Combining Semantic Annotation of Word Sense & Semantic Roles: A Novel Annotation Scheme for VerbNet Roles on German Language Data

Éva Mújdricza-Maydt[‡*], Silvana Hartmann[†], Iryna Gurevych[†*], Anette Frank[‡*]
Department of Computational Linguistics, Heidelberg University[‡]
Ubiquitous Knowledge Processing Lab (UKP-TUDA), Department of Computer Science, Technische Universität Darmstadt[†]
Research Training Group AIPHES, Heidelberg University and Technische Universität Darmstadt[*]
{mujdricz | frank}@cl.uni-heidelberg.de, {gurevych | hartmann}@ukp.tu-darmstadt.de

Abstract

We present a VerbNet-based annotation scheme for semantic roles which we explore in an annotation study on German language data that combines word sense and semantic role annotation. We reannotate a substantial portion of the SALSA corpus with GermaNet senses and a revised scheme of VerbNet roles. We provide a detailed evaluation of the interaction between sense and role annotation. The resulting corpus will allow us to compare VerbNet role annotation for German to FrameNet and PropBank annotation by mapping to existing role annotations on the SALSA corpus. We publish the annotated corpus and detailed guidelines for the new role annotation scheme.
Keywords: word sense annotation, semantic role annotation, GermaNet, VerbNet, SALSA corpus, guidelines, German

1. Introduction

Semantic annotation of predicate-argument structure is an important task in NLP. During decades, different frameworks for representing semantic predicate-argument structure have been established, notably *FrameNet*, *VerbNet* and *PropBank*, with accompanying sense and role inventories and annotated resources, primarily for English (Baker et al., 1998; Kipper-Schuler, 2005; Palmer et al., 2005). The resources are interoperable via *SemLink* and the Unified Verb Index (Loper et al., 2007).

Combining GermaNet sense and VerbNet role annotation Within these established frameworks, the characterization of predicates and roles differ in important ways: *FrameNet* defines semantic roles for verbs, nouns and adjectives evoking a frame. Due to its large frame inventory, and since *roles are specific to a frame*, there is a vast amount of roles to distinguish, and roles do not generalize across frames. *PropBank* defines six major roles (A0-A5); except for A0 and A1, their interpretation is not consistent across predicates, and role sets that characterize specific senses are predicate-specific. In our work we focus on *VerbNet*. *VerbNet* operates with a closed set of semantic roles that are not predicate-specific, and hence can be applied to any predicate to characterize the semantic roles that arguments take in relation to that predicate.

These differences in design have consequences for the scalability of automatic semantic role labeling (SRL) systems, and the semantic expressiveness of the resulting annotations. While FrameNet representations bear the most informative content, systems suffer from sparsity due to incomplete coverage. Frames also represent relatively coarse meanings, covering both synonyms and antonyms. PropBank has wide coverage and yields good automatic annotation quality (Merlo and van der Plas, 2009), but the roles bear less semantic content. VerbNet-style role labeling represents a middle ground between these frameworks, with a constrained role inventory of approximately 35 roles that offers interpretable role semantics and generalizability across predicates.

In this work we focus on VerbNet-style role labeling from several perspectives. First, we investigate **VerbNet-style role labeling** as an attractive role labeling scheme that is generalizable across predicates, is less prone to sparsity problems compared to FrameNet, and that has potential for generalizing across languages (cf. Taulé et al. (2010)). Second, our work aims at **filling an important resource gap for German**, which is lacking resources for VerbNet-style SRL. At the same time, we provide evidence that the VerbNet role inventory can be successfully applied to German. In doing so, we investigate two revisions of the VerbNet semantic role annotation scheme and establish detailed annotation guidelines. Finally, **combining word sense and semantic role annotation** raises several questions that we will address in our study: To what extent are word sense and role annotation dependent on each other? What kinds of preannotation for predicates and arguments are helpful to support role annotation (e.g., presenting dependency heads or full argument spans as role targets)? To answer these questions, we design contrastive annotation setups and evaluate their impact on annotation processes and results. In contrast to VerbNet, we use GermaNet 9.0 (Hamp and Feldweg, 1997; Henrich and Hinrichs, 2010) – the German counterpart of Princeton WordNet (Fellbaum, 1998) – as a fine-grained sense inventory for predicate labeling, and investigate its suitability for combined sense and VerbNet-style role labeling.

Our contributions We provide a novel adaptation of a VerbNet-style semantic role set for manual annotation on a German corpus. We identify best annotation practice for joint word sense and VerbNet-style role labeling, including analysis of IAA, and investigate systematic dependencies between predicate and role annotation. Using the combined annotation scheme, we conduct a large-scale annotation study on the SALSA corpus. We will publish annotations for about 3,500 predicate argument structures for 275 verb types, along with detailed annotation guidelines.[1]

[1] http://projects.cl.uni-heidelberg.de/GNVN_semanno

Outline Section 2 discusses related work. Section 3 provides a description of the annotation study for combined GermaNet sense and VerbNet role annotation on German language data. This includes general annotation principles and guidelines, data description, and the design of different annotation setups for joint word sense and semantic role annotation. We provide details about the adapted VerbNet role inventories and guidelines and analyze the obtained annotation results, with special focus on the interaction of sense and role annotation. Section 4 summarizes our findings and concludes with an outlook on future work.

2. Related Work

Relevant related work concerns the development of SRL schemata as well as sense and role annotation for German.

FrameNet offers a full-fledged semantic predicate-argument representation. Predicates trigger a prototypical situation, called frame, that defines the possible participants in the situation and their semantic roles in relation to that predicate. Due to a large number of frames, the role inventory is very large: FrameNet 1.5 contains 908 lexical frames, and 8,884 unique role labels. Despite its considerable size, FrameNet is still incomplete, lacking both frames for unattested predicate meanings and missing assignments of lemmas to existing frames. Sparsity at the level of predicate meanings, combined with the diversity of roles leads to lexicon coverage issues and sparsity problems when training automatic SRL systems. On the other hand, FrameNet provides very intuitive and highly expressive sense (frame) and role labels, as well as frame-to-frame relations, including inheritance, within the FrameNet hierarchy.

PropBank provides a small role inventory. Labels for obligatory arguments lack semantic transparency and are marked as A0 to A5. Adjuncts are tagged with a small set of labels, such as ArgM-LOC. The Agent-like A0 and the Patient-like A1 roles generalize over syntactic alternations. Their definition is based on Dowty's proto-roles (Dowty, 1991), capturing salient Agent- and Patient-like properties. Reisinger et al. (2015) attempt a characterization of **Proto-Roles** as distributions over properties, collected by crowdsourcing. In our study, we aim at a concise role inventory that covers the full range of arguments to a predicate.

VerbNet is located between FrameNet and PropBank on a continous scale between a fine-grained interpretable role inventory on one side and a compact, coarse-grained inventory on the other. Silberer and Frank (2012) identified a stronger generalization capacity of VerbNet roles compared to FrameNet roles in the task of non-local role binding. Merlo and van der Plas (2009) found that PropBank roles, being closer to syntax, are easier to assign than VerbNet roles, while the latter provide better semantic generalization. VerbNet defines a set of up to 35 roles, which are defined independently from specific verb senses. Next to VerbNet, the ISO-standard SemAF-SR based on LIRICS (ISO, 2014) proposed a semantic role inventory in close consideration of VerbNet roles. Bonial et al. (2011) and Hwang (2014) present revised VerbNet role hierarchies. They serve as a basis for proposed revisions to the VerbNet role inventory, as described in Sec. 3.2. and 3.6.

VerbNet is based on a set of **semantic verb classes**. Assigning a verb to its class is similar to a sense disambiguation act, with senses defined by the syntactic alternation behaviour of verbs (Levin, 1993). Prior work has investigated the multilingual applicability of VerbNet semantic classes (Falk et al., 2012). While Levin-style verb classes are well established, the classes are relatively coarse-grained and not all of them are semantically homogeneous.

German sense- and role-annotated corpora. There are only few sense- and role-annotated corpora for German.
Corpora with word sense annotations according to GermaNet have been created by Broscheit et al. (2010), Henrich et al. (2012), and Henrich and Hinrichs (2014). However, none of these provide semantic role annotation.
Schulte im Walde (2006) clustered German verbs to semantic verb classes inspired by Levin (1993), trying to match semantic verb 'fields' as defined by Schumacher (1986). These classes have not been employed for manual corpus labeling or automatic sense tagging.
For semantic role annotation, SALSA (Burchardt et al., 2009) is the only available corpus. It is annotated with FrameNet 1.3 frames and roles, and extended with proto-frames for predicates unknown to FrameNet 1.3. For the CoNLL 2009 SRL task (Hajič et al., 2009), the SALSA annotation was converted semi-automatically to PropBank role labels. By reannotating a substantial part of SALSA with VerbNet roles, we facilitate comparative research on semantic role inventories for the German language.

3. Combining GermaNet Sense and VerbNet Semantic Role Annotation

The present work is the first to apply the VerbNet role inventory to German language data. As part of this process we explore both revisions to the VerbNet role inventory, and we combine VerbNet semantic roles with fine-grained WordNet-style senses obtained from GermaNet.
This section describes our annotation studies and resulting resources in detail, beginning with an introduction of the sense inventory and the adapted VerbNet role inventory.

3.1. Sense Inventory

Given that VerbNet roles are defined independently of specific verb senses, the role inventory can be freely combined with alternative predicate sense inventories. Employing independent sense and role inventories in joint predicate and role-semantic labeling offers great flexibility for automated systems. WordNet, and its German counterpart GermaNet, offer more fine-grained senses with concise semantic distinctions compared to VerbNet semantic classes. We thus expect a more expressive semantic annotation from this combination and combine VerbNet role annotation with the sense inventory of GermaNet 9.0 for predicate labeling.

3.2. A VerbNet-style Role Inventory for German

As main starting points for our work we consider VerbNet roles (Kipper-Schuler, 2005; Hwang, 2014) (see Figure 1), and the unified semantic role inventory in the ISO standard SemAF-SR (ISO, 2014) which is based on the role set proposed in the LIRICS project and integrates elements

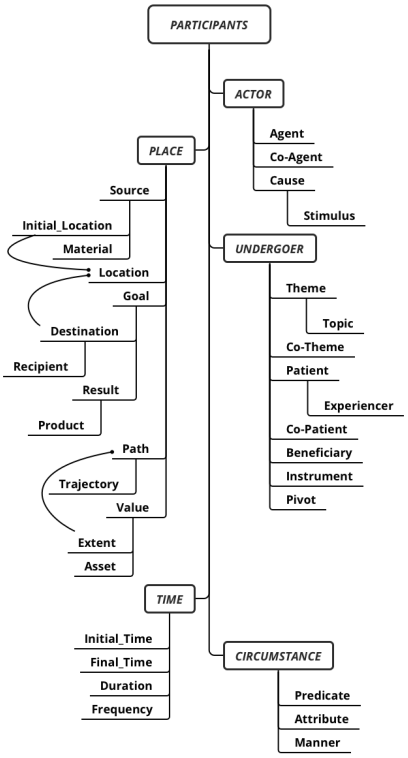

Figure 1: Recent VerbNet role inventory (Hwang, 2014)

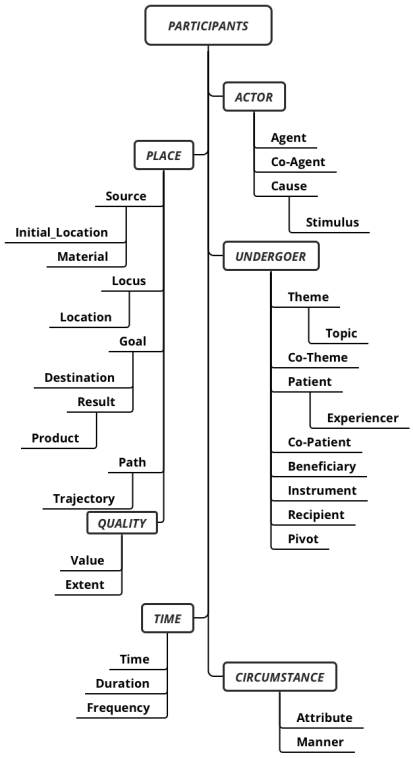

Figure 2: RI-II: Semantic role inventory for German.

of VerbNet, PropPank, and FrameNet. We take the most recent VerbNet role inventory in Hwang (2014) as a basis, and propose modifications or refinements, using SemAF-SR as a base for comparison. We also consider SemAF-SR to refine role definitions if needed. In contrast to SemAF-SR, we do not consider adjuncts for annotation. Additionally, we provide detailed role descriptions, examples, and annotation guides for German. It is important to note that our *guiding annotation principle* is not to start from a lexicon view, and find examples for predefined role sets and alternations for each predicate, but rather to provide solid semantic criteria for roles that allow for a purely data-driven annotation, based on the semantic role inventory. In this we follow the corpus-driven annotation perspective of SALSA, as opposed to the lexicon-driven practice in FrameNet or VerbNet.

Role descriptions, examples and guides. Since we annotate German data, we provide German formulations for role definitions, in some cases providing definitions where none were available (e.g. *Frequency* or *Trajectory*). In these cases we formulated the definition according to VerbNet annotations and the ISO-standard inventory description, if available. We added example sentences to the descriptions for illustration. Following Tremper and Frank (2013), we assist the annotators with questions as a guide for role selection, in the form of a 'decision tree' that asks questions about properties of predicates and arguments, and guides the annotator in discriminating between similar roles like, e.g., *Theme* or *Topic*: "Is the participant being communicated (by someone)?" – if so, it should be annotated with the more special role *Topic*. The formulated questions are similar to the proto-role properties of Reisinger et al. (2015), explicitely referring to salient features of the roles.

Adaptations to VerbNet hierarchy and role inventory. The VerbNet role hierarchy follows two main principles: lower-level roles are more specific, and restricted by semantic properties and constraints; consequently, roles in a parent-child relation tend to not co-occur. However, the hierarchy contains multiple inheritance links and is therefore difficult to conceptualize (cf. Figure 1).

Role Inventory I (RI-I): Inspired by SemAF-SR, we tried to flatten this hierarchy. We also borrow descriptions for roles not declared in detail within VerbNet: *Manner* and *Path*. Both roles are specific and rarely found in our data.

We further simplified the inventory by grouping similar roles into 'multi-roles', for instance, *Product* and *Result*; or *Value*, *Extent*, and *Asset*, respectively. The assumption was that decisions between coarse-grained roles are easier and yield more fluent and consistent annotation compared to the original inventory. All in all, we offered the annotators an inventory of 27 roles.

Role Inventory II (RI-II): During the first annotation round, using *Role Inventory I*, we observed difficulties in annotation. Feedback from, and discussions with the annotators culminated in a revised role inventory, role descriptions and guidelines, cf. Figure 2.

The changes we propose for the original VerbNet role inventory in Figure 1 are kept as small as possible. However, our aim is to strengthen the *semantic principles* that underlie the role hierarchy in a more systematic way. This process results in the redefined role inventory *RI-II*.

RI-II maintains the hierarchical structure of the original VerbNet role inventory, but we resolve multiple inheritance relations by selecting a single parent, according to the se-

mantically most prominent relation. This yields a transparent and semantically coherent *tree* structure.
Moreover, in *RI-II* a child role is systematically assumed to be either more *specific* (e.g. *Topic* to *Theme*), or more *concrete* (e.g. *Material* to *Source*) compared to its parent. We adapt existing definitions in case they need to be sharpened to clearly reflect this structuring principle, or we define new ones if necessary (*Frequency, Path, Trajectory, Manner*).
These assumptions led to further modifications of the inventory: 1) For *Recipient*, we identified a stronger semantic relatedness to the *Undergoer* roles, thus *Recipient* was rearranged from the group *Place* to *Undergoer*. 2) We introduce a new role *Locus* as an abstract counterpart for *Location*, to make it consistent with other abstract–concrete role pairs. The role is used in examples like 'to live in$_{Locus}$ a cruel world'. 3) We drop VerbNet's role *Predicate*, which is defined on purely syntactic criteria, to refer to embedded clauses. These clausal arguments can be successfully analyzed with existing roles, most often *Theme*. Thus, we obtain the same role set for 'He knew the answer' and 'He knew he was right'. 4) We introduced a new subgroup *Quality* for measurable entities, with roles *Value* and *Extent*. The former integrates the previous role *Asset* (a realization of a value as an object), to avoid too fine-grained distinctions that are difficult to differentiate. 5) Similarly we merge *Initial_Time* and *Final_Time* to a single role *Time*. This decision is supported by the fact that VerbNet does not provide any annotated instances for these roles.
This restructuring process yields overall 32 roles in a tree hierarchy that is based on clearly defined semantic criteria, with roles gouped into five semantic groups *Actor, Undergoer, Place, Time*, and *Circumstance*.

3.3. Guidelines for Semantic Annotation

We created annotation guidelines for both considered semantic annotation tasks that fulfill general desiderata: a) clear definition of annotation targets, b) instructions about the procedure, and c) clear definition of the annotation labels, including guides for deciding difficult and irregular cases, using guiding questions and examples.
We created new role annotation guidelines for German that are assumed to generalize to other languages (Petukhova and Bunt, 2008). The original VerbNet guidelines[2] do not provide detailed instructions. The UVI[3] web interface with VerbNet roles gives insight into how roles are applied for specific verbs or verb classes, but does not serve as a general guideline. SemAF-SR provides more specific definitions which we refer to in some of the role descriptions.
Beyond technical instructions for usage of the annotation tool, the annotation targets, 'predicate' and 'argument', have to been defined. Even though we provide preannotated targets, the annotators must be able to verify – and if needed, correct – the proposed targets. Thus, the introductory part of the guidelines declares the theoretical background and technical instructions of the annotation, including special cases, e.g. how to handle split verb targets (e.g. separable verb prefixes in German), or multi-word expressions.

3.4. General Annotation Setup

Task We divided the annotation process for combined word sense and semantic role annotation into two consecutive tasks, which are performed by trained annotators: 1) word sense annotation (WSD), and 2) semantic role annotation (SRL). For both tasks we designed several settings to work out 'best practices'.
The annotation is performed in two big phases:

1. In a **pilot study** on a subset of the data we evaluated different annotation configurations for joint sense and role annotation. The pilot data was annotated twice, using alternative role annotation schemata: RI-I/II.
2. The **main study** covers the complete data set, using the RI-II role annotation scheme, and following the 'best configuration' setting.

To evaluate the quality of the annotations in the different setups, we perform double annotation and compute agreement between the annotators.

Annotation configurations evaluated in pilot studies The pilot study implements a controlled experiment for establishing a *best practice* for combined GermaNet-style sense and semantic role annotation. Of particular interest are interactions between word senses and roles: (i) Given the fine-grained sense inventories in GermaNet, we are interested in whether specific meaning differences interact with role annotation differences. (ii) We also investigate whether verb senses need to be preannotated as a guide for role annotation, or whether annotators will intuitively resolve the predicate meaning when annotating roles in context. We also investigate what type of preannotation for arguments is helpful to support role annotation.
To answer these questions, we designed contrastive annotation settings and evaluated differences in annotation processes and results. The particular settings for sense and role annotation will be described in Sections 3.5 and 3.6.

Annotators In each setting, annotations are provided by two annotators and disagreements are resolved by an adjudicator in order to create a final gold standard. All annotators are (under)graduate students with expertise in linguistics or computational linguistics. Both groups were introduced to the task in a previous workshop. As we redefined the role inventory during our study, the role annotators were supported through discussion meetings to receive feedback and to clarify challenging cases.

Data Our data is based on the SALSA corpus (Burchardt et al., 2009). The corpus provides an appropriate base for several reasons: it is manually annotated with FrameNet frames and roles, and the annotations are available in a dependency-parsed, semi-automatic conversion to PropBank roles (Hajič et al., 2009). Our annotations will thus allow for systematic comparisons of role schemata.
For the *pilot study*, we select the first 18,000 tokens from the dependency-parsed CoNLL09 version. We split the data into four subsets containing 100 verbal predicate instances each, which we used in different annotation configurations.

[2] http://verbs.colorado.edu/verb-index/VerbNet_Guidelines.pdf
[3] http://verbs.colorado.edu/verb-index/

Annotation	Configuration
Predicate sense (WSD)	[*lemma*] predicate target [*best*] best GermaNet sense [*all*] all (fitting) GermaNet senses
Argument role (SRL)	[*span*] span [*head*] head [*span+head*] span with head
Role inventory (SRL)	[*RI-I*] flat multi-role set [*RI-II*] hierarchical, refined set

Table 1: Configurations for joint WSD and SRL annotation.

For the *main study*, we select the 210 most frequent polysemous verb lemmas as target predicates. For each predicate, we randomly sample up to 15 instances from the entire training section. Additionally, we annotate all instances of the target predicates in the CoNLL09 test section, so that the annotations can be employed for evaluation on the offical shared task test set. This results in overall 3,500 annotated predicate-argument structures for 275 predicate types.
We perform new annotations on the annotated predicates and arguments from the dependency-parsed CoNLL09 corpus, and present them to the annotators as unlabeled annotation targets. To support annotation consistency, predicate targets are grouped by lemma, so that the annotators can concentrate on all instances of one predicate type at a time. Each predicate target is surrounded by a 5-sentence context window to support disambiguation.

Annotation tool We employ WebAnno (Yimam et al., 2014).[4] The tool provides custom annotation layers and tagsets, which we extended with annotation layers for semantic predicates and semantic roles.

Annotation procedure For sense annotation, the preselected predicate instances and their context are presented to the annotators. The role annotators are provided with the adjudicated predicate sense annotations and the preselected arguments. The annotators choose appropriate roles for the arguments of *each annotated predicate sense*. The sense and role descriptions are provided in the same manner as those provided to the sense annotators.

3.5. Sense Annotation Study

For sense annotation, the target predicates are annotated with GermaNet 9.0 sense(s). For convenience, sense definitions, synonyms, hypernyms, hypernyms, antonyms (and other relations if existing), as well as example sentences in GermaNet 9.0 are presented in the annotation interface sorted and constrained by the predicate lemma. The annotators are instructed to first annotate *all senses* for a predicate which are appropriate in the presented context (configuration *[all]*), and then to select a *single best sense* (*[best]*), cf. Table 1. If GermaNet does not offer an appropriate sense, the annotators choose the label "XXX". With the fine-grained sense distinctions in GermaNet, 23.1% of the pilot targets receive multiple sense labels.

[4] https://webanno.github.io/webanno/

	Types		Tokens	
	lemma	sense	lemma	sense
pilot	168	265	397	499
main	239	571	3,085	3,639
all	275	633	3,482	4,138

Table 2: Corpus statistics: annotated GermaNet senses

Sense IAA	pilot	main
[best]	0.73	0.76
[all]	0.81	0.81

Table 3: IAA (Krippendorff's α) for sense annotation.

Pilot study Table 3 shows the inter-annotator agreement (IAA) between the annotators in terms of Krippendorff's α (Meyer et al., 2014), which is respectably high: We measure $\alpha = 0.73$ for the *[best]* configuration, and $\alpha = 0.81$ for the multi-label annotation *[all]*.

Main study The final sense-labeled corpus contains 2,606 validated sense annotations for 2,290 target verbs. The main data set contains up to 38 instances for 210 lemmas – see Table 2.
The majority of predicate targets receive a single unique sense label with only 16.9% of targets in the main study receiving multiple labels – most of them two labels, compared to 23.1% in the pilot. 66 tokens in the main study could not be labeled with a valid sense from GermaNet.
Table 3 shows that the IAA results for both the main and the pilot study are stable, and demonstrate the high quality of the sense annotations.

3.6. Role Annotation Study

In the pilot study, we investigate different configurations for role annotation (cf. Table 1).
First, we vary the presence of sense labels for the predicate: in the setting [*lemma*], only the predicate lemma is displayed; in the settings [*all*] and [*best*], GermaNet sense labels and their definitions are displayed to the annotators. While annotators found the additional sense information to be helpful, we found only few differences in the annotated labels between both settings. For the main study to provide a maximum amount of information, we chose the configuration where annotators see all GermaNet senses.
A second parameter contrasts different ways of presenting preannotated argument candidates:

a) [*head*]: the argument's head word is marked as annotation target (jointly with the dependency structure),

b) [*span*]: the whole argument span is presented as a candidate (inferred from the dependency structure),

c) [*span+head*]: presentation of span with head labeling.

Since WebAnno is able to display full argument spans according to the dependency structure, preannotation marking only the head word was found to be most convenient by our annotators.
Not only arguments, but also predicates can contain more than one word, e.g. with separable verb particles. In this

case, the particle is marked as part of the predicate.[5]

The choice of an appropriate and well-defined role inventory is crucial for the annotation. Our first semantic role inventory *RI-I* (see Section 3.2.) was designed to be as simple as possible. The VerbNet hierarchy was broken down to a flat list of roles, following the model of SemAF-SR. Further, we merged some of the roles to 'multi-roles' on the assumption that the reduced and coarser role inventory makes it easier to distinguish roles. However, this assumption was not confirmed, as we obtained low IAA for *RI-I*.

Pilot Study I The IAA for role annotation with *RI-I* was unexpectedly low, at Krippendorff's $\alpha = 0.46$ for all roles. Petukhova and Bunt (2008), for their similar-sized role inventory, reported Cohen's[6] $\kappa = 0.77$ for argument identification and $\kappa = 0.68$ for role labeling. We investigated the reasons and identified several aspects of importance.

First, some roles are rare, and annotators initially tended to avoid them. By restricting IAA evaluation to the roles occurring at least 20 times, we obtain $\alpha = 0.64$ for 24 roles, which is significantly better, though still not high.

Grouping roles to multi-roles did not always help the annotators to make decisions more easily. While grouping of *Value*, *Extent*, and *Asset* helped a lot, merging *Source* with *Material* was found to be difficult. Also, some role definitions were not clearly differentiated from others, and annotators found the example sentences not to generalize well to 'real world' instances. Finally, the annotators report that more training time and continuous feedback and discussions were extremely helpful.

Role Inventory *RI-II* Following this analysis, we refined the role inventory and enhanced role descriptions to provide a more stable annotation guide. The revised role inventory *RI-II*, displayed in Figure 2, comprises 32 roles and follows clear semantic structuring principles (cf. Section 3.2.). The hierarchical role topology helps to distinguish agent-like, patient-like, and further semantically motivated role types. The role definitions were considerably sharpened, and enhanced with further discriminating guiding questions. We also included more 'real world' examples for illustration.

Pilot Study II and Main Study Reannotation of the pilot data using *RI-II* yields much more consistent results. In addition we completed annotation of 30 predicate types out of the whole main data set ($main_{30}$). Agreement on this partial data set shows satisfactory results and is consistent with IAA measured for the pilot data. Results for the joint data set are summarized in Table 4. IAA reaches $\alpha = 0.73$ for individual roles (measuring agreement separately for each argument), and is at least $\alpha = 0.62$ on role sets (measuring IAA for the complete role set of a predicate). In average, 1.94 arguments belong to a predicate instance, with up to four arguments in a role set. We will publish the role annotations for the full dataset, which is expected to contain around 6,000 role instances.

[5]Further multiword expressions like phrasal verbs or the reflexive pronoun of a reflexive verb are not preannotated as part of the predicate, but need to be marked as such by the annotators.

[6]Cohen's κ doesn't handle non-pairable items.

Role IAA	Roles	Role sets
[best]	0.73	0.62
[all]	0.74	0.73

Table 4: Role IAA for RI-II on pilot and $main_{30}$ data.

Role IAA	Role
$\alpha \geq 0.8$	*Time, Duration, Agent, Patient, Topic*
$\alpha \geq 0.6$	*Goal, Recipient, Theme, Attribute, Result, Location, Value, Pivot, Experiencer, Frequency, Cause, Stimulus*
$\alpha \geq 0.4$	*Product, Manner, Locus, Beneficiary, Initial_Location, Instrument, Source*
$\alpha < 0.4$	*Trajectory, Extent, Destination, Co-Agent, Co-Theme, Path*

Table 5: IAA for role types on pilot and $main_{30}$ data.

3.7. Analysis of Annotation Results

Analysis of Role Annotation The most frequently annotated roles are *Agent*, *Theme*, *Topic*, and *Goal* occuring altogether in 68% of the annotations. The high number of *Topic* shows that the distinction from *Theme* is useful. By contrast, no *Co-Theme* or *Co-Patient* has been annotated. Both roles are specific to a small number of predicates. IAA on individual roles, shown in Table 5, shows high agreement of $\alpha \geq 0.6$ for more than half of the role types.

There are roles which tend to be confused by the annotators. A typical case is the rather general *Undergoer* role *Theme*, which needs to be distinguished from more specific roles such as *Source*, *Goal*, or *Beneficiary*.

Two roles prove difficult due to their definition: *Pivot* and *Goal*. *Pivot* is a vaguely defined role in VerbNet and SemAF-SR, and used for verbs that take two *Theme*-like roles: the more central of them receives the role *Pivot*. As 'centrality' in the event is not clearly defined, we define the 'logical subject' of the event to fill the *Pivot* role, as seen in (1). With this refinement we observe an IAA of 0.69.

(1) *Diese Länder*$_{Pivot}$ *brauchen*$_1$ *keine Alternative*$_{Theme}$.
'These countries do not need an alternative.'

The role *Goal* actually unites two readings: an 'abstract destination' and also 'aim, purpose', where the latter can be interpreted as an abstract destination. The difficulty lies in areas of unclear distinction between the two functions, as in example (2)[7]. Because of these difficulties we maintain the ambiguous interpretation of *Goal*, as opposed to splitting it into goal and purpose, as SemAF-SR does.[8]

(2) *Man*$_{Agent}$ *kann diese Position*$_{Goal}$ *über schulische Leistungen und Intelligenz erwerben*$_2$.
'One can reach this position through academic performance and intelligence.'

Co-occurrence of roles Some roles form patterns in the sense that they typically co-occur in an argument frame.

[7]Predicate senses are marked by a GermaNet sense index.

[8]VerbNet does not define a role for *purpose* if the object in question exists independently of the event.

Role set type	#RST (%)	#Lm	%Lm	Lemma example
all 1-role	18 (14.7)	94	100	
Agt-Thm		54	23.2	anbieten 'offer'
Agt-Top		28	12.0	klagen 'complain'
Attr-Thm		11	4.7	bleiben 'remain'
Piv-Thm		10	4.3	besitzen 'own'
Exp-Stim		9	3.8	spüren 'feel'
all 2-roles	60 (49.1)	232	100	
Agt-Rec-Top		7	10.0	bitten 'ask for'
Agt-Attr-Thm		7	10.0	nennen 'call'
Agt-Rec-Thm		5	7.1	anbieten 'offer'
all 3-roles	39 (31.9)	70	100	
all 4-roles	5 (4.0)	6	100	

Table 6: Frequent role sets (out of 122) in pilot and main$_{30}$ data, grouped by number of roles in a role set, with number and percentage of role set types and lemmas.

Such co-occurrences can be due to purely semantic reasons, in that one role implies the occurrence of another one, as in the following cases: 1) *Experiencer – Stimulus*: Both roles are special to perception predicates where the *Stimulus* is perceived by the *Experiencer*. 2) *Cause – Result/Goal*: The *Cause* represents the cause or precondition for an aim or result, which are captured by the *Goal* or *Result* role, respectively. 3) *Attribute – Theme*: The *Attribute* role refers to a property of another participant, typically a *Theme*.
A double assignment of the same role for a predicate is prevented by the following patterns: 4) *Co-Agent/-Theme/-Patient – Agent/Theme/Patient*: *Co-Roles* occur only with another argument of the corresponding base role. 5) *Pivot – Theme*: The *Pivot* role is assigned to a *Theme*-like role in case another participant is clearly interpreted as a *Theme*.

Role sets A *role set* is defined as the set of argument roles assigned to a given predicate in context. Table 6 shows frequent types of role sets found in the annotated data. Almost half of the role set types consist of two roles, about 30% are role triples. Some role set types are frequent: 5 role set types cover 48% of the lemma types annotated with two roles, and 3 role set types cover 27% of lemma types annotated with 3 roles, respectively. These patterns can be exploited via global constraints in automatic SRL models.

Interactions between role sets and GermaNet senses The *frequent role sets* shown in Table 6 typically occur with many predicate types. However, we also observe that *alternating role sets* can be assigned to a specific predicate sense, as illustrated in Table 7. The observed variations are in accordance with VerbNet's class-internal optional roles (e.g., *Recipient* for *investieren* or *Attribute* for *wissen*), or are due to syntactic alternations (e.g. active–passive, here involving *Agent* and *Cause*). Further analysis of the data is required to establish general patterns for VerbNet-like alternations that can be exploited in joint word sense and role assignment modeling.
Conversely, several senses of a single predicate can share a common role set. In examples (3) and (4), the verb *kontrollieren* 'control' occurs with the role set *Agent–Theme* for all of its three senses. In fact, the multiple senses assigned in (3) can be considered candidates for grouping them to a single sense, as discussed in Palmer et al. (2005).

Sense	Alternating Role sets
investieren$_1$ 'invest'	*Agt-Rec-Value, Agt-Value, Value*
wissen$_1$ 'know'	*Agt-Thm-Attr, Agt-Thm, Agt*
zwingen$_1$ 'force'	*Th-Result-Cause, Thm-Result*

Table 7: Alterntive role sets associated with single senses.

(3) *Athen$_{Agent}$ kontrollierte$_{1,3}$ etwa siebzig Prozent der Wirtschaft$_{Theme}$.*
'Athen controlled about 70 perc. of the economy'

(4) *Für$_{Agent}$ die Bürger ist dies die einzige Möglichkeit die Politik$_{Theme}$ zu kontrollieren$_2$*
'This is the only way for citizens to control politics'

Finally, almost half of the role set types *uniquely correspond to a single predicate sense*. Further investigation and additional volumes of annotated data will show to what extent such one-to-one correspondences can support automatic word sense disambiguation. Moreover, such correspondences can serve as a basis for linking our VerbNet-based role sets to other semantic role inventories, as in SemLink (Loper et al., 2007).

4. Conclusion and Outlook

We present a novel adaptation and description of a VerbNet-style role inventory, which we define on the basis of semantic structuring criteria. We designed a discriminative annotation study for combined GermaNet sense and VerbNet-style semantic role annotation. We examine factors that influence the annotation quality, and provide detailed IAA statistics proving the high quality of the annotations, as well as evaluation and interpretation of results. We specifically analyze properties of VerbNet role annotation and their interaction with word sense annotation. Future work will further explore these observations in relation to automatic sense and role annotation.
By enriching a substantial part of the German SALSA corpus with both semantic annotation layers, our corpus fills an important resource gap. The preexisting FrameNet annotation on the corpus will allow us to align VerbNet role labels with frame-semantic annotation for German. In combination with the GermaNet sense annotations, we will construct a linked lexical resource, as a German counterpart of SemLink (Loper et al., 2007).
The final corpus resource will provide 3,500 annotated instances for approximately 275 German verb types with joint annotation of GermaNet senses and VerbNet-style semantic role labels. We will publish the detailed annotation guidelines and the final corpus for further research.

5. Acknowledgements

This work has been supported by the BMBF CLARIN-D project "Semantic Annotation for Digital Humanities". We thank the sense and semantic role annotators: Elisa Starke, Annabel Köppel, Silke Kalmer, Christopher Tauchmann, Patrick Claus, Uli Paul Steinbach, and Simon Will; the WebAnno team for support with new features; Jennifer Sikos for comments and information regarding VerbNet.

6. References

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet Project. In *Proceedings of COLING-ACL*, pages 86–90, Montreal, Canada.

Bonial, C., Corvey, W., Palmer, M., Petukhova, V., and Bunt, H. (2011). A Hierarchical Unification of LIRICS and VerbNet Semantic Roles. In *Proceedings of ICSC Workshop on Semantic Annotation for Computational Linguistic Resources*.

Broscheit, S., Frank, A., Jehle, D., Ponzetto, S. P., Rehl, D., Summa, A., Suttner, K., and Vola, S. (2010). Rapid Bootstrapping of Word Sense Disambiguation Resources for German. In *Proceedings of the 10. Konferenz zur Verarbeitung Natürlicher Sprache*, pages 6–8.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Pado, S., and Pinkal, M. (2009). FrameNet for the semantic analysis of German: Annotation, representation and automation. In *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton De Gruyter.

Dowty, D. (1991). Proto-Roles and Argument Selection. *Language*, 67(3):547–619.

Falk, I., Gardent, C., and Lamirel, J.-C. (2012). Classifying French verbs using French and English lexical resources. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 854–863.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.

Hajič, J., Ciaramita, M., Johansson, R., Kawahara, D., Martí, M. A., Màrquez, L., Meyers, A., Joakim, N., Padó, S., Štěpánek, J., Straňák, P., Surdenau, M., Xue, N., and Zhang, Y. (2009). The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL): Shared Task*, pages 1–18, Boulder, Colorado.

Hamp, B. and Feldweg, H. (1997). GermaNet - a Lexical-Semantic Net for German. In *Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15, Madrid.

Henrich, V. and Hinrichs, E. (2010). GernEdiT - The GermaNet Editing Tool. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, pages 2228–2235, Valletta, Malta.

Henrich, V. and Hinrichs, E. (2014). Consistency of Manual Sense Annotation and Integration into the TüBa-D/Z Treebank. In *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories*, pages 62–74, Tübingen, Germany.

Henrich, V., Hinrichs, E., and Vodolazova, T. (2012). WebCAGe – A Web-Harvested Corpus Annotated with GermaNet Senses. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 387–396.

Hwang, J. D. (2014). *Identification and Representation of Caused Motion Constructions*. Ph.D. thesis, University of Colorado.

ISO. (2014). Language resource management – Semantic annotation framework (SemAF) – Part 4: Semantic roles (SemAF-SR).

Kipper-Schuler, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania, Pennsylvania, Philadelphia, PA.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Loper, E., Yi, S.-t., and Palmer, M. (2007). Combining lexical resources: Mapping between PropBank and VerbNet. In *Proceedings of the 7th International Workshop on Computational Linguistics*.

Merlo, P. and van der Plas, L. (2009). Abstraction and Generalisation in Semantic Role Labels: PropBank, VerbNet or both? In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pages 288–296, Suntec, Singapore.

Meyer, C. M., Mieskes, M., Stab, C., and Gurevych, I. (2014). DKPro Agreement: An Open-Source Java Library for Measuring Inter-Rater Agreement. In *Proceedings of COLING 2014: System Demonstrations*, pages 105–109, Dublin, Ireland.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: A Corpus Annotated with Semantic Roles. *Computational Linguistics*, 31(1):71–106.

Petukhova, V. and Bunt, H. (2008). LIRICS semantic role annotation: design and evaluation of a set of data categories. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, pages 39–45, Marrakech, May.

Reisinger, D., Rudinger, R., Ferraro, F., Harman, C., Rawlins, K., and Van Durme, B. (2015). Semantic Proto-Roles. *Transactions of the Association for Computational Linguistics*, 3(8):475–488.

Schulte im Walde, S. (2006). Experiments on the Automatic Induction of German Semantic Verb Classes. *Computational Linguistics*, 32(2):159–194, June.

Schumacher, H. (1986). *Verben in Feldern*. de Gruyter, Berlin.

Silberer, C. and Frank, A. (2012). Casting Implicit Role Linking as an Anaphora Resolution Task. In *First Joint Conference on Lexical and Computational Semantics (*SEM)*, pages 1–10, Montréal, Canada.

Taulé, M., Martí, M. A., and Borrega, O. (2010). AnCora-Net: Mapping the Spanish AnCora-Verb lexicon to VerbNet. In *Proceedings of the Interdisciplinary Workshop on Verbs. The Identifcation and Representation of Verb Features*.

Tremper, G. and Frank, A. (2013). A Discriminative Analysis of Fine-Grained Semantic Relations including Presupposition: Annotation and Classification. In *Dialogue and Discourse, Special Issue on Annotating Pragmatic and Discourse Phenomena*, pages 282–324.

Yimam, S. M., Eckart de Castilho, R., Gurevych, I., and Biemann, C. (2014). Automatic annotation suggestions and custom annotation layers in webanno. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 91–96, Baltimore, Maryland, USA.