Enriching Argumentative Texts with Implicit Knowledge

Maria Becker, Michael Staniek, Vivi Nastase, and Anette Frank
{mbecker,staniek,nastase,frank}@cl.uni-heidelberg.de

Institute for Computational Linguistics, Heidelberg University
Leibniz Science Campus Empirical Linguistics & Computational Language Modeling

Abstract. Retrieving information that is implicit in a text is difficult. For argument analysis, revealing implied knowledge could be useful to judge how solid an argument is and to construct concise arguments. We design a process for obtaining high-quality implied knowledge annotations for German argumentative microtexts, in the form of simple natural language statements. This process involves several steps to promote agreement and monitors its evolution using textual similarity computation. To further characterize the implied knowledge, we annotate the added sentences with semantic clause types and common sense knowledge relations. To test whether the knowledge could be retrieved automatically, we compare the inserted sentences to Wikipedia articles on similar topics. Analysis of the added knowledge shows that (i) it is characterized by a high proportion of generic sentences, (ii) a majority of it can be mapped to common sense knowledge relations, and (iii) it is similar to sentences found in Wikipedia.

Keywords: argumentation, implicit knowledge, annotation process, textual similarity, semantic clause types, common sense knowledge, Wikipedia

1 Introduction

It is agreed, at least since the work of Grice [7], that overt communication relies on a large body of knowledge that both the speaker and the listener share, such that only part of the message conveyed by the speaker needs to be expressed in words, while the rest can be filled in by the hearer. The assumption of mutually shared knowledge could be a source of misunderstanding, when the knowledge implied by the speaker is different from what the hearer fills in. Omitting weak supporting information could also be used as a manipulation device, to make a weak argument seem sound. In an argumentative framework, it would then be beneficial to reconstruct this implied information to be able to rate the soundness and validity of an argument.

Retrieving implied information that is not explicitly mentioned in a text is difficult – when asked, different people may have different ideas of what exactly and how detailed such information should be. In this work, we describe the process we designed to elicit high-quality annotations of implied knowledge in the form of simple natural language sentences, for a concise argumentation dataset – the Microtext corpus [10]. We structure the annotation process in such a way that annotators working in parallel on the same data set have to review each other's annotations, encouraging them to understand the other's point of view, and adjust their own to gradually reach agreement.

Apart from initial instructions we did not interfere in the annotation process, but we did monitor the *evolution of annotator agreement* by measuring the similarity between the added annotations.

To learn more about the nature and (linguistic) characteristics of the added information, we further annotate the data with two specific semantic information types: semantic clause types [6] and ConceptNet knowledge relations [8, 15]. Previous work [1] has shown that argumentative texts are characterized by a specific distribution of *abstract linguistic clause types* [6] that distinguish them from other text genres. We will test whether the provided implied information shares these characteristics, by having it labeled with these categories. To complete the picture in terms of the type of knowledge expressed by the inserted sentences, they will be annotated with *common sense knowledge relations*, using the inventory of 28 relation types of ConceptNet [8, 15]. In addition, we compare the inserted knowledge with sentences from Wikipedia, to assess the difficulty of automatically harvesting missing knowledge in natural language arguments.

In summary, the contributions of this work are: (i) high-quality annotations of implicit knowledge on the argumentative Microtext corpus, in terms of natural language sentences; (ii) characterization of the specific nature of these sentences in terms of semantic clause types and common sense knowledge relations; (iii) design of an annotation process for a difficult task that promotes agreement between annotators by having them review each other's work; (iv) an approach to monitor the annotation process – in particular, the increase in agreement – using textual similarity techniques.

This annotated data will be made public as an extension to the Microtext corpus [10], to support further research in argument analysis.

2 Related Work

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 [11]. [12] present a feasibility study on the automatic detection of enthymemes in real-world texts and find that specific discourse markers (e.g. *let alone, therefore, because*) can signal enthymemes. Using these as trigger words, they reconstruct enthymemes from the local context, while [11] retrieve and fill missing propositions in arguments from similar or related arguments. Another method is the utilization of shared knowledge [3], which is related to our approach. [1], [2] show that argumentative texts are rich in generic and generalizing sentences, which often express common knowledge. We will show that large portions of implied

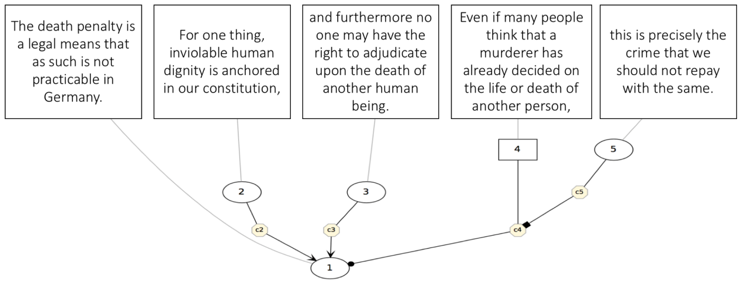

Fig. 1. Argumentation Graph from the Microtext Corpus (micro/b006)

knowledge in argumentative texts are naturally stated using these clause types. In their attempt to reconstruct implicit knowledge, [4] find that the claims that users make in online debate platforms often build on implicit knowledge and that the reconstruction of implicit premises supports claim detection. They release a 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, yielding a total of 500 gap-filling premises set. In contrast to our approach they asked the annotators to provide the premises without giving any further instructions, resulting in a substantial variance in the average number of premises and words in premises as well as a low word overlap (32%).

These studies suggest that a substantial amount of knowledge is needed for the interpretation and analysis of argumentative texts. In this work we report on an annotation method designed to encourage agreement, and through which we acquire high-quality annotations for implicit knowledge in arguments.

3 Annotating Implicit Knowledge in Arguments

3.1 The Microtext Corpus

The basis for our work is the argumentative Microtext corpus [10], which consists of 112 microtexts in German. Each microtext is a short, dense argument written in response to a question on a potentially controversial issue (e.g., *Should all universities in Germany charge tuition fees?*). 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, were written in German and professionally translated into English. An example together with its argument structure graph is given in Figure 1.

The produced microtexts were manually annotated according to a scheme based on Freeman's theory of the macro-structure of argumentation [5] for representing text-level argumentation structure. For this, each text was segmented into elementary units of argumentation which present either a conclusion or a premise. Each unit corresponds to a node in the argument graph. Nodes with

(a) *Alternative treatments should be subsidized just like conventional treatments*
(b) *since both methods can lead to the prevention, mitigation or cure of illness.*
(c) *Treatments are subsidized if they lead to the prevention, mitigation or cure of illness.*

Fig. 2. Example: Explicating implicit knowledge that connects related premises

outgoing pointed arrows are *proponent nodes*, while those with outgoing square-headed arrows mark *opponent nodes*. The arcs are labeled with *argumentative functions* [10]. The most frequent functions are: (a) *support*: a premise supports a conclusion or another premise; (b) *rebuttal*: a premise attacks a conclusion or premise by challenging its acceptability, or (c) *undercut*: a premise attacks the acceptability of an argumentative relation between two propositions.

3.2 Task I: Revealing Implicit Knowledge in Argumentative Texts

The aim of our annotation is to reveal implicit knowledge that connects superficially disconnected, but semantically coherent premises in argumentative texts. Being able to make this implicit information explicit could help assess the strength of an argument, apart from the benefit of making the underlying logics of the argument transparent for both humans and computational systems.

Figure 2 shows an example of the desired annotations, with a main claim *a* supported by statement *b*. While the knowledge underlying the argumentative function is not explicitly conveyed, we believe the reader has no trouble understanding why the argumentative relation holds: it is made explicit in *c*.

The difficulty of eliciting such implicit knowledge in an annotation task is that intuitions about what the implied connection is may be different between annotators. Even if their intuitions match, the phrasing chosen by the annotators may be different, structurally or in terms of lexical choice. This makes it hard to enforce agreement and to assess the quality of the annotations. We will address these challenges in two ways:
(i.) by designing an **multi-step annotation process** where annotators are asked to review and potentially revise each other's annotations and
(ii.) by **measuring the dynamic evolution of agreement** during this process using computational measures of **semantic textual similarity (STS)**.

Annotation process. The annotations are performed on sentence pairs from the Microtext corpus (the original German version)[1], that stand in an argumentative relation according to the argumentation graph. There are 464 such sentence pairs in the 112 texts in the corpus, i.e., approx. 4 pairs per microtext. The annotation process is illustrated in Figure 3. We use 5 human judges (H1 .. H5), all of them native German speakers with a linguistic background.

Step 1: $H1$ and $H2$ produce initial annotations **A1** and **A2**. The annotators are asked to add the minimal amount of information that makes the connection between the two sentences explicit by: (1) adding as few sentences as possible and (2) making the inserted sentences as simple as possible so that

[1] All examples are shown in English for convenience.

they ideally only contain one fact per sentence. Through these instructions we intend to avoid long and too detailed explanations, and in consequence, to support better agreement between the judges. If the annotators think that no information is missing (i.e., the connection between sentences is explicit), this is labeled correspondingly.

Step 2: $H1$ and $H2$ review each other's annotations, producing corrected versions: $H1(\mathbf{A2})$=$\mathbf{A2^c}$, $H2(\mathbf{A1})$=$\mathbf{A1^c}$.

Step 3: To avoid biases arising from the correction phase, two different judges $H3$ and $H4$ independently perform merges of $\mathbf{A1^c}$ and $\mathbf{A2^c}$, producing annotations **A3** and **A4**, respectively. They are allowed to select one annotation, combine them into a novel statement, or to produce a new annotation. Formally, annotator Hk produces for each sentence pair (i):

$$\mathbf{Ak_i} = merge(\mathbf{A1^c_{\mathit{i}}},\ \mathbf{A2^c_{\mathit{i}}}) = \begin{cases} \mathbf{A1^c_i} & \text{if } Hk \text{ confirms } \mathbf{A1}^c_i \\ \mathbf{A2^c_i} & \text{if } Hk \text{ confirms } \mathbf{A2}^c_i \\ \mathbf{A1^c_i}/\mathbf{A2^c_i} & \text{if } Hk \text{ combines parts of } \mathbf{A1^c_{\mathit{i}}} \text{ and } \mathbf{A2^{c_i}} \\ \mathbf{Ak_i} & \text{if } Hk \text{ produces a new annotation} \end{cases}$$

Step 4: A final annotator $H5$ produces the gold standard based on **A3** and **A4** following the merge process described above, with the difference that for this final step we allow two versions of the inserted information if both of them fill the gap. This decision was inspired by the observation that in many cases **A3** and **A4** provide the same information expressed slightly differently:
(1) *People of higher age have more experience.*
(2) *People in retirement age are considered more experienced.*

Measuring the evolution of agreement using semantic textual similarity. To trace the evolution of agreement between annotators we quantify the distance between their respective annotations – i.e. added sentences – using the Word Mover's Distance [9] as implemented in *gensim*[2]. The Word Mover's Distance (WMD) (Eq. 1) 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. For two documents d_1 and d_2 with vocabulary of size n, the WMD is computed using the embeddings for each word i (with embedding x_i) from document d_1 and word j (with embedding x_j) from d_2 as the solution of the optimization problem:

$$WMD(d_1, d_2) = min_{T\geq 0} \sum_{i,j=1}^{n} T^{\star}_{ij} \|x_i - x_j\|_2 \tag{1}$$

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

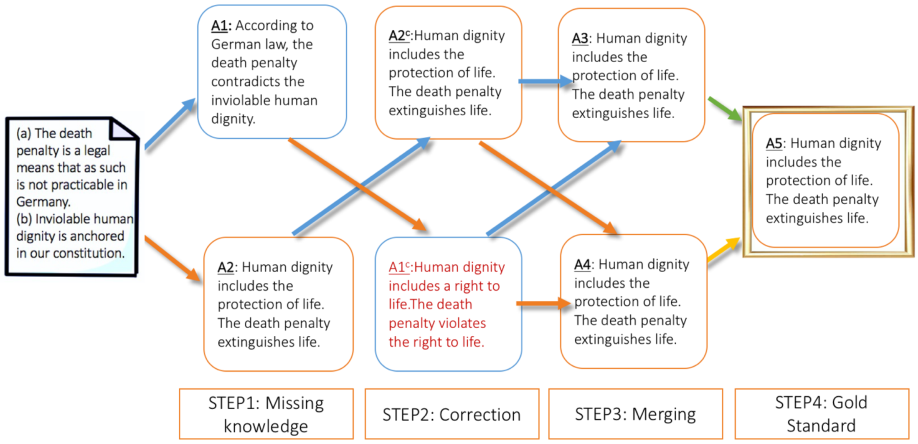

Fig. 3. Annotation Pipeline illustrated with an example

$$T^{\star}_{ij} = \begin{cases} \frac{freq(i)}{\sum_{x=1}^{n} freq(x)} & \text{if } j = argmin_j \|x_i - x_j\|_2 \\ 0 & \text{otherwise} \end{cases}$$

The WMD method has two main components: the word embeddings used and the scoring function, and we adjusted these to our language and task.

Word Embeddings. The WMD relies on word embeddings that map similar words to close regions in embedding space. From several pretrained word embeddings for German, we chose those that yielded the highest correlation with human scores on four word similarity datasets:[3] 100-dimensional word2vec embeddings trained on a German "meta-corpus" of 116 million sentences (that combines several German corpora) using Skip-Gram mode with 5 negative samples [13].

Scoring function. Generally, a short text subsumed by a larger text would warrant a low distance score. In our case, however, since we aim to quantify how strongly the annotators agree and to which degree their inputs are similar, differences in length should be penalized by increasing the distance score. We therefore add a *length difference penalty* score (LDP) (3) to the WMD (2):

$$WMD'(d_1, d_2) = WMD(d_1, d_2) + LDP(d_1, d_2) \tag{2}$$

$$LDP(d_1, d_2) = \frac{|length(d_1) - length(d_2)|}{\frac{length(d_1)+length(d_2)}{2}} \tag{3}$$

LDP increases the distance score if there is a large difference in length between the sentences. The length difference is normalized by the average of the

[3] Spearman correlation results on the German version of the MC30: 0.76; RG65: 0.79; wordsim353: 0.69; ZG222: 0.42. https://dkpro.github.io/dkpro-similarity/wordpairsimilarity/

sum of the sentence lengths, such that longer sentences with a variation in length are penalized less compared to shorter ones.

WMD' (Equation 2) will quantify the disagreement between sets of annotations A_{xk} and A_{yk} for sentence pairs k (cf. Section 4.1). A_{xk} and A_{yk} may contain a different number of sentences. We compare the complete annotation for a sentence pair k (as opposed to sentence by sentence), because the annotators may have split the information they added differently across the added sentences, as illustrated in the examples below from $H1$ (1.1, 1.2) and $H2$ (2.1, 2.2). The first example (1.1 and 2.1) has a distance score of 0.57, while the second is higher at 3.82:

(1.1) *Public broadcasters are financed by general broadcasting contributions.*
(2.1) *Public broadcasters are financed by broadcasting contributions.*

(1.2) *Since many mistakes have happened, because too long waited, the EU is to interfere very quickly this time.*
(2.2) *Interference can prevent war.*

3.3 Task IIa: Situation Entity types annotations

We asked the annotators to characterize the inserted sentences by labeling them with *situation entity types.* The distribution of these clause types is distinctive for argumentative texts compared to other genres [2], showing particularly high ratios of generic and generalizing sentences. For the inventory of SE types we adopt the most frequent types in [6]:

states describe specific properties of individuals:
(e1) *Waste separation is a form of environmental protection.*
events are things that happen or have happened:
(e2) *Edward Snowden revealed information.*
generic sentences are predicates over classes or kinds:
(e3) *Health insurance funds take over the payment of medicine.*
generalizing sentences describe regularly occurring events or habits:
(e4) *The broadcasting fee is paid by all citizens alike.*

The annotations are performed independently by two trained annotators. They assign SE type labels at the clause level. The segmentation is performed automatically with DiscourseSegmenter [14], a python package offering both rule-based and machine-learning based discourse segmenters.

3.4 Task IIb: Concept Net relations annotations

To gain further insight into the type of knowledge covered by the provided sentences, we annotate them with ConceptNet relation *types* such as *PartOf, Causes* or *IsA*. ConceptNet [8, 15] is a semantic network that contains common sense facts about the world. The knowledge is collected from volunteers over the Internet (via templates, free text, games etc.) and is represented as tuples

<left term,relation,right term>, terms (words/short phrases) are nodes, and the relations between them are edges – e.g. <dogs,*IsA*,animals>.

The annotation was performed by two annotators in parallel. They were asked to label each inserted sentence, if applicable (irrespective of whether or not the relation *instance* is covered in ConceptNet). Examples of annotated relation types are presented in Section 4.3.

3.5 Task III: Retrieving similar sentences from a Wikipedia Corpus

The annotations of implied knowledge represent a gold standard that should be obtainable automatically, whether in their exact form or as approximations. We test whether the implied knowledge that the annotators made explicit through the provided sentences can be found in a textual corpus.

We collect a corpus of Wikipedia sentences from articles that match the topics of the microtexts in the Microtext corpus. Each microtext was elicited with a query, e.g. *Should Germany introduce the death penalty?*. We match this query to German Wikipedia article titles and extract the introduction section of the article, if it exists, or the first 10 sentences. From all 18 queries used in to produce the corpus, we find matches for 50 related topics such as *tuition fees* or *waste separation*, resulting in a corpus of 874 sentences.

To test whether we can find sentences in Wikipedia that match sentences in the inserted information set, we use the distance formula WMD' (Equation 2).

4 Analysis of the Annotations

4.1 Task I: Data statistics and evolution of annotator agreement

The annotators were provided with 464 sentence pairs from the Microtext Corpus. The annotations of Situation Entity types and ConceptNet relations were done on the sentences inserted at step 2 of the annotation process ($A1^{\mathbf{c}}$-$A2^{\mathbf{c}}$). $A1^{\mathbf{c}}$ includes 750 sentences (1.62 sents/gap on average) and $A2^{\mathbf{c}}$ 720 sentences (1.55 sents/gap on average) in total.

Only 44 (9%) of the 464 sentence pairs were labeled as *no information missing* at step 2 of the annotation process, indicating that coherence among statements strongly relies on implicit knowledge.

Evolution of annotator agreement. To compare two complete annotations, we compute the average and standard deviation of their dissimilarity with WMD' (Table 1) and plot histograms of the disagreement scores (Fig. 4).

Columns 1 & 2 in Table 1 measure the amount of editing $H1$ and $H2$ performed on the other's annotations. A2 (3.69) was edited more than A1 (0.71). Columns 3, 4 & 5 show how inter annotator agreement improves after each annotation step – smaller numbers mean lower distance and therefore higher similarity. Mutual reviewing improves the agreement between $H1$ and $H2$. The distance decreases again with the third annotation step where $H3$ and $H4$ merge $A1^{\mathbf{c}}$ and $A2^{\mathbf{c}}$, producing A3 and A4 with a semantic distance of 1.91. The evolution towards agreement is illustrated in Figure 4, through the shift towards 0 in the distance between annotations from one step of the process to the next.

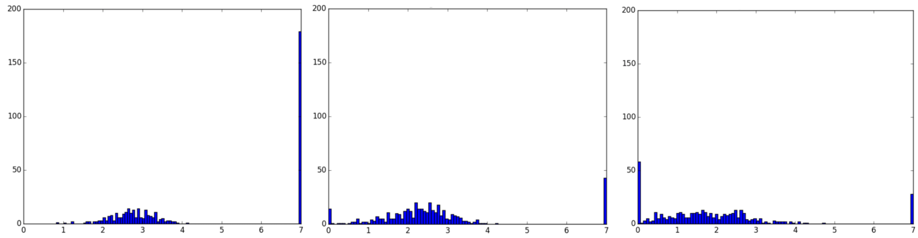

Fig. 4. WMD' histogram for (1) A1-A2, (2) A1$^{\mathbf{c}}$-A2$^{\mathbf{c}}$ and (3) A3-A4. A shift towards 0 from one step of the annotation process to the next indicates an increase in agreement. 7 is the maximum distance score and was assigned when one annotator labeled a sentence pair with *no information missing*, while the other inserted information.

	A1 ↓ A1$^{\mathbf{c}}$	A2 ↓ A2$^{\mathbf{c}}$	A1 ↓ A2	A1$^{\mathbf{c}}$ ↓ A2$^{\mathbf{c}}$	A3 ↓ A4	A1$^{\mathbf{c}}$ ↓ A3	A1$^{\mathbf{c}}$ ↓ A4	A2^{c} ↓ A3	A2^{c} ↓ A4	A3 ↓ A5	A4 ↓ A5
WMD	0.71	3.69	4.62	2.71	1.91	2.06	1.79	2.00	1.71	1.52	1.49
SD (σ)	1.90	8.03	4.67	1.69	1.69	1.76	1.95	1.77	2.01	1.47	1.32

Table 1. Evolution of annotator agreement measured by WMD'.

New solutions from annotators. High distance scores in step 3 are often assigned to annotations (i) that provide different information, (ii) where one annotator assigned *no information missing* while the other inserted missing information or (iii) when one annotator provides the information in finer-grained steps than the other. The reason for (i) is very often that one/both annotators provide a new solution to fill the gap that is different from the annotations A1$^{\mathbf{c}}$ and A2$^{\mathbf{c}}$ (and also from what the other annotator inserted at that step). Thus, $H3$ and $H4$ did not only select from existing annotations, but produced statements based on their own intuitions.

Word overlap. Compared to [4], who report a word overlap of 32% for the premises inserted by their annotators (they don't report which similarity coefficient they applied), we obtain an averaged word overlap score of 57% (Dice)/ 47% (Jaccard) for A3-A4, while only 36% (Dice)/ 26% (Jaccard) for A1$^{\mathbf{c}}$-A2$^{\mathbf{c}}$ and 10% (Dice)/ 6% (Jaccard) for A1-A2, showing again improved agreement of the annotations along the process.

Comparison of annotations across different annotation steps. The averaged distance scores between annotations produced in early vs. later stages show an evolution towards agreement: the averaged distance score from the second annotation step (A1$^{\mathbf{c}}$ and A2$^{\mathbf{c}}$) compared to those in the third step (A3 and A4) (columns 6-9 in Table 1) is 1.89; the averaged distance score between annotations from the third step to the fourth step (columns 10-11) is 1.51.

Gold Standard. In the last step an expert annotator ($H5$) merges A3 and A4. $H5$ is allowed to accept both A3 *and* A4 as part of the final gold standard if both provide the required information to fill the gap. In 68% of cases A3 and

Genre	GENERIC	GENERALIZ.	STATE	EVENT
Inserted Information	0.81	0.08	0.10	0.01
Microtexts	0.48	0.12	0.22	0.08
Report	0.03	0.04	0.54	0.39

Table 2. Distribution of SE Types among different genres (expressed as percentages)

A4 are accepted as equivalent, in 25% either A3 or A4 are picked, and in 7% of cases, $H5$ supplies a new solution, indicating high quality of the annotations.

4.2 Task IIa: Analysis of Situation Entity Types annotations

The annotator agreement is 0.44 (Cohen's Kappa) for the Situation Entity type annotations on the missing information set. The gold standard was obtained by reannotating the disputed segments. Table 2 shows the distribution of SE types (gold) on the missing information set compared to other genres [1]. The sentences inserted as missing information are characterized by a very high proportion of generics (81%) and very few events (1%), while reports, for example, contain a high proportion of states (54%) and events (39%) [1]. The proportion of generics within the inserted information is significantly higher than the already high one in microtexts (48%). This suggests that the knowledge captured by generic sentences plays an important role with respect to implicit information, and we can use this tendency for acquiring such missing information automatically.

4.3 Task IIb: Analysis of ConceptNet relation type annotations

ConceptNet provides an inventory of 28 relation types, from which our annotators used 20 relations to label the inserted sentences. We measure a relatively low annotator agreement of 0.30 (Cohen's Kappa) and produce a gold standard done by an expert annotator (one of the authors) which provides the basis of our final

CN Relation	Example Tuple	Example sentence
Causes	(penalties, change in behavior)	Penalties lead to a change in behavior.
HasProperty	(control data, expensive)	Control data are expensive.
CapableOf	(camera, snapshots)	The camera can take snapshots.
IsA	(Olympic disciplines, sports)	Olympic disciplines are sports.

Table 3. ConceptNet Relations, with examples from the inserted sentences

CN rel type		CN rel type		CN rel type		CN rel type	
N.A.	0.21	CapableOf	0.08	Desires	0.04	ReceivesAction	0.02
Causes	0.16	UsedFor	0.06	PartOf	0.03	MotivatedByGoal	0.02
HasProperty	0.12	HasSubevent	0.04	HasA	0.02	DefinedAs	0.02
IsA	0.08	HasPrerequisite	0.04	AtLocation	0.02	Others	0.04

Table 4. Proportion of ConceptNet Relation types in annotated data

analysis. Here, 1163 out of 1470 sentences (79%) are labeled with ConceptNet relations. Examples of the relations (taken from our data) and the distribution of relation types are shown in Tables 3 and 4. The most frequently occurring relation is *Causes* (16%) followed by *HasProperty* (12%), *IsA* (8%) and *CapableOf* (8%), while there is a relatively high amount of rarely annotated relation types. 21% of sentences could not be assigned an existing ConceptNet relation.

4.4 Task III: Aligning knowledge annotations with Wikipedia

To test whether we can find sentences in Wikipedia that match inserted information sentences, for each sentence in the inserted information set we find the most similar sentence in the Wikipedia corpus using WMD' (Equation 2) as distance score. For example, the most similar sentence for the inserted sentence *The death penalty extinguishes life* in Wikipedia is *The death penalty is the killing of a person as a punishment for a criminal offense.* The averaged distance between A1$^{\mathbf{c}}$ and Wikipedia is 2.60, very similar to A2$^{\mathbf{c}}$ and Wikipedia (2.66) or A5 and Wikipedia (2.58). These distance scores are also close to A1$^{\mathbf{c}}$-A2$^{\mathbf{c}}$ (cf. Table 1), the distance of the annotations by $H1$ and $H2$ after the first correction round.

We take this as a strong indication that Wikipedia can in fact be a useful source for retrieving information that is missing in arguments. While we did not perform a deeper analysis of the retrieved most similar sentences, we will release them, together with the extracted Wikipedia subcorpus as a background textual knowledge resource to the Microtext corpus.

5 Conclusion

In this paper we present a multi-step annotation method through which we acquire high-quality annotations for implicit knowledge in argumentative texts.

Eliciting implicit knowledge in argumentative texts is a highly complex and subjective task, and formulating such knowledge in natural language sentences adds to the challenge of assessing the quality of the data. We rely on views of 5 human judges who provide, review, select or revise annotations or state novel solutions. With this process we observe continuous evolution towards increased similarity of the annotations, using textual similarity computation, and confirm the high quality of the data set in the final annotation step.

The acquired sentences enrich the argumentative microtexts with carefully curated implicit information. Additional annotation of semantic clause types and common sense knowledge relations further characterize the elicited implicit knowledge: a majority of the inserted sentences are generic. This tendency could be deployed for acquiring such knowledge automatically. A large majority of the sentences can be mapped to common sense knowledge relations as defined in ConceptNet. Thus, knowledge repositories could play an important role in future work on argument analysis. We finally show that the inserted sentences are similar to sentences found in Wikipedia, which suggests that the missing knowledge can be found in textual sources. We thus consider Wikipedia as a valuable textual knowledge resource for automatically acquiring knowledge that is needed to fill gaps in arguments. Future research needs to investigate how the exact knowledge provided by humans can be extracted from such sources.

We release our data set as an extension to the Microtext corpus, to facilitate future research on argument analysis and implicit knowledge acquisition. While the data set size is small, we expect it to be useful for the community as a gold standard for automatically filling knowledge gaps in arguments.

Acknowledgments. This work has been conducted within the Leibniz Science-Campus "Empirical Linguistics and Computational Modeling", funded by the Leibniz Association under grant no. SAS-2015-IDS-LWC and by the Ministry of Science, Research and Art (MWK) of the state of Baden-Württemberg. We thank our annotators Sabrina Effenberger, Jesper Klein, Sarina Meyer and Rebekka Sons for their contribution.

References

1. Becker, M., Palmer, A., Frank, A.: Argumentative texts and clause types. In: Proceedings of the 3rd Workshop on Argument Mining. pp. 21–30 (2016)
2. Becker, M., Palmer, A., Frank, A.: Clause Types and Modality in Argumentative Microtexts. In: Workshop on Foundations of the Language of Argumentation (in conjunction with COMMA). pp. 1–9. Potsdam, Germany (2016)
3. Black, E., Hunter, A.: A relevance-theoretic framework for constructing and deconstructing enthymemes. Journal of Logic and Computation 1(22), 55—78 (2012)
4. Boltuzic, F., Snajder, J.: Fill the Gap! Analyzing Implicit Premises between Claims from Online Debates. In: Proceedings of the 3rd Workshop on Argument Mining. pp. 124–133 (2016)
5. Freeman, J.B.: Argument Structure: Representation and Theory. Springer (2011)
6. Friedrich, A., Palmer, A.: Automatic prediction of aspectual class of verbs in context. In: Proceedings of the ACL 2014 (2014)
7. Grice, H.P.: Logic and conversation. 1975 pp. 41–58 (1975)
8. Havasi, C., Speer, R., Pustejovsky, J., Lieberman, H.: Digital intuition: Applying common sense using dimensionality reduction. IEEE 4(24), 24–35 (2009)
9. Kusner, M.J., Sun, Y., Kolkin, N.I., Weinberger, K.Q.: From Word Embeddings To Document Distances. In: Proceedings of the 32nd International Conference on Machine Learning (ICML). pp. 957–966 (2015)
10. Peldszus, A., Stede, M.: An annotated corpus of argumentative microtexts. In: Proceedings of the First European Conference on Argumentation (2015)
11. Rajendran, P., Bollegala, B., Parsons, S.: Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews. In: Proceedings of the 3rd Workshop on Argument Mining (2016)
12. Razuvayevskaya, O., Teufel, S.: Recognising enthymemes in real-world texts. In: Workshop on Foundations of the Language of Argumentation (2016)
13. Reimers, N., Eckle-Kohler, J., Schnober, C., Kim, J., Gurevych, I.: Germeval-2014: Nested Named Entity Recognition with neural networks. In: Proceedings of the 12th Edition of the KONVENS Conference. p. 117–120 (2014)
14. Sidarenka, U., Peldszus, A., Stede, M.: Discourse Segmentation of German Texts. Journal for Language Technology and Computational Linguistics 30(1) (2015)
15. Speer, R., Havasi, C.: Representing General Relational Knowledge in ConceptNet. In: Proceedings of LREC (2012)