Superframes: A Schema for Lexicon-free Frame-semantic Annotation

Kilian Evang
Heinrich Heine University Düsseldorf, Germany
evang@hhu.de

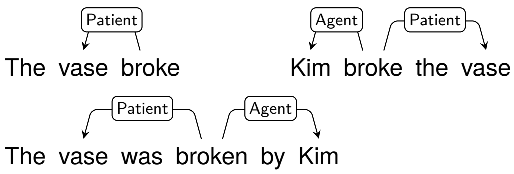
Abstract

Semantic frame, role, and relation labels are important parts of symbolic meaning representations. The mainstream approach is to define large language-specific lexicons that map predicate senses to their frame and argument labels, and to use separate label inventories for modifiers. Maintaining lexicons is very labor-intensive and scales poorly to the multilingual case. There are schemas that aim to simplify the annotation task using a more coarse-grained inventory of frames and roles, but suffer from a lack of systematicity and clear definitions. We present a schema that 1) uses a small inventory of frames and roles for lexicon-free annotation, 2) systematizes the frame inventory by factoring out aspect and mode, 3) has a unified vocabulary for arguments and modifiers, and 4) is designed to be annotated atop Universal Dependencies syntactic annotation. We argue for the adoption of such a schema for multilingual annotation and demonstrate promising results in annotation experiments on German.

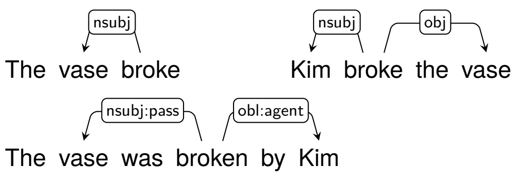
Keywords: frame semantics, semantic roles, annotation

1. Introduction

Semantic role labeling (SRL) is a classic task in natural language processing (NLP; Gildea and Jurafsky, 2002). Given a predicate and its arguments, SRL aims to answer the question "who does what to whom, where and when", i.e., which argument plays which role in the situation described. It abstracts away from syntactic alternations such as causative or passive, as exemplified in (1).

(1)


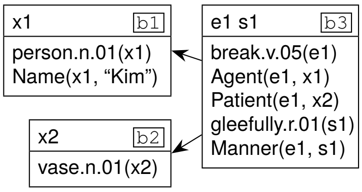
The situations described are similar, and consequently the semantic roles of all arguments (*Kim* and *the vase*) stay the same across alternations, while the syntactic roles differ, as seen in (2).

(2)


Because of this abstracting away from syntactic differences, SRL is an important upstream or auxiliary task for semantic applications such as relation extraction (e.g., Tsai et al., 2006; Christensen et al., 2010; Guo et al., 2016; Humphreys et al., 2020; Zhang et al., 2022), dialogue systems (e.g., Rachman et al., 2018; Xu et al., 2020; Wu et al., 2022), Natural Language Inference (Cengiz and Yuret, 2020), or Natural Language Understanding in general (Zhang et al., 2020). Finally, semantic roles are a central part of formal meaning representations in frameworks such as the Parallel Meaning Bank (PMB; Abzianidze et al., 2017) or Abstract Meaning Representations (AMR; Banarescu et al., 2013). Such meaning representations are useful for semantic parsing and providing structured input to downstream applications such as chatbots and information extraction systems, but can also be useful to corpus linguists who want to query for specific semantic structures. Different meaning representation frameworks use different semantic role inventories. For example, the PMB uses a small set of generic roles across all predicates and across arguments and modifiers. AMR, on the other hand, uses predicate-specific roles based on PropBank (Palmer et al., 2005) for arguments, and a separate set of relations for modifiers. Examples of both are shown in Figure 1.

Syntactic analyses of sentences are these days relatively easy to produce automatically. Software libraries like spaCy (Montani et al., 2023), Stanza (Qi et al., 2020), or UDPipe (Straka et al., 2016) come with readily trained syntactic parsing models for many languages. Part of the reason for this wide availability is the success of the Universal Dependencies project (UD; de Marneffe et al., 2021), from which the syntactic relations shown in (2) are drawn. UD emphasizes universality (language-independence) over fine-grained distinctions, using a relatively coarse inventory of syntactic relations, which enables easy annotation, a shared understanding between annotators of different languages, therefore a cost-effective way to produce large amounts of training data, and the ability for machine-learning models to generalize

Proceedings of The Seventh International Workshop on Designing Meaning Representations (DMR 2026) @ LREC 2026, pages 124–135
11 May 2026.




```
(b / break-01
   :ARG0 (k / Kim)
   :ARG1 (v / vase)
   :manner (g / gleeful))
```

Figure 1: A DRS and an AMR for the sentence "Kim broke the vase gleefully"

multilingually and cross-lingually (Kondratyuk and Straka, 2019; Samardžić et al., 2022). UD corpora (manually or automatically annotated) have been used for psycholinguistic and cross-linguistic corpus research, e.g., on word order (Futrell et al., 2015; Guzmán Naranjo and Becker, 2018; Levshina, 2019). But since UD only provides morphosyntactic relations, it is less well suited for similar corpus linguistic research on semantic phenomena, or for NLP tasks such as relation extraction and dialogue systems.

Given the advantages of semantic roles, one would think that SRL tools for many languages should be as readily available as UD parsers. However, this is not currently the case. None of the software libraries mentioned above comes with the capability to predict semantic roles. Freely available SRL systems are usually research prototypes and come with pre-trained models only for a single or few languages and domains (e.g., He et al., 2017; Strubell et al., 2018; Larionov et al., 2019). To a large part this gap is due to the lack of a language-independent annotation standard for semantic roles, comparable to the UD annotation guidelines for syntax. In Section 2, we review existing schemas for semantic role annotation and point out limitations that hamper their widespread adoption. We argue that bottlenecks stem mainly from either lack of frames or too many and fine-grained frames, and from restrictions of schemas to specific morphosyntactic word classes. We present a scheme that addresses these issues in Section 3, report on annotation experiments in Section 4, and conclude in Section 5.

2. Limitations of Existing SRL Annotation Schemes

Large frame inventories Most schemas define semantic roles via *frames*: A predicate is labeled as evoking a frame such as `break-01`, which defines frame-specific roles such as `ARG0` (breaker) and `ARG1` (thing broken). Some schemas have rather large and fine-grained inventories of frames. For example, PropBank has 10 687, SynSemClass has 1 993 (Urešová et al., 2022; Urešová et al., 2025), FrameNet has 1 224 (Baker et al., 1998), and VerbAtlas, 466 (Di Fabio et al., 2019). This creates a practical problem in annotation: for each predicate instance, annotators have to find the most appropriate frame, which is non-trivial, and fine-grained distinctions mean that the probability of disagreement is high. It may also be that an appropriate frame does not exist yet and has to be created. Frame lexicons mitigate this problem by creating explicit mappings from lexemes to frames. However, such lexicons are still forever incomplete, and they are language-specific. For a new language, a new lexicon has to be created, or annotators have to translate predicates, which introduces additional ambiguity and disagreement. There are ways to annotate non-English corpora with PropBank frames and roles automatically (Akbik et al., 2015; Jindal et al., 2022), but this also introduces errors.

Focus on specific morphosyntactic word classes In annotation, it is important to determine whether and when an annotation is complete. In UD, a syntactic annotation is complete when the dependency edges form a tree over the words of the sentence. A comprehensive semantic role annotation should label the semantic function of every argument and modifier edge between content words. Existing schemas, however, have a tendency to focus on a specific subset of natural-language predicates and their arguments and to some extent modifiers, typically by morphosyntactic word class. VerbNet (Kipper Schuler, 2005) and VerbAtlas (Di Fabio et al., 2019) focus on English verbs; so did PropBank, initially, although support for other words such as nouns and adjectives was later added (Pradhan et al., 2022). FrameNet focused on "nouns, adjectives, and verbs" (Baker et al., 1998) from the start. Modifier relations can be considered to evoke a frame on their own that is not necessarily evoked by a content word (although it is often associated with an adposition or conjunction). PropBank and AMR define different inventories of modifier relations that are separate from the inventory of frames for verbs and other frame-evoking content words. There are also schemas that specialize in other relations that sometimes coincide with modifier relations, such as adposition senses (Schneider et al., 2018), discourse relations (Carlson et al., 2003; Prasad et al., 2008), or compound relations (Tratz and Hovy, 2010; Pepper, 2022). FrameNet and SynSemClass comprehensively address arguments and modifiers of all content word classes, but there is no single *compact* set of frames that does this. As a result, treatment of predicates of different word classes and

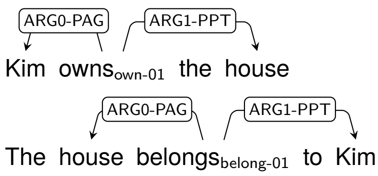
```
(o / own-01
   :ARG0 (k / Kim)
   :ARG1 (h / house))

(h / house
   :poss (k / Kim))
```

Figure 2: AMR annotations of the phrases *Kim owns the house* and *Kim's house*

especially between arguments and modifiers remains somewhat incoherent in many existing annotation frameworks. To illustrate this point, consider the phrases *Kim owns the house* and *Kim's house*. Both evoke a similar frame, namely a possessor-possessum relationship between Kim and a house. Now consider their AMR annotations in Figure 2. The similarity of the situations denoted is lost in this representation because the verb *own* and the possessive modifier draw from different inventories of labels.

Lack of precise definitions for roles Considering the problems with large numbers of frames discussed above, one might be tempted to do away with frames entirely and label each argument (and modifier) in isolation, according to a small set of role labels, each defined in some way. This approach is taken by schemas such as LIRICS (Petukhova and Bunt, 2008, used with some modifications in the PMB) or WISeR (Feng et al., 2023). In addition, several schemas do use some notion of frame, but attempt to give meaningful and frame-independent names to roles that are supposed to enable annotators to determine the correct role for each argument even without reference to the frames. Examples of this are VerbAtlas, the tectogrammatical level of the Prague Dependency Treebank (Hajič et al., 2024), and the "functions" decorating PropBank's frame-specific numbered argument roles. However, it is very hard to clearly define roles without reference to frames, because what makes different arguments of one predicate distinct from each other is their relation to each other. To illustrate this point, consider the sentences *Kim owns the house* and *The house belongs to Kim*. Both evoke a similar frame, namely a possessor-possessum relationship between Kim and a house. Now consider their PropBank annotations (showing both numbered argument roles and "functions") in (3).

(3)


In the first annotation, Kim is PAG (prototypical agent) and the house is PPT (prototypical patient) whereas in the second, it is the other way around. The similarity between the situations described is missed. Moreover, it is not clear how annotators could determine that these are the correct role labels without looking them up in the frame inventory, or in a language-independent way.

3. The Superframes Annotation Scheme

3.1. Design Principles

Superframes was developed to enable rapid, consistent annotation with semantic roles the way UD does for syntactic relations. To this end, Superframes is designed with the following key features:

(1) Small number of frames There are only a few dozen frames, shown in Table 1, all binary and indicating coarse semantic relations between two concrete or abstract entities. In annotation, predicates are sorted into these coarse semantic classes, similar to the sorting of word senses into coarse "supersenses" by Ciaramita and Altun (2006); Schneider et al. (2018). For annotation consistency, frames are kept as few and general as possible, but as many and specific as necessary to provide intuitive labels for each argument. For example, while the predicates *own* and *include* both have two arguments, the relation between both arguments in each case is different enough to warrant separate frames: POSSESSION with roles possessum and possessor, and PART-WHOLE with role part and whole, respectively.

(2) Frame composition Predicates with more than two arguments are annotated by composing two or more frames. For example, *include* with a causativ subject *include* evokes an additional CAUSATION frame where the core PART-WHOLE frame fills the result role.

(3) Aspect decomposition Frames denoting state changes are derived from frames denoting the corresponding states. For example, while *own* evokes POSSESSION, *buy* evokes POSSESSION-INIT.

(4) Unified frame inventory The same inventory of frames is used to annotate arguments of all classes of content words, and all modifiers.

(5) Constrained granularity Superframes is annotated atop Universal Dependencies. Content words receive frame labels, and the edges connecting them receive argument or modifier labels. In addition, secondary edges for nonlocal dependencies arising from coordination, control, raising,

SUPERFRAME	initial-arg2	arg1	arg2	transitory-arg2	target-arg2
SITUATION	initial-situator	theme	situator	transitory-situator	target-situator
└ ACCOMPANIMENT	initial-accompanier	accompanied	accompanier		target-accompanier
└ └ DEPICTIVE		has-depictive	depictive		
└ ASSET		has-asset	asset		
└ ATTRIBUTE		has-attribute	attribute		
└ COMPARISON		compared	reference		
└ └ CONCESSION		asserted	conceded		
└ EVENT		undergoer	event		
└ └ ACTIVITY		is-active	activity		
└ EXISTENCE	initial-exists	material	exists		target-exists
└ └ REPRODUCTION		original			copy
└ └ TRANSFORMATION-CREATION		material			created
└ EXPERIENCE		experiencer	experienced		
└ EXPLANATION		explained	explanation		
└ └ PURPOSE		has-purpoe	purpose		
└ IDENTIFICATION	initial-identifier	identified	identifier		target-identifier
└ LOCATION	initial-location	has-location	location	transitory-location	target-location
└ └ ADORNMENT-TARNISHMENT	initial-surface	ornament	surface		target-surface
└ └ EXCRETION	excreter	excreted		transitory-location	target-location
└ └ HITTING		hitting	hit		
└ └ INGESTION		ingested		transitory-location	ingester
└ └ UNANCHORED-MOTION		in-motion		transitory-location	
└ └ WRAPPING-WEARING	initial-wearer	worn	wearer		target-wearer
└ MEANS		has-means	means		
└ MESSAGE	initial-content	topic	content		target-content
└ MODE		has-mode	mode		
└ NONCOMP		has-noncomp	noncomp		
└ PART-WHOLE	initial-whole	part	whole		target-whole
└ POSSESSION	initial-possessor	possessed	possessor		target-possessor
└ QUANTITY	initial-quantity	has-quantity	quantity		target-quantity
└ RANK	initial-rank	has-rank	rank		target-rank
└ SCENE	initial-scene	participant	scene	transitory-scene	target-scene
└ STATE	initial-state	has-state	state		target-state
└ └ QUALITY	initial-quality	has-quality	quality		target-quality
└ └ └ CLASS	initial-class	has-class	class		target-class
└ └ DESTRUCTION		destroyed			
└ SENDING		sent	sender		
└ SEQUENCE		follows	followed		
└ └ CAUSATION		result	causer		
└ └ CONDITION		has-condition	condition		
└ └ EXCEPTION		has-exception	exception		
└ └ REACTION		reaction	trigger		
└ └ RESULTATIVE		has-resultative	resultative		
└ SOCIAL-RELATION	initial-social-relation	has-social-relation	social-relation		target-social-relation
└ TIME	initial-time	has-time	time		target-time

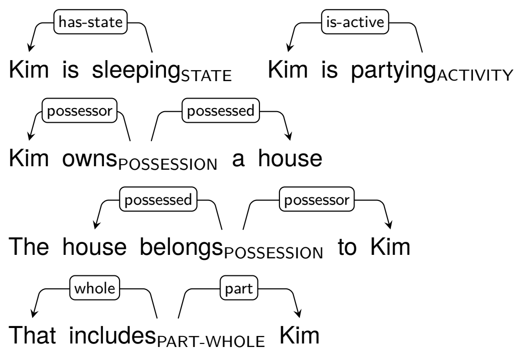
Table 1: Hierarchy of superframes and their roles

secondary predicates, etc., are added and labeled. No additional nodes are introduced.

3.2. Overview of the Annotation Guidelines

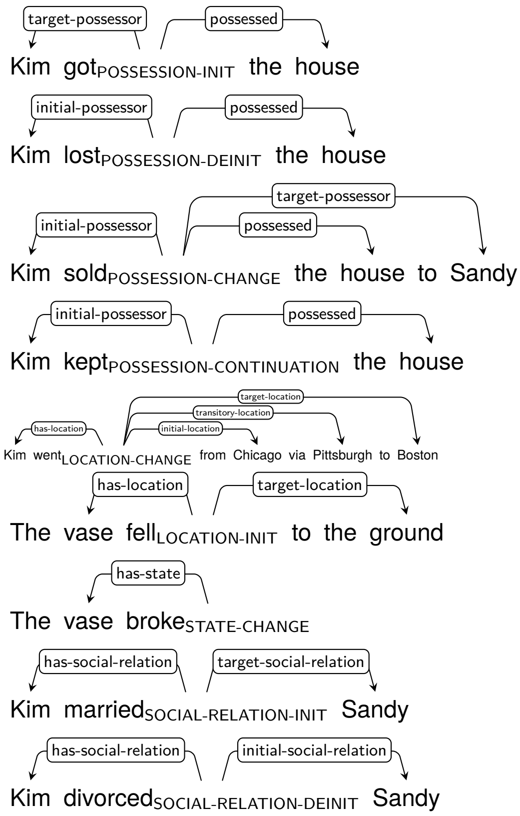
Superframes is annotated atop UD. Content words (tagged NOUN, VERB, ADJ, or ADJ in UD) receive a frame label, and dependency arcs between content words (labeled nsubj, dobj, nmod, etc.) receive a role label.

(4)



Stative verbal predicates An example of the annotation of some unary and binary verbal stative predicates are shown in (4). Role labels have to be drawn from the frame that the predicate is annotated with.

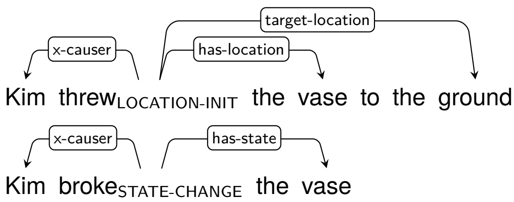
Aspect Predicates denoting events that can be framed in terms of an actual or hypothetical change of state are framed using one of the aspect operators -INIT, -DEINIT, -CHANGE, or -CONTINUATION. The second argument accordingly receives a role label prefixed with initial-, target-, or transitory-. Examples are shown in (5).

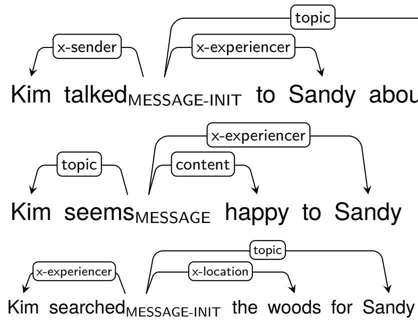
(5)


Mode In addition to aspect operators, we also use the modal operators -NECESSITY, -POSSIBILITY, and -NEG. An example are shown in (6).

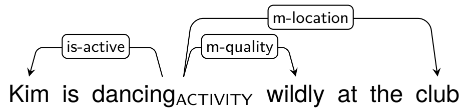
(6)


Frame composition Predicates can be annotated as evoking more than one frame, with a limited form of embedding. For example, a causative subject is considered to evoke an additional CAUSATION frame and fill its causer role while the core frame fills the result role. To avoid introducing an additional CAUSATION node into the annotation format, a special notation indicates this configuration, annotating the causative subject as x-causer. Analogously, arguments such as senders and perceivers are annotated with x-sender and x-experiencer, respectively. This frame composition mechanism is also used for predicate-specific arguments that do not fit into onto one of the core argument slots. Examples are shown in (7).

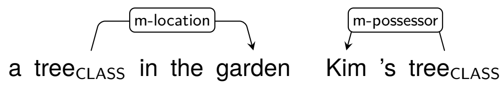
(7)




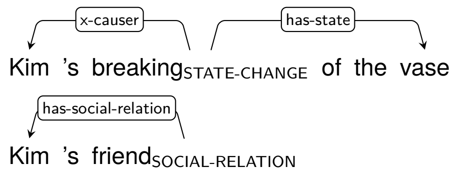
Modification Like non-core arguments, modifiers are assumed to evoke an additional frame and labeled with the role they fill in that frame, but with the prefix m- marking them as modifiers. Examples are shown in (8).

(8)


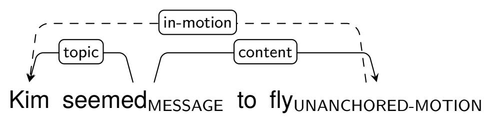
Nonverbal predicates We have so far shown only examples of *verbal* predicates, but everything applies to the annotation of nonverbal predicates as well. An ordinary noun like *tree* evokes the CLASS frame, marking the entity it refers to as a member of a class (in this case: the class of trees). There are no arguments here because the predicate itself doubles as a referent. However, it can be modified, as shown in (9).

(9)


Event nouns and relational nouns evoke frames and have arguments just like verbs, as shown in (10).

(10)


Nonlocal dependencies Many constructions introduce semantic predicate-dependent dependencies that do not correspond to (surface) syntactic dependencies. This includes constructions like control, raising, relative clauses, secondary predicates, coordination, etc. We add such links in the annotation beyond those provided by UD. Examples are shown in (11) with nonlocal dependencies dashed.

(11)


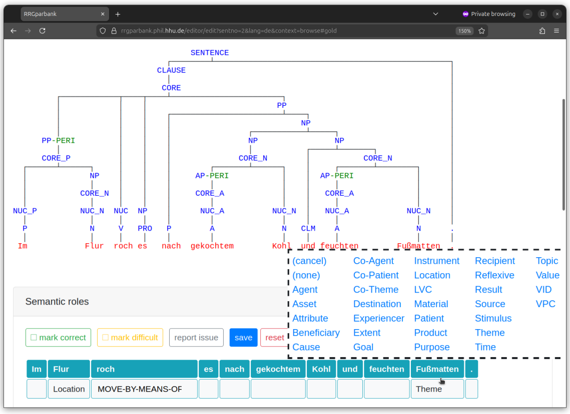


Figure 3: Browser-based VerbAtlas annotation interface showing the German translation of the sentence “The hallway smelt of boiled cabbage and old rag mats”

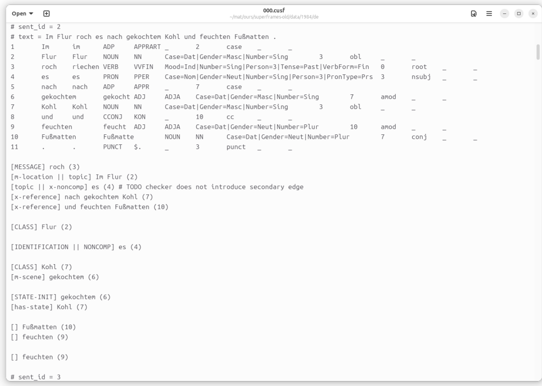
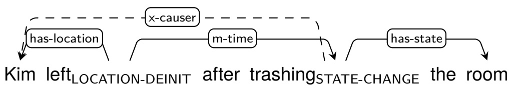
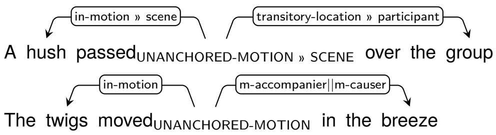


Figure 4: Text-based Superframes annotation interface. There is a UD syntax tree and 7 predicates, 5 of which have already been annotated with frames and roles.



Figurativity and uncertainty Figurative use of predicates is a pervasive phenomenon and often makes it hard to choose a single frame for a predicate because a more literal meaning and a more idiomatic reading are both salient. In such cases, annotators are instructed to provide annotations for both readings using labels of the form A » B where A is for the more literal reading and B is for the more figurative reading. When it is hard to choose a predicate for reasons other than figurativity, annotators can likewise provide two alternatives using the A || B notation. Examples are shown in (12).

(12)



4. Experiments

To demonstrate the viability of Superframes as an annotation scheme, we report the results of two pilot annotation projects, both carried out on a German translation of George Orwell’s novel *1984* (Orwell, 2003), one using the VerbAtlas annotation scheme and one using Superframes. Both annotation projects were carried out by paid student assistants who were either currently pursuing or had recently completed a bachelor’s degree in computational linguistics, and who were native or advanced speakers of German. Annotators worked independently but came together for weekly annotation meetings with the author, discussing difficult cases and producing consensus annotations for them.

VerbAtlas VerbAtlas (Di Fabio et al., 2019) is a frame inventory with full explicit coverage of the verb senses in WordNet 3.0 (Fellbaum, 1998). It contains 466 frames created by semiautomatically clustering WordNet synsets and 25 semantic role labels that are shared between frames, such as Agent, Patient, Source, Goal. Annotation was carried out between July 2021 and March 2024. Annotators were provided an annotation manual (Evang, 2024) and a graphical annotation interface that showed them syntactic trees for the sentences they had to annotate (taken from RRGparbank; Bladier et al., 2022) as well as preliminary annotations generated using InVeRo (Conia et al., 2020, 2021), which they had to correct (see Figure 3 for an example). Annotators were instructed to find the most approprate VerbAtlas frame for each verbal predicate, considering mainly meaning correctness (the frame should map to at least one English WordNet synset whose gloss fits the use of the predicate in this context) and role coverage (the frame should have an appropriate role for every argument present in the instance). They were also instructed to make sure the role labels matched their use in the chosen frame. They were given a searchable document containing for each VerbAtlas frame 1) its VerbNet synsets with lemmas, glosses, and example sentence 2) its PropBank frames with role descriptions and annotated example sentences. No modifiers ore nonverbal predicaets were annotated. The same annotators annotated the corresponding parallel English sentences in parallel. Statistics of the resulting annotations are shown in Table 2.

Superframes The annotation was carried out between April 2024 and March 2025. Annotators were provided an annotation manual (reference anonymized) and text files that contained UD anno-

Annotator(s)		Sentences	Predicates	Arguments
jh		2 192	4 784	13 475
sg		2 405	5 122	14 799
sk		2 175	4 606	12 445
jh+sg	Both annotated	893	1 863	4 669
	Agreement		.74	.75
jh+sk	Both annotated	566	1 113	2 466
	Agreement		.76	.80
sg+sk	Both annotated	765	1 604	3 548
	Agreement		.71	.81

Table 2: VerbAtlas annotation results. The lower part excludes sentences with consensus annotations.

Annotator(s)		Sentences	Predicates	Dependents
ab		350	4 168	4 219
sg		296	2 955	2 431
xs		300	3 213	3 404
ab+sg	Both annotated	147	476	308
	Agreement (strict)		.71	.63
	Agreement (lax)		.77	.68
	Agreement (superlax)		.81	.73
ab+xs	Both annotated	150	655	609
	Agreement (strict)		.78	.73
	Agreement (lax)		.84	.81
	Agreement (superlax)		.89	.86
sg+xs	Both annotated	99	885	869
	Agreement (strict)		.85	.79
	Agreement (lax)		.92	.84
	Agreement (superlax)		.93	.86

Table 3: Superframes annotation results. The lower part excludes sentences with consensus annotations.

tations of the sentences to annotate as well as blank templates for filling in frame and role labels (see Figure 4 for an example). The templates were generated automatically from the UD annotation. Annotators were provided a Python script with which they checked their annotations for validity (all blanks filled out, all frame and role labels defined, all role labels appropriate for the respective frame, etc.) before committing them to a shared Git repository. Statistics of the resulting annotations are shown in Table 3. Strict agreement is defined as both annotators choosing the exact same label for a predicate or dependency edge. Lax agreement is also satisified if they agree in at least one sub-label when using the uncertainty/figurativity mechanism. Superlax agreement is like lax agreement but ignores all prefixes or suffixes in labels such as -INIT, `m-`, or `target-`, focusing only on the main frame or role label.

Close comparison 69 sentences were doubly annotated with no consensus annotation under both schemas. We examine this intersection more closely and report overall agreement between the respective two annotators in Table 4.

Results Looking at the close comparison in Table 4, we see that with strict agreement, Superframes scores lower than VerbAtlas for frame labels and role labels. Agreement becomes higher than for VerbAtlas with lax agreement for frames and with superlax agreement for roles. When we look at the bigger but less comparable dataset in Tables 2 and 3, we see that agreement on frame labels from .71 to .76 for VerbAtlas (depending on the pair of annotators), but from .71 to .85 for Superframes even under strict agreement. On the other hand, agreement on role labels ranges from .75 to .81 for VerbAtlas but only from .63 to .79 for Superframes unders strict agreement, but from .73 to .86 under superlax agreement.

We take these numbers to hint at general viability of Superframes, but of course cannot draw very strong conclusions from them. The Superframes dataset is much smaller and the intersection of both datasets is smaller still. Annotations were made by (partially) different people and under different conditions. The numer of Superframes used is much lower than that of VerbAtlas frames whereas the Superframes roles are more diverse due to VerbAtlas's reuse of role labels across different frames. VerbAtlas focuses on verbs and their arguments whereas all content words and their relations are annotated in Superframes. Perhaps most importantly, VerbAtlas agreement profits from the anchoring bias due to pre-annotations whereas Superframes was annotated from scratch.

We provide additional statistics in Tables 5, 6, 7, 8, 9, 10, 11, and 12.

	Sentences	Predicates	Dependents
All annotated	69	98	190
VerbAtlas agreement		.67	.79
Superframes agreement (strict)		.62	.69
Superframes agreement (lax)		.71	.77
Superframes agreement (superlax)		.77	.81

Table 4: Average agreement between two annotators using VerbAtlas vs. Superframes on the same sentences. Sentences with consensus annotations are excluded.

Frame	Count
IDENTIFICATION	1946
CLASS	1060
MESSAGE	973
QUALITY	778
TIME	565
MODE	538
QUANTITY	407
PART-WHOLE	353
MESSAGE-INIT	322
LOCATION	258

Table 5: Most common frame labels in the Superframes annotation

Role	Count
m-scene	725
m-quality	642
m-time	523
m-mode	500
topic	476
x-experiencer	430
m-quantity	376
has-location	344
x-sender	310
participant	261

Table 6: Most common role labels in the Superframes annotation

Pair		Count
MESSAGE	MESSAGE-INIT	45
CLASS	IDENTIFICATION	40
CLASS	PART-WHOLE	34
IDENTIFICATION	NONCOMP	28
QUALITY	STATE	26

Table 7: Most common frame label disagreement pairs in the Superframes annotation

Pair		Count
m-quality	m-scene	35
m-scene	m-time	16
m-noncomp	x-noncomp	16
m-quality	m-state	14
target-content	topic	13

Table 8: Most common role label disagreement pairs in the Superframes annotation

Label	Count
QUALITY » NONCOMP	41
SOCIAL-RELATION » IDENTIFICATION	40
QUALITY-NEG » QUALITY	23
COMPARISON » QUALITY	15
RANK » TIME	15

Table 9: Most common figurative frame labels in the Superframes annotation

Label	Count
SOCIAL-RELATION || STATE	8
CLASS || LOCATION	7
STATE || QUALITY	5
STATE || SOCIAL-RELATION	5
MESSAGE || SEQUENCE	4

Table 10: Most common uncertain frame labels in the Superframes annotation

VerbAtlas frame	Superframe	Count
KNOW	MESSAGE	86
SPEAK	MESSAGE-INIT	77
SEEM	MESSAGE	75
SEE	MESSAGE	73
EXIST-WITH-FEATURE	SCENE	47
WRITE	MESSAGE-INIT	38
EXIST_LIVE	EXISTENCE	37
LIE	LOCATION	33
GO-FORWARD	LOCATION-INIT	32
EXIST_LIVE	SCENE	32

Table 11: Most common combinations of VerbAtlas and Superframes frames

VerbAtlas role	Superframes role	Count
Time	m-time	392
Theme	has-location	287
Agent	x-causer	257
Experiencer	x-experiencer	241
Agent	x-sender	227
Theme	topic	224
Agent	x-experiencer	145
Destination	target-location	139
Stimulus	topic	121
Theme	content	115

Table 12: Most common combinations of VerbAtlas and Superframes roles

5. Conclusions, Limitations, and Future Work

We have presented Superframes, a frame-semantic annotation scheme that aims to enable consistent annotation of predicates with frame labels and dependents with role labels across languages, achieved through a small number of frames, frame composition, aspect decomposition, and annotation atop Universal Dependencies. Non-local dependencies, figurativity, idiomaticity, and uncertainty are addressed through dedicated mechanisms. We have presented the results of a preliminary annotation study where Superframes is compared with a traditional annotation scheme with a larger number of frames. We were able to show higher agreement for Superframes under certain conditions, although it should be noted that the comparability is limited, especially due to the fact that annotators corrected pre-annotations for VerbAtlas but annotated from scratch for Superframes. Based on the insights from annotation discussion, we are currently in the process of revising the scheme and making the inventory of superframes more compact and more systematic, as well as preparing a larger-scale and more systematic annotation effort.

6. Acknowledgements

The author would like to thank the anonymous reviewer for important pointers. This work was funded by grant no. 560341082 (Superframes) of the German Research Foundation (DFG).

7. Bibliographical References

Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen, and Johan Bos. 2017. The Parallel Meaning Bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 242–247, Valencia, Spain. Association for Computational Linguistics.

Alan Akbik, Laura Chiticariu, Marina Danilevsky, Yunyao Li, Shivakumar Vaithyanathan, and Huaiyu Zhu. 2015. Generating high quality proposition Banks for multilingual semantic role labeling. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 397–407, Beijing, China. Association for Computational Linguistics.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Tatiana Bladier, Kilian Evang, Valeria Generalova, Zahra Ghane, Laura Kallmeyer, Robin Möllemann, Natalia Moors, Rainer Osswald, and Simon Petitjean. 2022. RRGparbank: A parallel role and reference grammar treebank. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4833–4841, Marseille, France. European Language Resources Association.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2003. Building a discourse-tagged corpus in the framework of rhetorical structure theory. pages 85–112.

Cemil Cengiz and Deniz Yuret. 2020. Joint training with semantic role labeling for better generalization in natural language inference. In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 78–88, Online. Association for Computational Linguistics.

Janara Christensen, Mausam, Stephen Soderland, and Oren Etzioni. 2010. Semantic role labeling for open information extraction. In *Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*, pages 52–60, Los Angeles, California. Association for Computational Linguistics.

Massimiliano Ciaramita and Yasemin Altun. 2006. Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 594–602, Sydney, Australia. Association for Computational Linguistics.

Simone Conia, Fabrizio Brignone, Davide Zanfardino, and Roberto Navigli. 2020. InVeRo: Making semantic role labeling accessible with intelligible verbs and roles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 77–84, Online. Association for Computational Linguistics.

Simone Conia, Riccardo Orlando, Fabrizio Brignone, Francesco Cecconi, and Roberto Navigli. 2021. InVeRo-XL: Making cross-lingual Semantic Role Labeling accessible with intelligible verbs and roles. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 319–328, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Andrea Di Fabio, Simone Conia, and Roberto Navigli. 2019. VerbAtlas: a novel large-scale verbal semantic resource and its application to semantic role labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 627–637, Hong Kong, China. Association for Computational Linguistics.

Kilian Evang. 2024. Superframes manual. Technical report, Heinrich Heine University Düsseldorf. `https://github.com/texttheater/superframes/blob/main/doc/manual/manual.pdf`, retrieved 2024-11-05.

Christiane Fellbaum, editor. 1998. *WordNet: An Eletronic Lexical Database*. MIT Press, Cambridge, MA.

Lydia Feng, Gregor Williamson, Han He, and Jinho D. Choi. 2023. Widely interpretable semantic representation: Frameless meaning representation for broader applicability. *ArXiv*, abs/2309.06460.

Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. 112:10336–10341.

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

Jiang Guo, Wanxiang Che, Haifeng Wang, Ting Liu, and Jun Xu. 2016. A unified architecture for semantic role labeling and relation classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1264–1274, Osaka, Japan. The COLING 2016 Organizing Committee.

Matías Guzmán Naranjo and Laura Becker. 2018. Quantitative word order typology with UD. In *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018)*, pages 91–104.

Jan Hajič, Eduard Bejček, Alevtina Bémová, Eva Buráňová, Eva Fučíková, Eva Hajičová, Jiří Havelka, Jaroslava Hlaváčová, Petr Homola, Pavel Ircing, Jiří Kárník, Václava Kettnerová, Natalia Klyueva, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, David Mareček, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Michal Novák, Petr Pajas, Jarmila Panevová, Nino Peterek, Lucie Poláková, Martin Popel, Jan Popelka, Jan Romportl, Magdaléna Rysová, Jiří Semecký, Petr Sgall, Johanka Spoustová, Milan Straka, Pavel Straňák, Pavlína Synková, Magda Ševčíková, Jana Šindlerová, Jan Štěpánek, Barbora Štěpánková, Josef Toman, Zdeňka Urešová, Barbora Vidová Hladká, Daniel Zeman, Šárka Zikánová, and Zdeněk Žabokrtský. 2024. Prague dependency treebank - consolidated 2.0 (PDT-c 2.0). LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL).

Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. Deep semantic role labeling: What works and what's next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 473–483, Vancouver, Canada. Association for Computational Linguistics.

Llio Humphreys, Guido Boella, Luigi Di Caro, Livio Robaldo, Leon van der Torre, Sepideh Ghanavati, and Robert Muthuri. 2020. Populating legal ontologies using semantic role labeling. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2157–2166, Marseille, France. European Language Resources Association.

Ishan Jindal, Alexandre Rademaker, Michał Ulewicz, Ha Linh, Huyen Nguyen, Khoi-Nguyen Tran, Huaiyu Zhu, and Yunyao Li. 2022. Universal Proposition Bank 2.0. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1700–1711, Marseille, France. European Language Resources Association.

Karin Kipper Schuler. 2005. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.

Dan Kondratyuk and Milan Straka. 2019. 75 languages, 1 model: Parsing Universal Dependencies universally. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2779–2795, Hong Kong, China. Association for Computational Linguistics.

Daniil Larionov, Artem Shelmanov, Elena Chistova, and Ivan Smirnov. 2019. Semantic role labeling with pretrained language models for known and unknown predicates. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 619–628, Varna, Bulgaria. INCOMA Ltd.

Natalia Levshina. 2019. Token-based typology and word order entropy: A study based on universal dependencies. 23:533–572.

Ines Montani, Matthew Honnibal, Matthew Honnibal, Adriane Boyd, Sofie Van Landeghem, Henning Peters, Paul O'Leary McCann, Jim Geovedi, and others. 2023. explosion/spacy: v3.7.1: Bug fix for ‘spacy.cli‘ module loading.

George Orwell. 2003. *1984*, 37th edition. Ullstein. German translation by Kurt Wagenseil (first published 1950 by Alfons Bürger Verlag).

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Steve Pepper. 2022. Hatcher-Bourque: Towards a reusable classification of semantic relations. In *Binominal Lexemes in Cross-Linguistic Perspective*, pages 303–354. De Gruyter Mouton.

Volha Petukhova and Harry Bunt. 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 (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Sameer Pradhan, Julia Bonn, Skatje Myers, Kathryn Conger, Tim O'gorman, James Gung, Kristin Wright-bettner, and Martha Palmer. 2022. PropBank comes of Age—Larger, smarter, and more diverse. In *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics*, pages 278–288, Seattle, Washington. Association for Computational Linguistics.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

Valdi Rachman, Rahmad Mahendra, Alfan Farizki Wicaksono, Ahmad Rizqi Meydiarso, and Fariz Ikhwantri. 2018. Semantic role labeling in conversational chat using deep bi-directional long short-term memory networks with attention mechanism. In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, Hong Kong. Association for Computational Linguistics.

Tanja Samardžić, Ximena Gutierrez-Vasques, Rob van der Goot, Max Müller-Eberstein, Olga Pelloni, and Barbara Plank. 2022. On language spaces, scales and cross-lingual transfer of UD parsers. In *Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL)*, pages 266–281, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Nathan Schneider, Jena D. Hwang, Vivek Srikumar, Jakob Prange, Austin Blodgett, Sarah R. Moeller, Aviram Stern, Adi Bitan, and Omri Abend. 2018. Comprehensive supersense disambiguation of English prepositions and possessives. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 185–196, Melbourne, Australia. Association for Computational Linguistics.

Milan Straka, Jan Hajič, and Jana Straková. 2016. UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4290–4297, Portorož, Slovenia. European Language Resources Association (ELRA).

Emma Strubell, Patrick Verga, Daniel Andor, David Weiss, and Andrew McCallum. 2018. Linguistically-informed self-attention for semantic role labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5027–5038, Brussels, Belgium. Association for Computational Linguistics.

Stephen Tratz and Eduard Hovy. 2010. A taxonomy, dataset, and classifier for automatic noun compound interpretation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 678–687, Uppsala, Sweden. Association for Computational Linguistics.

Richard Tzong-Han Tsai, Wen-Chi Chou, Yu-Chun Lin, Cheng-Lung Sung, Wei Ku, Ying-Shan Su, Ting-Yi Sung, and Wen-Lian Hsu. 2006. BIOSMILE: Adapting semantic role labeling for biomedical verbs:. In *Proceedings of the HLT-NAACL BioNLP Workshop on Linking Natural Language and Biology*, pages 57–64, New York, New York. Association for Computational Linguistics.

Zdeňka Urešová, Cristina Fernández Alcaina, Peter Bourgonje, Eva Fučíková, Jan Hajič, Eva Hajičová, Veronika Kolářová, Georg Rehm, Kateřina Rysová, and Karolina Zaczynska. 2025. SynSemClass 5.5. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL).

Zdeňka Urešová, Karolina Zaczynska, Peter Bourgonje, Eva Fučíková, Georg Rehm, and Jan Hajic. 2022. Making a semantic event-type ontology multilingual. In *Proceedings of the Thir-*

teenth Language Resources and Evaluation Conference, pages 1332–1343, Marseille, France. European Language Resources Association.

Shijie Wu, Benjamin Van Durme, and Mark Dredze. 2022. Zero-shot cross-lingual transfer is under-specified optimization. In *Proceedings of the 7th Workshop on Representation Learning for NLP*, pages 236–248, Dublin, Ireland. Association for Computational Linguistics.

Kun Xu, Haochen Tan, Linfeng Song, Han Wu, Haisong Zhang, Linqi Song, and Dong Yu. 2020. Semantic Role Labeling Guided Multi-turn Dialogue ReWriter. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6632–6639, Online. Association for Computational Linguistics.

Bo Zhang, Yue Zhang, Rui Wang, Zhenghua Li, and Min Zhang. 2020. Syntax-aware opinion role labeling with dependency graph convolutional networks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3249–3258, Online. Association for Computational Linguistics.

Zhisong Zhang, Emma Strubell, and Eduard Hovy. 2022. Transfer learning from semantic role labeling to event argument extraction with template-based slot querying. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2627–2647, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.