two meaning components, STATEMENT (*say*) and BECOMING_AWARE (*notice*), both of which apply to some extent:

(1) Kein Wunder, daš 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”.’

As an example for frame elements, consider the metonymic sentence (2): *the motion* describes the MEDIUM used to convey the demand, but metonymically it also refers to the SPEAKER.

(2)

Die nachhaltigste Korrektur fordert [**ein Antrag**]$_{\text{MEDIUM} \vee \text{SPEAKER}}$
The most radical change is demanded by **a motion**

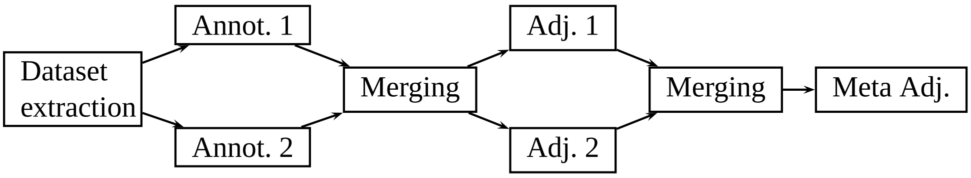

Figure 5. SALSA Annotation and Quality Control.

In cases like these, SALSA annotators can assign more than one frame (or more than one frame element), 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 it is impossible for annotators to decide reliably between the two (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, and between (transparent) metaphors and (no longer transparent) idioms.

2.4. Consistency Control

SALSA aims at guaranteeing quality by double, independent analysis of all data. Figure 5 shows the global structure of the annotation workflow: Each dataset for a given predicate is annotated independently by two annotators – trained undergraduate students – in changing pairs. Through this *double annotation* process, a fair number of annotation mistakes can be detected automatically, and resolved in a manual *adjudication step*: After annotation, the two annotated versions of a dataset are merged into a single copy in which annotation differences are marked. These conflicts are resolved independently by two senior SALSA members in a process we call *double adjudication*. Remaining differences are typically notoriously difficult cases which are then resolved jointly in a final *meta-adjudication* step, by merging the (independently adjudicated) datasets into a single copy, again (see below).

SALTO can be used to manage the whole workflow, as shown in Figure 5, including dataset extraction and merging. Merging means that SALTO uses

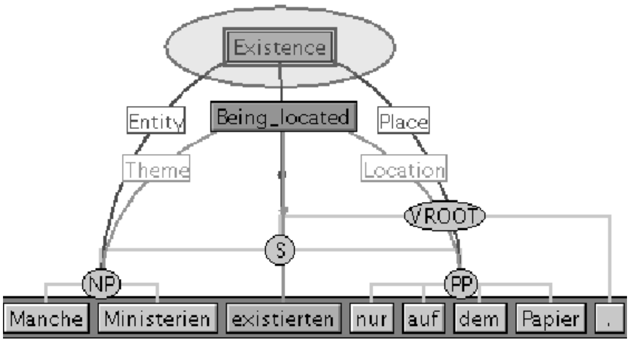

Figure 6. Inter-annotator Difference: EXISTENCE vs. BEING_LOCATED.

two independently annotated datasets to produce a single set in which disagreements are highlighted. In a special annotation mode, SALTO guides the user specifically through those differences to allow their manual inspection and correction. Figure 6 shows an example of an inter-annotator disagreement: the sentence *Manche Ministerien existieren nur auf dem Papier* (*Some ministries exist only on paper*). One annotator has tagged the word *existieren* (*exist*) with the semantic class EXISTENCE, while the other annotator has chosen BEING_LOCATED. The tool has circled EXISTENCE to show that this is the next annotation choice to be either confirmed or denied by the adjudicator.

Almost all disagreements which remain after adjudication are truly difficult cases. Many are *idiosyncractic problems*, i.e. problems with particular instances. Examples are referential ambiguities, which can lead to ambiguous role assignments. A second category consists of *conceptual problems* with respect to the FrameNet inventory. Examples are systematic problems in distinguishing roles, or usages which meet frame descriptions only partially, or else combine aspects of several frames. In cases where the adjudicators cannot reach a uniform decision, underspecification is used as a "last resort". This way, the annotation reflects the difficulty of decision, and uncertainty about the correct analysis.

2.4.1. Computing Agreement

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 role labelling individually, due to their different characteristics.

We are aware that it is best practice for annotation projects to report chance-corrected agreement, such as the kappa statistic (Siegel and Castellan, 1988). However, due to technical difficulties with applying kappa to our annotation task, which are discussed at length in Burchardt et al. (2006b), we do not report kappa, but percentage agreement according to a strict evaluation metric (labelled exact match). Under this scheme, inter-annotator agreement is 85% on frames and 86% on roles. Inter-adjudicator agreement is 97% on frames and 96% on roles. Informally speaking, annotators agree on more than 4/5ths of all instances; adjudication creates consensus for another 4/5ths of the disagreements. These numbers indicate substantial agreement, demonstrating that the task is well-defined.

2.4.2. Limits of Double Analysis.

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. This limits the effectiveness of quality control by inter-annotator agreement with regard to systematic mistakes.

For this reason, we draw random samples for all completely annotated predicate-frame-pairs, which are inspected for possible systematic annotation mistakes. We have also experimented with *intra*-annotator agreement, trying to detect errors by finding "outliers" with non-uniform behaviour. However, due to the highly lexicalised nature of semantic annotation, even correctly annotated datasets can show non-uniformities, which leads to false positives.

A currently unsolved problem is how consistency can be guaranteed across different predicates annotated with the same FrameNet frame, especially in the face of difficult distinctions, e.g. between frame elements.

2.5. From Corpus to Lexicon

2.5.0.1. Lexicon.

Generalisations over semantic structures and their linking properties as encoded in the corpus can, more generally, be represented in the form of a lexicon. SALSA is currently designing a German frame-based lexicon model in a description logic framework (Spohr et al., 2007). The model will include frame descriptions, their syntax-semantics linking patterns with frequency distributions, as well as further information, such as selectional preferences. The lexicon descriptions are extracted from the corpus annotations and at the same time provide back-references to the annotation instances, thus "grounding" the lexicon in the corpus.

Being formalised in description logics, the model is designed to allow for consistency control of the annotated data, both in terms of axioms defined over classes and properties, and by way of flexible queries that extract generalisations and frequency information from the annotations.

3. Cross-lingual Aspects of FrameNet

3.1. Using FrameNet for Exhaustive Semantic Annotation of German Text

The fact that our German corpus annotation is based on frames and roles that were created for English raises the question of the applicability of frame semantic descriptions to other languages, i.e. the multilingual dimension of Frame Semantics in general, and the FrameNet resource in particular. Moreover, applying the (still incomplete) FrameNet lexicon presents us with the challenge of gaps in the inventory of frames, and the problem of "grey areas" and productive usages usually not described in a lexicon. Both of these aspects will be discussed in this section.

In our experience, the vast majority of FrameNet frames can be used fortuitously to describe German predicate-argument structure. Nevertheless, some FrameNet frames required adaptation for SALSA annotation. We identified three classes of problems.

3.1.1. Missing Frame Elements

The use of dative objects is much less restricted in German than it is in English. This leads to problems when a frame fits a sense of a German predicate, but does not foresee a frame element that can be realised as a dative in German. An example is the frame TAKING, in which an AGENT takes possession of a THEME by removing it from a SOURCE. In English, the SOURCE, usually realised as a *from*-PP, can be either a source location or a former possessor;

both together can be expressed only clumsily. In contrast, the German verb *nehmen* can realise location and possessor simultaneously:

(3)

Er	nahm	[ihm]$_{\text{POSSESSOR}}$	[das Bier]$_{\text{THEME}}$	[aus der Hand]$_{\text{SOURCE}}$
He	took	him	the beer	out of the hand

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

3.1.2. *Differences in the Lexicalisation of Frames*

At times, German verbs run counter to the frame distinctions that were designed on the basis of English data. An example is the German verb *fahren*, which encompasses both English *drive* (frame OPERATE_VEHICLE, with a DRIVER role) and *ride* (frame RIDE_VEHICLE, with a PASSENGER role). In German, context often does not disambiguate between the two frames, which makes it impossible to make a decision between these alternative frames. As example, consider (4). Here, German *fahren* is fully unspecified as to whether the people referred to (*they*) were drivers or passengers of the 14 vehicles.

(4)

In 14 Armeefahrzeugen	fuhren	sie	von dem abgezäunten Gelände,
In 14 army vehicles	departed	they	from the enclosed area

"With 14 army vehicles they departed from the enclosed area, which

das	der Besatzungsmacht	28 Jahre lang	als Hauptquartier	gedient	hatte
which	the occupying forces	28 years long	as headquarter	had	served

had served the occupying forces as headquarter for more than 28 years."

In the case at hand, FrameNet has introduced the frame USE_VEHICLE, which subsumes both OPERATE_VEHICLE and RIDE_VEHICLE. While the frame is unlexicalised for English, it is the right level to describe the meaning of German *fahren*. In general, such cases need to be discussed from a multilingual perspective. In the ongoing annotation, we resort to underspecification (see Sec. 2.3.0.4) for such cases, while working towards a cross-lingually valid redefinition of problematic frames.

3.2. SALSA and FrameNet Projects for other Languages

We have indicated above that although SALSA uses the English FrameNet frames, there are considerable differences in annotation philosophy. This section expands on these issues.

3.2.1. Lexicographic vs. Corpus-driven Resource Creation

The aim of the Berkeley FrameNet project is primarily a lexicographic one: to create a linguistically structured network of frames and roles by exploring and documenting semantic classes and their linguistic realisation possibilities. To do so, the project proceeds *frame by frame*, to provide a complete, and potentially contrastive, description of classes. Part of this description is formed by a substantial, but not necessarily complete, list of frame-evoking elements for these frames, and a selection of more or less prototypical example sentences. However, as noticed above, in the case of polysemous predicates, additional readings may remain undescribed for the time being. This philosophy has been adopted by lexicographically oriented projects which "fill" FrameNets for new languages, such as the Spanish FrameNet (Subirats and Petruck, 2003) and Japanese FrameNet (Ohara et al., 2004) projects.

The SALSA project addresses the problem of lexical semantic resource creation from a different perspective, being primarily concerned in providing exhaustive corpus annotation, which we consider as a pre-requisite for obtaining large-scale NLP resources for realistic NLP tasks.

Thus, the primary goal of the annotation in SALSA is complete coverage of a given corpus, which we achieve by proceeding *predicate by predicate*. Since we regard ourselves more as users of the existing FrameNet resource than as creators of a German FrameNet, this releases us from the requirement of covering the entire semantic space, in a way FrameNet aspires to; on the other hand, our exhaustive annotation policy forces us to analyse all instances of a given lemma, which often requires creating proto-frames on the fly, as described in Section 2.2. Finally, exhaustive annotation requires us to address frequently occurring linguistic phenomena as described in Section 2.3, whereas FrameNet mainly considers predicates with a clean syntax-semantics mapping that display "core" conceptual meanings. As a further consequence, we encounter cases of systematic as well as idiosyncratic ambiguity and vagueness, as opposed to "chosen" examples that are intended to illustrate core meanings. In our scenario, therefore, annotators may assign more than one frame or semantic role and mark the occurrence as being 'underspecified'.

Dispite these differences, the two approaches are continuously converging in practice, in that FrameNet is starting to pursue corpus-driven annotation projects; SALSA on the other hand is starting to extract a general lexicon resource from corpus annotations, and spends considerable effort on "protoframing".

3.2.2. Interaction between FrameNets

While SALSA annotates a syntactically analysed corpus, other FrameNet projects are annotating examples on the basis of unparsed sentences, specifying the syntactic properties of annotated roles manually on the fly. This is mirrored on the technical level in the choice of storage format: FrameNet's "lexical unit report" XML files represent annotations one frame at a time, and characterise role spans by way of character spans of the sentence string. SALSA uses SALSA/TIGER XML, 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 6, for example), defining their span in terms of (sets of) syntactic constituents.

In spite of these differences, we have developed several ways of bridging the gap between SALSA and the other FrameNet projects.

Conversion of Annotation Formats

Our first goal was simply to allow the exchange of annotated data between projects. This is very desirable from a technical point of view: mutually convertible data formats make it possible to develop common toolboxes, e.g. for modelling, consistency checking, or simply visualisation using the SALTO tool (see Section 2). We determined that SALSA subcorpora and FrameNet lexical unit (LU) reports were the most appopriate level of granularity to exchange data: One SALSA subcorpus for a predicate corresponds to a set of LU reports, one for each reading of the predicate (i.e., frame). The direction SALSA $\rightarrow$ FrameNet is comparatively simple, since it only consists of removing most of the syntactic structure, retaining only the constituents labelled with semantic roles. The inverse direction (FrameNet $\rightarrow$ SALSA) is also fairly straightforward, in that the span-based characterisation of roles, in conjunction with categorial or functional information, can be used to define a partial syntactic and semantic structure in SALSA/TIGER XML which is restricted to the annotated target word and roles. In practice, this conversion direction was implemented in a different, pragmatically motivated way, in the context of developping a shallow semantic parser (see Section 4 for details): The conversion FrameNet $\rightarrow$ 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 coverts the character-based annotation into a constituent-based annotation. Even though the correctness of the automatic analysis cannot be guaranteed, the results are fairly good, and have made it possible to train the shallow semantic parser directly on FrameNet data and to inspect the output in SALTO.

Alignment of Multi-lingual Frame-annotated Data and Lexicons

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 an informed way. This goal has been realised in the lexicographical domain by FrameSQL, a database-oriented browser for the FrameNet database developed by Hiroaki Sato (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 lately also for English–German (see Figure 7). For example, it is possible to compare the lexical units of two languages for the same frame, and their valencies. This is an interesting application both in terms of translation and foreign language teaching, but more generally for the study of cross-lingual commonalities and divergences in the frame semantic paradigm.

Clearly, these applications only represent first steps in the convergence effects that can be obtained from cross-lingual FrameNet annotation. An important area for future research will be the development of a lexicon model that is modular and powerful enough to represent both SALSA-style and FrameNet-style representations, together with annotated examples and statistical generalisations. This would be a real novelty: a cross-lingual, declarative, computational lexicon with important application aspects. Our current

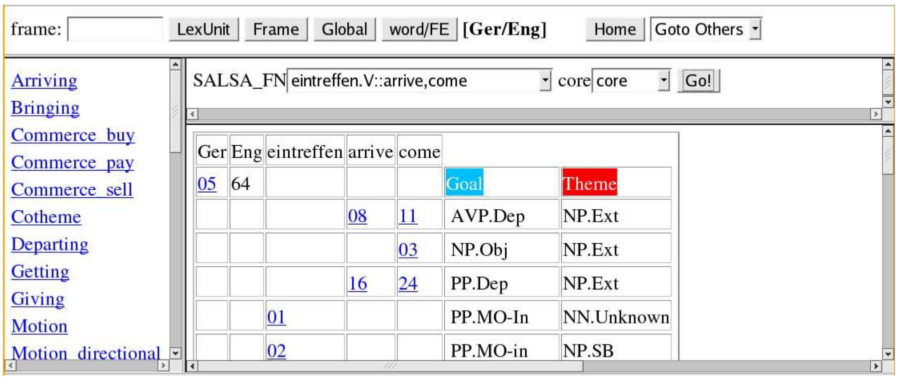

Figure 7. Sato Tool snapshot contrasting English *arrive* with German *eintreffen.*

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

Another direction we would like to pursue in this context is to close the feedback loop between annotation and lexicography, in making proto-frames available as input for full-scale, lexicographic frame construction.

3.3. Cross-lingual Projection for Resource Creation

SALSA, as well as annotation projects for other languages, has found that a large majority of the FrameNet frames, originally developed for English, were well suited to describe the predicate-argument structure of different languages. This observation suggests that there is a chance of automating the task of creating frame-semantic resources for new languages, at least when they are typologically similar to English. This can be done in different ways: a resource-driven strategy which uses dictionaries and ontologies is possible as well as a data-driven strategy which exploits correspondences in large, parallel corpora.

Within SALSA, we have pursued the data-driven approach: We used an instance of *annotation projection* to transfer the information from English FrameNet across word alignment links in a parallel corpus, resulting in comparable frame-semantic resources for French and German. We divided the task into two subproblems: (1), the induction of frame-semantic predicate classifications (i.e., lists of admissible frame-evoking elements for frames); and (2), the creation of a corpus of sentences with role annotation.

With regard to (1), we have developed a general, language-independent architecture to bootstrap frame-semantic predicate 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 non the basis of shallow linguistic knowledge, by exploiting the distributional profile of translation pairs in a large parallel corpus. In experiments on the EUROPARL corpus (Koehn, 2005), we have been able to construct FrameNet-sized predicate classifications for both German and French with a precision of between 65% and 70% at the same size of Berkeley FrameNet (Padó and Lapata, 2005b).

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 which constitute possible roles. This problem can be phrased as a graph optimisation problem, using word alignments to describe the pairwise cross-lingual similarity of phrases, and solved efficiently. In an experimental evaluation (Padó and Lapata, 2005a), we were able to show that roles can be projected with an accuracy of up to 69% F-Score (75% Precision) when English manual role annotation is used. When an imperfect state-of-the-art automatic shallow semantic parser is used to analyse the English text, the performance degrades to 57% F-Score. However, this 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.

4. Automation

4.1. Shallow Semantic Parsing

The last decade has seen immense successes in automatic syntactic analysis, with the availability of syntactically annotated corpora playing a pivotal role. The same development is currently gaining momentum in the area of wide-scale automatic semantic analysis. In particular, automatic predicate-argument structure analysis – the automatic assignment of word senses to predicates and the identification of semantic roles – is important for all NLP applications that benefit from deeper text understanding, such as the applications that Manning (2006) calls “Information Retrieval++”: question answer-

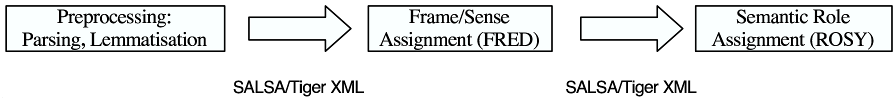

Figure 8. SHALMANESER: A loosely coupled toolchain

ing, information extraction, and customer response systems.

The task of automatic predicate-argument structure analysis, commonly known as *shallow semantic parsing*, can be divided into *Word Sense Disambiguation (WSD)* (in the FrameNet setting: an assignment of frames to target expressions) and *Semantic Role Labeling (SRL)*. While WSD is one of the oldest NLP tasks of all (Ide and Véronis, 1998), SRL has only recently become a task of considerable interest in the computational linguistics community, beginning with the seminal paper of Gildea and Jurafsky (2002).

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 a “black box” to obtain semantic role analyses for text without having to considering the intricacies of shallow semantic parsing, much like it is the case for syntactic parsers today.

SHALMANESER is realised as a loosely coupled toolchain, as shown in Figure 8. It takes plain text as input, which is first lemmatised, part-of-speech tagged, and syntactically analysed. Semantic analysis is then added in two consecutive steps, WSD and SRL: First the frame disambiguation system assigns semantic classes (senses) to predicates; then the role assignment system adds semantic roles to surrounding constituents. Both sense and role assignment are modeled as supervised learning tasks. Sense assignment is decided on the basis of the lexical context and syntactic properties of predicates (Erk, 2005); for role assignment, we rely both on syntactic features (e.g., path from FEE to constituent) and lexical features, which, although sparse, provide crucial information (Erk and Padó, 2005).

The interchange format used in the SHALMANESER system, SALSA/TIGER XML (Erk and Padó, 2004), is a very general format designed for the representation of multi-level annotation. Other applications can be integrated into the toolchain simply by making them SALSA/TIGER XML-compliant. Most importantly, the SALTO annotation tool (cf. Section 2) reads and writes

SALSA/TIGER XML and can therefore 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 that SHALMANESER offers allows for a modular view on 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; furthermore, the last years have seen impressive progress in the accurate computation of individual kinds of semantic information. These range from lexical information (ontological status, lexical relations, polarity) and structural information (scope, modality, anaphoric and discourse structure) to proposition-level information (factivity).

Currently, there is no comprehensive theoretical account of interaction between different kinds of information, even less a theory of processing. Therefore, we believe that the best-suited architecture for semantic processing is a loosely coupled toolchain architecture with a flexible number of individual modules which work more or less independently to solve particular subproblems of the task. Which modules are necessary or helpful is very much a matter of the application.

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. In a "research scenario", the modular architecture enables the integration of additional processing modules; furthermore, we have kept the processing components encapsulated to make them easily adaptable to new features, parsers, languages, or classification algorithms.

For researchers primarily interested in a robust system for shallow semantic analysis, SHALMANESER comes with pre-trained classifiers for English and German. A single command starts the complete 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 predicates. For German, the training data is a portion of the SALSA corpus (Erk et al., 2003), 17,743 annotated instances covering 485 predicates.

One aim of SHALMANESER is to allow research in semantic role assign-

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

Table 3. SRL evaluation results

ment 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 and other conditions are specified declaratively in *experiment files.*

SHALMANESER is freely available for research and can be obtained from the Salsa webpage (see Section 7).

4.2. Evaluation

Both the WSD and the SRL system 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, which 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 a constituent a role or not?) and argument labelling (if it is a role, which role is it?). The results are summarised in Table 3.

4.3. Coverage extension

Coverage is a problem for any lexical resource, and it is even more a problem with FrameNet, which is still a growing resource. Aggravating the problem is the fact that FrameNet is growing one frame (sense) at a time, with the result that some lemmas are missing some of their senses. This causes a problem for automatic semantic analysis: In cases where one of the senses of a target word is missing from the lexicon, standard WSD will *always* wrongly assign one of the existing senses, because it assumes that it knows all applicable sense labels for a target word. Figure 9 shows an example, a sentence from the *Hound of the Baskervilles* analyzed by SHALMANESER. FrameNet is lacking a sense of “expectation” or “being mentally prepared” for the verb *prepare*, so *prepared* has been assigned the sense COOKING_CREATION, a possible but improbable analysis. 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 have developed an approach to detecting occurrences of unknown senses (Erk, 2006), based on outlier 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. If a new occurrence of the word is classified as an outlier, it is considered an occurrence of a previously unseen sense. In an evaluation on FrameNet 1.2 data, using one sense of each lemma as pseudo-unknown, the best parameter set achieved a precision of 0.77 and a recall of 0.81 in detecting occurrences of the unknown sense.

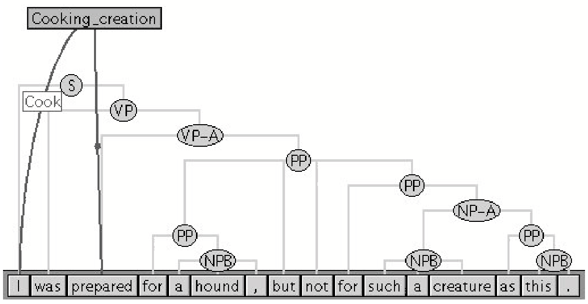

Figure 9. Wrong assignment due to missing sense: from the Hound of the Baskervilles

After predicates not covered by the lexical resource have been identified, they should be given some approximate semantic analysis. To that end, we have developed the “Detour to FrameNet” system (Burchardt et al., 2005a) which exploits the larger coverage of WordNet for heuristically assigning frames to words not yet covered by FrameNet. The Detour system is used in the SALSA system that participated in the Recognizing Textual Entailment (RTE) challenge (see Section 5).

5. Applications

One of the aims of the SALSA project is to explore the usefulness of frame semantic descriptions in practical NLP tasks.

FrameNet differs from alternative lexical semantic descriptions, e.g. those found in PropBank, in that it combines 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 semantic roles, and finally, (iii) semantic relations between frames, in terms of FrameNet's frame hierarchy (Fillmore et al., 2004). On the other hand, as a lexical-semantic framework, it crucially differs from truth-conditional semantic frameworks such as DRT or DPL, in disregarding finer-grained, structural-semantic categories of semantic interpretation, such as tense, modality, or quantificational force and scope.

It is therefore interesting to explore the specific types of uses that frame semantic analysis offers for diverse NLP tasks, as well as novel types of NLP architectures that combine Frame Semantics with truth-conditional semantic frameworks. In SALSA, we designed a system architecture that combines deep syntactic and frame semantic parsing with further semantic and ontology resources, to explore the usefulness of frame semantic analysis in a practical NLP task, namely Recognising Textual Entailment.

5.1. The SALSA Contribution to the RTE Challenge

5.1.1. Textual Entailment

The observation that *entailment* can be taken as semantic constraint for many information access tasks – such as Information Extraction, Question Answering, Information Retrieval or Summarisation – has instigated the community to introduce the pre-theoretical notion of *Textual Entailment*, as a relation holding between a text (T) and a hypothesis (H): a hypothesis H is *textually entailed* by a text T "if the meaning of H can be inferred from the meaning of T, as would typically be interpreted by people." (Dagan et al., 2005) In the PASCAL Recognizing Textual Entailment (RTE) Challenge,[1] training data in terms of Text-Hypothesis pairs is provided together with human judgments as to 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.

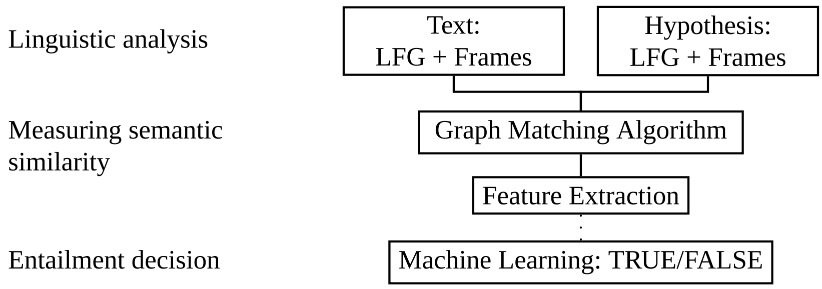

Figure 10. SALSA RTE Architecture

5.1.2. The SALSA Approach to Textual Entailment

For the task of determining semantic entailment between pairs of text fragments, FrameNet's coarse-grained conceptual classification and role-semantic analysis offers a useful abstraction layer with significant degrees of normalisation across lexical predicates, parts-of-speech and syntactic argument realisation, i.e. diathesis variations. Moreover, similar to WordNet, FrameNet allows us to determine different types of *semantic similarity measures*, based on the FrameNet hierarchy (cf. Burchardt et al. (2005a)).

SALSA has participated in the last RTE challenge with a system (Burchardt and Frank, 2006) that is centered around a frame-semantic projection on top of a symbolic LFG grammar (Frank and Erk, 2004, Frank and Semecky, 2004). As can be seen in Figure 10, linguistic analyses of H and T in terms of LFG and frame semantic structures 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, 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 similarity measures serve as input to a statistically trained model which decides whether entailment holds or not.

However, for the task of recognising textual entailment, frame semantic analysis on its own is not sufficient. More fine-grained lexical information is needed, e.g., that *rise* and *fall* are antonyms, and sentence semantic phenomena like negation and modality have to be treated. Moreover, additional world knowledge is sometimes required. Another practical issue is coverage: As of today, we cannot expect to always get full analysis on free text. In order to provide diverse kinds of information and as a fall-back for missing or par-

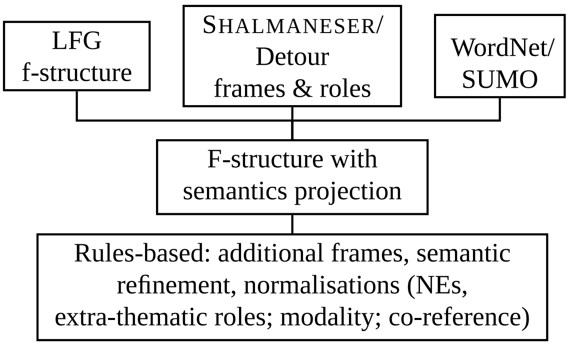

Figure 11. Linguistic Analysis Component of the SALSA RTE System.

tial anlyses, we combine different resources in a layered approach. We integrate the automatic frame-semantic annotations produced by SHALMANESER and the Detour system (Burchardt et al., 2005a) with LFG analysis, and extend the semantic projection with WordNet and SUMO ontologies, as well as rule-based semantic refinements to capture phenomena such as co-reference, modality, etc. The linguistic analysis module is detailed in Figure 11.

As resources we are using the English LFG grammar developed at Parc (cf. Riezler et al., 2002). LFG analysis serves two purposes: First, the f-structure trees serve as anchor for all information provided by the other resources. Second, we model phenomena like negation and modality on the basis of respective f-structure information. In addition we use a WordNet-based Word Sense Disambiguation system (Banerjee and Pedersen, 2003) and mappings from WordNet to SUMO (Niles and Pease, 2003) to assign WordNet synsets and SUMO ontological classes to main predicates.

After the f-structure with the semantic projection has been generated, we apply heuristic rules for normalization and further semantic refinement, e.g., from the output of the LFG Named Entity Recognizer (NER) and analyses of dates and places, additional frames and non-core roles are added to the analysis. Moreover, e.g., anaphoric binding, negation and modality are marked to have this information accessible in the subsequent step computating semantic similarity. For more details, we refer to (Burchardt and Frank, 2006).

The following example from the RTE-2 dataset illustrates how the system operates.

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

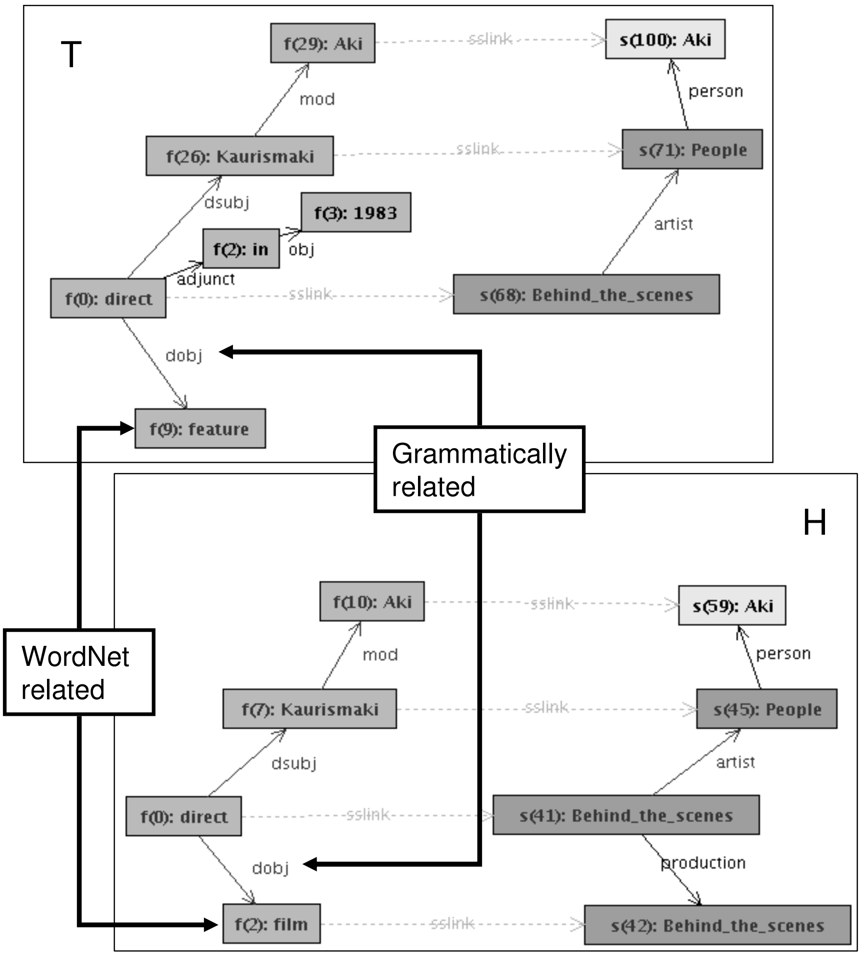

Figure 12. View for example (1).

Figure 12 shows the LFG and frame semantic analysis of T and H in the two boxes. The LFG information is displayed on the left and the frame semantic projection on the right. The frame BEHIND_THE_SCENES has been assigned to *direct* and *film* by the automatic frame and role assignment system. Based on LFG's NER, the PEOPLE frame has been assigned in the rule-based refinement step. Due to a dismbiguation problem, *feature* had not been framed correctly. However, both *feature* and *film* are recognized as a deep syntactic object (`dobj`) of the main predicate, and a defeasible match based on WordNet has been found to relate both predicates. So, the semantic similarity between T and H is very high, in fact, H is fully covered by T and the statistical model successfully confirms entailment here.

With 59% accurracy, the SALSA RTE system scored in the middle ranges, in fact, the exact median, of the participating systems' results. We take this as a piece of evidence that frame semantic analysis integrated with other syntactic, lexical, and knowledge resources is a promising basis for large-scale semantic processing.

5.2. Further Explorations

As a specialisation of the more general problem of textual entailment recognition, frame-based processing has also been applied for textual question answering (QA) in (Fliedner, 2006, Kaisser, 2005).

The cross-linguistic nature of FrameNet has been exploited in an NLP architecture for cross-linguistic question answering from structured knowledge bases (Frank et al., 2006). Here, HPSG-based semantic analysis in terms of Minimal Recursion Semantics (Copestake et al., 2005) is extended with a frame semantic layer, to enable automatic translation of (multi-lingual) natural language questions to structured queries over (language-independent) domain ontologies. In this architecture, the combination of frame semantics with fine-grained truth-conditional semantics successfully accounts for the treatment of complex quantificational questions.

The FrameNet hierarchy, with its diverse inventory of frame-to-frame relations has further been subject to investigations that study the interactions of frame semantic structures with discourse phenomena. As Burchardt et al. (2005b) have shown, frame semantic structures are tightly interrelated with discourse phenomena, and thus may serve as an informative component in models of discourse structure.

6. Conclusions

Acknowledgements.

The research reported here was funded by the German Research Foundation, DFG under Grant PI 154/9-2.

7. Appendix: SALSA Resources

SALTO

The SALTO tool was implemented by a team at CLT Sprachtechnologie GmbH3 under the direction of Daniel Bobbert. It is implemented in Java using the Swing library for the GUI. The system was tested successfully under Windows, Linux, SunOS and Mac OS X. SALTO is available free of charge for academic research. It can be downloaded from the SALSA project homepage, at `http://www.coli.uni-saarland.de/projects/salsa/page.php?id=software`.

SHALMANESER

SHALMANESER is written in Ruby, an object-oriented scripting language. The requirements for using it are as follows: For preprocessing, it necessitates external NLP tools for linguistic analysis. FRED is self-contained. For ROSY, an installed MySQL database server for data storage, and one of the supported classification toolkits. The complete system is free for research and 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 freely available for academic research and can be downloaded from the CPAN archive at `http://search.cpan.org/~reiter/FrameNet-WordNet-Detour/`.

SALSA Release 1.0

A first SALSA release, planned for 2007, will consist of a portion of the frame-annotated SALSA/TIGER corpus, together with FrameNet-style documentation of the applied FrameNet frame inventory as well as proto-frames. The release will include a queryable lexicon model that stores the corpus-extracted lexicon data. The release will be made accessible from the SALSA homepages, at `http://www.coli.uni-saarland.de/projects/salsa/page.php?id=release1.0`.

Bibliography

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet project. In *Proceedings of the joint Annual Meeting of the Association for Computational Linguistics and International Conference on Computational Linguistics*, Montreal, QC, 1998.

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

Sabine Brants, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, and George Smith. The TIGER treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*, Sozopol, Bulgaria, 2002.

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

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

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

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

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

Miriam Butt, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. The parallel grammar project. In *Proceedings of the COLING Workshop on Grammar Engineering and Evaluation*, pages 1–7, 2002.

Anne Copestake, Dan Flickinger, Ivan Sag, and Carl Pollard. Minimal Recursion Semantics. To appear in: *Research on Language and Computation*, 2005.

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

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

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

Katrin Erk. 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*, New York City, NY, 2006.

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

Katrin Erk and Sebastian Padó. Analysing 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*, Vancouver, BC, 2005.

Katrin Erk, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 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*, pages 537–544, Sapporo, Japan, 2003.

Charles J. Fillmore. Frames and the semantics of understanding. *Quaderni di Semantica*, IV(2):222–254, 1985.

Charles J. Fillmore, Collin F. Baker, and Hiroaki Sato. Framenet as a "net". In *Proceedings of LREC 2004*, pages 61–65, Lisbon, Portugal, 2004.

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

Anette Frank and Katrin Erk. Towards an LFG syntax-semantics interface for frame semantics annotation. In *Proceedings of CICLing*, 2004.

Anette Frank and J. Semecky. Corpus-based Induction of an LFG Syntax-Semantics Interface for Frame Semantic Processing. In *Proceedings of LINC 2004*, Geneva, Switzerland, 2004.

Anette Frank, Hans-Ulrich Krieger, Feiyu Xu, Hans Uszkoreit, Berthold Crysmann, Brigitte Jörg, and Ulrich Schäfer. Question answering from structured knowledge sources. *Journal of Applied Logic, Special Issue on Questions and Answers: Theoretical and Applied Perspectives*, 2006. to appear.

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

Nancy Ide and Jean Véronis. Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics*, 24:1–40, 1998.

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

Adam Kilgarriff and Joseph Rosenzweig. Framework and results for English Senseval. *Computers and the Humanities*, 34(1-2), 2000.

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

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

Christoper D. Manning. Local textual inference: It's hard to circumscribe, but you know it when you see it - and nlp needs it. MS, Stanford University. `http://nlp.stanford.edu/~manning/papers/LocalTextualInference.pdf`, 2006.

Ian Niles and Adam Pease. Linking lexicons and ontologies: Mapping wordnet to the suggested upper merged ontology. In Hamid R. Arabnia, editor, *Proceedings of the International Conference on Information and Knowledge Engineering*, pages 412–416, Las Vegas, Nevada, 2003. CSREA Press. ISBN 1-932415-08-4.

K. Ohara, S. Fujii, T. Ohori, R. Suzuki, and H. Saito an d S. Ishizaki. The Japanese FrameNet project: An introduction. In *Proceedings of the Workshop on Building Lexical Resources from S emantically Annotated Corpora at LREC 2004*, 2004.

Sebastian Padó and Mirella Lapata. 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*, Vancouver, BC, 2005a.

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

Martha Palmer, Dan Gildea, and Paul Kingsbury. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1), 2005.

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

Josef Ruppenhofer, Michael Ellsworth, Miriam R.L. Petruck, and Christopher R. Johnson. FrameNet: Theory and Practice. `http://www.icsi.berkeley.edu/~framenet/book/book.html`, 2005.

H. Sato. FrameSQL: A software tool for the FrameNet database. In *Proceedings of the 3rd Conference of the Asian Association for Lexico graphy*, Tokyo, Japan, 2003.

S. Siegel and N.J. Castellan. *Nonparametric statistics for the Behavorial Sciences*. McGraw-Hill, 2nd edition, 1988.

Dennis Spohr, Aljoscha Burchardt, Sebastian Pado, Anette Frank, and Ulrich Heid. Inducing a Computational Lexicon from a Corpus with Syntactic and Semantic Information. In *Proceedings of the 7th International Workshop on Computational Semantics, IWCS-7*, Tilburg, The Netherlands, 2007.

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

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