Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 3942–3949
Marseille, 20-25 June 2022

A Framenet and Frame Annotator for German Social Media

Eckhard Bick
Institute of Language and Communication, University of Southern Denmark
Campusvej 55, 5230 Odense M, Denmark
eckhard.bick@mail.dk

Abstract

This paper presents PFN-DE, a new, parsing- and annotation-oriented framenet for German, with almost 15,000 frames, covering 11,300 verb lemmas. The resource was developed in the context of a Danish/German social-media study on hate speech and has a strong focus on coverage, robustness and cross-language comparability. A simple annotation scheme for argument roles meshes directly with the output of a syntactic parser, facilitating frame disambiguation through slot-filler conditions based on valency, syntactic function and semantic noun class. We discuss design principles for the framenet and the frame tagger using it, and present statistics for frame and role distribution at both the lexicon (type) and corpus (token) levels. In an evaluation run on Twitter data, the parser-based frame annotator achieved an overall F-score for frame senses of 93.6%.

Keywords: German Framenet, Automatic frame annotation, Semantic roles, Constraint Grammar

1. Introduction

In corpus linguistics, semantic annotation is a valuable addition to ordinary, morphosyntactic tagging, lemmatization and dependency relations. This is true in particular for corpus research transcending lexicographical issues, language variation and usage statistics (in our case, German/Danish hate speech). Still, with the possible exception of named-entity recognition (NER), automatic semantic annotation tools are still difficult to find for most languages. Ontologies like WordNet (Fellbaum 1998) have a good coverage and effective classification principle for nouns (hyponymy), but usually do not provide the structural-relational information necessary to disambiguate senses/synsets. And though troponymy goes part of the way, semantically bundling and annotating verb senses is even harder, because verb semantics is closely related to clause structure and argument relations. Thus, for English, VerbNet (Kipper et al., 2006), Berkeley FrameNet (Baker et al., 1998, Johnson & Fillmore, 2000, Ruppenhofer et al., 2010) and PropBank[1] (Palmer et al., 2005) address entire predications or frames, assigning a semantic/frame class to the core lexeme (typically but not necessarily a verb) and semantic roles (also called case/thematic roles, Fillmore 1968) to its arguments and possibly adjuncts. A prototypical frame consists of a full verb and its nominal, adverbial or subclause complements. However, for a comprehensive semantic annotation, a frame annotator must also address noun- or adjective-based frames and semantically transparent or verb-incorporated material.

In his comparison of the WordNet and FrameNet approaches, Boas (2005) stresses the added level of abstraction provided by the latter (e.g. cross-POS frames), as well as the systematic link between semantic information and lexico-syntactic patterns. Thus, both FrameNet and PropBank provide morphosyntactic restrictions, and the former in addition adds ontological information on slot fillers. As will be discussed in Sections 3 and 4, such a link between frame structures and the underlying morphosyntax is paramount for robust automatic frame annotation.

Historically, the link between framenet creation and automatic frame annotation, as well as a focus on corpora, have become more prominent over time. For German, the SALSA project[2] (Burchardt et al., 2006) used and appended the English Berkeley FrameNet in a corpus-driven fashion, using over 600 different frames, more than half newly added for German. In an interesting cross-resource approach, Burchardt et al. (2009) developed a heuristic frame assignment system based on WordNet synsets, in order to address lexical FrameNet gaps and to improve live annotation with their SHALMANESER parser. Still, the authors note that SALSA is not a lexicographical project. Therefore, heuristics aside, SALSA does not claim to systematically cover the German lexicon, and coverage is limited to what is found in the hand-annotated corpus.

The new German "parser framenet" (PFN-DE) described here takes the opposite, unabridged lexical approach, aiming to trade depth for breadth with a simple, parsing-oriented scheme meant to support robust automatic annotation. The resource is being developed in the context of a bilingual[3] social-media hate speech project[4] (XPEROHS, Baumgarten et al., 2019) examining Danish and German data harvested from Twitter and Facebook. In order to support semantically informed corpus searches in its 2 billion token corpus, the project uses the Constraint Grammar (CG) annotation framework (Bick & Didriksen, 2015), where each token is assigned morphosyntactic, semantic and relational tags through lexically and contextually informed manual rules. The available CG parsers for Danish and German, DanGram[5] and

[1] Methodologically, VerbNet and the English Berkeley FrameNet were conceived as lexicographical projects, one frame at a time, while PropBank and the German SALSA framenet depart from corpus data, one sentence at a time. Coverage problems are therefore different in nature: In the former, a common sense may be missing in a verb with several rare senses assigned. In the latter, common senses are registered first, but rare lemmas may be missing entirely.

[2] http://www.coli.uni-saarland.de/projects/salsa/page.php?id=overview (last publication 2010)

[3] Bilingual in the sense that the corpus contains comparable data for the two languages, but not mixed or code-switching data.

[4] The project has received support from the Velux Foundation.

[5] https://visl.sdu.dk/visl/da/parsing/automatic/

GerGram[6], use a unified tagset (VISL) for POS, morphology, syntax and semantic ontologies, but only the Danish parser offered frame and role annotation. What was needed to fill this gap, was a German framenet scheme honouring the following conditions:

1. high lexicon coverage, including rare lexemes, special focus on social media usage
2. syntactic/semantic slot-filler information for roles/arguments compatible with existing lower-level corpus annotation
3. same category set for Danish/German, both in terms of the supporting morphosyntactic annotation and the actual frame and role tag sets

Since no existing resources fulfilled all three conditions, we decided to achieve notational compatibility by adopting the existing Danish frame templates (Bick, 2011) already used in XPEROHS, while adding German-specific features such as case and separable verbs.

In a first round of bootstrapping, we identified Danish-German verb sense matches by harvesting the machine-translation (MT) dictionary used in the Danish-German version of the GramTrans[7] MT system (Bick, 2007), where polysemy is handled by providing syntactic argument and semantic slot-filler information in much the same way a frame entry would. Rather than using MT to match existing framenets in two languages (Gilardi & Baker, 2018), we use it to match valency patterns as an anchor for frame transfer. For instance, if a German verb is allowed four different translations depending on the semantics of its subject and object, we would look up the translations in the Danish FrameNet and choose a frame with the same slot-filler conditions. Similarly, prepositional complements and idioms with incorporated material could be harvested and heuristically matched to a Danish frame. All in all, the method came up with frame suggestions and slot-filler semantics for over 6000 German verb lemmas. In a second round of manual revision, all these frame candidates were checked and, if necessary, corrected. In this process, the valency patterns listed in GerGram's parser lexicon were used to check semi-automatically for missing frames in existing entries, and as skeletons for completely new manual entries for those verbs without a match in the MT dictionary. The third round of frame lexicography was corpus-driven, based on a preliminary annotation (cf. Section 4) of the XPEROHS social media corpus. Verbs with no frames were added in the order of token frequency, and a number of frequent construction verbs systematically checked for idiomatic senses/constructions.

2. Lexicon size and granularity

The verb part of PFN-DE contains 11,333 verb lemmas, with manually revised or assigned frames. All in all, there are 14,695 different lemma+frame combinations. On average this amounts to 1.297 frames per lemma, with a Zipfian distribution, where some frequent construction verbs and polysemic verbs needed dozens of frames, while the vast majority of rarer verbs only received one single frame each. At the level of semantic frame class, i.e. not counting differences in role types and slot-filler information, the average was 1.237 semantic types per verb lemma. With regard to condition (1) – coverage -, it is important to stress that virtually all verb lemmas in the parser lexicon (as well as all valency-marked nouns and adjectives) were assigned at least one frame and that corpus evaluation (Section 5) indicates a raw lexical failure rate as low as 1-2%[8]. By comparison, the above-mentioned SALSA resource, albeit more refined, more revised and less "bootstrapped", only contains about 1,000 unique lemma-frame types (Rehbein et al. 2014).

Out of the 494 different verb frame categories[9] available in the Danish framenet scheme, almost all (483) also ended up being used for German. In addition, we introduced about 1,700 different combinations of these "atomic" frames, in order to capture additional lexical information such as aspect/aktionsart, directionality or urgency. Often, but not always, the secondary "aspect" frame is triggered by a productive German prefix. For instance, 'continue', 'start' and 'stop' are all frames in their own right, but are used as secondary frames too:

weiter*laufen* (run on) – fn[10]:run&continue
los*laufen* (start running) – fn:run&start
ver*glimmen* (stop burning) – fn:burn&stop

Other examples of prefix-triggered secondary frames are *ent-* (&rid), *zer-* (&split) and *über-* (&exaggerate). Since the actual frame patterns in the lexicon may have the same semantic frame class, but different syntactic realizations and transitivity, optional roles etc., the number of distinct role/complement-specified "syntactic" frames is higher – 7,316 – and even higher (10,460), when also counting differences in semantic slot-fillers.

In addition to verbal predicates, we also handle nominal predicates, albeit at a more experimental (and unevaluated) level. Thus, about 1,400 nouns and 400 adjectives were assigned frames, mostly based on prepositional valency. Also, morphological analysis of productive nominalization is used to systematically transfer frames from verbs to deverbal nouns on the fly. For instance, *Erkrank**ung*** (falling ill) is derived from the verb *erkranken* and inherits the 'sick' frame, including information about the preposition trigger (*Erkrankung* ***an***) for its argument role, §CAU (cause). Similarly, participle adjectives can be referred to a corresponding verb base.

[6] https://visl.sdu.dk/visl/de/parsing/automatic/

[7] https://gramtrans.com (QuickDict)

[8] Lexical coverage is an issue even for English. Thus, Palmer & Sporleder (2010), comparing SemEval data with Framenet data, found that 3.4% of lexical units and 12.1% of frames from the former were not found in the latter. In terms of training data, the gaps were even more pronounced, 6.9% missing senses and 26.0% missing verbs.

[9] A smaller set of 200 frame senses is also available, with a hypernym-mapping from the more fine-grained set. This smaller set was meant to facilitate cross-language comparisons and parser training.

[10] In this article, we use lower-case 'fn:...' for framenet category names, and upper-case 'FN:...' for fully specified frame "recipies" where argument, roles and slot-filler restrictions are included.

3. Lexical support for the frame annotator: syntactic and semantic slot restrictions

In order to help the frame annotator assign frames to syntactic structures in running text, frame senses are linked to the valency patterns associated with each verb, for instance <vt> (monotransitive), <vdt> (ditransitive) or <für^vrp> (reflexive verb with a prepositional argument headed by the preposition *für*). For a given lemma, each of these valency patterns is assigned at least one, and possibly more[11], verb senses, corresponding to different semantic frames. While some senses exhibit syntactic variation and may manifest with e.g. different argument arities (number of obligatory arguments), it is a reliable rule-of-thumb that two different verb senses will almost always differ with regard to one or more syntactic or semantic slot fillers. Therefore, the frame annotator will in most cases be able to disambiguate word senses (frame senses) solely based on distinctions provided by the morphosyntactic parsing stage (lexical tags, syntactic function and dependency arcs).

Though the frame inventory and role granularity of our German framenet is modeled on the Danish framenet, we decided to make an important change in notational conventions, extending the shorthand system suggested by Bick (2017) for noun frames to cover the main, verbal lexicon too. Thus, for each of the almost 15,000 verb sense frames, a list of arguments is provided in a single, composite tag ready to be used by CG rules. The verb *bestehen*, for instance, has five different meanings, each with its only valency and frame patterns (<FN:...>).

1. meaning: *'pass' (an exam)*
   valency: accusative-monotransitive
   <FN:succeed/S§AG'H/O§TH'occ>
2. meaning: *'consist of'* (PP-monotransitive)
   <FN:consist/S§HOL'cc/P-aus§PART'cc|H>
3. meaning: *'insist on'* (PP-monotransitive)
   <FN:demand/S§SP'H/P-auf§TH'cc|act>
4. meaning: *'be'* (PP-monotransitive)
   <FN:be_copula/S§TH'ac|act/P-in§ATR'ac|act>
5. meaning: *'persist'* (intransitive)
   <FN:persist/S§PAT'conv|build|inst>

In this scheme, frames match valencies in terms of arity, so a monotransitive frame gets two role arguments. Arguments are slash-separated ('/') and contain themselves three information fields:

1. Syntactic function (S, O, D etc.)
2. Thematic role (e.g. §AG, §TH)
3. Semantic slot fillers (e.g. 'H', 'food', 'act') for nominal material

Possible lexical restrictions are added to field 1. Thus, all prepositional objects are sub-specified for the preposition(s) in question (e.g. *P-auf* for *bestehen auf* – insist on). For non-nominal arguments, syntactic form or POS can be used in field 3 (e.g. 'fcl' – finite clause, 'icl' – non-finite clause, 'num' – numeral).

In only 93 cases, frames had the same verb sense and valency, but different role and semantic-slot assignments. In 927 cases, valency patterns were ambiguous with regard to verb frames, with a maximum of 7 frames for one valency pattern. Hence 927-93=834 valency patterns were verb-sense ambiguous, with only about 30 verbs being affected by more than 1 such ambiguity. This means that 92-93% of all verbs could in theory be sense-disambiguated using syntactic argument structure and verb incorporates alone. For the rest, semantic slot-filler clues are needed. The most common valency patterns in the lexicon were monotransitive-accusative (48.5%), intransitive (15.9%), PP-monotransitive (7.9%), accusative+PP (7.1%) and reflexive (4.1%).

Our German framenet uses 44 atomic semantic roles (or case/thematic roles - Fillmore, 1968)[12]. Where necessary, these roles can be combined, e.g. §AG-EXP (agent & experiencer) for the subject of *zuhören* ('listen'). All in all, 88 such combinations occur in the lexicon. In addition, the frame tagger rules can add a number of adverbial roles that can only occur as free adverbials (such as §COND for conditional subclauses). The 44 roles are far from evenly distributed in running text. Table 1 shows role frequencies at the token level, for German Twitter data.

	Thematic Role	**surface verb args %**	**secon-dary v-args %**	**all surface args %**
§TH	Theme	18.88	20.67	**36.17**
§ATR	Attribute	8.13	1.32	8.91
§LOC-TMP	Point in time	8.64	0	6.33
§MNR	Manner	8.13	0	5.49
§LOC	Location	6.24	2.90	5.30
§AG	Agent	**9.29**	**38.63**	5.72
§EXT	Extension	1.89	0.05	3.04
§META	Meta adverbial	3.95	0	2.59
§COG	Cognizer	4.01	**8.21**	2.27
§DES	Destination	2.13	0.76	1.82
§BEN	Beneficiary	2.56	1.32	1.81
§PAT	Patient	2.44	**4.33**	1.61
§REFL	Reflexive	2.48	0	1.44
§ID	Identity	0.01	0	1.21
§SP	Speaker	1.95	**6.57**	1.17
§CAU	Cause	1.49	1.24	1.02
§ACT	Action	1.34	1.40	2.19
§REC	Recipient	0.94	0.94	1.75
§EV	Event	1.18	1.61	1.56
§EXP	Experiencer	1.32	2.63	1.31
§DON	Donor	0.12	0.31	0.07

Table 1: Semantic roles[13]

[11] In 780 cases, multiple verb senses share the same valency frame – in other words, in 6.5% of cases, verb senses cannot be disambiguated on syntactic function and form alone, but need help from semantic (noun) classes.

[12] This number reflects the desire to strike a balance between true semantic distinctive values on the one hand (not too few) and robustness and the possibility of automatic disambiguation on the other (not too many).

Overall (5th column), i.e. counting arguments of both verbs and nouns, as well as free adverbials, §TH (theme) was the most frequent role, followed by §ATR (attribute) and roles for time, place and manner. The statistics looks quite different, however, if only verb argument roles are considered. For primary dependencies, i.e. with verb and argument in the same clause (column 3), §AG (agent) jumps to rank 2, and prototypical noun roles like §COG (cognizer), §BEN (beneficiary) and §PAT (patient) get higher percentages, but adverbial (adjunct) roles still dominate in quantitative terms. For secondary, cross clause, dependencies (4th column), e.g. subjects of infinitives or relative antecedents, §AG, §COG and §SP together make up over half of all roles, with most adverbial roles now being irrelevant, with the notable exception of §LOC (location).

Even in a language with only four cases, we found some clear likelihood relations between thematic roles and syntactic functions (Table 2). For instance, agents (§AG, §COG, §SP) and experiencer (§EXP) are typical subject roles, while theme (§TH), patient (§PAT), result (§RES) and actions/events (§ACT, §EV) are typical direct object roles.

	Function	**most likely roles, ordered**
@SUBJ	Subject	AG, TH, COG, SP, EXP, PAT, POSS, EVT
@ACC	Accusative object	TH, REFL, PAT, RES, ACT, SOA, MES, BEN, ASS, STI
@DAT	Dative object	BEN, REFL, REC, COG, EXP, TH, ORI
@PIV	Prepositional object	TH, TP, CAU, EV, DES, FIN, ACT, ORI, REC, BEN, COMP
@ADVL	Adverbial (PP only)	LOC, DES, CIRC, TMP-LOC, BEN, COMP, COM, ROLE
@GEN	Genitive obj.	REFL, EXP, COG
@SC	Subj. compl.	ATR, ATR-RES
@OC	Obj. compl.	ATR-RES, ATR
@ADVL	Adverbial	LOC-TMP, MNR, META, EXT, LOC

Table 2: Syntactic functions

In addition to ordinary argument slots, some German verbs allow incorporated material (phrasal verbs or support verb constructions). While these incorporations trigger important sense distinctions, they are not independent complements and need not be semantically transparent. A special case are German verb prefixes that can occur either as a morphological part of the verb or as a separate (and not necessarily adjacent) token. In these cases, the frame annotator will reconstitute the full, prefixed lemma, but not assign a semantic role to the particle, as there is already an incorporation tag at the syntactic level (@MV<)[14]:

er schläft ***ein*** (he falls asleep)
lemma: "einschlafen"

sie machte das Licht ***aus*** (she turned off the light),
lemma: "ausmachen"

Support verb constructions with nominal material are different. Here, the incorporated word (noun or adjective) carries semantic weight - often more than the verb itself - and may also be the valency anchor for further complements. Syntactically, such nominal incorporates fill an existing argument slot, i.e. that of direct or prepositional object, and are tagged as such. Thus, it is only the frame annotator that makes the distinction, assigning a special "incorporation" role (§INC):

jmd. ***Hilfe*** *leisten*
(help sb., 'perform help')

Verdacht *schöpfen gegen jmd.*
(become suspicious of sb., 'scoop suspicion')

When used without a support verb, the frame annotator would treat *Hilfe für* ('help for') and *Verdacht gegen* ('suspicion against') as nominal frame carries and link the argument of the prepositions to these nouns as frame heads. Given their semantic weight, such nouns could in principle be used as frame heads in the support verb constructions as well, linking e.g. the beneficiary of *Hilfe leisten* to the §INC constituent *(Hilfe)* rather than the verb *(leisten).* However, for consistency reasons and in order to allow easy exploitation of verbal features such as tense and person, the current version of the frame annotator marks the frame on the support verbs themselves.

A third type of incorporate are PP incorporates:

auf der Strecke *bleiben* (be lost, 'stay on the road')

in Kraft *treten* (come into effect, 'step into power')

For these, both the preposition and its argument will be listed in the frame pattern. Because of the above-mentioned principle of semantic dependencies, the frame annotator will assign the §INC role to the nominal part of the PP, blocking assignment of other, real adverbial roles.

4. The frame annotator

We perform frame annotation by matching the elements of potential frames against the feature sets of a verb's arguments, exploiting morphosyntactic and semantic information already assigned and disambiguated by the GerGram parser. For instance, if a verb allows three different frames, only one of which – a "tell" frame – has a slot for a finite clause complement (fcl), the presence of an object clause will trigger this frame, harvesting a speaker role (§SP) for the subject (S), a message role (§MES) for the subclause and a receiver role (§REC) for a possible dative object (D), as specified by the frame template example below:

[13] Other roles: §STI Stimulus; §PATH Path; §ORI Origin; §VAL Value; §EXT-TMP Duration; §MES Message, §TP Topic; §SOA State of Affairs; §CAU Cause; §ROLE Role; §INS Instrument, §MNR Manner; §FIN Purpose; §COMP Comparison; §HOL Whole, §PART Part; §POSS Possessor, §ASS Asset; §CONT Content; §COM Co-role; §INC Incorporated

[14] We are here taking into account the (syntactic) annotation performed by GerGram, the parser used as input for our frame annotation system.

<FN:**tell**/S**§SP**'H/D**§REC**'H/O**§MES**'fact|sem-s|fcl>

This annotation mechanism was implemented using the same grammar formalism used by the GerGram parser, Constraint Grammar (CG). The frame/role annotation grammar is run as an additional module after the existing GerGram parser pipe, or on corpus data that has already been morphosyntactically annotated. In a first step, all possible frames for a given verb lemma are mapped as template tags on the verb (cf. <FN:...> tag above). Note that this already implies a first disambiguation step, namely in the case of German separable verbs, where the parser creates different lemmas for each prefix:

er nahm den Bus (he took the bus)
lemma: nehmen <FN:take/...>

er nahm 5 kg zu (he put on 5 kg)
lemma: zu|nehmen <FN:increase/...>

er nahm ihr die Aufgabe ab (he relieved her of the task)
lemma: ab|nehmen <FN:rid/...>

The annotator grammar has four main rule types:

1. frame template tag selection
2. frame template tag removal
3. role instantiation
4. mapping of free roles

Template removal is a simpler task than template selection, because a single mismatch is enough to trigger the former, while as many frame elements as possible should match for the latter. Safest are lexical matches, where word forms (e.g. verb incorporates or prepositions) are mentioned as such in a frame template. For instance, for the highly ambiguous verb *legen* (put), the "minding" frame can be safely removed or selected depending on the presence of an object *Wert* (O-Wert) and a PP argument with the preposition *auf* (P-auf):

<FN:**mind**/S**§COG**'H/O-Wert**§INC**/P-auf**§TH**'all>

Another relatively safe method are syntactic mismatches – at least as long as the parser gets syntactic functions right, with correct dependency links[15]. Thus, impersonal frames (e.g. fn:exist for *'es gibt'* – 'there is'), can be discarded in the absence of a situative subject, copula frames need subject complements and frames with dative (e.g. fn:give) or genitive slots can be removed, if the verb in question has no object dependent with that particular case. Because some complements are not always expressed, and utterances can be incomplete, such rules usually have a fail-safe condition asking for the existence of a competing, similar frame with lower valency (e.g monotransitive instead of ditransitive).

The most important rules, however, are the ones capable of differentiating frame templates with identical syntactic/valency skeletons. In this scenario, semantic slot filler information is used, exploiting the so-called semantic prototype tags that GerGram assigns to nouns and proper nouns. Although these are lexical tags, and only partially disambiguated by the parser, ambiguities rarely overlapped with frame ambiguities found in a given verb. The noun ontology has about 200 categories, e.g. <tool>, <food>, <event> or <mon> (money). Categories are organized in a shallow hierarchy, with lower-case or hyphenated subcategories. 'H', for instance, signals +HUM: <Hprof> (profession), <Hfam> (family member), <Hideo> (ideological) etc., while <sem-r> (readable), <sem-c> (concept), <sem-s> (sayable) etc. share the semantic 'sem-' prefix.

In the frame templates, all non-trivial function/role pairs (e.g. S§AG for subject agent) have at least one semantic slot filler category. The matching mechanism proceeds from safe to unsafe by first trying rules with more, and more specific, conditions, leaving underspecified matches for last. To account for category fuzziness, overlaps, creative language use or just incomplete slot filler lists, the grammar uses "umbrella category" matches, by defining 15-20 semantic hypernym sets, e.g. 'HUMAN', 'THING', 'PLACE' etc. At this intermediate level, two categories (from a frame template and a sentence token, respectively) will be considered a match, if they share their hypernym ("umbrella") category. The full progression of the matching algorithm is the following:

- > 2 syntactic slots with a full semantic match
- 1 slot: full match, 1 slot: "umbrella" match
- 2 or more "umbrella" matches
- 1 slot with a full match
- 1 slot with an "umbrella" match
- syntactic match, slot(s) marked <all> or <cc>
- longest syntactic match

As long as roles manifest as surface constituents with a direct dependency link to the frame-evoking verb, the method is quite robust, since it can simply draw on the existing dependency parse. However, in about 45% of cases, the role-carrying constituent has no, or only pronominal, surface representation in the clause itself. In German, this is the case for e.g. subjects of infinitive clauses and antecedents of relative clauses. Thus, in the annotated example sentence in the following paragraph, the word *Eltern* ('parents') functions – in two different subclauses – as both an object-theme in an 'exist' frame and as a subject-cognizer in an 'allow' frame. Here, to check slot-filler conditions and assign roles, we first introduce secondary dependency relations using special, relation-mapping CG rules. These additional dependency relations can then be drawn upon by variants of the ordinary frame-matching rules.

In a vertical, one-word-per-line CG notation, the frame-tagger adds <fn:sense> tags on verbs (red), and §ROLE tags on arguments (blue). The example contains three ordinary main verb frames with several role dependents each, one noun frame with a single dependent, and a naked frame without roles (an auxiliary). All dependency links have been "raised" from syntactic to semantic links (e.g. bypassing prepositions). Primary dependency arcs

[15] In fact, this is not always a given with Twitter data, because of its non-standard jargon, orthographic or grammatical errors and incomplete utterances. Therefore, improving parser robustness was part of the project plan.

are shown as #n->m ID-links, secondary dependencies are marked as R:c- (child) and R:p (parent) relations.

Ich [ich] PERS @SUBJ> **§COG** #1->2 *(I)*
verstehe [verstehen] <mv> **<FN:comprehend>** V FIN @FS-STA #2->0 *(understand)*
nicht [=] ADV @<ADVL #3->2 *(not)*
warum [=] <clb> <interr> ADV @ADVL> **§CAU** #4->7 *(why)*
es [=] PERS @S-SUBJ> **§TH-NIL** #5->7 *(there)*
Eltern [=] <HH> N @ACC> **§TH** #6->7 ID:6 R:c-subj:17 R:sd-**COG**:17 *(parents)*
gibt [geben] **<FN:exist>** <mv> V FIN @FS-<ACC **§TH** #7->2 *(are)*
$, [,] PU @PU #8->0
die [=] <clb **<rel>** INDP nG P NOM @SUBJ> #9->17 *(that)*
die [=] ART @>N #10->11 *(the)*
Erziehung [=] **<FN:teach>** **<V:erziehen>** N @ACC> **§ACT** #11->17 *(education)*
ihrer [sie] <poss> DET @>N #12->13 *(their)*
Kinder [Kind] <H> N @N< **§BEN** #13->11 *(children)*
möglichst [möglich] <jcan> ADV SUP @>A #14->15 *(as possible)*
früh [=] <atemp> ADV @ADVL> **§LOC-TMP** #15->17 *(early)*
Fremden [fremd] <jsoc> <Q-> <nadj> ADJ @DAT> **§REC** #16->17 *(strangers)*
überlassen [=] **<FN:allow>** <mv> V INF @FS-N< **§ATR** #17->6 ID:17 R:p-subj:6 *(leave)*
wollen [=] **<FN:wish>** <aux> V FIN @AUX #18->17 *(want)*

(*I don't understand why there are parents who want to leave the education of their children to strangers as early as possible)*

5. Evaluation

About two years' worth of German Twitter were annotated at both the morphological, syntactic and semantic levels, the latter assigning ontology classes for nouns and adjectives, and frame structures for verbal and other predications. For evaluative purposes we extracted all main verb-lemmas and their semantic class frame (i.e. without argument structure) with a frequency >= 1,000. The frequency threshold helped remove a lot of noise, in particular words from non-German tweets that had received heuristic German verb readings. There were 8894 lemma-frame combinations above the threshold, representing 202.4 million tokens. These were manually checked, removing 1268 non-German words and POS errors, retaining 7,726 real German verb frames representing 6,127 verb lemmas and 193.4 million tokens. There were on average 1.245 frame classes per verb lemma type, matching almost exactly the distribution in the lexicon (1.237) and indicating that even with a frequency cut-off, our sample is representative of the corpus (and, to a degree, the language) as a whole. Also, the number of lemmas represented amounts to half the German verb lexicon, and thus constitutes a very high type percentage, equivalent to a more than 99.9% token coverage according to Zipf's law.

As expected, ambiguity at the token level was much higher than at the type level, with 3.126 frame senses per verb. The same holds for coverage: At the token level, only 1.11% of verbs had no frame in the lexicon, with a further 0.25%, where there were frames in the lexicon, but none survived disambiguation. At the type level, due to the impact of very rare verbs, frame failures were five times higher (5.88%).

Frame sense ambiguity is very unevenly distributed, with a lot of ambiguity at the token level, while 78.6% of verb types in the corpus had only one frame sense. For the 10 most frequent verbs, together accounting for 10.36% of all verb tokens in the corpus, frame sense ambiguity is shown in Table 3:

verb lemma	**token count**	**frame senses**
lassen	2824239	11
geben	2455458	10
machen	2124256	34
spielen	1457122	4
nehmen	1416502	24
sehen	1414451	5
kommen	1251055	13
bleiben	1250034	3
haben	1237781	8
halten	1226771	17

Table 3: Frame sense ambiguity

Since frame sense disamiguation relies on slot-filler matches, performance may suffer if verb arguments can not be recovered in the sentence, either because of missing dependency links, passives, infinitive clauses or the incomplete clauses seen in Twitter jargon. As can be seen in Table 4, surface expression is high for subject complements, and low for the usually optional dative arguments. About 1/4 of subject and direct object slots remained unmatched, even when adding secondary dependencies, underlining the need for robustness in the face of input anomalies and parse errors, as well as the importance of progressive match relaxation for frame assignment.

	filled slots (incl. secondary dep.)	**filled slots (primary dep. only)**
SUBJ	74.5 %	72.7 %
ACC	73.1 %	72.9 %
DAT	60.3 %	60.3 %
SC	97.7 %	97.7 %

Table 4: Surface expression of case arguments

To judge the performance of the frame tagger, we evaluated a random set of tweets from the minority-filtered subcorpus, with 9,054 tokens.[16] All in all, there were 884 main verb tags, 22 of which (2.5%) had been wrongly tagged for POS (20), lemma (1) or auxiliarity (1), often due to spelling errors in the word itself (7 cases) or in the rest of the sentence. 8 verbs were not recognized as

[16] Because 1% of these tokens contain spaces, e.g. names like "Dalai Lama", and foreign nouns like "Task Force", a space-delimited token count would a bit higher.

such. Our frame tagger found a correct frame sense for 95.2% of the correctly tagged verbs (Table 5), with 21 spurious frames (97.4% precision). There were only 9 lexical coverage failures (1.0%), with 1 verb, "tackern" (staple) getting no frame entry at all, and 8 constructions, where the correct frame was not among the ones listed in the lexicon.

	R	P	F-score
total incl. POS errors	90.7%	96.5%	93.6
ignoring POS errors	93.0%	97.4%	95.2

Table 5: Recall and precision

These figures are an encouraging result. Although the “weak” (inspection-based) evaluation method makes a direct comparison problematic, performance compares favourably with e.g. an early English baseline for rule-based frame tagging, Shi & Mihalcea (2004), with an F-score of 74.5%. More recently, Hermann et al. (2014) report F=70.1% for predicate frame identification, using an ML model exploiting syntactic context features. Specifically for (English) Twitter, with out-of-domain training data, Hartmann et al. (2017) achieved 62.17% full frame identification. For the correct/gold frames, semantic role labeling (SRL) had F=76.74%. For German, the SHALMANESER system (Burchardt et al. 2009) achieved 79% correct word sense disambiguation (WSD). Cai & Lapata (2019) performed (in-domain) SRL with an F-score of 82.7%, using syntax-aware neural networks. Without using linguistic features, Do et al. (2018) reached a somewhat lower F=73.5 for SRL on the same German test data (CoNLL 2009).

6. Conclusion and future work

We have presented and evaluated PFN-DE, a new, unabridged framenet for German with over 11,000 verb entries, discussing the advantages and challenges of a parsing-oriented, valency-based "framenet light" approach when faced with an annotation task of bilingual social-media data. A CG-based frame tagger, built using the resource, demonstrated a good coverage at both the type and token levels, and achieved an F-score of 93.6 for frame senses. However, recall was lower than precision, and based on error distribution, future work should improve granularity by adding missing senses to existing verb entries. Evaluation also showed that frame tagging is very dependent on the quality of the underlying parser at both the morphological and syntactic levels. In particular, more work is needed to minimize POS and dependency errors in the face of non-standard orthography and clause structure in Twitter data. Conversely, it is reasonable to assume that the frame tagger could be successfully used for domains with *less* orthographical problems, such as news and literature, given the fact that the underlying (rule-based) morphosyntactic analysis was developed for general use and does not depend on the existence of a training corpus.

The task of compatible bilingual corpus annotation and the interoperability of parser methodology, as well as the bootstrapping benefits when modeling a German framenet on a Danish one, raise the question whether a similar approach would work for less related languages. Ongoing work with Portuguese, using a similar method and an equivalent Constraint Grammar parser (PALAVRAS), presents an attempt to breach the Germanic-Romance language divide and should allow comparable frame annotation of mixed-language corpora for all three languages in the near future.

7. Bibliographical References

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

Baumgarten, N., Bick, E., Geyer, K., Iversen, D.A., Kleene, A., Lindø, A.V., Neitsch, J., Niebuhr, O., Nielsen, R., and Petersen, E.N. (2019). Towards Balance and Boundaries in Public Discourse: Expressing and Perceiving Online Hate Speech (XPEROHS). In Mey, J., Holsting, A., Johannessen, Ch. (eds.): *RASK – International Journal of Language and Communication*. Vol. 50, pp. 87–108. University of Southern Denmark

Bick, Eckhard. (2007). Dan2eng: Wide-Coverage Danish-English Machine Translation, In Maegaard, B. (ed.), *Proceedings of Machine Translation Summit XI* (10-14. Sept. 2007, Copenhagen, Denmark). pp. 37-43

Bick, E. (2011). A FrameNet for Danish. In *Proceedings of NODALIDA 2011* (May 11-13, Riga, Latvia). NEALT Proceedings Series, Vol 11, pp. 34–41. Tartu: Tartu University Library.

Bick, E. (2014). Swedish-Danish Machine Translation in a Constraint Grammar Framework. In Adam Przepiórkowski, A. and Ogrodniczuk, M. (eds.): *Advances in Natural Language Processing, Proceedings of 9th International Conference on NLP* (PolTAL 2014, Warsaw, Poland, September 17–19, 2014), pp. 216-227. Heidelberg:Springer.

Bick, E. and Didriksen, T. (2015). CG-3 – Beyond Classical Constraint Grammar. In Megyesi, B. (ed.): *Proceedings of NODALIDA 2015* (May 11-13, 2015, Vilnius, Lithuania), pp. 31–39. Linköping: LiU Electronic Press.

Bick, E. (2017). From Treebank to Propbank: A Semantic-Role and VerbNet Corpus for Danish. In Tiedemann, J. (ed.): *Proceedings of the 21st Nordic Conference on Computational Linguistics* (NoDaLiDa 2017, Göteborg). NEALT Proceedings Series Vol. 29. pp. 202–210. Linköping University Electronic Press.

Boas, H.C. (2005). From theory to practice: Frame semantics and the design of FrameNet. In Langer, S. and Schnorbusch, D. (ed.): *Semantik im Lexikon*. pp. 129–159. Tübingen: Gunter Narr Verlag.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006). The SALSA Corpus: A German Corpus Resource for Lexical Semantics. In *Proceedings of LREC-06* (Genoa, Italy). pp. 969–974

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2009). Using FrameNet for the semantic analysis of German: Annotation, representation and automation. In Boas, H. C. (ed.): *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. pp. 209–244. Mouton de Guyter.

Cai, R., and Lapata, M. (2019). Syntax-aware Semantic Role Labeling without Parsing. In *Transactions of the Association for Computational Linguistics*. 7. pp 343–356.

Do, Q.N., Leeuwenberg, A., Heyman, G., and Moens, M. (2018). A Flexible and Easy-to-use Semantic Role Labeling Framework for Different Languages. In *Proceedings of COLING 2018* (Demos). pp. 161–165

Fellbaum, C. (ed.). (1998). *WordNet: An Eletronic Lexical Database*. Language, Speech and Communications. MIT Press: Cambridge, Massachusetts.

Fillmore, C.J. (1968). The case for case. In Bach and Harms (Ed.): *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston. pp. 1–88.

Gilardi, L. & Baker, C.F. (2018). Learning to Align across Languages: Toward Multilingual FrameNet. In *International FrameNet Workshop 2018: Multilingual FrameNets and Constructions* (Miyaki, Japan). pp. 13–22

Hermann, K.M; Das, D., Weston, J., and Ganchev, K. (2014). Semantic Frame Identification with distributed Word Representations. In *Proceedings of the 52nd Annual Meeting of ACL* (Baltimore, Mariland). pp. 1448–1458

Hartmann, S., Kuznetsov, I., Martín-Valdivia, M.T., and Gurevych, I. (2017). Out-of-domain framenet semantic role labeling. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Vol. 1, Long Papers. pp. 471–482.

Johnson, C.R. and Fillmore, C.J. (2000). The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics* (ANLP-NAACL 2000, Seattle WA). pp. 56–62.

Kipper, K., Korhonen, A., Ryant, N., and Palmer, M. (2006). A Large-Scale Extension of VerbNet with Novel Verb Classes. In *Proceedings of the 12th EURALEX International Congress* (Turin, Italy, September 2006). pp. 173–184

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. In *Computational Linguistics*, 31:1. pp. 71–105

Palmer, A., and Sporleder, C. (2010). Evaluating FrameNet-style semantic parsing: the role of coverage gaps in FrameNet. In *Proceedings of the 23rd International Conference on Computational Linguistics: posters*. pp. 928-936. Association for Computational Linguistics.

Rehbein, I, Ruppenhofer, J., Sporleder, C. and Pinkal, M. Adding nominal spice to SALSA – frame-semantic annotation of German nouns and verbs. In *Proceedings of KONVENS 2012* (Vienna, Austria). pp. 89–97

Ruppenhofer, J., Ellsworth, M., Petruck. M.R.L., Johnson, C.R., and Scheffczyk, J. (2010). *FrameNet II: Extended Theory and Practice.* At: http://framenet.icsi.berkeley.edu/

Shi, Lei and Mihalcea, R. (2004). Open Text Semantic Parsing Using FrameNet and WordNet. In *Proceedings of HLT-NAACL 2004, Demonstration Papers*. pp. 19–22