Corpus-based Induction of an LFG Syntax-Semantics Interface for Frame Semantic Processing

Abstract

We present a method for corpus-based induction of an LFG syntax-semantics interface for frame semantics, porting frame annotations from a manually annotated corpus to a computational LFG parsing architecture. We show how to model frame semantic annotations in an LFG projection architecture, including special phenomena that involve non-isomorphic mapping between levels.

1 Introduction

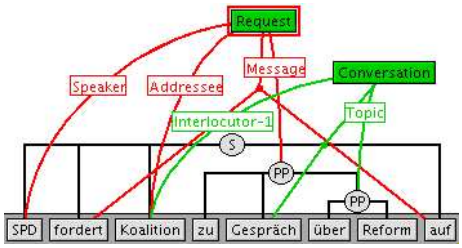
There is a growing insight that high-quality NLP applications for information access are in need of deeper, in particular, semantic analysis. A bottleneck for semantic processing is the lack of large domain-independent lexical semantic resources. While WordNets provide paratactic semantic relations, we are lacking large-scale lexical semantics resources for predicate-argument structure.

There are now efforts for the creation of large lexical semantic resources that provide information on predicate-argument structure. The FrameNet project (Baker et al., 1998), building on Fillmore's theory of frame semantics, has a lexicographic character, providing definitions of frames and their semantic roles, with manual annotation of selected example sentences. A strictly corpus-based approach is carried out with "PropBank" (Kingsbury et al., 2002) – a manual predicate-argument annotation on top of the Penn treebank.

There are first approaches for learning stochastic models for semantic role assignment from annotated corpora, e.g. (Gildea and Jurafsky, 2002; Fleischman et al., 2003). While current competitions explore the potential of shallow parsing for role labelling, (Gildea and Palmer, 2002) have emphasised the role of deeper syntactic analysis for semantic role labelling. We follow (Gildea and Palmer, 2002), and explore the potential of deep syntactic analysis for semantic role labelling, choosing Lexical Functional Grammar as underlying syntactic framework. A computational syntax-semantics interface for semantic role assignment can be used for semi-automatic annotation of unparsed corpora, and thus provide larger training sets for stochastic frame assignment than the current manually annotated corpora.

We discuss advantages of semantic role assignment on the basis of functional syntactic analyses as provided by LFG parsing, and present an LFG syntax-semantics interface for frame semantics, building on a first study in (Anonymous, 2004). In the present paper we focus on the corpus-based induction of a computational LFG interface for frame semantics from a semantically annotated corpus. We describe the methods used to derive an LFG-based frame semantic lexicon, and discuss the treatment of special (since non-isomorphic) mappings in the syntax-semantics interface. Finally, we apply the acquired frame assignment rules in a computational parsing architecture, using a wide-coverage LFG grammar of German.

The paper is structured as follows. Section 2 gives background information on the semantically annotated corpus we are using, and the LFG resources that provide the basis for automatic frame



SPD requests that coalition talk about reform

Figure 1: SALSA/TIGER frame annotation

assignment. In Section 3 we discuss advantages of deeper syntactic analysis for a general, principle-based syntax-semantics interface for semantic role labelling. We present an LFG interface for frame semantics which we realise in a modular description-by-analysis architecture. In Section 4 we describe the method we apply to derive generic frame assignment rules from corpus annotations: we port the frame semantic annotations to a "parallel" LFG corpus (Section 4.1); the resulting semantically enriched LFG corpus is used to induce LFG frame assignment rules, by extracting syntactic descriptions for the frame constituting elements (Section 4.2). We make use of LFG's functional representations to distinguish local and non-local assignment rules. The derived frame assignment rules are reapplied to the original syntactic LFG *corpus* to measure the ambiguity rate imported by the obtained frame assignment rules (Section 4.3). In Section 5 we apply and evaluate the frame projection rules in an LFG parsing architecture, using a wide-coverage LFG grammar of German. In Section 6 we summarise our results and discuss future directions.

2 Corpus and Grammar Resources

Frame Semantics Corpus Annotations The basis for our work is a corpus of manual frame annotations, which follow the definitions of frames and their semantic roles in the FrameNet database[1] (the SALSA/TIGER corpus, (Erk et al., 2003)). The underlying corpus is a syntactically annotated corpus of German newspaper text, the TIGER treebank (Brants et al., 2002). The TIGER syntactic annotations consist of relatively flat constituent graph representations, with edge labels that indicate functional information, such as head (HD), subject (SB), etc., cf. Fig. 1.

Frame annotations are flat graphs connected to constituents in the syntactic annotation. Fig. 1 displays frame annotations where the REQUEST frame is triggered by the (discontinuous) frame evoking element (FEE) *fordert..auf* (requests). The semantic roles (or frame elements, FEs) are represented as labelled edges pointing to constituents in the syntactic analysis: the noun *SPD* for SPEAKER, *Koalition* for ADDRESSEE, and the PP *zu Gespräch über Reform* for the MESSAGE.

LFG Grammar Resources We aim at a computational syntax-semantics interface for frame semantics that can be used for (semi-)automatic corpus annotation, to extend the size of current training corpora, and ultimately as a basis for automatic frame assignment, using the acquired stochastic models. As a grammar resource we chose a wide-coverage computational LFG grammar for German (developed at IMS, University of Stuttgart). The grammar runs on an efficient processing platform that further provides stochastic training and online disambiguation packages. The German LFG grammar was used for semi-automatic syntactic annotation of the TIGER corpus, reporting coverage of 50%, and 70% precision (Brants et al., 2002). The grammar is currently being further extended, and will be enhanced with stochastic disambiguation, following (Riezler et al., 2002).

LFG Corpus Resource Next to the German LFG grammar, (Forst, 2003) has derived a "parallel" LFG f-structure corpus from the TIGER treebank, by applying a transfer method for treebank conversion. We make use of this "parallel" corpus to induce LFG-based frame annotation rules from the SALSA/TIGER annotations (cf. Section 4).

3 LFG for Frame Semantics

Lexical Functional Grammar (Bresnan, 2001) assumes multiple levels of representation. Most prominent are the syntactic representations of c(onstituent)- and f(unctional)-structure. The correspondence between c- and f-structure is defined by functional annotations of rules and lexical entries. This architecture can be extended to semantics projection (Halvorsen and Kaplan, 1995).

[1] http://www.icsi.berkeley.edu/~framenet

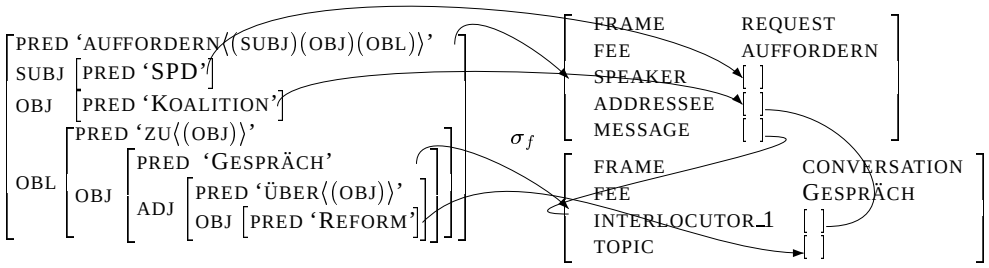


Figure 2: LFG projection architecture for Frame Annotation

auffordern V,
(↑PRED)='AUFFORDERN⟨(↑SUBJ)(↑OBJ)(↑OBL)⟩'
...
(σ_f(↑) FRAME) = REQUEST
(σ_f(↑) FEE) = (↑ PRED FN)
(σ_f(↑) SPEAKER) = σ_f(↑ SUBJ)
(σ_f(↑) ADDRESSEE) = σ_f(↑ OBJ)
(σ_f(↑) MESSAGE) = σ_f(↑ OBL OBJ)

Figure 3: Frame projection by co-description

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

Figure 4: Frame projection by DBA (via transfer)

LFG f-structure representations abstract away from surface-syntactic properties, by localising arguments in mid- and long-distance constructions, and therefore allow for uniform reference to syntactic dependents in diverse syntactic configurations. This is important for the task of frame annotation, as it abstracts away from aspects of syntax that are irrelevant to frame (element) assignment.

In (1), for example, the SELLER role can be uniformly associated with the *local* SUBJect of *sell*, even though it is realized as (a.) a relative pronoun of *come* that controls the SUBJect of *sell*, (b.) an implicit second person SUBJ, (c.) a non-overt SUBJ controlled by the OBLique object of *hard*, and (d.) a SUBJ (*we*) in VP coordination.

(1) a. *The woman who* had come in to *sell* flowers overheard their conversation.
b. Don't *sell* the factory to another company.
c. It would be hard for *him* to *sell* newmont shares.
d. .. *we* decided to sink some of our capital, buy a car, and *sell* it again before leaving.

LFG Semantics Projection for Frames As in a standard LFG projection architecture, we define a frame semantics projection σ_f from the level of f-structure. We define the σ_f–projection to introduce elementary frame structures, with attributes FRAME, FEE (frame-evoking element), and frame-specific role attributes. Fig. 2 displays the σ_f–projection for the sentence in Fig. 1.[2]

Fig. 3 states the lexical entry for the REQUEST frame. σ_f is defined as a function of f-structure. The verb *auffordern* introduces a node σ_f(↑) in the semantics projection of ↑, its local f-structure, and defines its attributes FRAME and FEE. The frame elements are defined as σ_f–projections of the verb's SUBJ, OBJ, and OBL OBJ functions. E.g. the SPEAKER role, referred to as (σ_f(↑) SPEAKER), the SPEAKER attribute in the frame projection σ_f(↑) of ↑, is defined as identical to the σ_f–projection of the verb's SUBJ, σ_f(↑ SUBJ).

Frames in Context The projection of frames in context can yield partially connected frame structures. In Fig.2, *Gespräch* projects to the MESSAGE role of REQUEST, but it also introduces a frame of its own, CONVERSATION. Thus, the CONVERSATION frame, by coindexation, is an instantiation, in context, of the MESSAGE of REQUEST.

Co-description vs. description-by-analysis In the so-called co-description architecture we just presented, f- and s-structure equations jointly determine the valid analyses of a sentence. Analyses that do not satisfy *both* f- and s-structure constraints are inconsistent and ruled out.

[2]The MESSAGE role is coindexed with a lower frame, the frame projection introduced by the noun *Gespr¨ach*.

An alternative to co-description is semantics construction via description-by-analysis (DBA) (Halvorsen and Kaplan, 1995). Here, semantics is built on top of fully resolved (disjunctive) f-structures. F-structures that are consistent with semantic mapping constraints are semantically enriched – remaining analyses are left untouched.

Both models are equally powerful – yet while co-description integrates the semantics projection into the grammar and parsing process, DBA keeps it as a separate module. Thus, with DBA, semantics does not interfere with grammar design and can be developed separately. The DBA approach also facilitates the integration of external semantic knowledge sources (word senses, named entities).

DBA by transfer We realise the DBA approach by way of a term-rewriting transfer system.[3] The system represents f-structures as sets of predicates which take as arguments variables for f-structure nodes or atomic values. Transfer is defined as a sequence of ordered rules. If a rule applies to an input set of predicates, it defines a new output set. This output set is input to the next rule in the cascade. A rule applies if all terms on its left-hand side match some term in the input set. The terms on the right hand side (prefixed '+') are added to the input set. There are obligatory (==>) and optional (?=>) rules. Optional rules introduce two output sets: one results from application of the rule, the other is equal to the input set.

Fig. 4 displays a transfer rule that corresponds to the co-descriptive lexicon entry in Fig. 3. For matched f-structure nodes (predicate, subject, object and oblique object) we define a σ_f–projection (predicate 's::$_f$') with new s-structure nodes. For these, we define the frame information (FRAME and FEE) and the linking of semantic roles (e.g., the σ_f–projection SemA of the SUBJ is defined as the SPEAKER role of the head's projection SemX).

4 Corpus-based induction of an LFG frame semantics interface

4.1 Porting SALSA annotations to LFG

A challenge for corpus-based induction of a syntax-semantics interface for frame assignment is the transposition of manual corpus annotations from a given syntactic annotation scheme to the target syntactic framework. The basis for our work are semantic annotations of the SALSA/TIGER corpus (Erk et al., 2003), encoded in an XML annotation scheme that extends the syntactic TIGER XML annotation scheme (Erk and Pado, 2004).

The TIGER treebank has been converted to a 'parallel' LFG corpus (Forst, 2003). The TIGER/SALSA and TIGER-LFG corpora could be used to learn corresponding path descriptions within the respective syntactic structures – due to the fact that they consist of identical sentences. That is, we could establish the syntactic path descriptions for each frame constituting element in the SALSA/TIGER corpus, and port the annotations to the corresponding path in the TIGER LFG corpus.

We employed a simpler method, exploiting the fact that the TIGER-LFG corpus preserves the original TIGER constituent identifiers, as f-structure features TI-ID (cf. Fig. 7). We make use of these 'anchors' to port the SALSA annotations to the corresponding LFG TIGER treebank. Thus, in a first step we extend the latter to an LFG *corpus* with a frame semantics projection. From this extended corpus we induce general LFG-based frame assignment rules (cf. Section 4.2).

Porting annotations by transfer For each sentence we extract the constituent identifiers of frame constituting elements in the SALSA XML annotations (cf. Fig. 5). This information is coded into generic transfer rules, where we refer to the

Frame	FeeID	Role(s)	RoleID(s)
Request	2 (from {2, 8})	Speaker	1
		Addressee	3
		Message	501

Figure 5: Core frame information for ex. in Fig. 1

% template: get node N for which TI_ID contains ID
get_n(N,ID) :: ti_id(N,TI_set), in_set(ID,TI_set).

get_n(X,FeeID), pred(X,Pred) ==>
+'s::'(X,S_X), +frame(S_X,Frame), +fee(S_X,Pred).

get_n(X,FeeID), 's::'(X,S_X), frame(S_X,Frame),
get_n(Y,RoleID), pred(Y,Pred) ==>
+'s::'(Y,S_Y), +Role(S_X,S_Y), +rel(S_Y,Pred).

Figure 6: SALSA-2-TIGER-LFG transfer

[3] The system operates on packed (disjunctive) f-structures.

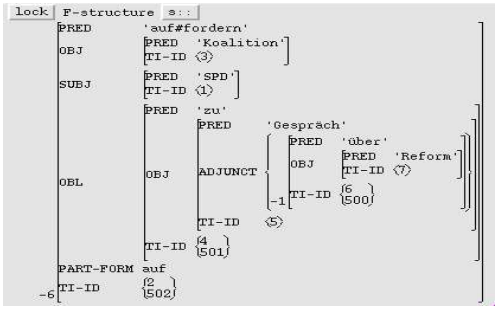


Figure 7: TIGER-LFG f-structure (w/ TI-ID)

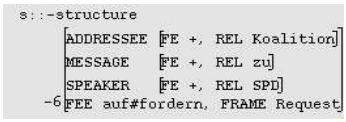


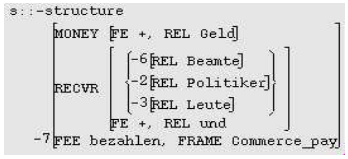
Figure 8: Frame projection from f-str of Fig. 7

corresponding TI-ID features in the f-structure as anchors to project the frame information of a given frame annotation instance. The first rule of Fig. 6 defines the semantic projection of the FEE. Subsequent rules – one for each role to be assigned – define the given semantic role as an argument of the FEE's semantic projection.

We generate a set of frame projection rules for each sentence in the SALSA/TIGER corpus, and apply them to the corresponding f-structure in the LFG-TIGER corpus. The result is an LFG corpus with frame semantic anntations (see Figs. 7, 8).

The basic structure of frame-inducing rules in Fig. 6 was refined to account for **special cases**:

Coordination For frame elements that correspond to coordinated constituents, as in Fig. 9, we project a semantic role that records a *set* of semantic predicates (REL), one for each of the corresponding syntactic conjuncts. The rules in Fig. 6 are extended to track coordinated frame elements, and introduce semantic projections for each of the conjuncts in a set-valued frame role.



Beamten, Politikern und Gesch¨aftsleuten wird Schmiergeld bezahlt–*Clerks, politicians and businessmen* are payed bribes

Figure 9: Frame with coordinated RECVR role

Underspecification The SALSA annotation scheme allows for *underspecification*, to represent unresolved word sense ambiguities or optionality (Erk et al., 2003). In a given context, a predicate may evoque alternative frames (i.e. word senses), where it is impossible to decide between them. E.g. the verb *verlangen* (demand) may convey the meaning of REQUEST, but also COMMERCIAL TRANSACTION. Such cases are annotated with alternative frames, which are marked as elements of an 'underspecification group'. Underspecification may also affect frame elements of a single frame. A motion (*Antrag*), e.g. , may be both MEDIUM and SPEAKER of a REQUEST. Finally, a constituent may or may not be interpreted as a frame element of a given frame. It is then represented as a single element of an underspecification group.

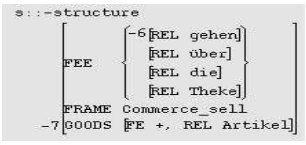
We model underspecification as disjunction, which in the transfer is encoded by optional rules. An optional rule creates two alternative contexts, where we take care to prevent application of subsequent (alternative) frame projection rules to a context that has been enriched by a previous rule. True optionality is modeled by a single optional rule. Fig. 10 displays the result of underspecified frame element assignment in an f-structure chart representation (Maxwell and Kaplan, 1989).

Figure 10: Underspecification as disjunction

In a symbolic account, disjunction does not correctly model the intended meaning of underspecification. However, a stochastic model for frame assignment should render the vagueness involved in underspecification by way of (close) stochastic weights. Thus, annotation instances of underspecification will provide alternative frames in the training data, and can also be used for fine-grained evaluation of statistical frame assignment models.

Multiword Expressions The treatment of multiword expressions (such as idioms or support constructions) requires special care. For idioms, the constituting elements are annotated as multiple frame evoking elements (cf. Fig. 11 for *über die Ladentheke gehen* (go over the counter, i.e. being sold)). We define semantic projections for the in-

dividual components, which are recorded in a set-based FEE value. Otherwise, idioms are treated like ordinary main verbs. Like *sell*, the expression triggers a COMMERCE_SELL frame with the appropriate semantic roles, here GOODS.



Vier Artikel gingen über die Ladentheke–Four items were sold

Figure 11: Multiword expressions

Asymmetric Embedding Another type of non-isomorphism between syntactic and semantic representation occurs in cases where distinct syntactic constituents are annotated as instantiation of a single semantic role. In (2), PP and NP are annotated as the MESSAGE of a STATEMENT, since they jointly convey its content. Projecting distinct constituents to a single semantic node can, however, lead to inconsistencies in cases where both constituents independently project semantic frames.

(2) Der Geschäftsführer gab [$_{PP-MO}$ als Grund für die Absage] [$_{NP-OBJ}$ Terminnöte] an
The director mentioned [time conflicts] [as a reason for cancelling the appointment]

In the SALSA annotations *asymmetric* embedding at the semantic level is the typical pattern for such double-constituent annotations. I.e., for (2), we assume a target frame structure where the MESSAGE of STATEMENT points to the PP – which itself projects a frame REASON, with semantic roles CAUSE for *Terminnöte*, and EFFECT for *Absage*.

Multiple-constituent annotations of this kind arise in cases where frame annotations are *partial*: since corpus annotation proceeds frame-wise, in examples like (2) the REASON frame may not have been treated yet. Moreover, annotators are in general not shown complete(d) sentence annotations.

We account for these cases by a simulation of *functional uncertainty* equations, which accommodate for a potential embedded frame within either one of the otherwise re-entrant constituents. We apply a transfer rule set that embeds one (or the other) of the two constituent projections as an *embedded* role of an unknown frame, to be evoked by the respective 'dominating' node. We introduce an 'unknown' role ROLE* for the embedded constituent, which is to be interpreted as a functional uncertainty path over variable semantic roles.

Fig. 12 displays the alternative (hypothetical) frame structures for (2), where the second one – with FRAME instantiated to REASON and ROLE* to CAUSE – corresponds to the actual reading.

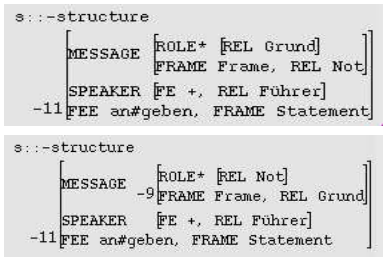


Figure 12: Asymmetric embedding

Overview of data Our current data set comprises 5076 frame annotations. Table 1 gives frequency figures for the special phenomena. Besides the standard cases, our transfer routines for porting SALSA to LFG cover underspecification and coordination. We deferred the treatment of multiword expressions (mwe)[4] and double constituent annotations (asym)[5] – which leaves 4096 frames. Yet, we had to filter out 1569 sentences from the LFG corpus that contained deficient f-structures.[6] This reduced our data set to 2837 frames (55.89%), with 178 (6.27%) coordination and 168 (5.74%) underspecification instances.

	coord	usp	mwe	asym	>dbl	all
abs	267	231	676	240	64	5076
in %	5.26	4.55	13.3	4.72	1.26	100

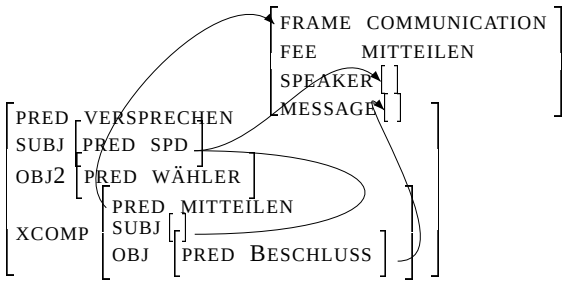
Table 1: Overview of special annotation types

399 of the enriched output structures again suffer from inconsistencies caused by the transfer component problem. 31 contain 'true' inconsistencies in the semantics projection, which are caused by incorrect f-structures in the TIGER-LFG corpus.

[4] A recent revision of annotation guidelines caused a (temporary) inconsistency in the data that makes automatic processing and evaluation difficult.

[5] We haven't yet implemented automated procedures to reliably distinguish asymmetric constituent annotations from other cases of multiple constituent annotations (>dbl), which constitute a rather disparate set of data we do not try to cover.

[6] They were caused by a bug in the employed transfer component that is currently being fixed. We will be able to report on a more comprehensive data set in the final paper.



SPD verspricht Wählern, Beschlüsse mitzuteilen
SPD promises voters to report decisions

Figure 13: Local and non-local frame elements

4.2 Inducing frame projection rules

From the enriched corpus with frame projections, we extract lexical frame assignment rules that – instead of node identifiers – use f-structure descriptions to identify constituents and map them to frame semantic roles. These rules can then be applied to f-structures resulting from LFG parsing.

We designed an algorithm for extracting f-structure paths between pairs of f-structure nodes that correspond to the s-structure of the frame evoking element and one of its semantic roles, respectively. Fig. 13 gives an example, with an absolute f-structure path (f-path) for (the f-structure that projects to) the FEE. From this, we extract f-paths leading to the semantic roles MESSAGE and SPEAKER. The f-path for the MESSAGE (↑OBJ) is *local* to the f-structure that projects to the FEE. For the SPEAKER we identify two paths: one local, the other non-local. The *local* f-path (↑SUBJ) leads to the local SUBJ of *mitteilen* in Fig. 13. By co-indexation with the SUBJ of *versprechen* we find an alternative *non-local* path, which we render as an inside-out functional equation ((XCOMP↑) SUBJ).

Since f-structures are directed acyclic graphs, we make use of graph accessibility to distinguish local from non-local f-paths. In case of alternative local and non-local paths, we choose the local one. In case of alternative non-local paths, we chose the one(s) with the shortest inside-out subexpression.

Generating frame assignment rules We extracted f-paths from 2584 frame assignments of the semantically enriched corpus. These were reduced and compiled to 1095 lexicalised frame assignment rules in the format of Fig. 4. Table 2 gives figures of the average ambiguity per FEE, and per Frame (abstracting from the specific FEE).

	absolute f-path	relative f-path	
FEE	XCOMP PRED	↑	
MSG	XCOMP OBJ	↑OBJ	local
SPKR	SUBJ	(XCOMP↑)SUBJ	nonlocal
	XCOMP SUBJ	↑SUBJ	local

Figure 14: Local and nonlocal path equations

rules per FEE			rules per Frame		
avg.	min	max	avg	min	max
6.33	1	65	8.73	1	78

Table 2: Ambiguity of frame projection rules

Due to the surface-oriented SALSA/TIGER annotation format, the original annotations contain a high number of non-local frame element assignments that are localised in the corresponding LFG f-structures. The assignment paths extracted from the LFG-based corpus yield 4.7% non-local vs. 95.3% local frame element assignments.[7]

4.3 Reapplying frame assignment rules

We reapplied the abstracted frame assignment rules to the original syntactic LFG-TIGER corpus, to control the results. We evaluated the generated frame annotations against the frame-enriched LFG-TIGER corpus that was created by explicit node anchoring (Section 4.1). We obtain 97.93% recall for the target annotations, and 30.87% precision. The low precision is due to the overgeneration of the more general abstracted rules, which are not yet controlled by statistical selection. We measured an ambiguity rate of 3.27 frames per annotation instance.

5 Applying frame assignment rules in an LFG parsing architecture

We finally apply the obtained frame assignment rules to original LFG *parses* of the German LFG grammar. The grammar produces f-structures that are compatible with the LFG-TIGER corpus, thus the syntactic constraints can match the parser's f-structure output. In contrast to the LFG-TIGER

[7] Since non-local frame elements are tied to very specific syntactic contexts, they will ultimately be defined as optional. In combination with stochastic modelling, we will further experiment with frame assignment rules that are not conditioned to specific FEEs, but to frame-specific syntactic descriptions, to assign frames to 'unknown' lexical items.

treebank, the grammar delivers f-structures for alternative syntactic analyses. We do therefore not expect frame projections for all syntactic readings. On the other hand, we now apply the entire set of rules to any given sentence. The projection rules can thus apply to new annotation instances, and create more ambiguity in the semantics projection.

We applied the frame assignment rules to the parses of 1399 sentences. Against the frame-enriched LFG-TIGER corpus we obtain 69.97% recall and 16.23% precision of annotation targets. The average ambiguity induced by frame assignment is 4.3 frame assignments per sentence. The higher ambiguity is expected, since we apply the entire rule set to each sentence. The lower recall is probably due to divergences in the f-structure encodings of the corpus vs. grammar.

6 Summary and Future Directions

We presented an automated method for corpus-based induction of an LFG syntax-semantics interface for frame semantic processing. We port frame annotations from a manually annotated syntactic corpus to an LFG parsing architecture that allows us to process unparsed text. We show how to model frame semantic annotations in an LFG projection architecture, including phenomena that involve non-isomorphic mapping between levels.

As the semantic corpus is under construction, our results are small-scale. Yet, we give a proof-of-concept for how to build a uniform computational semantics interface for frame assignment that can be used to process unparsed corpora.

In future work we will train stochastic models for disambiguation of the assigned frame semantic structures. We are especially interested in exploring the potential of deeper, functional syntactic analyses, in conjunction with additional semantic knowledge (e.g. word senses, named entity typing), using methods along the lines of (Riezler et al., 2003). We intend to set up a bootstrapping cycle for learning increasingly refined stochastic models from growing training corpora, using semi-supervised learning methods.

Finally, we will explore multi-lingual aspects of frame assignment and learning from parallel corpora, using English FrameNet data and an English LFG grammar with comparable f-structure output.

References

Anonymous. 2004.

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

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

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

K. Erk and S. Pado. 2004. A powerful and versatile xml format for semantic annotation. In *Proceedings of LREC'04*, Lisbon, Portugal.

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

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

M. Forst. 2003. Treebank Conversion – Establishing a testsuite for a broad-coverage LFG from the TIGER treebank. In *Proceedings of LINC '03*, Budapest.

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

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

P.-K. Halvorsen and R.M. Kaplan. 1995. Projections and Semantic Description in Lexical-Functional Grammar. In M. Dalrymple, R.M. Kaplan, J.T. Maxwell, and A. Zaenen, eds, *Formal Issues in Lexical-Functional Grammar*, CSLI Publications.

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

J. T. III Maxwell and R. M. Kaplan. 1989. An overview of disjunctive constraint satisfaction. In *Proceedings of IWPT*, pages 18–17.

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

S. Riezler, T. H. King, R. Crouch, and A. Zaenen. 2003. Statistical sentence condensation using ambiguity packing and stochastic disambiguation methods for Lexical-Functional Grammar. In *Proceedings of HLT-NAACL'03*, Edmonton, Canada.