Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022)

Semantic Role Labeling for Sentiment Inference: A Case Study

Manfred Klenner, Anne Göhring
Department of Computational Linguistics
University of Zurich
{klenner,goehring}@cl.uzh.ch

Abstract

In this paper, we evaluate in a case study whether semantic role labelling (SRL) can be reliably used for verb-based sentiment inference (SI). SI strives to identify polar relations (against, in-favour-of) between discourse entities. We took 300 sentences with 10 different verbs that show verb alternations or are ambiguous in order to find out if current SRL systems actually can assign the correct semantic roles and find the correct underlying predicates. Since in SI each verb reading comes with a particular polar profile, SRL is useful only if its analyses are consistent and reliable. We found that this is not (yet) given for German.

1 Introduction

Sentiment Inference (SI) is the task of predicting opponents and proponents given a text. SI reveals how the writer conceptualises the world and how she perceives the discourse entities she refers to. Take for instance the sentence *This government cheats the world*. The writer tries to convey that the government is against the world and that it is - in the perspective of the writer - a negative actor and the world is the victim, which means that there is a negative effect on the world. We, thus, can talk about positive and negative actors, positive and negative effects, about negative (opponents) and positive (proponents) relations. We call these specifications the polar profile of a verb.

In (Klenner et al., 2017), we introduced a verb-based SI system that uses dependency labels in order to express such polar profiles. For instance, the subject of the verb *cheat* - if used in a factual sentence - is identified as indicating a negative actor, the filler of the direct object receives a negative effect, and a negative relation (against) between the two is casted. Even after normalization of dependency trees, e.g. by resolving passive voice, some problems remain, namely verb alternations and verb ambiguity. It certainly will lead to false analyses. Verb alternation, among others, is given if a semantic role changes its syntactic host. As an example of an instrument-subject verb alternation, compare *The police man killed the aggressor with a knife* versus *The knife killed the aggressor*. For a dependency-based approach the police man and the knife are both the subjects although the police man is the agent and the knife is the instrument. There should be a negative polar relation between police man and aggressor, but not between knife and aggressor (a knife cannot be against somebody). If SRL was used instead of dependency parsing, the agent role would indicate the against relation while the instrument role would block such an inference[1] and thus might be a means to provide a general solution to this problem.

SRL could also be useful for verb sense disambiguation. Part of SRL is a step called predicate identification (Conia et al., 2021b), where a verb is mapped to a predicate frame covering the semantic roles of the underlying verb reading. Take as an example German *bedauern* which has a subject and a direct object. It could mean either *feel sorry for* as in *Ich bedauere diese Menschen* (I feel sorry for these people), or *regret* as illustrated by *Ich bedauere den Vorfall* (I regret the incident). In the first case, there is a in-favour-of relation while in the second one the relation is against. In this example, it is not the semantic role that makes the difference in the first place, but the predicate identification (*feel sorry for* versus *regret*).

In this paper, we describe a case study applying SRL to cases of verb alternations and verb ambiguity. For SRL to be applicable, it must hold that the identification of semantic roles is consistent given some verb and that predicate identification is reliable. We found both requirements are currently

[1] The SRL approach InVeRo using VerbAtlas actually produces this result, see https://verbatlas.org

not given for German.

2 Verb Alternations and Verb Ambiguity

As a first step, we identified 10 German verbs[2] from our verb lexicon (Klenner and Amsler, 2016) that have verb alternations or are ambiguous. We focused on challenging cases where a verb has at least two semantic frames given a **single** dependency frame. Take the transitive (i.e. subject,object) and ambiguous verb *verbessern* which might mean *improve* or *correct*. In a dependency setting we just have the subjects and objects of the particular verb *verbessern*. In our current system we cannot distinguish the readings and, thus, only have one polar profile. But in fact we'd need two: for both readings. So either verb disambiguation (which is not available for German) or SRL might do the trick.

As an example of verb alternation take *drohen* (threaten), which has an instrument alternation:

(1) Er droht ihm mit Vergeltung
subject verb object oblique
He threatens him with retaliation

(2) Ihm droht Vergeltung
object verb **subject**
He is threatened with retribution

Only in (1) there is a polar relation (against) between the agent (He) and the recipient (him). In our case study we looked at the transitive versions of such cases: *Er droht ihm* versus *Vergeltung droht ihm* (a bit unusal word order, but correct). Again, in the dependency setting we have a single transitive verb with two unaccesible readings (*threaten* versus *face*).

We semi-automatically extracted 300 sentences from a newspaper corpus where for each verb at least two different semantic frames were given. For instance for the verb *drohen,* we found 5 sentences with an actor as subject (one reading) and 8 with a theme as subject (the second reading).

We applied InVeRo in the PropBank and the VerbAtlas mode and manually analysed the results. We will now introduce these tools.

3 Semantic Role Labeling for German

We have tried to find SRL systems for German, but only InVeRo (Conia et al., 2021b) using VerbAtlas (Di Fabio et al., 2019) was available. It was not possible to install SRL-S2S[3] (Daza and Frank, 2019), and the DameSRL[4] system described in (Do et al., 2018a,b) has no predicate identification model for German which is needed for a proper SRL. Another option was to train our own model. However after we have analysed the available resources, the CoNLL shared task description and data (Hajič et al., 2009), and the Universal Proposition Bank (Akbik et al., 2015), we skipped this idea. The German data from CoNLL is derived from Salsa (Erk et al., 2003), the German version of FrameNet. It came into existence by mapping FrameNet roles, which are very fine-grained, to more coarse-grained PropBank semantic roles (Palmer et al., 2005). However, the mapping procedure is hardly described and no quality control is reported. We do not know how much noise was introduced by this mapping. In a footnote, Daza and Frank (2020) reflect on the difficulty of using heterogeneous SRL styles, above all for a cross-lingual comparison, and comment that "annotations for German use a role inventory with roles A0-A9, and a one-to-one mapping to all English labels is not available". Also, after we analysed a few entries in the German Universal Propositions Bank[5], we had to recognise that this semi-automatically generated resource is too noisy. Training our own SRL model no longer was an option. We, thus, carried out our experiments with InVeRo (Conia et al., 2021a).

InVeRo is a multi-lingual SRL model that was trained on various languages including German. Given a (German) sentence, predicate identification yields an English (predicate) frame and the corresponding semantic roles. The frames are from VerbAtlas, a hand-crafted lexical-semantic resource that uses the verb synsets of BabelNet (Navigli and Ponzetto, 2010), a multilingual encyclopedic dictionary that covers 500 languages (actually the synsets of WordNet are used via BabelNet which integrates Wordnet). VerbAtlas frames specify a prototypical argument structure including implicit and so-called shadowed arguments (Conia et al., 2021a). Such a frame clusters verb meanings having similar semantics. Also selectional preferences (not restrictions) are formulated on the basis of WordNet synsets.

[2] See the appendix for the full verb list.

[3] https://github.com/Heidelberg-NLP/SRL-S2S

[4] https://liir.cs.kuleuven.be/software_pages/damesrl.php

[5] http://alanakbik.github.io/UniversalPropositions_German

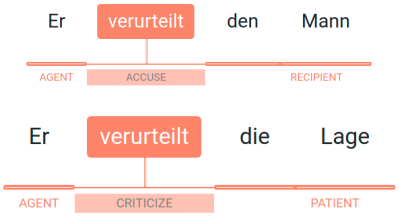

Figure 1: InVero's predicate identification for two German sentences with the verb *verurteilen*, and their corresponding semantic role frames ('He accuses the man' versus 'He criticizes the situation').

Semantic roles are either in PropBank style or following VerbNet nomenclature (25 roles like agent, patient, etc.) (Kipper Schuler et al., 2009).

In Figure 1 predicate identification maps the verb *verurteilen* to *accuse* and *criticize*. As a consequence, two different roles for the direct object become available, namely *recipient* and *patient*. The selectional preferences for the patient role of *criticize* are *individual* and *social group*. Although *situation* is not subsumed under neither restriction, we get a result. The system is robust, thus. However sometimes restrictions seem to be taken seriously and no result appears. The sentence *Sie kämpft für mehr Geld* (She fights for more money) is correctly analysed. If we substitute *Gerechtigkeit* (justice) for *Geld* (money), no result is given, presumably since *Gerechtigkeit* is not subsumed under the restriction which is *entity*.

4 Empirical Evaluation

We manually analysed the output of InVeRo for the 300 sentences. Three types of errors or problems can be distinguished:

- predicate identification (disambiguation) fails
- assigning different semantic roles given a single predicate
- assigning a particular semantic role to syntactically different phrases for the same verb (under a particular reading)

Why are these three points problematic in SI? As we have discussed on various examples, each verb reading has its own polar profile, thus it is crucial to find the right reading (problem 1). A polar profile assigns a directed polar relation (against, in-favour-of) to a verb as well as a holder role (e.g. the agent) and a target role (e.g. theme). That is, in order to specify these relations, the semantic roles of the holder and target roles must be known and they must be stable (not assigned to different roles), otherwise no lexical entry is possible (problem 2). If SRL assigns for a verb reading different roles and role pairings, it is unclear how to anchor the relation correctly. Finally, SRL is syntax-agnostic (problem 3): the same semantic role of a verb might be assigned to different syntactic phrases thereby possibly collapsing verb readings. In the examples (3) and (4) both sentences (according to VerbAtlas[6] have a theme role. In sentence (3) it is realized as a to-infinitive, in sentence (4) as a prepositional phrase (PP).

(3) Er droht zu scheitern
agent verb to-infinitive-**theme**
He is in danger to fail

(4) Er droht mit Konsequenzen
agent verb PP-**theme**
He threatens consequences

As a consequence, these two verb readings would have the same semantic role frame. However, their polar profiles differ. Sentence (3) casts a negative effect on the experiencer (He), while in (4) there is a negative actor, but no negative effect. SRL is not helpful in these cases, it also collapses readings (*danger*, *threatens*).

Predicate identification failure is most problematic. In the examples above, both (3) and (4) get the same predicate assigned: *guarantee/ensure/promise*[7]. However, only sentence (4) is an instance of this predicate.

This problem becomes clearer, in our case study, if we quantify the number of predicates and predicate frames[8] that were chosen by InVeRo per verb (see the last line of Table 2 in the appendix). For PropBank a verb is, in the mean, mapped to 1.55 predicates, and 3.7 different frames, i.e. pairing of semantic roles, per predicate are used. For VerbAtlas it is 2.75 and 4.5, respectively. Ideally, only one mapping would be given: a verb maps to one or more predicates, each predicate has a stable subcategorization frame (expressed with semantic roles). If this was the case, we could assign a single polar profile to a particular verb reading.

Table 1 shows the mappings for *bedauern*. In the first column the *feel-sorry-for* reading is given.

[6] https://verbatlas.org, accessed 2022-06-03.

[7] Predicates in VerbAtlas are sometimes specified with reference to more than one label.

[8] *frame* here refers to role pairings.

	feel-sorry-for	regret
DE	bedauern.2 (A0,A1) [1]	bedauern.1 (A0,A1) [4] (A0,A3) [11] bedauern.2 (A0,A1) [15] (A0,A3) [8]
VA	DISLIKE (Agent,Theme) [1]	DISLIKE (Agent,Theme) [2] (Exp.,Stimulus) [4] REGRET_SORRY (Agent,Theme) [25] (Exp.,Stimulus) [1] (Agent,Attribute) [1] CRITICIZE (Agent,Theme) [5]

Table 1: Different predicates and roles for the verb 'bedauern' according to two readings: *feel-sorry-for* and *regret*. In square brackets are the numbers of sentences labeled with the given semantic roles.

Here we have a single mapping, both with respect to PropBank (DE) and VerbAtlas style (VA). However in the second column, the *regret* reading, PropBank mode shows a variation in the assignment of semantic roles (A0,A1 versus A0, A3). The VerbAtlas analysis is even more confusing. Here three predicates are identified and within the same predicate (e.g. REGRET_SORRY), different roles and role pairings are present. We carried out an error analysis in order to find out how many of the 38 sentences with *bedauern* are wrongly analysed either by choosing the wrong predicate or the wrong semantic role pairing (the subcategorization frame): 7 cases (18.5%) are clearly wrong, 8 cases are hard to decide. Not in every case does the usage of *bedauern* actually involve a (real) regret. Sometimes it is used in more formal way in order to express dislike (as suggested by InVeRo): without context this cannot be resolved reliably (some of the 8 cases are of that type). But nevertheless, even if InVeRo sometimes is right to map a verb to more than one predicate, the diversity of suggested solutions makes it impossible to carry out SI in a lexicon-based way: the necessary mapping from a single polar profile of a verb to some VerbAtlas representation in a one-to-many fashion is bound to produce errors, as our little error analysis with *bedauern* reveals.

Also, although in principle assigning semantic roles depending on the filler object is a desirable solution, if it comes in such an unpredictable diverse way, a lexicon-based approach cannot make use of it. The problem is not neglectable, since the distribution of semantic role pairings for different VerbAtlas predicates is high. The numbers at the end of the roles pairings (in square brackets) in Table 1 indicate the frequency of a pairing. For instance, DISLIKE (Agent,Theme) was assigned 2 times, DISLIKE (Experiencer,Stimulus) 4 times.

The statistics we have gathered on the diversity of predicate and frame mappings coming with InVeRo makes it superfluous to have a full-fledged error analysis for all 300 sentences (like we did for *bedauern*). The InVeRo results are just too diverse to be useful (see Table 2 in the appendix).

In the course of our case study, we have noticed that there is a correlation between the (non)animacy of role fillers and different verb readings. Actually, all examples in this paper could be analysed correctly by taking (non)animacy into account: compare e.g. *er bedauert sie* (he feels sorry for her) with *er bedauert den Vorfall* (he regrets the incident). We have trained an animacy classifier (Klenner and Göhring, 2022) and are about to apply it to the small data set of 300 sentences. To sketch the idea: depending on the animacy of the filler of a dependency label of a verb, different polar profiles become available.

5 Related Work

Sentiment inference is sometimes called sentiment propagation (Deng and Wiebe, 2014) and opinion implicature. It also shares similarities with fine-grained opinion analysis (Marasović and Frank, 2018a). Our positive/negative effects are comparable to the GoodFor/BadFor distinction of (Choi and Wiebe, 2014). However, we also distinguish positive/negative actors. In (Wiebe and Deng, 2014) a sophisticated rule-based system was introduced that specifies general inference rules on the basis of GoodFor/BadFor effects.

Approaches exist that claim that the combination of SRL and Opinion Role Labeling, i.e. the identification of opinion holder and target, is beneficial, e.g. in (Marasović and Frank, 2018b) a multi-task learning-based joint model is introduced.

6 Conclusion

German Semantic Role Labeling does not provide a suitable solution for our task: German sentiment inference based on polar profiles of verb readings.

With InVeRo, lexicon design is difficult since (too) many verb-predicate mappings and role pairings occur. InVeRo is only partially able to deal with the - admittedly - difficult cases of verb alternations and verb ambiguity. Instead of SRL, a combination of dependency parsing and animacy detection might be useful for the task at hand. We are currently evaluating such a disambiguation strategy for sentiment inference.

References

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*, pages 397–407, Beijing, China. ACL.

Yoonjung Choi and Janyce Wiebe. 2014. +/-effectwordnet: Sense-level lexicon acquisition for opinion inference. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, SIGDAT*, pages 1181–1191.

Simone Conia, Andrea Bacciu, and Roberto Navigli. 2021a. Unifying cross-lingual semantic role labeling with heterogeneous linguistic resources. In *Proceedings of the 2021 Conference of the North American Chapter of the ACL: Human Language Technologies*, pages 338–351.

Simone Conia, Riccardo Orlando, Fabrizio Brignone, Francesco Cecconi, and Roberto Navigli. 2021b. 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. ACL.

Angel Daza and Anette Frank. 2019. Translate and label! an encoder-decoder approach for cross-lingual 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*, pages 603–615, Hong Kong, China.

Angel Daza and Anette Frank. 2020. X-SRL: A parallel cross-lingual semantic role labeling dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3904–3914, Online. ACL.

Lingjia Deng and Janyce Wiebe. 2014. Sentiment propagation via implicature constraints. *Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2014)*.

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*, pages 627–637, Hong Kong, China. ACL.

Quynh Ngoc Thi Do, Artuur Leeuwenberg, Geert Heyman, and Marie-Francine Moens. 2018a. A flexible and easy-to-use semantic role labeling framework for different languages. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 161–165, Santa Fe, New Mexico. ACL.

Quynh Ngoc Thi Do, Artuur Leeuwenberg, Geert Heyman, and Marie-Francine Moens. 2018b. How to use damesrl: A framework for deep multilingual semantic role labeling. In *Proceedings of the CLARIN Annual Conference*, pages 159–162, Pisa, Italy.

Katrin Erk, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 2003. Towards a resource for lexical semantics: A large German corpus with extensive semantic annotation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 537–544, Sapporo, Japan. ACL.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the 13th. Conference on Computational Natural Language Learning (CoNLL 2009)*, pages 1–18, Boulder, Colorado. ACL.

Karin Kipper Schuler, Anna Korhonen, and Susan Brown. 2009. VerbNet overview, extensions, mappings and applications. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Tutorial Abstracts*, pages 13–14, Boulder, Colorado. ACL.

Manfred Klenner and Michael Amsler. 2016. Sentiframes: A resource for verb-centered German sentiment inference. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA).

Manfred Klenner, Simon Clematide, and Don Tuggener. 2017. Verb-mediated composition of attitude relations comprising reader and writer perspective. In *18th International Conference on Computational Linguistics and Intelligent Text Processing*. ResearchBib.

Manfred Klenner and Anne Göhring. 2022. Animacy denoting german nouns: Annotation and classification. In *Proceedings of the Language Resources and*

Evaluation Conference, pages 1360–1364, Marseille, France. European Language Resources Association (ELRA).

Ana Marasović and Anette Frank. 2018a. SRL4ORL: Improving opinion role labeling using multi-task learning with semantic role labeling. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 583–594, New Orleans, Louisiana. ACL.

Ana Marasović and Anette Frank. 2018b. SRL4ORL: Improving opinion role labeling using multi-task learning with semantic role labeling. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics, Volume 1*, pages 583–594, New Orleans, Louisiana. ACL.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the ACL*, pages 216–225, Uppsala, Sweden.

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

Janyce Wiebe and Lingjia Deng. 2014. An account of opinion implicatures.

Appendix

	DE			VA		
verb	pr	fr	fr/pr	pr	fr	fr/pr
akzeptieren	1	1	1.00	4	4	1.00
	1	2	2.00	9	11	1.22
bedauern	1	1	1.00	1	1	1.00
	2	8	4.00	3	7	2.33
bedrohen	2	5	2.50	7	11	1.57
	3	8	2.67	7	13	1.86
belastern	2	2	1.00	1	2	2.00
	2	3	1.50	4	9	2.25
blockieren	3	6	2.00	1	2	2.00
	3	6	2.00	1	3	3.00
schaden	1	3	3.00	2	3	1.50
	1	2	2.00	2	2	1.00
töten	1	5	5.00	1	5	5.00
	1	5	5.00	1	3	3.00
unterstützen	1	1	1.00	1	1	1.00
	2	6	3.00	2	5	2.50
verbessern	1	1	1.00	1	1	1.00
	1	3	3.00	3	3	1.00
vergewaltigen	1	5	5.00	3	3	1.00
	1	1	1.00	1	1	1.00
avg	1.55	3.70	**2.43**	**2.75**	4.50	1.81

Table 2: Number of predicates (pr), frames (fr) and frames per predicate (fr/pr) the SRL assigned to example sentences of the listed 10 pairs of verb profiles (each verb has 2 profiles). Average (avg) over all profiles (macro = micro). The German PropBank scheme (DE) seems to assign less different predicates per verb profile than the VerbAtlas (VA) scheme (1.55 compared to 2.75), though with proportionally more frames (fr/pr= 2.43).