

Argumentative texts and clause types

Maria Becker, Alexis Palmer, and Anette Frank

Leibniz ScienceCampus *Empirical Linguistics and Computational Language Modeling*

Department of Computational Linguistics, Heidelberg University

Institute of German Language, Mannheim

{mbecker, palmer, frank}@cl.uni-heidelberg.de

Abstract

Argumentative texts have been thoroughly analyzed for their argumentative structure, and recent efforts aim at their automatic classification. This work investigates linguistic properties of argumentative texts and text passages in terms of their semantic clause types. We annotate argumentative texts with *Situation Entity* (SE) classes, which combine notions from lexical aspect (states, events) with genericity and habituality of clauses. We analyse the correlation of SE classes with argumentative text genres, components of argument structures, and some functions of those components. Our analysis reveals interesting relations between the distribution of SE types and the argumentative text genre, compared to other genres like fiction or report. We also see tendencies in the correlations between argument components (such as premises and conclusions) and SE types, as well as between argumentative functions (such as support and rebuttal) and SE types. The observed tendencies can be deployed for automatic recognition and fine-grained classification of argumentative text passages.

1 Introduction

In this study we annotate a corpus of short argumentative texts with semantic clause types drawn from work in theoretical linguistics on modes of discourse. The aim is to better understand the linguistic characteristics of argumentative text passages. Our study suggests that these clause types, as linguistic features of argumentative text passages, could be useful for automatic argument mining – for identifying argumentative regions of text, for identifying premises and conclusions/claims, or

for classifying the argumentative functions served by premises.

Specifically, this is an empirical investigation of the semantic types of clauses found in **argumentative text passages**, using the inventory of clause types developed by Smith (2003) and extended in later work (Palmer et al., 2007; Friedrich and Palmer, 2014). **Situation entity (SE) types** describe how clauses behave in discourse, and as such they are aspectual rather than ontological categories. Individual clauses of text evoke different types of situations (for example, states, events, generics, or habituals), and the situations evoked in a text passage are linked to the text type of the passage. For more detail see Section 2. Furthermore, SE types are recognizable (and annotatable) through a combination of linguistic features of the clause and its main verb, and models have recently been released for their automatic classification (Friedrich et al., 2016).

Our approach is the first we know of to link clause type to argumentative structure, although features of the verb have been widely used in previous work for classifying argumentative vs. non-argumentative sentences. For example, Moens et al. (2007) include verb lemmas and modal auxiliaries as features, and Florou et al. (2013) find that, for Greek web texts related to public policy issues, tense and mood features of verbal constructions are helpful for determining the role of the sentences within argumentative structures.

Our analysis is performed on German texts. Taking the argumentative microtext corpus (Peldszus and Stede, 2015a) as a set of prototypical argumentative text passages, we annotated each clause with its SE type (Section 3).¹ In this way we are able to investigate which SE types are most prevalent in argumentative texts and, further, to link the clause

¹The segmentation of microtexts into clauses is discussed in Section 3.2.

type (via SE label) to the argumentation graphs provided with the microtext corpus (Section 4). We additionally provide SE annotations for a number of non-argumentative (or at least only partially argumentative) German texts, in order to contrast SE type distributions across these text types.

This annotation case study addresses the following questions:

1. Do argumentative text passages differ from non-argumentative text passages with respect to clause type?
2. Do particular clause types correlate with particular elements in the argumentation graphs?
3. Do particular clause types correlate with particular functions of argument components?

Recently, systems have been developed for automatically deriving full argumentative structures from text. Peldszus and Stede (2015b) present a system for automatic argument structure prediction, the first for the microtext corpus. The linguistic features used by the system include discourse connectives, lemmas and parts-of-speech, verb morphology, and dependency patterns. Stab and Gurevych (2016) develop an end-to-end argument structure parser for persuasive texts. The parser performs the entire task pipeline, including segmentation of texts into argumentative vs. non-argumentative components, using sequence labeling at the token level in order to accurately model sentences containing multiple argument components. The features used vary with the particular task, with structural, syntactic, lexical, and lexico-syntactic, as well as discourse connective, features.

None of these systems have investigated **semantic features from the perspective of the clause**. We propose, based on the outcomes from our annotation and analysis, that SE types are worth exploring as features for argument mining.

2 Theoretical background

The phrase *situation entity* refers to the fact that clauses of text evoke situations within a discourse. For example, the previous sentence describes two situations: (i) the meaning of *situation entity*, and (ii) what clauses of text do, in general. The second situation is embedded as part of the first. Notions related to SE type have been widely studied in theoretical linguistics (Vendler, 1957; Verkuyl, 1972; Dowty, 1979; Smith, 1991; Asher, 1993; Carlson and Pelletier, 1995, among others) and have

seen growing interest in computational linguistics (Siegel and McKeown, 2000; Zarcone and Lenci, 2008; Herbelot and Copestake, 2009; Reiter and Frank, 2010; Costa and Branco, 2012; Nedoluzhko, 2013; Friedrich and Pinkal, 2015, for example).

2.1 Situation entity types

We directly follow Friedrich and Palmer (2014) for the inventory of SE types, described below.

The inventory of SE types starts with states and events, including a subtype of events for attributional statements. REPORT-type clauses such as (3) do not necessarily refer to an actual event of speaking but rather indicate a source of information.

1. STATE (S) *Armin has brown eyes.*
2. EVENT (E) *Bonnie ate three tacos.*
3. REPORT (R) *The agency said applications had increased.*

Phenomena such as modality, negation, future tense, and conditionality, when coupled with an EVENT-type clause, cause a coercion to STATE. In brief, such phenomena refer to actual or potential states of the world rather than actual events. Several examples appear below.

- E \rightarrow S: *Carlo should get the job.*
- E \rightarrow S: *Darya did not answer.*
- E \rightarrow S: *If he wins the race, ...*

An important distinction for argumentative texts, as we will see, is between generic sentences and generalizing sentences, or habituals. While the former predicate over classes or kinds, the latter describe regularly-occurring events, such as habits of individuals.

4. GENERIC SENTENCE (GEN): *Birds can fly. / Scientific papers make arguments.*
5. GENERALIZING SENTENCE (GS): *Fei travels to India every year.*

The next category of SE types is broadly referred to as *Abstract Entities*. This type of clause presents semantic content in a manner that draws attention to its epistemic status. We focus primarily on a small subset of constructions - factive and propositional predicates with clausal complements. Of course a wide range of linguistic constructions can be used to convey such information, and to address them all would require a comprehensive treatment of subjective language. In the examples

below, the matrix clause is in both cases a STATE, and the embedded clause (in italics) is both an EVENT and either a FACT, PROPOSITION, or RESEMBLANCE-type situation entity. Note that in the RESEMBLANCE case, it is unclear whether or not the embedded event took place.

6. FACT (F): Georg knows that *Reza won the competition*.
7. PROPOSITION (P): Georg thinks that *Reza won the competition*.
8. RESEMBLANCE (RES): Reza looks like *he won the competition*.

The first sentence of this section, for example, would be segmented and labeled as shown below.

- (i) The phrase *situation entity* refers to the fact [S]
- (ii) that clauses of text evoke situations within a discourse. [F, GEN]

The first segment is a STATE, and the embedded clause is both a FACT, by virtue of being the complement clause following *the fact that*, and a GENERIC SENTENCE, by virtue of predicating over a class of entities (*clauses of text*).

Finally, the labels QUESTION and IMPERATIVE are added to allow complete annotation of texts. Although they fulfill important and varied rhetorical functions, neither of these clause types directly evoke situations.

9. QUESTION (Q): *Why do you torment me so?*
10. IMPERATIVE (IMP): *Listen to this.*

2.2 Discourse modes

The inventory of SE types is motivated by theoretical work (Smith, 2003; Smith, 2005) which aims to determine which specific linguistic features of text passages allow human readers to distinguish, e.g. narrative passages from argumentative passages. Smith identifies five **modes of discourse**: Narrative, Description, Report, Information, and Argument/Commentary. Each mode is linked to linguistic characteristics of the *clauses* which compose the passages.

The set of SE types outlined above allows a complete, comprehensive description of these five discourse modes. The discourse modes approach is related to that of Argumentative Zoning (Teufel, 2000; O’Seaghdha and Teufel, 2014), in which linguistic features of scientific texts are used to distinguish genre-specific types of text passages, such

as Methods or Results. Going back to discourse modes, Smith’s claim, supported by manual textual analysis, is that different modes show different characteristic distributions of SE types.

Previous work shows that this claim is supported at the level of genre (Palmer and Friedrich, 2014), taking genre as an approximation of discourse mode. This approximation is problematic, though, because texts of any genre are in fact composed of multiple text passages which instantiate different discourse modes. In this study we focus down to the level of the text passage, offering the first empirical analysis of argumentative text passages from the perspective of clause types. The analysis looks at German texts, both contrasting purely argumentative texts with mixed texts and, further, analyzing the correlations between SE annotations and argument structure graphs.

3 Data

In this section we describe the corpus of texts we annotate with SE types and the process used for annotating the texts. Section 3.3 presents the first step of the analysis, comparing the distribution of SE types in the argumentative microtexts to those for texts from other genres.

3.1 Argumentative microtext corpus

Our data includes the whole argumentative microtext corpus (Peldszus and Stede, 2015a) which consists of 112 German texts.² Each microtext is a short, dense argument written in response to a question on some potentially controversial topic (e.g. “Should intelligence services be regulated more tightly by parliament?”). The writers were asked to include a direct statement of their main claim as well as at least one objection to that claim. The texts, each of which contains roughly 5 argumentative segments (and no irrelevant segments), were first written in German and professionally translated into English.

An example text from the corpus appears in Table 1, segmented into **argumentative discourse units (ADUs)** as in the original corpus version.

The texts in the corpus are manually annotated according to a scheme based on Freeman’s theory of the macro-structure of argumentation (Freeman, 1991; Freeman, 2011) for representing text-level argumentation structure. This scheme represents

²<https://github.com/peldszus/arg-microtexts>

SE	German/English
GS	Die Geheimdienst müssen dringend stärker vom Parlament kontrolliert werden, <i>Intelligence services must urgently be regulated more tightly by parliament;</i>
S	das sollte jedem nach den Enthüllungen von Edward Snowden klar sein. <i>this should be clear to everyone after the disclosures of Edward Snowden.</i>
S	Die betreffen zwar vor allem die britischen und amerikanischen Geheimdienste, <i>Granted, those concern primarily the British and American intelligence services,</i>
GEN	aber mit denen arbeiten die Deutschen Dienste bekanntermaßen eng zusammen. <i>but the German services evidently do collaborate with them closely.</i>
GS	Deren Werkzeuge, Daten und Knowhow wird schon lange zu unserer Überwachung genutzt. <i>Their tools, data and expertise have been used to keep us under surveillance for a long time.</i>

Table 1: Sample microtext (`micro_b005`), both German and English versions, with SE labels.

an **argument**, consisting of one conclusion and one or more premises, as a “hypothetical exchange” (Peldszus and Stede, 2013a) between the **proponent** and the **opponent**, who respectively defend or question a specific claim. Note that Peldszus and Stede (2013a) use the term “argument” to describe the complex of premises put forward in favor of a conclusion, while premises and conclusions are propositions expressed in the text.

The microtexts are segmented into “elementary units of argumentation” (sometimes called ADUs as above), which consist of either the conclusion or a single premise. Each ADU corresponds to a structural element in an argument graph (Figure 1). In these texts, **conclusions** are only associated with the proponent, but **premises** can be associated with either the proponent or the opponent.

Round nodes in the graph are proponent nodes (premises or conclusions) while square ones are opponent nodes (premises only). In addition, several different supporting and attacking moves (also called “argumentative functions” of a segment (Peldszus and Stede, 2013b)) are labelled and represented by the arcs connecting the nodes. The most frequent argumentative functions are:

- **support:** a premise supporting a conclusion
- **rebuttal:** a premise which attacks a conclusion or premise by challenging the acceptability of the proposition being attacked
- **undercut:** a premise which attacks by challenging the acceptability of an inference between two propositions

The argument graph for our sample text appears in Figure 1. Here we see that the second segment

supports the conclusion (first segment), while the third segment undercuts segment two. This undercutting move itself is then undercut by the fourth segment (a proponent node), which is in turn supported by segment five.

3.2 Annotation process

The granularity of situation-evoking clauses is different from that of ADUs, requiring that the texts be re-segmented prior to SE annotation. Table 3 illustrates a microtext with more SE segments than ADUs. For segmentation we use DiscourseSegmenter (Sidarenka et al., 2015), a python package offering both rule-based and machine-learning based discourse segmenters.³ After preprocessing texts with the Mate Tools,⁴ we use DiscourseSegmenter’s rule-based segmenter (`edseg`), which employs German-specific rules to determine the boundaries of elementary discourse units in texts. Because DiscourseSegmenter occasionally over-split segments, we did a small amount of post-processing. On average, one ADU contains 1.16 SE segments. Table 2 shows the number of segments of each type, as well as the distribution of both argument components and SE types over the segments.

In addition to labeling each segment with its SE type, we also annotate three important verb- or clause-level features: (a) lexical aspect (dynamic/stative) of the main verb, (b) whether the main referent of the clause is generic or non-

³<https://github.com/WladimirSidorenko/DiscourseSegmenter>

⁴<http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/matetools.html>

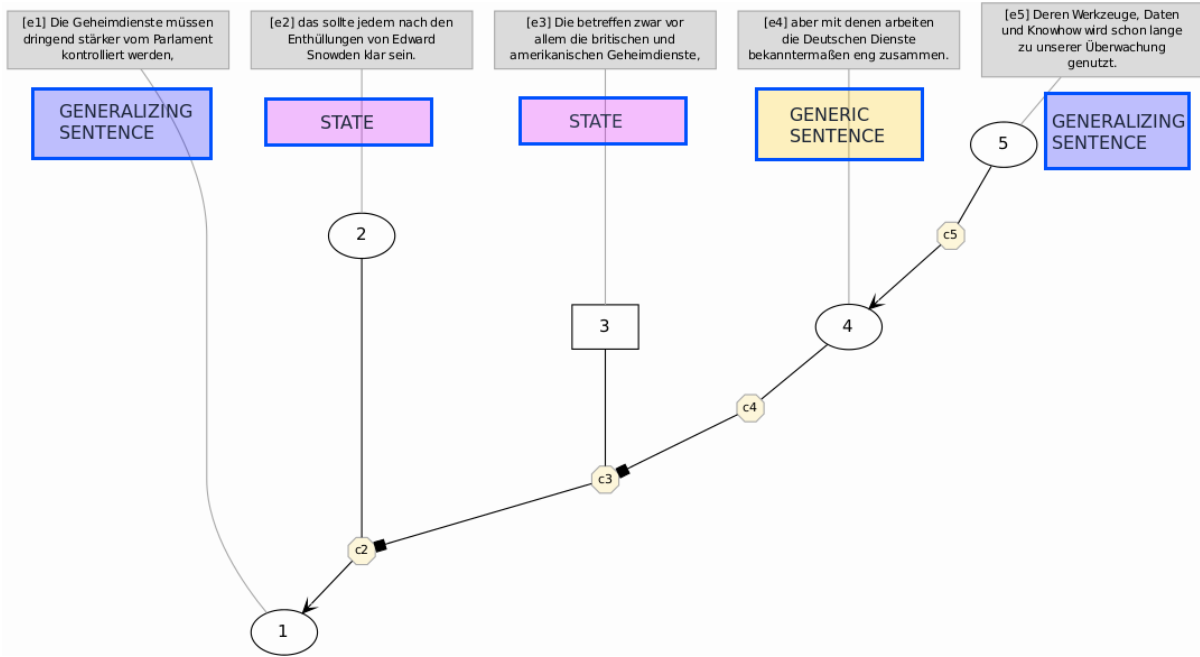


Figure 1: Argument graph for sample microtext (`micro_b005`), with SE labels added.

#texts=112	#ADUs=576	#SE Segs=668					
concl./premise	112 / 464	state	212 (0.32)	generic	320 (0.48)	question	10 (0.01)
prop./opp.	451 / 125	event	57 (0.09)	generalizing	67 (0.10)	resemblance	2 (0.00)

Table 2: Microtext corpus used for this analysis (percentages in brackets). ADUs are the argument component segments in the original data; SE Segs are situation entity segments.

	German/English (+SE-Labels)
CONCL.	Alternative Heilpraktiken sollten genauso wie ärztliche Behandlungen bezuschusst werden, (GEN) <i>Alternative treatments should be subsidized in the same way as conventional treatments,</i>
PREMISE	weil beide Wege zur Verhinderung, Minderung oder Heilung einer Krankheit führen könnten. (GEN) <i>since both methods can lead to the prevention, mitigation or cure of an illness.</i>
PREMISE	Zudem müsste es im Sinne der Krankenkassen liegen, (GEN) \\ Naturheilpraktiken als ärztliche Behandlung anzuerkennen, (GEN) <i>Besides it should be in the interest of the health insurers \\ to recognize alternative medicine as treatment,</i>
PREMISE	da eine Chance der Genesung besteht. (ST) <i>since there is a chance of recovery.</i>
PREMISE	Es spielt dabei doch keine Rolle, (GS) \\ dass die Behandelnden nicht den 'Ärzte-Status' tragen. (GEN) <i>It doesn't matter after all \\ that those who administer the treatment don't have 'doctor status'.</i>

Table 3: Sample microtext (`micro_b010`), both German and English versions, with SE labels.

generic, and (c) whether the clause is habitual (a pattern of occurrences), episodic (a fixed number of occurrences), or static (an attribute, characteristic, or quality). Using these feature values in a decision tree has been shown to improve human agreement on the SE type annotation task (Friedrich and Palmer, 2014).

Annotators and annotator training. Each text was annotated by two trained (but novice) student

annotators and one expert annotator. Following Mavridou et al. (2015), with a modified and translated version of an existing SE annotation manual,⁵ student annotators were trained on a set of longer texts from different genres, automatically segmented as described above: fiction (47 segments), reports (42 segments), TED talks (50 seg-

⁵www.coli.uni-saarland.de/projects/sitent/page.php?id=resources

ments), and commentary (127 segments).

Inter-annotator agreement. We compute agreement separately for SE type and for the three features introduced above, as shown in Table 4 (reported as Cohen’s Kappa). Numbers in brackets represent IAA for the training texts.

level	Observed Agreement	Chance Agreement	Cohen’s Kappa
Main Referent	0.62 (0.76)	0.50 (0.57)	0.23 (0.45)
Aspect	0.84 (0.91)	0.50 (0.52)	0.69 (0.81)
Habituality	0.68 (0.70)	0.26 (0.27)	0.57 (0.58)
Situation Entity	0.52 (0.61)	0.21 (0.22)	0.40 (0.50)

Table 4: Inter-annotator agreement on microtexts (and training texts).

The table shows that the IAA for our training set is (slightly) better than the IAA we gained for the microtexts. The best results we obtained in the category Aspect (0.69 K), the worst in the category Main Referent (0.23 K).

Error Analysis. As reported by several studies (Peldszus and Stede, 2013b), the annotation of argumentative texts is a difficult task for humans. Our annotators often disagreed about the genericity of the Main Referent (binary: generic/non-generic). The SE-labels that caused most disagreement among the annotators are STATE vs. GENERIC SENTENCE (39% of all disagreements regarding SE-type), followed by STATE vs. GENERALIZING SENTENCE (22%).

This first annotation round suggests that we might benefit from:

1. adapting the SE annotation scheme for (purely) argumentative texts; and
2. training annotators specifically to deal with (purely) argumentative texts.

Given the relatively low agreement, for the analysis we produce a gold standard annotation by re-annotating all segments for which the two student annotators disagree about the SE type label. This third annotation is done by an expert annotator (one of the authors). For the segments which needed re-annotation, the expert annotator agreed with one of the two student annotators regarding SE label 87% of the time.

3.3 Distributions

One key claim of Smith (2003) is that text passages in different discourse modes have different

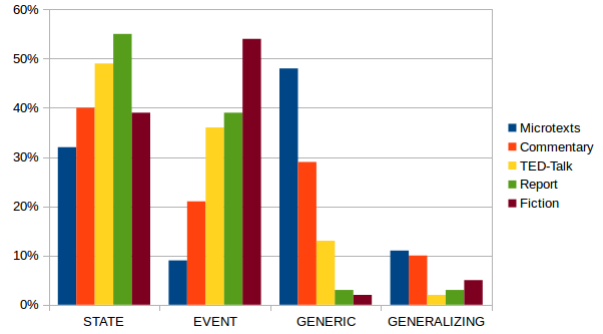


Figure 2: Distribution of SE types in all genres.

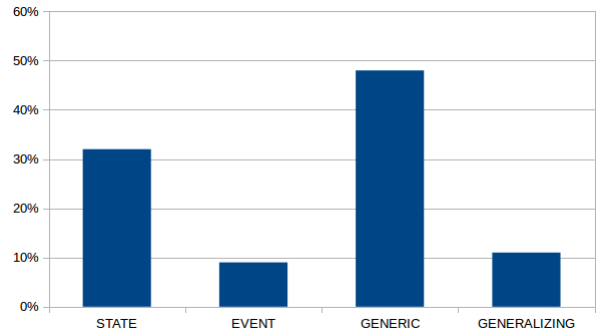


Figure 3: Distribution of SE types in microtexts.

predominant SE types. We investigate this claim by comparing the distribution of SE types in the microtext subcorpus to those in the training texts (see Figure 2). The microtexts (Figure 3), which can be described as ‘purely’ argumentative texts, are characterized by a high proportion of generic and generalizing sentences and very few events, while reports and talks, for example, contain a high proportion of states. Fiction is characterized by a high number of events. The genre commentary, which contains many argumentative passages but is not as ‘purely’ argumentative as the microtexts, is most likely to be comparable to the microtexts. This finding suggests that SE types could be helpful for modeling argumentative regions of text.

4 Analysis

Extraction of argument graphs. As a next step, we look at the correspondences between SE type labels and various aspects of the argument graph components (ADUs):

- conclusion vs. premise
- proponent premise vs. opponent premise
- support vs. attack by rebuttal vs. attack by undercut

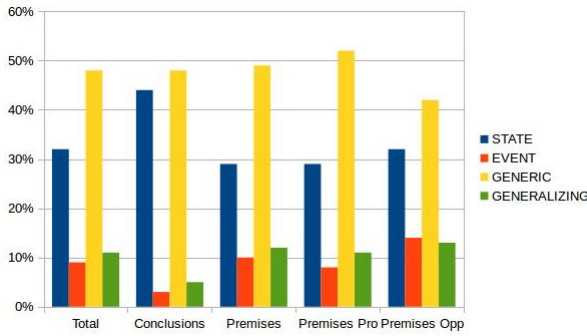


Figure 4: Correlations between SE types and argument components.

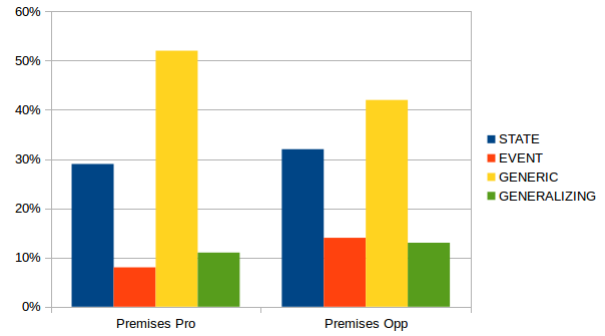


Figure 6: Correlations between SE types and proponent/opponent argument components.

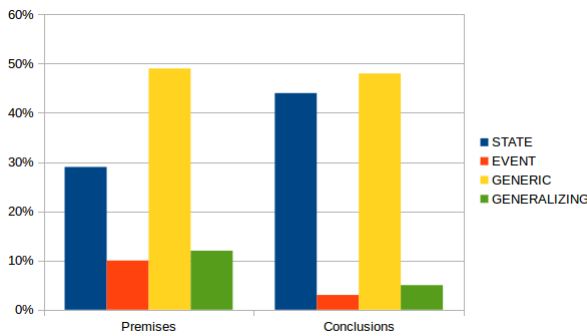


Figure 5: Correlations between SE types and conclusions/premises.

For this analysis, we match the ADU segments extracted from the argument graphs to the situation segments produced by our own segmentation routine. Because the segments of the microtext corpus (ADUs) sometimes contain several situation segments, it can happen that several SE labels are assigned to one ADU. The results of these analyses are presented below.

4.1 Correlations between SE types and argument components

The mapping of the two sets of annotations reveals interesting correlations between SE types and argument components. Figure 4 shows the overall distribution of SE types over specific argument components. First, conclusions are almost exclusively either GENERIC SENTENCES (48%) or STATES (44%), while premises also consist of GENERALIZING SENTENCES (12%) and EVENTS (10%).

An example of a generic conclusion is given below:

GEN: *Nicht jeder sollte verpflichtet sein, den Rundfunkbeitrag zu bezahlen.* (Translation: *Not everyone should be obliged to pay the TV & radio*

licence.)

Figure 5, where all premises are contrasted with all conclusions, shows this difference even more clearly. The STATE label is more frequent for conclusions (44%) than for premises (29%). On the other hand, premises contain 10% EVENT labels, while there are only 3% EVENT labels in the set of conclusions.

The SE types also correspond to the distinction between premises that support a conclusion (“pro”-label) and premises that attack a conclusion (“opp”-label), as Figure 6 indicates. First, proponent ADUs contain more GENERIC SENTENCE labels (52%) than opponent units (42%). Furthermore, the STATE and EVENT labels are more frequent within opponent units (32% and 14%), while they make up only 29% and 8% in proponent units.

To illustrate this, below we show an extract of an argument consisting of a stative opponent premise (in bold face) attacking a conclusion:

*Die Krankenkassen sollten Behandlungen beim Natur- oder Heilpraktiker nicht zahlen, **es sei denn der versprochene Effekt und dessen medizinischer Nutzen sind handfest nachgewiesen.***

(Translation: *Health insurance companies should not cover treatment in complementary medicine **unless the promised effect and its medical benefit have been concretely proven.***)

Of course, the corpus on which these investigations are carried out is quite small, and the phenomena we observe can be interpreted solely as tendencies. Nevertheless, we hypothesize that the prevalence of the GENERIC SENTENCE and GENERALIZING SENTENCE labels, as well as the absence (or rareness) of the EVENT label are indicators for

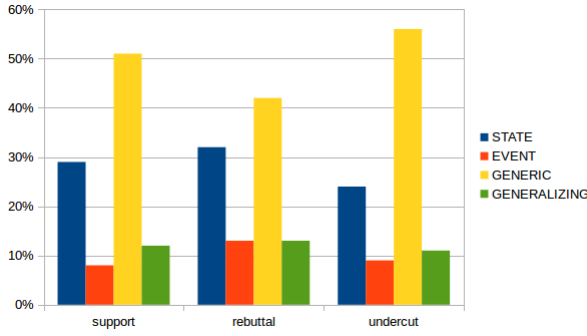


Figure 7: Correlations between SE types and argumentative functions

argumentative text passages. Furthermore, such prevalences (or also rarenesses) can be interpreted as indicators for particular argument components such as proponent premises or conclusions.

4.2 Correlations between SE types and argumentative functions

There are also some interesting correlations between SE types and the argumentative functions of premises. In this study we focus on the three most frequent functions in our data, namely support, rebuttal and undercut. Table 5 gives an overview of their distribution within the microtexts.

function	pro/opp	n
support	proponent	250
support	opponent	13
rebut	proponent	12
rebut	opponent	96
undercut	proponent	51
undercut	opponent	12

Table 5: Distribution of argumentative functions within the microtexts.

It is worth mentioning that the rebuttals in the microtext corpus seem to differ slightly from all of the other segments in terms of the frequency of the label GENERIC SENTENCE. While this is by far the most frequent label in the microtexts (overall frequency of 48%), here it has only a frequency of 42%, followed by the label STATE with a frequency of 32% (see Figure 7). Rebuttals are less strongly biased toward GENERIC SENTENCES, compared to other argumentative functions. In contrast, supporting and undercutting moves are characterized by an above-average number of GENERIC SENTENCES

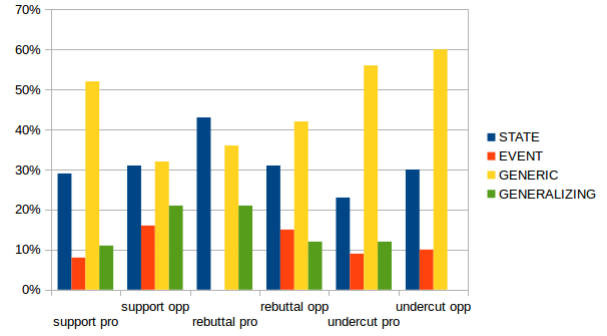


Figure 8: Correlations between SE types and argumentative functions with regard to pro/opp

(51% and 56%).

To give a better idea of how rebuttals of different SE types look, below we show two examples, one a STATE and the other one GENERIC SENTENCE:

Die Krankenkassen sollten Behandlungen beim Natur- oder Heilpraktiker nicht zahlen, es sei denn der versprochene Effekt und dessen medizinischer Nutzen sind handfest nachgewiesen. (STATE)

(Translation: *Health insurance companies should not cover treatment in complementary medicine unless the promised effect and its medical benefit have been concretely proven.*)

Ja, seinen Müll immer ordentlich zu trennen ist nervig und mühselig. [...] Aber immer noch wird in Deutschland viel zu viel Müll produziert. (GENERIC SENTENCE)

(Translation: *Yes, it's annoying and cumbersome to separate your rubbish properly all the time. [...] But still Germany produces way too much rubbish.*)

Finally, Figure 8 shows the distribution of SE types among the different argumentative functions, separated into proponent and opponent premises.

According to this distribution, support moves, for example, can be distinguished by SE type into premises supporting the opponent and those supporting the proponent. While proponent support premises contain clearly more GENERIC SENTENCES (52%) than STATES (29%), the number of GENS and STATES within opponent support premises is almost equal (32% and 31%).

The following is a generic premise that supports the proponent, contrasted with a STATE premise supporting the opponent:

*Mit der BA-Arbeit kann man jedoch die Interessen und die Fachkenntnisse besonders gut zeigen. **Schließlich ist man nicht in jedem Fach sehr gut.***
(GENERIC SENTENCE)

(Translation: *With a BA dissertation one can, however, demonstrate interests and subject matter expertise particularly well. **After all one doesn't excel in every subject.***)

*Es bleibt jedoch fragwürdig, ob die tatsächliche Durchführung derartiger Kontrollen auch gleichzeitig zur stärkeren Einhaltung von Gesetzen führt, **denn jetztendlich liegt diese Entscheidung in den Händen des jeweiligen Regierungschefs.***
(STATE)

(Translation: *Yet it remains questionable whether the actual implementation of such supervision would at the same time lead to a stronger observance of laws, **as ultimately this decision is in the hands of the respective government leaders.***)

These results suggest that SE types could be helpful even for a finer-grained analysis of argumentative functions. Nonetheless we would reiterate here that our results are based on a small dataset only and need to be confirmed by further experiments. Therefore, annotations of larger texts with a mixture of argumentative and non-argumentative passages are already underway.

5 Conclusion

This paper presents an annotation study whose goal is to dig into the usefulness of semantic clause types (in the form of situation entity labels) for mining argumentative passages and modeling argumentative regions of text. This is the first study to label argumentative texts (microtexts, in this case) with situation entity types at the clause level. We have explored the correlation of SE classes with argumentative text genres, particularly in comparison to non-argumentative texts. We have also looked at the correlations between SE types and both specific argument components (premise vs. conclusion, opponent vs. supporter) and specific argumentative functions (support, rebuttal, undercut).

We do our analysis on German texts, but we fully expect that the results should carry over to other languages, as the link between SE type distribution and discourse mode has been shown to hold cross-

linguistically. Due to the small dataset, our results can be interpreted solely as tendencies which have to be confirmed by more extensive studies in the future. Nonetheless there is some evidence that the observed tendencies can be deployed for automatic recognition and fine-grained classification of argumentative text passages.

In addition to the ongoing annotation work which will give us more data to analyze, we intend to cross-match SE types with the newly-available discourse structure annotations for the microtext corpus (Stede et al., 2016). We would additionally explore the role of modal verbs within this intersection of SE type and argument structure status. The end goal of this investigation, of course, is to deploy automatically-labeled SE types as features for argument mining.

Acknowledgments

We want to thank Michael Staniek for building up the annotation tool and preprocessing the texts. Sabrina Effenberger and Rebekka Sons we thank for the annotations and their helpful feedback on the annotation manual. We would also like to thank the reviewers for their insightful comments and suggestions on the paper.

This research is funded by the Leibniz Science-Campus *Empirical Linguistics and Computational Language Modeling*, supported by Leibniz Association grant no. SAS-2015-IDS-LWC and by the Ministry of Science, Research, and Art of Baden-Württemberg.

References

- Nicholas Asher. 1993. *Reference to Abstract objects in Discourse*. Kluwer Academic Publishers.
- Gregory N. Carlson and Francis Jeffrey Pelletier, editors. 1995. *The Generic Book*. University of Chicago Press.
- Francisco Costa and António Branco. 2012. Aspectual type and temporal relation classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 266–275. Association for Computational Linguistics.
- David Dowty. 1979. *Word Meaning and Montague Grammar*. Reidel.
- Eirini Florou, Stasinou Konstantopoulos, Antonis Kukurikos, and Pythagoras Karampiperis. 2013. Argument extraction for supporting public policy formulation. In *Proceedings of the 7th Workshop on*

- Language Technology for Cultural Heritage, Social Sciences, and Humanities.*
- James B. Freeman. 1991. *Dialectics and the Macrostructure of Argument*. Foris.
- James B. Freeman. 2011. *Argument Structure: Representation and Theory*, volume 18 of *Argumentation Library*. Springer.
- Annemarie Friedrich and Alexis Palmer. 2014. Situation entity annotation. In *Proceedings of the Linguistic Annotation Workshop VIII*.
- Annemarie Friedrich and Manfred Pinkal. 2015. Discourse-sensitive Automatic Identification of Generic Expressions. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*, Beijing, China, July.
- Annemarie Friedrich, Alexis Palmer, and Manfred Pinkal. 2016. Situation entity types: automatic classification of clause-level aspect. In *Proceedings of ACL 2016*.
- Aurelie Herbelot and Ann Copestake. 2009. Annotating genericity: How do humans decide? (A case study in ontology extraction). *Studies in Generative Grammar 101*, page 103.
- Kleio-Isidora Mavridou, Annemarie Friedrich, Melissa Peate Sorensen, Alexis Palmer, and Manfred Pinkal. 2015. Linking discourse modes and situation entities in a cross-linguistic corpus study. In *Proceedings of the EMNLP Workshop LSDSem 2015: Linking Models of Lexical, Sentential and Discourse-level Semantics*.
- Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau, and Chris Reed. 2007. Automatic detection of arguments in legal texts. In *Proceedings of ICAIL 2007*.
- Anna Nedoluzhko. 2013. Generic noun phrases and annotation of coreference and bridging relations in the Prague Dependency Treebank. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 103–111.
- Diarmuid O’Searghda and Simone Teufel. 2014. Unsupervised learning of rhetorical structure with untopic models. In *Proceedings of COLING*.
- Alexis Palmer and Annemarie Friedrich. 2014. Genre distinctions and discourse modes: Text types differ in their situation type distributions. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and NLP*.
- Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. 2007. A sequencing model for situation entity classification. In *Proceedings of ACL*.
- Andreas Peldszus and Manfred Stede. 2013a. From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1):1–31.
- Andreas Peldszus and Manfred Stede. 2013b. Ranking the annotators: An agreement study on argumentation structure. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 196–204, Sofia, Bulgaria, August. Association for Computational Linguistics.
- Andreas Peldszus and Manfred Stede. 2015a. An annotated corpus of argumentative microtexts. In *Proceedings of the First European Conference on Argumentation: Argumentation and Reasoned Action*.
- Andreas Peldszus and Manfred Stede. 2015b. Joint prediction in MST-style discourse parsing for argumentation mining. In *Proceedings of EMNLP*.
- Nils Reiter and Anette Frank. 2010. Identifying Generic Noun Phrases. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 40–49, Uppsala, Sweden, July. Association for Computational Linguistics.
- Uladzimir Sidarenka, Andreas Peldszus, and Manfred Stede. 2015. Discourse Segmentation of German Texts. *Journal for Language Technology and Computational Linguistics*, 30(1):71–98.
- Eric V. Siegel and Kathleen R. McKeown. 2000. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–628.
- Carlota S. Smith. 1991. *The Parameter of Aspect*. Kluwer.
- Carlota S Smith. 2003. *Modes of discourse: The local structure of texts*, volume 103. Cambridge University Press.
- Carlota S Smith. 2005. Aspectual entities and tense in discourse. In *Aspectual inquiries*, pages 223–237. Springer.
- Christian Stab and Iryna Gurevych. 2016. Parsing argumentation structures in persuasive essays. arXiv preprint, under review, April.
- Manfred Stede, Stergos Afantenos, Andreas Peldszus, Nicholas Asher, and Jeremy Perret. 2016. Parallel discourse annotations on a corpus of short texts. In *Proceedings of LREC*.
- Simone Teufel. 2000. *Argumentative zoning: Information extraction from scientific text*. Ph.D. thesis, University of Edinburgh.
- Zeno Vendler, 1957. *Linguistics in Philosophy*, chapter Verbs and Times, pages 97–121. Cornell University Press, Ithaca, New York.
- H.J. Verkuyl. 1972. *On the Compositional Nature of the Aspects*. Reidel.
- Alessandra Zarcone and Alessandro Lenci. 2008. Computational models of event type classification in context. In *Proceedings of LREC2008*.