Proposal for the establishment and funding of the Collaborative Research Centre 991

“The Structure of Representations in Language, Cognition, and Science”

for


Heinrich-Heine-Universität Düsseldorf

Coordinator
Prof. Dr. Sebastian Löbner
Institut für Sprache und Information
Universitätsstr. 1, 40225 Düsseldorf
0211 8113399
loebner@phil.uni-duesseldorf.de

Secretary
Frau Vera Aydt
Institut für Sprache und Information
Universitätsstr. 1, 40225 Düsseldorf
0211 8112959
aydt@phil.uni-duesseldorf.de

http://www.sfb991.uni-duesseldorf.de/sfb991/
3.1 General information about the project A02

3.1.1 Title:
Argument Linking and Extended Locality. A Frame-Based Implementation

3.1.2 Research areas:
Computational Linguistics, Syntax-Semantics Interface

3.1.3 Principal Investigator:
Kallmeyer, Laura, Prof. Dr., 06.10.1969, German
Abteilung für Computerlinguistik
Institut für Sprache und Information
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1
D-40225 Düsseldorf

Tel.: 07071 2977513 (Tübingen)
Fax: 07071 295830 (Tübingen)
E-mail: laura.kallmeyer@googlemail.com

Do the above mentioned persons hold permanent positions? ☒ Yes ☐ No

3.1.4 Legal issues
This project includes

<table>
<thead>
<tr>
<th>1. Research on human subjects or human material.</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>If applicable: A copy of the required approval of the responsible ethics committee is included with the proposal.</td>
<td>–</td>
</tr>
<tr>
<td>2. Clinical studies.</td>
<td>No</td>
</tr>
<tr>
<td>3. Experiments involving vertebrates.</td>
<td>No</td>
</tr>
<tr>
<td>4. Experiments involving recombinant DNA.</td>
<td>No</td>
</tr>
<tr>
<td>5. Research involving human embryonic stem cells.</td>
<td>No</td>
</tr>
<tr>
<td>If applicable: Legal authorization has been obtained.</td>
<td>–</td>
</tr>
<tr>
<td>6. Research concerning the Convention on Biological Diversity.</td>
<td>No</td>
</tr>
</tbody>
</table>

3.2 Summary
This project relates semantic frames to semantic decomposition and composition in lexicalized tree adjoining grammars (LTAG).

In grammar formalisms that are characterized by an extended domain of locality, such as LTAG, elementary syntactic structures already represent syntactic configurations that correspond to entire subcategorization frames. Therefore, the extended domain of locality considerably facilitates argument linking within the syntax-semantics interface. This observation is the starting point for this project.

During this project, we plan to incorporate frame-based lexical semantics into an LTAG grammar. The hypothesis is that the explicit semantic decomposition available in rich semantic frames leads in an immediate way to syntactic elementary structures equipped with an equally rich semantics. The latter can then be composed at sentence level using a semantics framework already available in LTAG. We expect that the argument linking that relates argument specifications within the frame of a predicate to nonterminal leaves (argument slots) in elementary TAG trees is rather immediate, due to the extended domain of
locality that characterizes LTAG. This is the main reason for choosing LTAG. A further advantage of LTAG is that it provides a certain degree of factorization concerning elementary constructions. Specifically, it makes it possible to pair templates of constructions with fragments of semantic frames. Furthermore, it allows the expression of generalizations about relations between different constructions that are available for the same lexical head. This property, in combination with the extended domain of locality, relates LTAG to ideas and principles underlying construction grammar. We plan to explore this relation in detail.

The project concentrates on verb semantics in English. A particular focus will be on the effects certain constructions have on the semantic contribution of an elementary structure. The availability of an extended domain of locality in LTAG is expected to facilitate the modeling of relations between certain parts of the syntactic tree and certain parts of the semantic frame.

We plan to implement our results, using the XMG tool (Nancy, Orléans) for the incorporation of frames into the lexicon and for the interface between frame-based lexicon and LTAG elementary trees with unification-based semantics. For parsing, we will use TuLiPA, a parser that has been developed in a previous project of the principal investigator. TuLiPA produces a syntactic analysis while computing an underspecified semantic representation.

3.3 Starting point of the project

3.3.1 State of the art and preliminary work

Frames and valency: As a starting point, we take the meaning of verbs to be related to conceptual frames, that basically define a set of roles or participants in a simple Fillmorean sense (Fillmore, 1982). Following Barsalou (1992) and Petersen (2007), such a kind of frame is a typed feature structure that can be represented by an attribute-value matrix (AVM), as can be seen in Figure 1 for the concept giving. It contains three attributes (DONOR, THEME, and RECIPIENT) with exactly one value each. We use 1, 2, 3 as variables over the set of values.

Applied to conceptual frames, the valency of a verb specifies how the syntactic realization of the frame roles has to look and whether the syntactic argument corresponding to a specific frame role is obligatory or not (see, for instance, Zifonun, 2003). In other words, valency determines the syntactic arguments of a verb and, furthermore, the relation between the syntactic arguments and the grid of frame roles specified by the semantic frame, the latter also being known as argument linking.

Construction Grammar: Construction grammar (CG) (Goldberg, 1995) is a grammatical framework that assumes that the basic parts of syntactic structures are constituency trees (or, rather, tree fragments) that represent complete valency frames. Such basic constituency trees are called constructions. CG displays an extended domain of locality which means that syntactic dependencies can be stated locally. This, together with the fact that semantic frames encode (among others) the semantic roles of the syntactic arguments, makes CG a suitable candidate for developing a frame-based syntax-semantics interface.

Originally, the development of CG was motivated by the desire to have the same type of syntactic structures for rather idiomatic expressions as well as more regular syntactic constructions. The idea is to extend the domain of locality beyond that of standard phrase structure grammar. The elements of the grammar, the constructions, are trees that can contain more than just a mother and its immediate daughters. Specifically, they encode entire valency frames.

Constructions are fragments of constituent structures whose nodes are equipped with feature structures. In other words, they represent more or less underspecified constituent structures (that can be formalized using tree descriptions, see Kay, 2002). The combination of constructions in order to obtain lar-

```
giving
  DONOR  1
  THEME   2
  RECIPIENT  3
```

Figure 1: AVM for semantic frame of give
ger constituent structures is done via unification.

The encoding of constructions in the lexicon is achieved via an inheritance hierarchy of constructions. For instance, a construction HEADED-COMP inherits from a construction HEADED-PHRASE. Only the leaves in this hierarchy (the most specific classes) represent the constructions in the grammar.

CG adopts a lexicalist perspective on grammar in the sense that constructions are assumed to have a lexical head. However, a further degree of factorization is achieved by separating the construction template (without the lexical head) from the lexical head itself. Oftentimes, the term ‘construction’ denotes only the former, i.e., only the template.

CG is frequently used in combination with semantic frames (Goldberg, 1995; Goldberg, 2010) since, as already mentioned, the linkage between semantic roles specified in the frames and syntactic arguments specified in the constructions can be done locally due to the extended domain of locality available in CG.

**LTAG and valency:** Tree adjoining grammar (TAG, Joshi and Schabes, 1997) is a tree rewriting formalism that shares some crucial underlying ideas with CG but that is, in contrast to CG, thoroughly formalized and has been shown to be computationally tractable (Kallmeyer, 2010). In a TAG, so-called elementary trees are combined into larger trees by the use of two rewriting operations, substitution and adjunction. In the simple example in Figure 2, the elementary trees anchored by John, laughs and sometimes are used to compose the derived tree on the right. Substitution takes place at the nonterminal NP-node of the tree for the verb laughs, that is replaced by the elementary tree for John. In contrast, adjunction always applies to inner nodes of trees, as can be observed with the elementary tree for sometimes, which replaces the inner VP-node of the laughs tree.

After adjunction, the subtree below the replaced inner node is then the subtree below the foot node, i.e. the nonterminal frontier node marked with an asterisk. Only elementary trees with a foot node, also called auxiliary trees, qualify for adjunction, whereas substitution is reserved for elementary trees without a foot node, the so-called initial trees. A lexicalized tree adjoining grammar (LTAG) furthermore complies with the following restriction concerning the shape of elementary trees: elementary trees always include a terminal frontier node, known as the lexical anchor.

Taking elementary trees as lexical entities that enter into syntactic derivation has far reaching consequences. Compared to elementary structures in CFG-based grammar formalisms, elementary trees in LTAG come with an extended domain of locality that makes it possible to express every syntactic dependency locally within an elementary tree (see the fundamental TAG hypothesis in Frank, 2002). This even holds for allegedly nonlocal dependencies, such as some cases of wh-movement. Furthermore, due to lexicalization, exactly the syntactic dependencies governed by the lexical anchor are represented in an elementary tree. Since valency relations are seen as syntactic dependencies, they have to be indicated properly as well, namely through nonterminal frontier nodes. To give an example, the elementary tree for laughs in Figure 2 therefore includes a nonterminal frontier node (labeled here with ‘NP’) indicating the subject. The label of the valency-governed node usually specifies the morphological features of the syntactic argument. However, it cannot be seen from a single elementary tree, whether the realization of the syntactic argument is obligatory or optional. For this it is necessary to consider the set of elementary trees anchored by the same terminal and belonging to the same valency frame. Optional syntactic arguments, such as the at-PP of laughs, emerge only in some of these alternative elementary trees.

In previous work, we investigated ways to adopt the fundamental TAG hypothesis even in so-called
free word order languages such as German (Lichte, 2007; Lichte and Kallmeyer, 2008; Kallmeyer, 2005). **Metagrammar and syntactic factorization:** In order to produce and maintain a consistent LTAG with substantial coverage it is advisable to fall back on a sophisticated grammar management system. In previous work (Kallmeyer et al., 2008) we therefore made use of eXtensible MetaGrammar (XMG, (Duchier, Leroux, and Parmentier, 2004)) which makes it possible to set up factorized descriptions of elementary tree templates (i.e. not yet anchored elementary trees). Factorization here roughly means that tree fragments can be defined and reused in other descriptions of tree fragments or tree templates. An example is shown in Figure 3 where the elementary tree templates for intransitive verbs and verbs with a PP-object are assembled from smaller tree fragments. Based on these tree descriptions and a supplemental lexical resource (including the lexical anchors and e.g. morphological and valency information), XMG compiles a full-fledged LTAG.

Besides the rather technical virtues of XMG with respect to grammar engineering, XMG also offers powerful means to express and implement lexical generalizations. For example, it is possible to express inheritance hierarchies of tree fragments along the inheritance hierarchies proposed for constructions in CG. Furthermore, the specific contributions of an elementary tree template (i.e. a construction template) and a lexical anchor to syntax and semantics can be represented separately.

**LTAG syntax semantics interface:** There have been several proposals for the framework underlying the syntax-semantics interface in LTAG. In the following, we will present the architecture developed in Gardent and Kallmeyer (2003) and Kallmeyer and Romero (2008). Besides the fact that this constitutes previous work by the project’s principal investigator, this architecture shares several important aspects with the perspective on semantic composition present in CG in combination with frame semantics.

Gardent and Kallmeyer (2003) and Kallmeyer and Romero (2008) propose a unification-based semantic composition for LTAG. More precisely, each elementary tree is paired with a set of typed predicate logic expressions and of scope constraints (i.e., constraints on subterm relations) and a feature structure that characterizes a) which arguments need to be filled, b) which elements are available as arguments for other elementary trees and c) the scope behavior. The features are linked to positions in the elementary tree.

A simple example is shown in Figure 4. The elementary tree for *laughs* for example contains one non-terminal leaf, the argument slot for the subject. This encodes the syntactic valency information. As far as semantics is concerned, the predicate *laugh* requires one individual argument (metavariable $\Box$). This metavariable appears as the value of an attribute $I$ (for individual) at the NP substitution slot which indicates that the value of $\Box$ will be obtained via substitution at that node.

There is no proper functional application. Instead, depending on substitutions and adjunctions, we perform unifications that lead to assignments for the metavariables in the semantic representations. In Figure 4 for instance, the substitution yields a unification of $\Box$ and $x$ while the adjunction yields unifications between $\Box$ and $l_2$ and between $\Box$ and $l_1$. As a result one obtains the semantic representation at the bottom of Figure 4.

---

**Figure 3:** Tree fragments leading to tree templates for intransitive verbs or verbs with a PP-object

---

**Figure 4:** A simple example of unification-based semantic composition for LTAG.
The result can be underspecified, i.e., it is a set of predicate logical expressions eventually containing holes and of scope constraints along the lines of Hole Semantics (Bos, 1995). Therefore, a further disambiguation is needed in order to obtain the different readings.

(1)  
   a. everybody laughs

   \[
   l_1 : \text{laugh}(x), \quad l_2 : \text{every}(x, 4, 5), \quad l_3 : \text{person}(x) \\
   4 \geq l_3, \quad 5 \geq l_1
   \]

   b. every(x, person(x), laugh(x))

Figure 5 sketches the idea of how to treat quantifiers within this framework. The verb \textit{laughs} contributes a proposition (labeled \( l_1 \)) that is provided as minimal scope (attribute MINS) to the quantifier substituting into the NP argument slot. The scope constraints tell us that a) the restriction of the quantifier (4) contains the term \( l_3 : \text{person}(x) \) and b) the nuclear scope (5) contains the term \( l_1 : \text{laugh}(x) \). As a result, we obtain the underspecified semantic representation in (1b). After disambiguation, this yields the semantics in (1c).

\textbf{Feature structures and predicate logic:} In the above-mentioned LTAG semantics approach, the feature structures used for semantics do not represent the meaning itself, i.e., they are not semantic frames. They only specify how to combine meanings in semantic composition. In other words, they are a kind of glue that directs semantic unification.
There have been different proposals for encoding entire predicate logical expressions within feature structures, mainly in the context of head-driven phrase structure grammar (HPSG). Pollard and Sag (1994) adopt a Cooper storage approach to quantifier scope (Cooper, 1983) and express this within a feature structure. The semantics for (2a) according to this approach is shown in Figure 6. This feature structure encodes the predicate logic expression in (2b). Note that this feature structure allows for lists and sets as attribute values, a notion which extends the notion of frames as feature structures from Petersen (2007).

(2)  

a. I know a poem  

b. exists(x, poem(x), know(y, x))

Lexical resource semantics (LRS) (Richter and Sailer, 2004) specifies the subterms contributed by a lexical item to the logical form (LF) of the entire sentence in a more explicit way. As an example, consider the LRS lexical entry for the quantifier everybody in Figure 7. The contribution to the logical form is described in the parts attribute and the scope constraints that come with the feature structure specify subterm relations between the elements in parts. Here, the restriction of the quantifier, $\alpha$, contains the term person($x$) (constraint $\exists x < \alpha$).

This way of dealing with scope in LRS is close to the LTAG approach exemplified in Figure 5. In previous work (Kallmeyer and Richter, 2006; Richter and Kallmeyer, 2007), we provide a detailed comparison of the use of feature structures for semantic composition within LTAG and within LRS.

(For general references, see end of § 3.6. References in square brackets refer to pdf files on the CD attached to the proposal, see directory ‘3. Obligatorliche Zusatzdokumente\Projekt A02\’.)
3.3.2 Project-related publications by the principal investigator(s)

Reviewed papers in journals and conference proceedings


3.4 Project outline

3.4.1 Aim

The aim of this project is to incorporate a frame-based theory of lexical semantics into an architecture for a syntax-semantics interface. The syntactic formalism chosen for the latter, tree adjoining grammar, is thoroughly formalized and computationally tractable. Conceptually, its main features are an extended domain of locality and, related to this, the possibility of a local realization of valency frames. In this respect it shares some characteristic properties with construction grammar, a theory of syntactic components that also displays an extended domain of locality and has a similar local perspective on valency realization. The combination of a frame-based semantics and a construction-based syntax, that has already been proposed in the context of CG, will be developed in a much more formal and concrete way in this project.

3.4.2 Methods

The project has to address different questions concerning the use of frame semantics within an LTAG-based syntax. We will develop a general architecture for a frame-based LTAG semantics. Within this architecture, we plan to concentrate on the problem of establishing the correct linking between information on semantic roles encoded in the lexical semantic frames and predicate-argument relations in the sentential semantics. For this reason, elements with a richer valency frame are more interesting. We will concentrate on verb frames in English.

Another focus of the project will be the distinction between lexical semantic information (that has to come from the semantic frames in the lexicon) and, in contrast to this, construction-specific semantic aspects. The latter originate from the form of the elementary trees (i.e., from the constructions) and not from the lexical items themselves.

Finally, in order to compute a sentence-level truth-conditional semantics from lexical semantic frames, we will integrate quantification and scope into semantic frames.

A frame-based syntax-semantics interface: This part of the project can be considered its core component. Its starting point will be the unification-based semantics proposed in Gardent and Kallmeyer (2003)
and Kallmeyer and Romero (2008). The idea is to enrich the predicate logical expressions used within their semantics with semantic frames, or maybe even to replace them completely with frames. These frames can of course share some of their feature values with the features that decorate the syntactic tree. There is a single general semantic frame linked to an entire elementary tree, not linked to a single node in an elementary tree. Ideally, semantic composition will be done via unification in more or less the same way as it is done in the current LTAG semantics framework. When using frames instead of predicate logical expressions, this seems even more natural to do. The result would then be a semantic frame describing the semantics of an entire sentence.

One hypothesis is that the extended domain of locality available in LTAG facilitates the linking between syntax and semantics.

(3) Bill gives Mary the book.

As an example, consider the derivation of (3) that is sketched in Figure 8. Formally, semantic frames can be considered as generalized typed feature structures (Carpenter, 1992; Petersen, 2007). The substitution of the three syntactic arguments into the corresponding slots causes the feature structures of the roots of the NP trees and of the corresponding NP substitution nodes to unify. As a result, the semantic frame of Bill (5) unifies with the frame of the first argument in the semantics of gives (2), the frame of Mary (6) unifies with the recipient (4) and the book frame (7) unifies with (3). This leads to the semantic frame on the bottom of Figure 8 for the entire sentence.

This is only a rough sketch of how to encode the semantics of a lexical item within its semantic frame. It mainly serves to illustrate the relation between LTAG operations and unifications of semantic frames. Whether we really want to encode not only semantic roles but also the truth-conditional aspects of semantics within frames, along the lines of lexical resource semantics (Richter and Sailer, 2004), is an open question that we plan to investigate in the project. If we adopt such a purely frame-based semantics, then semantic frames must contain much more information that is relevant for sentence-level semantic composition. In particular, we need to find adequate ways to encode quantification and scope constraints (see below).

As we have seen, the semantic composition (which is unification on the semantic frames) is driven by the substitutions and adjunctions performed in the syntax. Because of this, we probably do not need to distinguish between those frame nodes that are open arguments and those that are not, in contrast to [Petersen & Osswald a]. Furthermore, the link between semantic arguments and nodes in the syntax specifies, of course, the concrete mapping from semantic roles to syntactic arguments. In other words, it encodes the syntactic valency. This is one of the major advantages of using a syntactic formalism such as LTAG that has an extended domain of locality.

Besides encoding semantics within frames, we could even move towards an entire frame-based approach by also expressing the syntactic structures in the form of feature structures, along the lines of HPSG (Pollard and Sag, 1994). In contrast to HPSG, TAG would, however, still keep its extended domain of locality and, consequently, we would have to define a nonmonotonic operation on frames that corresponds to adjunction. (Substitution can be achieved by a monotonic unification operation.) In this project, we will probably not pursue this idea, mainly for practical reasons. The parser we are using (TuLiPA) assumes a TAG in the standard tree-based format, and the same holds for the metagrammar compiler XMG. Therefore, we will keep the separation between syntactic trees, decorated with feature structures, and semantic frames.

As far as the overall semantic architecture and the semantics of specific verbs are concerned, we plan to cooperate with Maribel Romero (Universität Konstanz), besides cooperation with other CRC projects.

Construction-based aspects of frame semantics: As already mentioned, LTAG allows for a certain degree of factorization concerning the definition of elementary trees. This facilitates the separation between lexical semantic properties and semantic properties of specific types of constructions.

Consider for example the contrast between (4) and (5). Even though the argument structures of the ex-
amples in (a) and (b) are the same, the ditransitive construction requires the dative object to be animate. This is, according to Goldberg (1995), a property of the construction and not of the lexical item *bring*.

(4)  a. I brought Pat a glass of water.
    b. I brought a glass of water to Pat.

(5)  a. *I brought the table a glass of water.
    b. I brought a glass of water to the table.

Goldberg (1995) examines further examples of ditransitive constructions (with *give, pass, hand, earn, permit, ...*) and concludes that the semantics of such constructions is that an agent successfully causes a recipient to receive a patient.

In an LTAG grammar, we distinguish between unanchored elementary trees (the constructions) and lexical anchors. Putting the two together yields the set of elementary trees that compose syntactic structures. This factorization enables a distinction between construction-specific semantic properties and lexical semantic properties. For illustration, Figure 9 shows how the unanchored tree for the ditransitive construction and its semantics could appear. The diamond ◊ on the V node marks the position of the lexical anchor. The semantic frame tells us that the head of this construction denotes a causal event involving an agent and a receive event. The receiver (the first argument of the latter) must be animate.

The further decomposition of the tree in Figure 9 into metagrammar tree fragments could be such that the frame fragment \[\text{ANIMATE yes}\] is paired with the tree fragment for the dative NP argument.

The semantic frame in Figure 9 will have to unify with the frame contributed by the lexical anchor. This unification will further determine the semantic predicate and the roles of its arguments. The linking, however, is already part of the construction, i.e., the unanchored elementary tree (see the \(i\) features), and so is the constraint that the dative argument must be animate.

Once a construction-meaning pair – as illustrated in Figure 9 – gets linked to a concrete lexical item, the lexical item is inserted as the anchor, its syntactic features are unified with those of the anchor, and its semantic frame is unified with the semantic frame of the construction. The latter ensures, for instance,
that the semantic roles attributed to the argument of the lexical item match with the semantic roles required by the constructions. What this means in detail has to be investigated in the project. We probably need a much more fine-grained frame-based definition of semantic roles than the one sketched in Fig. 9. In cooperation with (and building on results from) project B01, we will also inspect verbal argument alternations (Levin, 1993; Frense and Bennett, 1996). These phenomena display modifications of semantic frames that are triggered by specific constructions. We expect that the above-mentioned properties of LTAG facilitate the modeling of argument alternations. More precisely, the factorization present within a metagrammar makes it possible to state generalizations about relations between tree fragments and their contributions to the semantic frame used within different elementary trees for the same type of verb.

The close relation of this overall architecture to the leading ideas of CG is rather obvious. The two approaches both display extended domains of locality and both assume a certain factorization concerning the specification of elementary structures that allows the identification of the properties of (fragments of) elementary trees/ constructions. We therefore think that the development of a frame-based LTAG syntax-semantics interface will show that LTAG actually yields an adequate formalization of many of the central ideas of CG.

Quantification, scope and underspecification: If we want to capture all aspects of sentence-level semantics within semantic frames, we have to incorporate quantification and scope information in addition to the specification of semantic roles. Furthermore, as in previous LTAG semantics approaches, we want to provide a way to express scope constraints and to generate underspecified scope representations. This is desirable for reasons of computational complexity and it is more or less the state of the art in computational semantics.

In LTAG, we do not need a mechanism such as Cooper storage for the storage and later retrieval of quantifiers since the extended domain of locality in LTAG, in combination with underspecification, allows for an immediate determination of scope possibilities depending on the syntactic derivation (Kallmeyer and Romero, 2008; Kallmeyer and Romero, 2006). Therefore, we will rather adopt representations along the lines of LRS (Richter and Sailer, 2004), though probably in a more restricted way.

Figure 10 gives an idea of how to encode the semantic representations (including scope constraints) from Kallmeyer and Romero (2008) within frames. The attribute LF contains the semantic representation that consists of a set (here a list) of propositions (attribute PROPOSITIONS) and a set of scope constraints (attribute CONSTRAINTS). Formally, the latter could also be specified as relations between frame elements, instead of explicitly listing them within an attribute value. Such additional relations between frame elements are needed in other cases as well (Barsalou, 1992).

As far as the specification of the form of our semantic frames and the encoding of quantification and scope is concerned, we will cooperate with Frank Richter (Universität Tübingen), Manfred Sailer (Universität Göttingen) and Maribel Romero (Universität Konstanz).

Implementation within TuLiPA and XMG: In addition to theoretical considerations concerning the particular properties of verb frames, we also plan to implement a small grammar fragment, as a way to test and demonstrate our results. This is, however, not intended to be a large coverage grammar.

For the implementation of the elementary trees and the corresponding semantic frames we will adopt a meta-grammar approach using the tool XMG described above. XMG allows for the specification of tree fragments whose nodes are labeled with feature structures. As far as semantics is concerned, it provides the possibility to define semantic representations as used in the LTAG semantics approach of Gardent and Kallmeyer (2003) and Kallmeyer and Romero (2008). In previous work, we have already used XMG (Kallmeyer et al., 2008) and it has proved reliable and comparatively easy to use.
In order to use frames as semantic representations, we need an extension of XMG that allows for the specification of (fragments of) semantic frames, linked to the syntactic fragments in the metagrammar. We expect this to be rather straightforward since the specification of feature structures is already possible within the syntactic layer of XMG. For this extension of XMG, we plan to cooperate with Denys Duchier and Yannick Parmentier (Université d’Orléans) who have participated in the implementation of XMG.

As a resource for semantic frames (in addition to the frames developed in other CRC projects), we plan to also use FrameNet (Ruppenhofer et al., 2010), a lexicon of semantic frames that encode valency information.

Since XMG allows for the definition of small tree fragments and of hierarchical relations among them, we expect that it permits a direct implementation of central CG principles concerning grammar organization and grammar factorization (Kay, 2002).

In addition to the metagrammar specification and the TAG that results from compiling it, we also need a TAG parser in order to test our analyses. For this purpose, we will use TuLiPA (Tübingen Linguistic Parsing Architecture), a tool that has been developed by the principal investigator and her Emmy Noether group at the University of Tübingen (Parmentier et al., 2008; Kallmeyer et al., 2010). TuLiPA supports the parsing of TAG and of a variety of other related formalisms, and it performs a semantic computation along the lines of Gardent and Kallmeyer (2003) and Kallmeyer and Romero (2008), together with a subsequent disambiguation of the underspecified representation.

In order to use TuLiPA for a frame-based syntax-semantics interface, we have to integrate a unification-based semantic composition. The complexity of this task depends heavily on the choice of feature structures made to encode semantic frames. We hope that the unification currently used in TuLiPA for the syntactic features can be adapted to our semantic needs. For this part of the project, we also plan to cooperate with Yannick Parmentier who took part in the development of TuLiPA while visiting as a postdoc in Laura Kallmeyer’s Emmy Noether group.

---

2 [http://sourcesup.cru.fr/tulipa/](http://sourcesup.cru.fr/tulipa/)
3.4.3 Work program

<table>
<thead>
<tr>
<th>Year</th>
<th>Acquisition of knowledge on verb frames;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General comparison of CG and LTAG;</td>
</tr>
<tr>
<td></td>
<td>Identification of construction-specific semantic constraints;</td>
</tr>
<tr>
<td>Year</td>
<td>Integrating quantification and underspecification into semantic verb frames;</td>
</tr>
<tr>
<td>2</td>
<td>Decision on a specific formal model for frames;</td>
</tr>
<tr>
<td></td>
<td>Definition of the specific form of semantic frames for lexical items and for verbal constructions;</td>
</tr>
<tr>
<td>Year</td>
<td>Implementation of selected constructions within XMG;</td>
</tr>
<tr>
<td>3</td>
<td>Development of a syntax-driven semantic frame composition within LTAG;</td>
</tr>
<tr>
<td></td>
<td>Implementation of semantic unification and integration in TuLiPA;</td>
</tr>
<tr>
<td>Year</td>
<td>Further work on specific semantic frames;</td>
</tr>
<tr>
<td>4</td>
<td>Implementation of semantic frames within the grammar fragment;</td>
</tr>
<tr>
<td></td>
<td>Implementation of semantic unification and integration in TuLiPA (continued)</td>
</tr>
</tbody>
</table>

3.4.4 Conclusion

By way of conclusion, let us summarize the principal contributions of this project.

- **Contribution to the CRC**: This project extends the work on lexical semantic frames by developing a further theory of sentence-level frame-based semantic composition. Furthermore, it adds construction-based aspects of sentential meaning to research on lexical semantics. Finally, it adds the new aspects of quantification and scope to the research on semantic frames conducted in the CRC.

- **Contribution to research on LTAG**: So far, lexical semantics has not been seriously investigated in the context of LTAG. This project deals with this issue in a thorough and extensive way and thereby considerably advances the research on LTAG semantics.

- **Contribution to construction grammar**: By using LTAG, a formalism with well-known mathematical properties, this project constitutes an important step towards a formalization of the principles and ideas underlying construction grammar, in particular in connection with frame semantics.

3.5 Role within the Collaborative Research Centre

As already mentioned, the project contributes a theory of sentence-based semantic composition to the research on semantic frames pursued within the CRC. It therefore enriches the CRC in an important way by adding a construction-based perspective to the lexical aspects the CRC concentrates on.

Besides this general contribution to the CRC topic, project A02 also has specific links to the following other projects in the CRC:

**Project A01** (Wiebke Petersen): Common research topics with the project A01 are the formal properties of semantic frames and the mechanisms that are used to combine them and, furthermore, the specific form of verb frames. We expect a close cooperation concerning these aspects.

**Project B01** (Van Valin): B01 is concerned with verb frames at the syntax-semantics interface and adopts the framework of role and reference grammar (RRG). We plan to integrate (some of) the verb frames resulting from B01 within our syntax-semantics interface. Furthermore, with B01 we will compare LTAG and RRG and discuss ideas for formalizing the latter.

**Project B02** (Löbner, Geisler): Project B02 deals with verb frames, a topic that it shares with our project. Even though the languages and the class of verbs considered are different, a collaboration will reveal general aspects of verbal meaning and the way they can be encoded within frames.

**Project C05** (Löbner): Project C05 deals, among other things, with the frames of deverbal nouns and therefore has a close connection to the topic of verb frames. This aspect constitutes (as for B02) an immediate link to our project A02.
3.6 Demarcation from other funded projects of the principal investigator(s)

Laura Kallmeyer has submitted a proposal to the DFG (Sachbeihilfeantrag KA 1665/4-1) for a project entitled “Grammar Formalisms beyond Context-Free Grammars and their use for Machine Learning Tasks”. The proposal will probably get accepted and the project will start more or less at the same time as the CRC. It is concerned with statistical NLP, in particular parsing and machine translation. It concentrates mainly on syntax, uses only statistical approaches and deals with grammar formalisms such as linear context-free rewriting systems and range concatenation grammars that are more powerful than LTAG. It therefore differs from the project A02 that is concerned with the symbolic treatment of a frame-based syntax-semantics interface within LTAG.

References

(References in square brackets refer to pdf files on the CD attached to the proposal, see directory ‘3. Obligatorische Zusatzdokumente\Projekt A02’.)


Kallmeyer, Laura and Maribel Romero (2006). Quantifier scope in German: An MCTAG analysis. In Proceedings of The Eighth International Workshop on Tree Adjoining Grammar and Related Formal-
isms (TAG+8), pages 73–80, Sydney, Australia, July.
Kallmeyer, Laura and Maribel Romero (2008). Scope and situation binding in LTAG using semantic
1–19.
the 12th Conference on Formal Grammar 2007, Dublin, Ireland.
In Proceedings of the Ninth International Workshop on Tree Adjoining Grammars and Related For-
malisms (TAG+9), pages 57–64, Tübingen, June.
Parmentier, Yannick, Laura Kallmeyer, Wolfgang Maier, Timm Lichte, and Johannes Dellert (2008).
TuLiPA: A syntax-semantics parsing environment for mildly context-sensitive formalisms. In Pro-
cceedings of the Ninth International Workshop on Tree Adjoining Grammars and Related Formalisms
(TAG+9), pages 121–128, Tübingen, June.
Petersen, Wiebke (2007). Representation of concepts as frames. In The Baltic International Yearbook of
between LRS and LTAG. In Anders Søgaard and Petter Haugereid, editors, Typed Feature Structure
Lang.
Richter, Frank and Manfred Sailer (2004). Basic concepts of lexical resource semantics. In Arnold Beck-
mann and Norbert Preining, editors, ESSLLI 2003 – Course Material I, (= Collegium Logicum, 5).
Kurt Gödel Society, Wien, pages 87–143.
Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, and Jan
Eroms, Hans Jürgen Heringer, and Henning Lobin, editors, Dependenz und Valenz. Ein internationa-
les Handbuch der zeitgenössischen Forschung, volume 25.1 of Handbücher zur Sprach- und Kommu-