Implicit Knowledge in Argumentative Texts: An Annotated Corpus

Maria Becker, Katharina Korfhage, Anette Frank
Institute of Computational Linguistics Heidelberg University
(mbecker|korfhage|frank)@cl.uni-heidelberg.de

Abstract

When speaking or writing, people omit information that seems clear and evident, such that only part of the message is expressed in words. Especially in argumentative texts it is very common that (important) parts of the argument are implied and omitted. We hypothesize that for argument analysis it will be beneficial to reconstruct this implied information. As a starting point for filling such knowledge gaps, we build a corpus consisting of high-quality human annotations of missing and implied information in argumentative texts. To learn more about the characteristics of both the argumentative texts and the added information, we further annotate the data with semantic clause types and commonsense knowledge relations. The outcome of our work is a carefully designed and richly annotated dataset, for which we then provide an in-depth analysis by investigating characteristic distributions and correlations of the assigned labels. We reveal interesting patterns and intersections between the annotation categories and properties of our dataset, which enable insights into the characteristics of both argumentative texts and implicit knowledge in terms of structural features and semantic information. The results of our analysis can help to assist automated argument analysis and can guide the process of revealing implicit information in argumentative texts automatically.

Keywords: implicit knowledge, argumentation, annotation, semantic clauses, commonsense knowledge relations, ConceptNet

1. Introduction

In everyday communication as well as in written texts people omit information that seems clear and evident, such that only part of the message needs to be expressed in words (Grice, 1975). While this information can easily be filled in by the hearer, a computational system typically does not possess commonsense or domain-specific knowledge that is needed to reconstruct the implied information. Especially in argumentative texts it is very common that (important) parts of the argument such as warrants are implied and omitted (Rajendran et al., 2016; Becker et al., 2017b; Hulpus et al., 2019). This leads us to the assumption that the logic of an argument is in general not fully recoverable from what is explicitly said, and that for argument analysis it will be beneficial to reconstruct such implied information.

We aim to fill such gaps by identifying and inserting knowledge that connects given statements. To perform this, we want to learn from human-generated data of missing and implied information. This motivates the current work, in which we gather high-quality annotations of implied knowledge in the form of simple natural language sentences in English. The annotations are performed on pairs of argumentative units from the Microtexts Corpus (Peldszus and Stede, 2015), a very concise and focused argumentation dataset which is already annotated with argumentative components and relations such as *support, rebuttal* or *undercut*. For all unit pairs they are presented with, annotators are asked to add the information that makes the connection between the units explicit, using short and simple sentences. To learn more about the nature and characteristics of both the argumentative texts and the added information, we further annotate the data with two specific semantic information types: semantic clause types (Friedrich and Palmer, 2014) and ConceptNet knowledge relations (Speer and Havasi, 2012; Havasi et al., 2009), which were both found to be characteristic for argumentative texts (Becker et al., 2016a; Becker et al., 2017b). The outcome of our work is a carefully designed and richly annotated dataset,[1] for which we provide an in-depth analysis by investigating characteristic distributions and correlations between the assigned labels.

The contributions of this work are: (i) high-quality annotations of implicit knowledge on the argumentative Microtext corpus, (ii) characterization of the argumentative units from the Microtext corpus and the inserted sentences in terms of semantic clause types and commonsense knowledge relations; and (iii) an in-depth study of properties and correlations of the assigned labels. The dataset will be made public as an extension to the Microtext corpus (Peldszus and Stede, 2015) to support further research in argument analysis.

2. Related Work

Finding and Adding Implicit Knowledge in Arguments. Relatively little attention has been devoted so far to the task of finding and adding implicit knowledge in arguments, which is closely related to the task of **enthymeme** reconstruction. Enthymemes – arguments with missing propositions – are common in natural language and particularly in argumentative texts (Rajendran et al., 2016). Razuvayevskaya and Teufel (2016) present a feasibility study on the automatic detection of enthymemes in real-world texts and find that specific discourse markers (e.g. *let alone, because*) can signal enthymemes. Using these as trigger words, they reconstruct enthymemes from the local context, while Rajendran et al. (2016) retrieve and fill missing propositions in arguments from similar or related arguments.

Becker et al. (2016a, 2016b) show that argumentative texts are rich in **generic and generalizing sentences**, which are semantic clause types (Friedrich and Palmer, 2014) that often express commonsense knowledge. We will show that large portions of implied knowledge in argumentative texts are naturally stated using these clause types.

In their attempt to reconstruct implicit knowledge, Boltuzic and Snajder (2016) find that the **claims** that users make in

[1] The data is available at `https://github.com/maria-becker/IKAT-EN`.

online debate platforms often build on implicit knowledge. They show that the amount of implicitness is dependent on genre and register and point out that the reconstruction of implicit premises can be helpful for claim detection.

Recently, Hulpus et al. (2019) point out the relevance of reconstructing implicit knowledge for understanding arguments in a computational setting, by proposing the task of **argument explicitation**, which they define as a task that makes explicit both (i) the structure of a natural language argument, as well as (ii) the background knowledge the argument is built on, in the form of implicit premises or contextual knowledge.

These studies reinforce the view that a substantial amount of knowledge is needed for the correct interpretation and analysis of argumentative texts, thus filling knowledge gaps in argumentative texts will be beneficial for argument analysis.

Related Datasets. Boltuzic and Snajder (2016) release a small dataset with human-provided **implicit premises** based on data from **online debate platforms**, consisting of 125 claim pairs annotated with the premises that connect them. In contrast to our approach they asked the annotators to provide the premises that bridge the gap between the two claims without giving any further instructions, resulting in a substantial variance in both the wording and the average number of premises.

Becker et al. (2017b) design a process for obtaining high-quality **implied knowledge annotations** for **German argumentative microtexts** (Peldszus and Stede, 2015), in the form of simple natural language statements which are then characterized with semantic clause types and commonsense knowledge relations. Since the decision of what exactly is missing and how detailed such information should be can be subjective, they propose several steps to promote agreement among the annotators and monitor its evolution using textual similarity computation. The implicit knowledge annotations we present in this paper are also based on argumentative microtexts (Peldszus and Stede, 2015), thus our corpus can be seen as an extension of the corpus published by Becker et al. (2017b). The main differences are that (i) our data is in English (as opposed to German), (ii) the semantic clause types and commonsense knowledge relations are not only annotated for the inserted sentences, but also for the argumentative texts themselves, (iii) our corpus includes more annotated unit pairs, and that (iv) in our corpus all annotations are conducted by expert annotators.

Habernal et al. (2018) present the **argument reasoning comprehension task**, where given an argument with a claim and a premise, the goal is to choose the correct implicit warrant from two options. Both warrants are plausible and lexically close, but lead to contradicting claims. They provide a dataset where Amazon Turkers added warrants for 2k arguments from news comments. As opposed to our dataset, the annotators were only supposed to fill in the gap between a pair of claim and premise, while we consider larger arguments consisting of a claim and *several* premises. Furthermore, we annotate implicit information not only between claim and premises, but between all adjacent argument units and all argument units that stand in a direct argumentative relation (cf. Sec. 3.2). The second

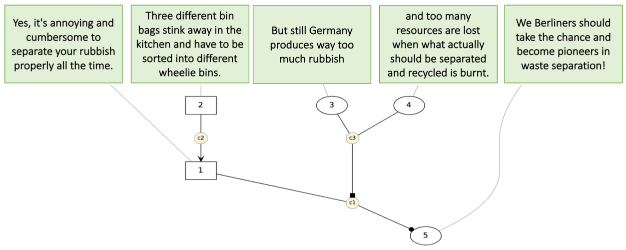

Figure 1: Example of a microtext (argument graph).

major difference is that in Habernal et al. (2018) the annotators were only asked to add *one* warrant (one sentence) per argument, while we assume that more than one sentence might be needed to fill a knowledge gap in an argument.

3. Enriching Argumentative Texts with Implicit Knowledge

3.1. General Annotation Procedure

The main goal of our annotation project is to uncover and characterize implicit knowledge that connects a given pair of argumentative units. This overall objective is subdivided into two consecutive annotation tasks:

i. First, we ask the annotators to detect missing knowledge that connects a pair of argumentative units, and to express this knowledge in terms of simple natural language statements.

ii. In the next step (cf. Sec. 4), the annotators are tasked with labeling both the inserted sentences and the given argumentative text units with characterizing semantic information. The annotation types that we select are *semantic clause types* (Friedrich and Palmer, 2014) and *commonsense knowledge relations*, following the ConceptNet relation inventory (Havasi et al., 2009).

3.2. The Microtext Corpus

As basis for our annotations we use the argumentative Microtext corpus (Peldszus and Stede, 2015), which consists of 112 microtexts. The corpus was created in German and has been translated to English. In this work we use only the English version (for annotations on the German version, cf. Becker et al. (2017b)). Each microtext is a short, dense argument consisting of roughly five elementary units of argumentation, so called argumentative units (Peldszus and Stede, 2015). The texts are written in response to a question on a potentially controversial issue (e.g. *Should there be a cap on rent increases for a change of tenant?*). Writers were asked to include a direct statement of their main claim as well as at least one objection to that claim. The generated arguments were then manually annotated with argumentation graphs (one graph per microtext, cf. Fig. 1 for an example) according to a scheme based on Freeman's theory of the macro-structure of argumentation (Freeman, 2011). The nodes in the graph are argumentative units and the edges are argumentative relations between them. The most frequent relations are *support* (premise which supports a conclusion or another premise), *rebuttal* (premise that attacks a conclusion or premise by challenging its acceptability), or *undercut* (premise which attacks the acceptability of an argumentative relation between two propositions).

(1-a) *BER should be re-conceptualized from scratch*
(1-b) *even if billions of Euros have already been invested in the existing airport project.*
(1-c) *BER is an airport.*

(2-a) *Capital punishment is not a solution*
(2-b) *as it cannot be ruled out that the judicial process may make mistakes.*
(2-c-I) *In a judicial process it is decided about capital punishment.*
(2-c-II) *Mistakes don't lead to solutions.*

Figure 2: Example annotations for explicating implicit knowledge (c) that connects argumentative units (a&b).

For our work we extract pairs of argumentative units that either stand in a direct argumentative relation (that means units that are directly connected in the argument graph), or units that are adjacent to each other, or both. In sum we extract 719 pairs of argumentative units which we then provide to our annotators to perform several different annotation tasks that we describe in the following sections.

3.3. Annotating Implicit Knowledge

We asked our annotators to detect whether the connection between the pair of units is made fully explicit by the text, and if this is not the case, to explain the missing connection by providing one or more sentences that make this connection explicit. Our annotators were supposed to add as few sentences as possible and to make these sentences very simple (if possible one fact per sentence) in order to retrieve the minimal amount of information that is needed to connect the two units and to avoid too detailed explanations. Since some unit pairs only make sense within a larger context, we also displayed the full microtext for every pair. Figure 2 shows two examples of the annotations, where in the first one the main claim *1-a* is attacked by statement *1-b*, while in the second one the premise *2-b* supports the main claim *2-a*. The knowledge underlying the connection between the main claim and the premise in both cases is made explicit in *c* respectively, whereby for the first example one and for the second example two sentences have been inserted.

The difficulty of eliciting such implicit knowledge in an annotation task is that intuitions about which knowledge exactly is missing may be different between annotators, and even if their intuitions match, the phrasing may be different, structurally or in terms of lexical choice. In order to enforce agreement and to assess the quality of the annotations, Becker et al. (2017b) design a multi-step annotation process where annotators are asked to review and revise each other's annotations, whereby the evolution of agreement during this process is monitored using computational measures of semantic textual similarity. Becker et al. (2017b) use five annotators for each argumentative unit pair, while we train two expert annotators with a linguistic background who produce two versions of the implicit knowledge, which then serve as the basis for the final gold standard produced by another expert annotator (one of the authors). This final adjudicated corpus provides the basis for the second annotation step (Sec. 4) and our analyses in Sec. 5.

3.4. Annotation Statistics

Annotator Agreement. Building on the insights of Becker et al. (2017b), we calculate the semantic similarity of the two initial annotations in order to evaluate the agreement between the annotators and compare it to the similarity scores reported in Becker et al. (2017b). Following Becker et al. (2017b), we quantify the distance between the annotations using the Word Mover's Distance (Kusner et al., 2015) as implemented in *gensim*[2]. The Word Mover's Distance (WMD) measures the dissimilarity between two documents as the aggregated minimum distance in an embedding space that the (non-stopword) words of one document need to "travel" to reach the (non-stopword) word of another document. As embeddings, we use 300-dimensional skip-gram word2vec embeddings trained on part of the Google News dataset (100 billion words, Mikolov et al. (2013)).
We compare the complete annotation for each argumentative unit pair (as opposed to sentence by sentence) and measure a WMD distance score of 1.97. Becker et al. (2017b) compare distance scores between implicit knowledge annotations produced in early vs. later stages of their multi-step annotation procedure. In their first two rounds of annotations which include initial annotations and mutual editing and correcting, they compute a WMD of 2.2 and 3.08, and in the third round where the corrected annotations are merged by new annotators, the WMD decreases to 1.89, demonstrating the evolution of annotator agreement. Our score of 1.97 is closest to the score reported for the third round, which we interpret as sufficient agreement between the annotators.

4. Annotating Argumentative Texts and Implicit Knowledge with Additional Information

Learning from Human Annotations. We hypothesize that the more we know about the knowledge that is needed to establish links between (argumentative) sentences, the easier it will be to reconstruct them automatically. All of the following tasks are therefore designed with the ultimate goal of learning more about the properties of the sentences that were stated by our annotators to make the missing information explicit, *within* their surrounding explicit context.
We expect semantic clause types to be useful features for characterizing argumentative texts, implicit knowledge *and* their interaction, since clause types have shown to be relevant for interpreting semantics at the clause level and discourse structure (cf. Friedrich and Palmer (2014)). Furthermore, Becker et al. (2016b) showed that the distribution of these clause types is distinctive for argumentative texts compared to other genres in terms of particularly high ratios of generic and generalizing sentences.
We furthermore expect ConceptNet to be a useful resource for finding and characterizing implicit sentences, since implied information is usually commonsense knowledge that seems clear and evident and is for that reason omitted. ConceptNet provides exactly that kind of information, since it contains commonsense facts about the world and everyday

[2] `https://radimrehurek.com/gensim`

Genre	GEN	GNZ	STA	EVT
Impl. Information	0.84	0.02	0.13	0.01
Microtexts	0.64	0.05	0.24	0.02
Report	0.03	0.04	0.54	0.39
TED Talk	0.12	0.03	0.49	0.36
Fiction	0.02	0.05	0.39	0.54

Table 1: Distribution of the most frequent Semantic Clause Types among different genres (expressed as percentages)

life (cf. Sec. 4.2). Also, the relation inventory of ConceptNet is targeted for capturing commonsense knowledge, and we therefore expect it to be appropriate for labeling and characterizing implicit knowledge.

What additionally makes clause types and commonsense relations attractive features for analyzing and characterizing argumentative texts and implicit knowledge is that recently for both – semantic clause types (Becker et al. (2017a)) and commonsense relations (Becker et al. (2019)) – automated classification models have been published, which can be used for pre-labeling the given texts and therefore facilitate the automatic analysis of arguments and implicit knowledge.

4.1. Annotating Semantic Clause Types

Inventory and Annotation Process. We asked the annotators to characterize both the argumentative units from the microtexts and the gold standard of the inserted sentences by labeling them with *Semantic Clause Types*. For the inventory we adopt the most frequent types in Friedrich and Palmer (2014) and give examples from our dataset:

States (STA) describe specific properties of individuals: *The Mayor of Berlin has an interest in Berlin's coffers.*

Events (EVT) are things that happen or have happened: *Edward Snowden revealed information.*

Generic Sentences (GEN) are predicates over classes or kinds: *Supermarkets should open on Sundays.*

Generalizing Sentences (GNZ) describe regularly occurring events/habits: *Germany produces much rubbish.*

The annotations are performed independently by two trained annotators who assign labels at the clause level, whereby one sentence may contain more than one clause.

Statistics. We measure a fair annotator agreement of 34.02% (Cohen's Kappa) and produce a gold standard done by an expert annotator (one of the authors) that provides the basis of our final analysis. Table 1 displays the distribution of semantic clause types within the implicit information annotations and the argumentative units from the microtexts, which we then compare to the numbers reported for other genres (Becker et al., 2016b). We find a high proportion of GENERICS within the Microtexts (64%) and an even higher amount within the implicit information annotations (84%), while the other genres (reports, speeches, fiction) rather contain mostly STATES and EVENTS. This indicates the relevance of knowledge captured by GENERIC SENTENCES within the added implicit information, and we can use this finding for acquiring such missing information automatically.

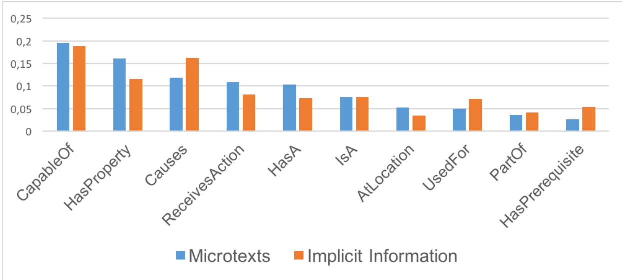

Figure 3: Distribution of ConceptNet relations within Microtexts and Implicit Information Annotations (in %)

4.2. Annotating Commonsense Knowledge Relations

Inventory and Annotation Process. In addition to clause types we annotate the argumentative units and the inserted sentences with ConceptNet relation types. ConceptNet (Havasi et al., 2009; Speer and Havasi, 2012) is a semantic network that contains commonsense facts about the world collected from volunteers over the Internet. Nodes in the network represent concepts in the form of words or phrases, and edges the knowledge relations holding between them (e.g., *health insurance* CAPABLEOF *cover ambulance transportation*). The inventory covers 37 relations, some of which are commonly used in other resources like WordNet (e.g., ISA, PARTOF) while most others are targeted for capturing commonsense knowledge and as such are particular to ConceptNet (e.g., HASPREREQUISITE, MOTIVATEDBYGOAL). The annotation was performed by two annotators in parallel who were asked to label all argumentative units and inserted sentences with ConceptNet relations (irrespective of whether or not the relation *instance* is covered in ConceptNet). The annotators labeled the complete relation triple by (I) selecting and marking two concepts (from the same argumentative unit/inserted sentence), and (II) the ConceptNet relation that they judge to hold between them. Two examples from our dataset are given in Fig. 4. Note that we didn't mark the concepts beforehand, but let our annotators label both: the concepts *and* the relation between them. This sometimes led to disagreements between annotators regarding the span they selected for the same concept [3], which was harmonized in the gold version by the expert annotator.

In preliminary annotation experiments we observed that in many cases several relations can be suitable for the same sentence (cf. example 2 in Fig. 4), therefore we allowed for more than one relation per sentence/argumentative unit. If none of the relations covered by the ConceptNet relation inventory was fitting, our annotators inserted *NONE* and collected suggestions for additional relations (such as REQUIRES). We release these suggestions along with examples from our data together with our dataset.

Statistics. Our annotators used 25 relations from the in-

[3] E.g. for the sentence *Sophisticated programmes should be financed by the licence fee*, A annotated the triple *sophisticated programme, financed by licence fee* (RECEIVESACTION), and B *sophisticated programme, financed* (RECEIVESACTION)

(I) *Fees result in longer durations of studies.*
Annotation: *fees, longer durations of studies* (CAUSES)

(II) *Dog dirt is disgusting and a hygiene problem.*
Annotation: *dog dirt, disgusting* (HASPROPERTY)/
dog dirt, hygiene problem (ISA)

Figure 4: Sentences from our dataset annotated with ConceptNet relations.

ventory of 37 relation types provided by ConceptNet to label the argumentative units from the Microtexts and the inserted sentences. We measure annotator agreement for (I) the marked concepts in order to evaluate if our annotators agree on the spans of texts selected as concepts, and for (II) the assigned relations (separately). (I) we measure in terms of word overlap and obtain high averaged word overlap scores of 76.98% (Jaccard) and 84.87% (Dice), indicating solid agreement between the selected concepts. (II) we measure in terms of Cohen's Kappa and achieve a moderate agreement of 45.05%. We produce a gold standard done by an expert annotator which provides the basis of our final analysis. In this gold standard, on average 3.58 relation triples were assigned per argumentative unit and 3.01 relation triples per inserted sentence. The distribution of the 10 most frequent relation types is shown in Fig. 3.
The most frequently occurring relation is CAPABLEOF (19% within argumentative units and 20% within inserted sentences) followed by HASPROPERTY (16/12%) and CAUSES (12/16%). The largest differences between relations assigned to argumentative unit vs. inserted sentence we observe for HASPROPERTY (4.6pp) and RECEIVESACTION (2.8pp), both more prominent in microtexts, and CAUSES (4.4pp), more prominent in implicit knowledge annotations. We find that only 9 of 576 argumentative units (1.56%) and 24 of 1295 inserted sentences (1.85%) were identified as cases where none of the relations covered by the relation inventory fits, which points to the fact that knowledge repositories such as ConceptNet can play an important role in argument analysis and the retrieval of implicit knowledge.

5. Analysis of the Annotations: Visualizing Correlations

In this section, we analyse correlations between the labels and properties annotated for our dataset. In addition to the analysis of the statistics and distribution of the labels annotated in our corpus (cf. Sec. 4), we want to reveal patterns and intersections between the annotation categories and properties of our dataset, with the goal of learning more about the characteristics of both argumentative texts and implicit knowledge in terms of structural features and semantic information. We expect the results of our analysis to be helpful for guiding and enhancing the process of automated argument analysis as well as of the automatic reconstruction of implicit knowledge in argumentative texts.

5.1. Number of Hops

Hops - Adjacency and Relatedness of Argument Units. The gold version of our dataset contains 719 pairs of argumentative units. 1295 sentences were inserted, that is on average 1.8 sentences per argument pair. The pairs of argumentative units either stand in a direct argumentative relation, which means that they are directly connected in the argument graph (like e_1 and e_2 in Fig. 1), or the units are adjacent to each other (e_1 and e_2, e_2 and e_3 ...), or both (e_1 and e_2). We expect that more inserted sentences are needed to connect argument pairs that stand in an argumentative relation but are not adjacent, since the missing information could be included in the intermediate argument units (e.g., what is missing between e_1 and e_5 in Fig. 1 could be expressed in e_2, e_3 or e_4). We also hypothesize that more implicit information is needed to connect argument pairs that don't stand in a direct argumentative relation, since argument units that aren't related can come from different chains of the argument and might therefore require more explications than directly related argument units (cf. Fig. 1, e_4 and e_5). Since – by our annotation design – the inserted sentences contain the minimal amount of information that makes the connection between two argumentative units explicit, we interpret each inserted sentence as one hop that is needed to connect the given argument pair.

	adjacent	not adjacent
nb. of pairs	464	255
percentage	0.65	0.35
nb. of inserted sentences	881	414
inserted sentences (avg)	1.9	1.62

Table 2: Adjacency of argument pairs and number of inserted sentences, Gold.

We find only a relatively small difference in the average number of sentences inserted between adjacent (1.9) and non-adjacent units (1.62) (cf. Table 2), indicating that it is not the case that more hops (inserted sentences) are needed when units are not adjacent. Interestingly, on the other hand we observe a remarkable difference between the number of sentences inserted between argumentatively related (1.6, Table 3) and non-related units (2.14, Table 3). This indicates that more hops are needed when there is no direct argumentative relation between the argument units.
Hops - Argumentative Relations. Next, we are interested whether there are argumentative relations for which more hops are needed than for others. Our dataset contains 5 argumentative relations, with *support* being the most frequent one (37%) followed by *rebuttal* (15%) and *undercut* (8%) (cf. Table 3). We find that for *undercut* relations most sentences are inserted on average (1.84). This makes sense since *undercuts* challenge the acceptability of an inference between two propositions and can therefore be seen as a very complex relation that requires more explications than others. The least sentences are inserted for *example* relations (1.11), indicating that they usually don't need multi-hop connections of implicit knowledge.
Hops - Commonsense Relations. Additionally, we want to know whether there are co-occurrences between the number of hops and commonsense relation types. We want to investigate whether specific commonsense relation types appear more often in single (one inserted sentence) vs. multiple hops (more than one inserted sentence). Therefore, for all commonsense relations within inserted sentences, we count

	argument relation		inserted sentences	
	total	percentage	total	per relation
support	263	37	423	1.61
rebuttal	108	15	165	1.53
undercut	61	8	112	1.84
addition	21	3	34	1.62
example	9	1	10	1.11
relations total	462	100	744	1.6
non-related units	257	36	551	2.14
TOTAL	719	100	1295	1.8

Table 3: Correlation between argumentative relations and number of hops (inserted sentences).

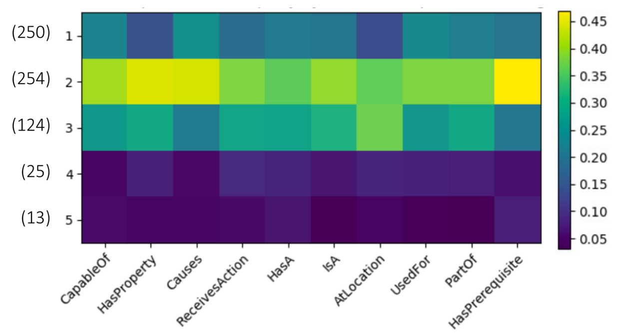

Figure 5: Distribution of commonsense relations within inserted sentences among hops, rel. freq. by relation type, with total number of hops given on the left side.

how often they occur in one hop connections (when one sentence was inserted as missing information), in two hop connections and so on. The resulting heatmap is displayed in Fig. 5. We observe that all relations occur most often within a set of two inserted sentences, which corresponds to the average number of inserted sentences (1.8, cf. Table 3). Interestingly, HASPROPERTY and ATLOCATION are relations which occur only rarely within one hop connections, the latter being most often used in sets of three inserted sentences. Those relations seem to mark information units that require other pieces of information to connect an argument pair.

Hops - Semantic Clause Types. Similarly, we want to investigate co-occurrences between the number of hops and semantic clause types. Again, for all clause types within inserted sentences, we count how often they occur in each set of inserted sentences (1-5), Fig. 6 shows the resulting heatmap. We find that STATES, EVENTS and GENERIC SENTENCES occur most often within two hop connections, while GENERALIZING SENTENCES are most often used within sets of three inserted sentences and rarely when only one sentence was inserted. GENERALIZING SENTENCES therefore can be interpreted as markers of information units that stand-alone are not able to connect argument pairs, but rather co-occur with other pieces of information for filling knowledge gaps in argumentative texts.

5.2. Adjacency and Argumentative Relatedness

When filling knowledge gaps in argumentative texts automatically, it might be useful to leverage the structure of an

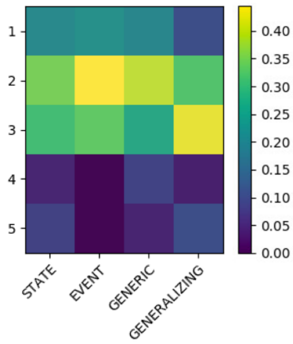

Figure 6: Distribution of Semantic Clause Types among hops, relative frequency by clause type.

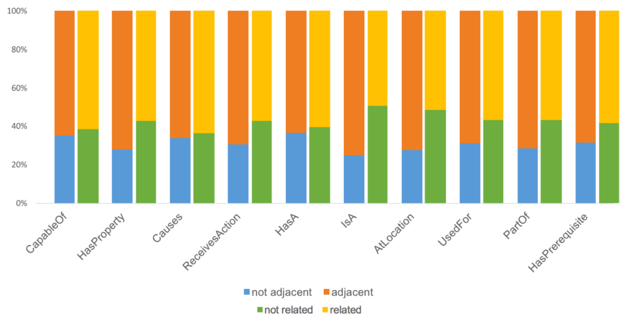

Figure 7: Adjacency and Argumentative Relatedness for Commonsense Relations (in %).

argument and to determine which type of knowledge exactly is missing for which pair of argument units. Knowing the semantic properties of the knowledge that is needed to connect argument units that are – for example – adjacent vs. those that are not, can guide the process of extracting knowledge for filling these gaps. Therefore, we want to investigate whether the distribution of the semantic properties we annotated for the inserted sentences – commonsense relation types and semantic clause types – respectively differs depending on the internal structure of an argument. In our case this is whether (i) the arguments from a given pair are adjacent or not, and/or whether (ii) the arguments from a given pair are argumentatively related or not.

Commonsense Relations. Fig. 7 (blue/orange bars) shows that the distribution of commonsense relation types only slightly differs between adjacent and non-adjacent units. We find that ISA (75%), ATLOCATION and HASPROPERTY (both 72%) occur most often within sentences inserted between adjacent units, while HASA and CAPABLEOF are relations that occur more often in sentences inserted between non-adjacent units (36% and 35%). We also observe only slight variations regarding the distribution of commonsense relations between units that are argumentatively related and those that are not (Fig. 7, green/yellow bars). While CAUSES (64%), CAPABLEOF (61%) and HASA (61%) are often assigned to sentences inserted between related units and therefore can be interpreted as argumentatively relevant, ISA and ATLOCATION are typical labels for implicit information between units that don't stand in a direct argumentative relation (51% and 49% for unrelated units).

Semantic Clause Types. We also want to investigate

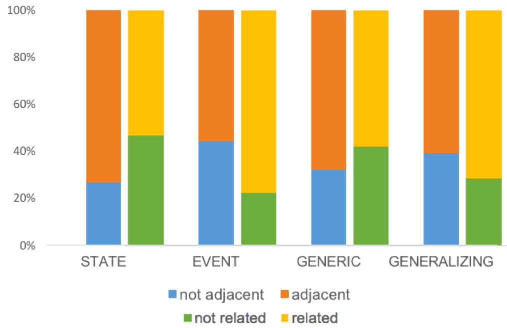

Figure 8: Adjacency and Argumentative Relatedness for Semantic Clause Types (in %).

whether the distribution of Semantic Clause Types differs between adjacent and not adjacent units, and/or between argumentatively related and unrelated units. Fig. 8 (blue/orange bars) shows that STATES occur most often between units that are adjacent (73%), while EVENTS show the lowest proportion for adjacent units (56%). Regarding the distribution of semantic clause types assigned to sentences between argumentatively related and unrelated units (Fig. 8, green/yellow bars), we find a large difference for EVENTS (78% between related and 22% between unrelated units) and GENERALIZING SENTENCES (71%/29%), while STATES (53%/47%) and GENERIC SENTENCES (58%/42%) are more equally distributed.

5.3. Correlations between Assigned Labels

In this section, we analyse correlations between argumentative relations, commonsense relations and semantic clause types. We want to reveal patterns and intersections between the annotation categories in order to learn more about the structural features and semantic properties of both argumentative texts and implicit knowledge. For all analyses reported in this section, we measure correlations using the Matthews correlation coefficient (MCC) (Matthews, 1975), which assigns correlation coefficient values between -1 and +1 to pairs of labels (here e.g. *support* and CAUSES). A coefficient of +1 represents a perfect correlation, 0 an average random prediction, and -1 an inverse correlation.

Argumentative Relation - Commonsense Relation. First, we look at correlations between argumentative relations and commonsense relations. We want to investigate if specific commonsense relations express specific argumentative relations, and if specific argumentative relations are more characteristic for specific commonsense relations than others. Fig. 10 shows that the relation CAUSES is very dominant within sentences inserted between argument units that stand in a *support* relation, which reveals the importance of causal explanations for filling knowledge gaps between supporting argument units. An example of our dataset is given in Fig. 9. The relations RECEIVESACTION and HASA correlate negatively with *support* but positively with *rebuttals*, underlining the difference in distributions of commonsense relation types between these two contrary argumentative relations. We also observe that *rebuttals* correlate negatively with CAUSES, indicating that causal explanations are not typical for connecting argument units that rebut each other.

Argumentative Relation - Semantic Clause Type. Next,

(e_2) *The developments in that conflict should not be left to former Cold War opponents alone,*
(e_3) *for that course can only lead to escalation in some form.*

Implicit Information: *A conflict may lead to escalation.*
Commonsense Relation: *conflict, escalation* (CAUSES)

Figure 9: Example of a causal explication for a support relation (e_3 supports e_2).

we are interested in the correlations between argumentative relations and semantic clause types. We analyse if specific argumentative relations are more characteristic for specific clause types, and vice versa, if specific clause types correlate with specific argumentative relations. Fig. 10 shows that *examples* differ from the rest of argumentative relations regarding the correlations with clause types: while we find high positive correlations with STATES and EVENTS, GENERICS very infrequently co-occur with *examples*. This makes sense since examples usually express knowledge about individuals rather than generic knowledge (cf. Becker et al. (2016a)). Our correlation analysis also reveals interesting patterns regarding the *support* relation: here we find a negative correlation with STATES and a positive correlation with GENERIC SENTENCES, indicating the importance of generic knowledge for sentences that connect two argument units which support each other. Interestingly, when looking at the correlations between GENERIC SENTENCES and argument relations, we find that this is the only positive correlation, while all others are negative. This underlines the finding that GENERICS can be seen as an important feature of sentences inserted between supporting argument units.

Commonsense Relation - Semantic Clause Type. There are also some interesting correlations between commonsense relations and semantic clause types which we display in Fig. 12. We aim to discover whether (i) specific clause types are indicators for certain commonsense relations or vice versa, and whether (ii) the distribution of clause types among certain commonsense relations differs between microtexts and inserted sentences. Fig. 12 (left) shows that within the microtexts, GENERIC SENTENCES correlate negatively with ISA, ATLOCATION and PARTOF, and positively with RECEIVESACTION and HASPREREQUISITE. For the three

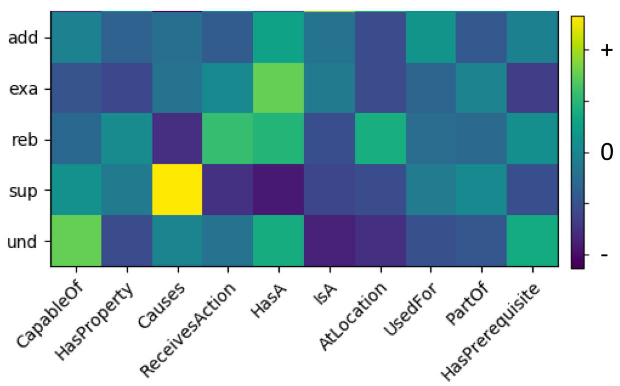

Figure 10: Correlations between Argumentative Relations and Commonsense Relations, MCC correlation matrix. Bright colours indicate positive and dark colours negative correlation from a scale between +1 and -1.

relations IsA, AtLocation and PartOf we find a positive correlation with States and Events and a negative correlation with Generic Sentences, indicating that these relations typically express individual rather than generic knowledge. Fig. 12 (right) shows that the correlations within the inserted sentences are not as strong as in the microtexts, but still we can see that similar to the microtexts, Generic Sentences correlate negatively with IsA and PartOf. We also find a high positive correlation between IsA and States, indicating that within the inserted sentences (as well as within the microtexts), IsA relations typically describe specific properties of individuals (cf. Sec. 4.1).

6. Conclusion and Outlook

In this paper, we presented a carefully designed dataset consisting of high-quality human annotations of implicit knowledge in argumentative texts. To learn more about the characteristics of both the argumentative texts and the added information, we further annotated the data with semantic clause types and commonsense knowledge relations. We then provided an in-depth analysis of our annotated dataset with the goal of revealing characteristic distributions and correlations, co-occurring patterns and intersections between the annotation categories. This helped us to gain insights into the properties of both argumentative texts and implicit knowledge in terms of structural features and semantic information: We found for example that Generic Sentences play a dominant role within the inserted sentences, indicating the relevance of generic knowledge within implicit information. Almost all sentences in our dataset – from both the microtexts and the inserted information – could be mapped to commonsense knowledge relations, pointing to the fact that knowledge repositories such as ConceptNet can play an important role in argument analysis and are an important source for the retrieval of implicit knowledge.

When analyzing correlations between the labels and structural properties of our dataset, we could furthermore reveal patterns and intersections between the annotation categories and structures of our dataset: We found for example that more inserted sentences are needed when there is no direct argumentative relation between the argumentative units, and that complex argumentative relations such as *undercut* require more explications than other relations. Our correlation analysis further demonstrated the benefit of leveraging the structure of an argument and the type of knowledge that is needed to connect argument pairs. We investigated whether the distribution of the semantic properties we annotated for the inserted sentences differs depending on the internal structure of an argument and revealed for instance that States occur most often between units that are adjacent, while Events are frequently used for connecting argumentatively related units. Finally, when investigating correlations between argumentative relations, commonsense relations and semantic clause types, we could for example reveal the importance of causal explanations for filling knowledge gaps between supporting argument units. Generics also turned out to be an important feature of sentences inserted between supporting argument units.

The knowledge we gained about the properties of argumentative texts and implicit knowledge, and our observations on their interaction can assist automated argument analysis, e.g., it can be beneficial for assessing the strength of an argument, apart from the benefit of making the underlying logics of the argument transparent for both humans and computational systems. The results from our in-depth analysis can furthermore guide the process of revealing implicit information in argumentative texts automatically, e.g. by utilizing the revealed properties of implicit information and the observed relations between implicit information and the surrounding argument units.

We release our dataset as an extension to the Microtext corpus. We expect it to be a useful starting point for automatically filling knowledge gaps in arguments, and we hope that it will inspire future research on argument analysis and implicit knowledge acquisition.

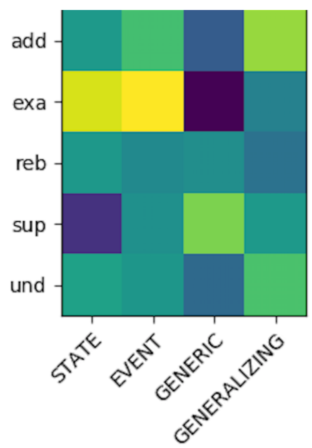

Figure 11: Correlations between Argumentative Relations and Semantic Clause Types, MCC correlation matrix.

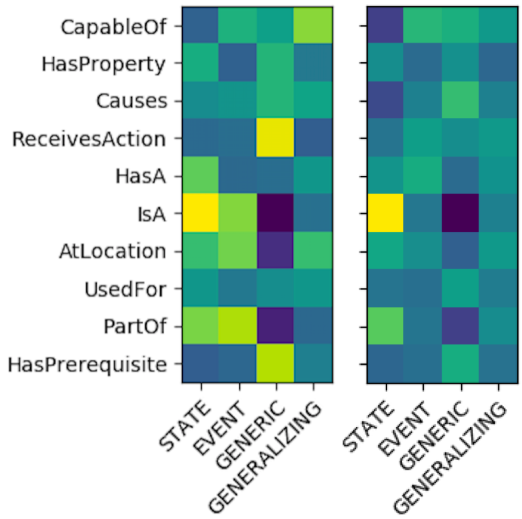

Figure 12: Correlations between Commonsense Relations and Semantic Clause Types in Microtexts (left) vs. Inserted Sentences (right), MCC correlation matrix.

Acknowledgements

This work has been funded by the DFG within the project ExpLAIN as part of the Priority Program "Robust Argumentation Machines" (SPP-1999). We thank our annotators Anna Becker and Katharina Korfhage for their contribution.

7. References

Becker, M., Palmer, A., and Frank, A. (2016a). Argumentative texts and clause types. In Proceedings of the 3rd Workshop on Argument Mining (ArgMining2016), pages 21–30, Berlin, Germany.

Becker, M., Palmer, A., and Frank, A. (2016b). Clause Types and Modality in Argumentative Microtexts. In Workshop on Foundations of the Language of Argumentation (in conjunction with COMMA), pages 1–9, Potsdam, Germany.

Becker, M., Staniek, M., Nastase, V., Palmer, A., and Frank, A. (2017a). Classifying Semantic Clause Types: Modeling Context and Genre Characteristics with Recurrent Neural Networks and Attention. In Proceedings of *SEM (Joint Conference on Lexical and Computational Semantics).

Becker, M., Staniek, M., Nastase, V., and Frank, A. (2017b). Enriching Argumentative Texts with Implicit Knowledge. In Flavius Frasinca, et al., editors, Applications of Natural Language to Data Bases (NLDB) - Natural Language Processing and Information Systems, Lecture Notes in Computer Science. Springer.

Becker, M., Staniek, M., Nastase, V., and Frank, A. (2019). Assessing the Difficulty of Classifying ConceptNet Relations in a Multi-Label Classification Setting. In RELATIONS - Workshop on meaning relations between phrases and sentences (co-located with IWCS).

Boltuzic, F. and Snajder, J. (2016). Fill the Gap! Analyzing Implicit Premises between Claims from Online Debates. In Proceedings of the 3rd Workshop on Argument Mining, pages 124–133.

Freeman, J. B. (2011). Argument Structure: Representation and Theory, volume 18 of *Argumentation Library*. Springer.

Friedrich, A. and Palmer, A. (2014). Automatic prediction of aspectual class of verbs in context. In Proceedings of the ACL 2014.

Grice, H. P. (1975). Logic and conversation. *1975*, pages 41–58.

Habernal, I., Wachsmuth, H., Gurevych, I., and Stein, B. (2018). The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. In Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1930–1940.

Havasi, C., Speer, R., Pustejovsky, J., and Lieberman, H. (2009). Digital intuition: Applying common sense using dimensionality reduction. *IEEE Intelligent systems*, 4(24):24–35.

Hulpus, I., Jonathan Kobbe, Maria Becker, J. O. G. H. C. M. V. N. H. S., and Frank, A. (2019). Towards Explaining Natural Language Arguments with Background Knowledge. In Proceedings of the 1st Workshop on Semantic Explainability – co-located with the 18th International Semantic Web Conference (ISWC 2019).

Kusner, M. J., Sun, Y., Kolkin, N. I., and Weinberger, K. Q. (2015). From Word Embeddings To Document Distances. In Proceedings of the 32nd International Conference on Machine Learning (ICML), pages 957–966.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 2(405):442—451.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13, pages 3111–3119, USA. Curran Associates Inc.

Peldszus, A. and Stede, M. (2015). An annotated corpus of argumentative microtexts. In Proceedings of the First European Conference on Argumentation.

Rajendran, P., Bollegala, B., and Parsons, S. (2016). Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews. In Proceedings of the 3rd Workshop on Argument Mining.

Razuvayevskaya, O. and Teufel, S. (2016). Recognising enthymemes in real-world texts: a feasibility study. In Workshop on Foundations of the Language of Argumentation (in conjunction with COMMA).

Speer, R. and Havasi, C. (2012). Representing General Relational Knowledge in ConceptNet. In Proceedings of LREC.