A Multi-layer Annotated Corpus of Argumentative Text: From Argument Schemes to Discourse Relations

Elena Musi[§], Tariq Alhindi[§], Manfred Stede [†], Leonard Kriese [†], Smaranda Muresan [§], Andrea Rocci
Columbia University[§], Potsdam University[†], Universita della Svizzera italiana
475 Riverside Drive New York (USA) [§], Karl-Liebknecht-Str. 24-25 Potsdam (Germany) [†], Via Buffi 13 Lugano (Switzerland)
em3202/ta2509/sm761@columbia.edu [§], stede@uni-potsdam.de, leonard.kriese@hotmail.de, andrea.rocci@usi.ch

Abstract

We present a multi-layer annotated corpus of 112 argumentative microtexts encompassing not only argument structure and discourse relations (Stede et al., 2016), but also argument schemes — the inferential relations linking premises to claims. We propose a set of guidelines for the annotation of argument schemes both for support and attack relations, and a new user-friendly annotation tool. The multi-layer annotated corpus allows us to conduct an initial study of dependencies between discourse relations (according to Rhetorical Structure Theory (Mann and Thompson, 1988)) and argument schemes. Our main contribution is that of offering the first resource for the combined study of (argumentative) discourse relations and inferential moves.

Keywords: argumentation mining, argument schemes, discourse relations

1. Introduction

Recent interest in Argumentation Mining (e.g., (Lippi and Torroni, 2016; Habernal and Gurevych, 2017) has brought to the fore the need for corpora annotated with argument information, which can be used as training data. Generally, the automatic search for arguments encompasses the following steps (Peldszus and Stede, 2013):

1. the segmentation of texts into *argumentative discourse units*(ADUs);
2. the classification of the role (e.g., claim, premise) played by each ADU;
3. the analysis of the relations linking ADUs (e.g., support, attack); and
4. the identification argument schemes, namely the implicit and explicit inferential relations within and across ADUs.

Annotation efforts have so far provided corpora that focus on the first three steps, with genres ranging from persuasive essays and scientific articles to online debates (e.g., (Kirschner et al., 2015; Ghosh et al., 2014; Stab and Gurevych, 2014; Walker et al., 2012). Step 4, which builds a bridge to reasoning, has not received nearly as much attention as the others. A notable exception is the *Araucaria* corpus (Reed and Rowe, 2004), which provides annotations based on Walton et al. (2008a) argument schemes. This annotated corpus has led to work on automatically classifying argument schemes focusing on the five most frequent ones (Feng and Hirst, 2011). Other projects looked at restricted subsets of argument schemes (Green, 2017; Schneider et al., 2013) and have not yet led to publicly-available data.

In this paper, we report on an annotation project that adds information about inferential rules, in the shape of argument schemes, to an existing corpus that already holds annotations of argumentation structure as well as discourse structure based on Rhetorical Structure Theory (RST) and Segmented Discourse Representation Theory(SDRT) (Stede et al., 2016). Our emphasis thus is on a *multi-layer* resource that allows for correlating different levels and for studying dependencies between discourse relations and argument structure, or — enabled by our new result — discourse relations and underlying inferential moves.

Specifically, we propose guidelines for the annotation of argument schemes for both SUPPORT and ATTACK relations using the *Argumentum Model of Topics* framework (Rigotti and Morasso, 2010) (Section 2). These guidelines, based on semantic principles, are scalable to other text genres as well as languages. In addition, we present a new annotation tool for argument schemes with a user-friendly interface to support the annotation process (Section 3). The reliability of the guidelines is tested through a pilot annotation project on top of 40 microtexts, obtaining moderate agreement (Section 4). Finally, we report on initial experiments on the mapping of rhetorical discourse relations and types of argument schemes (Section 5). The annotation guidelines, the corpus and the annotation tool are available at $http://angcl.ling.uni-potsdam.de/resources/argmicro.html$.

2. Data and methods

2.1. Corpus

The 'argumentative microtext' corpus includes 112 short texts that have been produced by means of a text generation experiment: 23 subjects were asked to choose a controversial topic from a fixed list and write a short argumentative text that should contain a clear major claim and arguments supporting or attacking it (Peldszus and Stede, 2016). The texts were written in German by native speakers and later professionally translated into English. The argument structure was annotated according to the scheme proposed by Peldszus and Stede (2013), which builds on the ideas of Freeman (2011). Briefly, the texts are segmented into argumentative units, each unit has an argumentative role and is related to another unit, except for the single main claim (resulting overall in a tree structure). Furthermore, the corpus was enriched with discourse structure information based on

RST and SDRT theories by Stede et al. (2016), and Becker et al. (2016) have provided the additional layer of situation entity types (Becker et al., 2016). Annotating the argument schemes, covering the underlying inferential moves, will provide a valuable annotation layer for studying the mechanics of argumentation from a theoretical, yet empirically grounded perspective, and for argumentation mining.

2.2. Annotation guidelines

Our annotation of argument schemes is grounded in the *Argumentum Model of Topics* (Rigotti and Morasso, 2010). Unlike other approaches (Walton et al., 2008b; Kienpointner and others, 1986; Van Eemeren and Houtlosser, 2006), it offers a *taxonomic hierarchy* of argument schemes with a limited number of schemes that are based on semantic, mutually exclusive distinctive criteria (Figure 1).

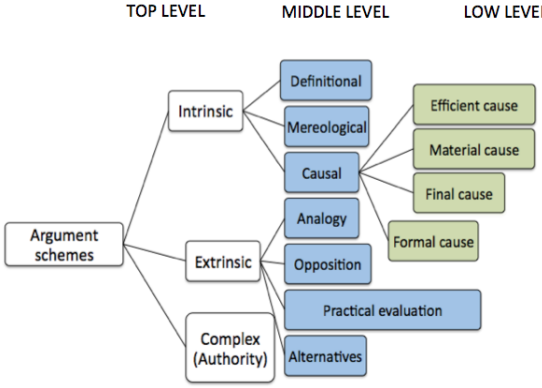


Figure 1: the AMT taxonomy of argument schemes

The annotation consists of two subtasks: 1) given a SUPPORT or REBUT relation, identify the argument scheme among the 8 middle level schemes (DEFINITIONAL, CAUSAL, MEREOLOGICAL, ANALOGY, OPPOSITION, PRACTICAL_EVALUATION, AUTHORITY) or NONE if no reasoning is present; and 2) identify the associated inference rule.

An early pilot annotation testing the first version of guidelines had been carried out on top of 30 persuasive essays from the corpus of Stab and Gurevych (2014), obtaining a fair inter-annotator agreement with trained but non-expert annotators (Musi et al., 2016).

The guidelines contain identification questions, linguistic clues and inferential rules for each argument scheme. Annotators are asked to first browse the identification yes/no questions and check whether inferential rules apply and linguistic clues are indeed present. The description of CAUSAL argument schemes contains, for example, the following information:

- *Identification Question*: is x a cause/effect of y or is it a means to obtain y?
- *Other clues*: Evaluations about actions play a role as common ground knowledge but do not constitute the premise textually expressed.
- *Inferential rules:*
  - if the cause is the case, the effect is the case
  - if the effect is the case, the cause is probably the case
  - if a quality characterizes the cause, then such quality characterizes the effect too
  - if the realization of the goal necessitates the means x, x must be adopted
  - if an action does not allow to achieve the goal, it should not be undertaken
  - if somebody has the means to achieve a certain goal, he will achieve that goal

Compared to the previous version, although the taxonomy and related theoretical insights have been maintained, the definition of the argument schemes that gave rise to the highest disagreement in the previous study have been clarified pointing to distinctive features. In particular, the CAUSAL argument schemes (from means to end) that we mentioned above have been frequently annotated by others as PRACTICAL EVALUATION, described as follows:

- *Identification Question*: does x express an evaluation about a state of affairs and does y express an advice/a recommendation about stopping/continuing the state of affairs the premise refer to?
- *Other clues*: the premise contains adjectives or other linguistic items which qualify something as more or less good
- *Inferential rules:*
  - if something is of important value, it should not be terminated
  - if something has a positive value, it should be supported/continued/promoted/maintained
  - if something has positive effects, it should be supported/continued/promoted/maintained
  - if something has a negative effect it should be terminated

These two argument schemes are perceived as similar due to the common reference to intentionality and to the frame of human action. To overcome this ambiguity, the guidelines have been refined to stress that the distinctive presence of evaluative propositions working as premises and of recommendation speech acts working as claims indicate a PRACTICAL EVALUATION scheme, but not a CAUSAL argument schemes. The reference to the linguistic clues has been retained only when they are argument scheme specific, and the list of inferential rules per argument scheme has been enriched with more examples (e.g. "if the state of affairs x shows a set of features which are also present in the state of affairs y and z holds for x, then z holds for y too" for ANALOGY argument schemes). A complete list of the inferential rules is available in the annotation guidelines. Furthermore, guidelines for the description of attack relations of the rebutt type have been added. Rebuttals are counterarguments which directly challenge the truth of a statement. Therefore, they have been conceived as arguments which support the negation of the proposition functioning as claim.

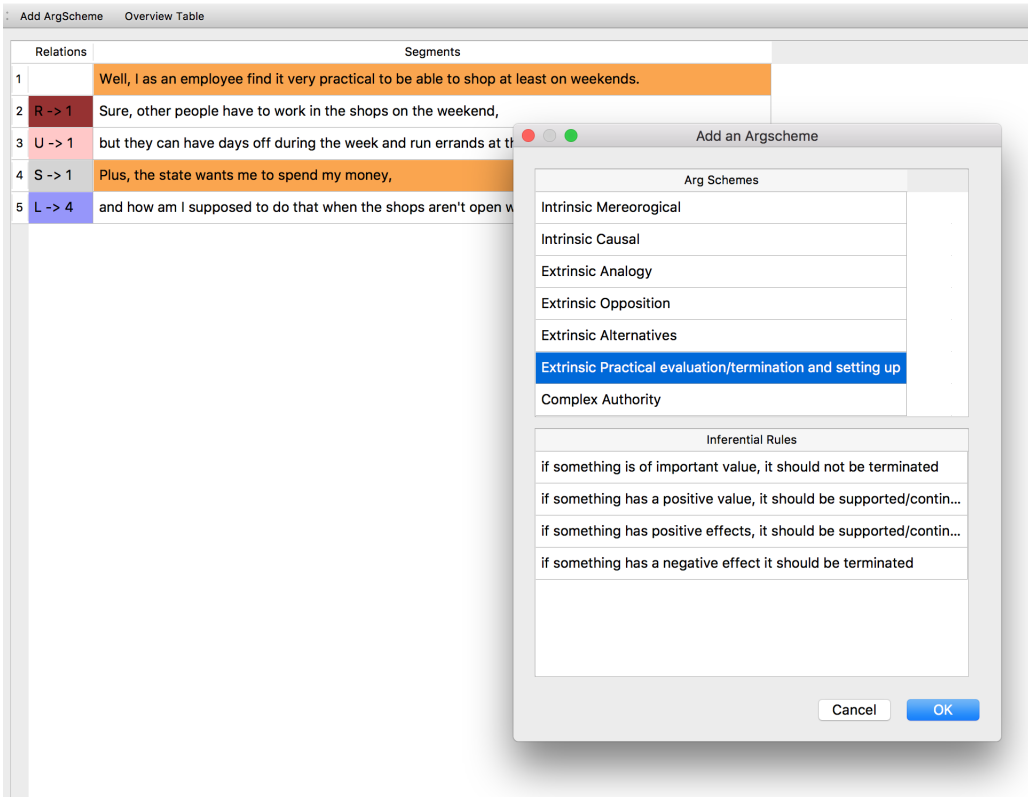


Figure 2: The ArgScheme Annotation Tool

3. The ArgScheme annotation tool

Human annotation is made most efficient when it is supported by an easy-to-use annotation tool. Since we are adding a new layer to the already-existing annotation of argument trees, none of the standard tools for text annotation can be used for our purposes. Therefore, we built the *ArgScheme* Annotator Tool, designed to provide a user-friendly interface for the labeling of support and attack relations with argument schemes. The annotation tool is written in C++ and was built within the Qt framework, which allows for building cross-platform GUI applications. We make the tool available for Linux, Windows, and Mac OS X.

Once a microtext is loaded, the tool shows each proposition paired with the type of relation it instantiates, as well as the index of the proposition it is connected to. After clicking on a relation, a window containing the list of argument schemes opens up. Once they have chosen the right argument scheme, annotators are shown a set of scheme-specific inference rules and are asked to select one. See Figure 2 for illustration. The result of the annotation is stored as additional attributes on the original XML format used in the microtext corpus, allowing for an easy mapping between all different annotation layers.

4. Results

4.1. Inter-Annotator Agreement

To measure inter-annotator agreement, two sets of 20 microtexts have been annotated by three annotators each (4 relations per microtext on average). The six annotators involved included four students with background in Linguistics and Argumentation, and 2 PhD students in NLP. Each annotator underwent a one-hour training session during which the guidelines have been explained and an exemplary analysis collaboratively discussed. To evaluate the reliability of the guidelines, we measured the inter-annotators agreement (IAA) using Fleiss κ to account for multiple annotators (Fleiss and Cohen, 1973). We obtained an IAA (k=0.296), which corresponds to fair agreement. To verify whether the idiosyncratic behavior of some annotators plays a role, we calculated the inter-annotator agreement pairwise for both sets. The IAAs for the first and second sets of microtexts (20 microtexts each) is given in Table 1. The set of three annotators is different for the 2 sets of 20 microtexts.

Annotators	κ **(1st set)**	κ **(2nd set)**
1,2	0.404	0.213
2,3	0.231	0.260
1,3	0.231	0.409

Table 1: pairwise IAA.

As can be noticed in Table 1, the 3rd annotator in the first set and the 2nd annotator in the second set are outliers. Zooming into their annotations, we noticed that annotator 3 has a tendency to signal the lack of argumentation schemes: s(he) annotated NONE in 21 cases, while the other two annotators always picked an argument scheme. Due to these individual tendencies, we have chosen majority voting as gold annotation. A manual check by an expert analyst has revealed the consistency of the choices made by the anno-

	Definitional	Mereorogical	Causal	Analogy	Opposition	Practical_evaluation	Alternatives	Authority	None
Definitional	8	1	10	0	2	14	1	0	16
Mereorogical		42	6	2	2	15	0	0	6
Causal			60	1	12	42	2	1	8
Analogy				12	0	3	0	0	2
opposition					8	5	3	3	3
Practical_evaluation						218	10	8	29
alternatives							4	0	4
Authority								24	0
none									6

Table 2: Confusion Matrix (since the matrix is symmetric we only show its upper triangular part)

tators who were in agreement. Considering the inherent difficulty of the task, which necessarily gives rise to ambiguities and disagreement (even among argumentation analysts) and the presence of outliers, this IAA score can be considered a positive index of the guidelines' reliability. On this basis, we have asked an additional set of seven highly expert annotators (researchers in Linguistics and Argumentation Theory) to annotate a set of 10 microtexts each. As a result, we release to the community a set of 112 microtexts fully annotated as to argument schemes.

4.2. Analysis of the disagreement space

With the aim of understanding the reasons underlying disagreements among annotators, we have built a confusion matrix as shown in Table 2.

We notice that CAUSAL scheme is still confused with PRACTICAL_EVALUATION, while DEFINITIONAL scheme is often confused with PRACTICAL_EVALUATION and NONE (i.e., no argument). The confusion between CAUSAL and PRACTICAL_EVALUATION has already been attested in the previous annotation project (Musi et al., 2016), as a confirmation of the perceived closeness of the two types of argument schemes.

From the qualitative analysis it emerges that DEFINITIONAL argument schemes have not been recognized by some annotators and have been annotated as NONE in particular for rebuttal relations (e.g., "Anti-virus software protects the users of a computer from dangers from the internet". "Admittedly they do not generally prevent you from catching a virus"). A possible explanation is that in those cases, the presence of category in the claim is less salient since it has to be recognized after the claim has been negated.

Finally, DEFINITIONAL argument schemes tend to be confused with PRACTICAL_EVALUATION argument schemes since the evaluative propositions featuring as premises in PRACTICAL_EVALUATION have been conceived as instances of categorizations. The premise-claim pair "Supermarkets and shopping centers should be allowed to open on any Sundays and holidays"."Considering the growing digitization of society, the traditional model of the weekend will soon be obsolete" constitutes an instance of PRACTICAL_EVALUATION argument scheme: a recommendation about carrying out an action is proposed on the basis of the positive/negative evaluation (e.g., "obsolete") of another state of affairs. The annotator that has chosen DEFINITIONAL has most likely interpreted the premise as a categorization. However, the categorization process is not at the basis of the inference that allows to support the conclusion. Even though the available data are not enough to draw generalizations, we consider them as hints to be used to refine the guidelines. In particular, we plan to stress on the differences between the propositional and the inferential levels of analysis.

5. Argument schemes and rhetorical relations

RST discourse relations (Mann and Thompson, 1988) overlap with argumentative discourse relations, particularly the subset of *Presentational* relations meant to increase the reader's positive inclination for the proposition functioning as nucleus clearly performing an argumentative act. The mapping of rhetorical discourse relations onto argumentative relations carried out by Stede et al. (2016) confirms this pragmatic similarity: the rhetorical relation REASON co-occurs for example 99 times with SUPPORT relations. Attack relations of the REBUT type tend, instead, to co-occur with the relation ANTITHESIS, while those of the UNDERCUT type with CONCESSIONS. In addition, Stede et al. (2016) showed that some *Subject-matter* relations, meant to allow the reader to recognize a semantic relation, may instantiate an argumentative relation in specific contexts. In order to extend this investigation by analyzing the correlations among rhetorical relations and argument schemes, we have mapped the annotation of argument schemes on top of RST relations and observed as well as explained attested overlaps or mismatches.

5.1. Mapping argument schemes to discourse relations

The mapping between RST discourse relations and argument schemes is done using an approach similar to the one introduced by Stede et al. (2016). Specifically, we convert both annotations to a common dependency structure where each relation has exactly one ADU in both its source and target. Argument scheme annotations follow a dependency structure for all relations except for the UNDERCUT relation where the target is another relation between ADUs,

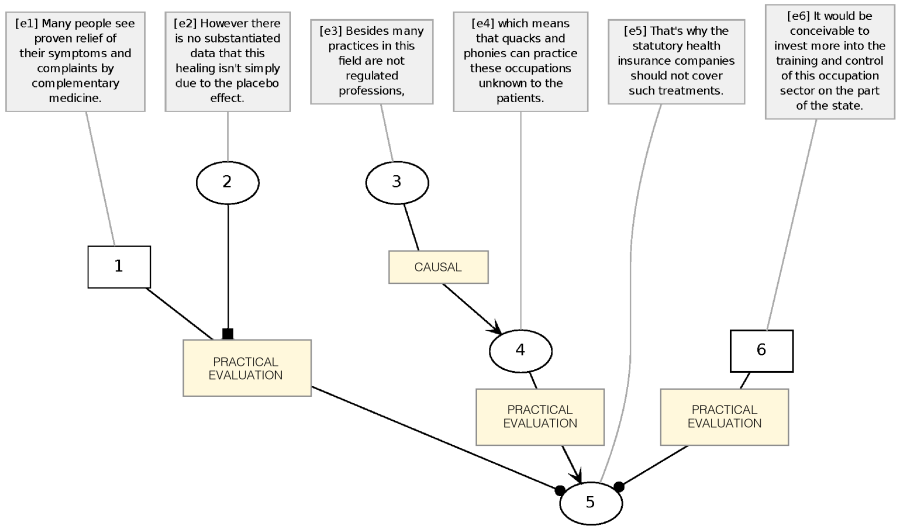


Figure 3: Argumentation structure and argument schemes of an example microtext

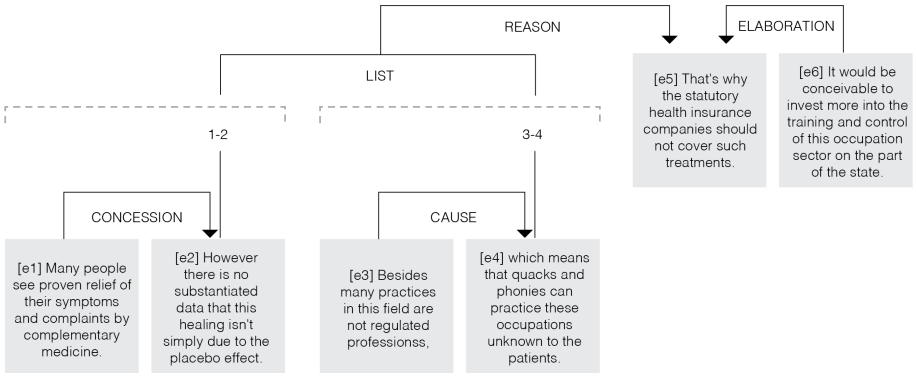
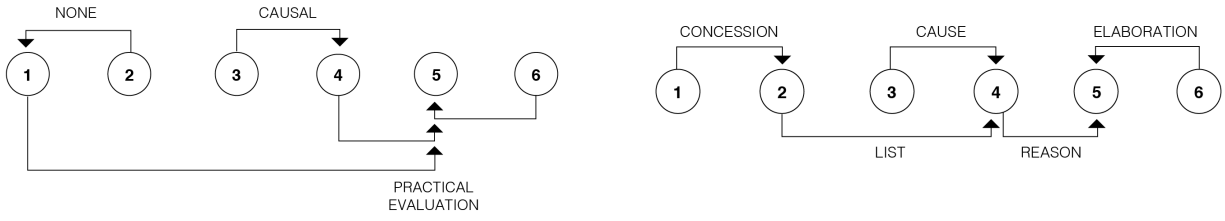


Figure 4: RST discourse structure of the same microtext given in Figure 3



(a) Argument Schemes

(b) RST

Figure 5: Conversion of argument schemes and RST relations to dependency structures

which is converted to the source of that relation (Stede et al., 2016). For RST, converting the relations to dependency structure is done in two steps. First, each *elementary discourse unit* (EDU) in RST is linked to its corresponding ADU in the argumentation structure. This is a one-to-one mapping when there is no join relation. If there is a join relation, multiple EDUs are mapped to a single ADU. Second, multi-nuclear nodes in RST are mapped with the ADU of its leftmost child as done in (Stede et al., 2016).

To illustrate, Figure 3 shows an example of argument schemes annotation, where the argument structure of the essay has six ADUs. Four out of the six ADUs represent the proponent's stances, who presents and defends the claim (round nodes), and the other two give voice to an imaginary opponent, who questions the claim (square nodes). The example has two REBUT relations, two SUPPORT relations and one UNDERCUT relation shown by the end symbol of each edge (circle for REBUT, arrow-head for SUPPORT an square for UNDERCUT). The argument schemes, which are available only for SUPPORT and REBUT relations, are shown as

the labels on each edge.

The corresponding RST structure for the same microtext is shown in Figure 4. It has the same six EDUs (there is a one-to-one mapping between EDUs are ADUs in this example) with five different discourse relations (REASON, LIST, CONCESSION, CAUSE, ELABORATION). It is worth noting that multi-nuclear discourse relations do not have an EDU as their target. They target another relation instead, which is the case for the LIST relation in the illustrated example.

The argumentation structure of the example shown in Figure 3 has an UNDERCUT relation. Therefore, the dependency structure of the argument scheme shown in Figure 5(a) is equivalent to the argument structure in Figure 3 except for the UNDERCUT relation, which maintains the same source and changes the target to the source of its target relation (i.e., $2 \rightarrow (1 \rightarrow 5)$ is changed to $2 \rightarrow 1$). To convert the RST structure of the microtext in Figure 4 to its dependency structure shown in Figure 5(b), the following steps are done:

1. All EDUs are labeled with their corresponding ADU using the argument structure of the microtext as a reference.

2. Converting multi-nuclear relations, LIST in this example, to the two following relations:

   (a) A relation with the source as the leftmost child of LIST and the target as the target of the next relation (i.e., target of REASON) and label as REASON.

   (b) A relation with the source as rightmost child of LIST and the target as its leftmost child and label as LIST.

The resulting common dependency structure of the argument schemes and RST discourse relations (Figure 5), allows us to analyze the overlap between the two annotations. We can see that three relations ($3 \rightarrow 4$, $4 \rightarrow 5$ and $6 \rightarrow 5$) exist in both RST and argument schemes, while one relation in each annotation does not exist in the other ($1 \rightarrow 5$ in argument schemes and $2 \rightarrow 4$ in RST). A study of the overlap between RST and argument schemes across the whole dataset is presented in the following section.

5.2. Analysis of the correlations

The overlap between argument schemes for SUPPORT and REBUT relations and RST discourse relations in our corpus is visualized in Table 3 and Table 4, respectively.

As a premise, it has to be stated that, with respect to the traditional taxonomy of RST relations (Mann and Thompson, 1988), two relations have been added by (Stede et al., 2016) according to Stede and Taboada (2017). Among *Presentational* relations, the new label REASON shows as nucleus "a subjective statement/thesis/claim, which R might not accept or might not regard as sufficiently important or positive" and as satellite "a subjective statement/thesis/claim"; the link between nucleus and satellite implies that "understanding S makes it easier for R to accept N, or to share the particular viewpoint of W". While sharing with REASON the same constraints on the nucleus, as well as the same effect ("R's belief in N is increased"), the relations EVIDENCE and JUSTIFY imply different constraints on the satellite, that has to state an objective description of a fact —EVIDENCE — or a fundamental attitude of the acting per-

	AUTHORITY	ALTERNATIVES	ANALOGY	OPPOSITION	PRACTICAL_EVALUATION	CAUSAL	DEFINITIONAL	MEREOROGICAL	NO SCHEME	NONE
ANTITHESIS	0	1	0	0	0	0	0	0	0	0
BACKGROUND	1	0	0	1	1	3	0	0	0	0
CAUSE	0	0	0	0	1	8	2	1	0	0
CIRCUMSTANCE	0	0	0	0	0	0	0	0	0	0
CONCESSION	0	0	0	0	0	0	0	0	0	0
CONDITION	0	0	0	0	0	1	0	0	0	0
CONJUNCTION	0	0	0	0	0	0	0	0	0	0
CONTRAST	0	0	0	0	0	0	0	0	0	0
DISJUNCTION	0	0	0	0	0	0	0	0	0	0
E-ELABORATION	0	0	0	0	0	0	0	0	0	0
ELABORATION	0	0	0	0	0	1	0	3	0	0
EVIDENCE	0	0	0	0	1	2	1	4	0	0
INTERPRETATION	0	0	0	0	0	0	0	0	0	0
JOINT	0	1	0	0	2	1	0	0	0	0
JUSTIFY	0	0	0	0	4	0	0	0	0	0
LIST	0	0	0	0	1	1	0	0	0	0
MOTIVATION	0	0	0	0	2	0	0	0	0	0
REASON	3	6	4	4	52	29	5	4	1	0
RESTATEMENT	0	0	0	0	0	1	1	0	0	0
RESULT	0	0	0	0	1	0	0	0	0	0
SAMEUNIT	0	0	0	0	1	0	0	0	0	0
SOLUTIONHOOD	0	0	0	0	0	0	0	0	0	0
UNLESS	0	0	0	0	0	0	0	0	0	0
NONE	3	1	3	0	45	17	6	4	1	1

Table 3: Overlap between RST relations and argument schemes for SUPPORT relations

	AUTHORITY	ALTERNATIVES	ANALOGY	OPPOSITION	PRACTICAL_EVALUATION	CAUSAL	DEFINITIONAL	MEREOROGICAL	NO SCHEME	NONE
ANTITHESIS	1	0	0	1	2	3	2	1	0	1
BACKGROUND	0	0	0	0	0	0	0	0	0	0
CAUSE	0	0	0	0	0	0	0	0	0	0
CIRCUMSTANCE	0	0	0	0	0	0	0	0	0	0
CONCESSION	1	0	0	1	3	0	1	1	0	0
CONDITION	0	0	0	0	1	0	0	0	0	0
CONJUNCTION	0	0	0	0	0	0	0	0	0	0
CONTRAST	0	0	0	1	0	0	0	0	0	0
DISJUNCTION	0	0	0	0	0	0	0	0	0	0
E-ELABORATION	0	0	0	0	0	0	0	0	0	0
ELABORATION	0	1	0	0	1	0	0	0	0	0
EVIDENCE	0	0	0	0	0	0	0	0	0	0
INTERPRETATION	0	0	0	0	0	0	0	0	0	0
JOINT	1	0	0	0	0	0	0	0	0	0
JUSTIFY	0	0	0	0	0	0	0	0	0	0
LIST	0	0	0	0	1	0	0	0	0	0
MOTIVATION	0	0	0	0	0	0	0	0	0	0
REASON	0	0	0	0	3	0	0	0	0	0
RESTATEMENT	0	0	0	0	0	0	0	0	0	0
RESULT	0	0	0	0	0	0	0	0	0	0
SAMEUNIT	0	0	0	0	0	1	0	0	0	0
SOLUTIONHOOD	0	0	0	0	0	0	0	0	0	0
UNLESS	0	0	0	0	1	0	0	0	0	0
NONE	2	3	3	6	33	11	6	5	1	0

Table 4: Overlap between RST relations and argument schemes for REBUT relations

son — JUSTIFY —. Among *Subject Matter* RST relations, the label E-ELABORATION differs from simple ELABORATION since the satellite provides more information about a single entity mentioned in the nucleus, rather than about the described state of affairs.
Comparing Table 3 and Table 4 it is clear that some RST relations overlap with argument schemes only in the presence of REBUT relations or SUPPORT relations, respectively.

Presentational RST Relations and Argument Schemes. As far as *Presentational* RST relations are concerned, ANTITHESIS and CONCESSION instantiate inferences only with REBUT relations, but do not seem to constrain the type of argument scheme at issue. Both relations entail a hierarchy of importance between the content of the nucleus (considered more important) and that of the satellite. The two relations differ in the type of comparison that is established: with ANTITHESIS, the propositions forming the nucleus and the satellite are deemed incompatible while with CONCESSIONS they are presented as both valid, while the nucleus is emphasized (e.g. "Although IBM s numbers haven't been staggering recently. You should buy IBM shares if you want to invest."). In both cases the inference underlying the creation of such a hierarchy can vary from context to context. However, it cannot serve e support function: the opposition relation signaled by the two relations makes the satellite support the negation of the nucleus and viceversa. In terms of frequency, the most frequent RST relation is REASON (92), followed by EVIDENCE (8), JUSTIFY (4) and MOTIVATION (2). While there is no attested overlap between argument schemes and EVIDENCE, JUSTIFY and MOTIVATION with REBUT relations, there is no incompatibility from a theoretical point of view.
The RST relation REASON overlaps with the widest range of argument schemes, with a preference for PRACTICAL EVALUATIONS — more than half of overlapping cases (e.g. "Actually only those people should pay a TV and radio licence fee who really watch ARD, ZDF, Arte etc. It is in fact good to support sophisticated programming through fees") — and CAUSAL argument schemes — ca. one third of overlapping cases. The subjectivity constraint imposed on the satellite well matches the semantic type of premises occurring with the argument scheme PRACTICAL_EVALUATION: evaluations, containing sentiment, presuppose a subjective perspective. CAUSAL argument argument schemes, coinciding with the REASON relation, show premises which are not descriptions of states of affairs, but speakers' interpretations (e.g."The universities in Germany should not under any circumstances charge tuition fees. The anticipated objectives of tuition fees can be achieved by other means."). Premises encoding factual states of affairs are, instead, present when CAUSAL argument schemes overlap with the RST relation EVIDENCE, due to the objectivity constraint imposed by this discourse relation on the satellite (e.g., "Our society is in danger of overheating due to 'never clocking off'.The last 20 years have seen the Sunday off work increasingly sacrificed to commerce."). From a semantic point of view, the most suitable argument scheme to overlap with EVIDENCE is MEREOLOGICAL: premises express states of affairs whom realization is attestable and, thus, objective ("Especially the home games have seen the team stay behind expectations. A sad highlight: Losing 60:87 to the weakish Mannheimers in a sold-out stadium."). However, due to the scarcity of data, it is not possible to statistically evaluate the significance of this preferential association.
The RST relations JUSTIFICATION and MOTIVATION overlap in our corpus with the argument scheme PRACTICAL EVALUATION only. This is probably due to the fact that the recommendation about setting up or terminating a certain action featuring as a claim with this type of argument scheme mirrors the agentive constraint imposed by the relations JUSTIFY (e.g. "Today I will delete my Facebook account.[Not only because] I'm constantly astonished about who wants to be my friend") and MOTIVATION (e.g "[But] still Germany produces way too much rubbish.We Berliners should take the chance and become pioneers in waste separation!") on their nuclei.

Subject-matter RST Relations and Argument Schemes. Turning to *Subject-matter* relations, BACKGROUND and RESTATEMENT overlap only with arguments schemes used in SUPPORT relations. These two relations are *per definition* incompatible with REBUT relations since the BACKGROUND relation has as effect that of increasing the reader's ability to comprehend the nucleus, while RESTATEMENT links a nucleus and a satellite with almost equivalent propositional contents. Although RESTATEMENT is primarily a textual and not a pragmatic relation, serving mainly to organize the text, it can acquire an argumentative function when the satellite is a rephrase that my help the reader in understanding the nucleus ("Intelligence services apparently enjoy indiscriminate liberties. No one can follow their activities in detail."). The inferential link between the satellite and nucleus is necessarily belonging to the INTRINSIC class since the state of affairs encoded by a rephrase necessarily belong to the same frame of that expressed by the rephrased proposition.
The RST relation CAUSE recurrently correlates with argument schemes (11 cases of overlap) that are for the majority, with no surprise, of the CAUSAL type. Although not constituting an inherently argumentative effect, making the reader aware of a cause/effect relation might have a persuasive outcome, especially when the effect/cause works as an argument in support of a general claim: for instance, the claim "Fees result in longer durations of studies" supported by the premise "That's costly!" through a causal relation, works as an argument for the standpoint "Tuition fees should not be charged in Germany."

6. Conclusion

We presented a *multi-layer* annotated corpus of 112 short argumentative texts that allows for correlating different levels of annotation and for studying dependencies between discourse relations and the underlying inferential moves. We presented a theoretically grounded annotation study of argument schemes both for support and attack relations together with a novel annotation tool. In the pilot annotation project to test the guidelines' reliability we obtained fair agreement. The qualitative analysis has shown the presence of outliers among the annotators and it has shed light on the main reasons underlying disagreement. We

plan to refine the guidelines accordingly. We have shown initial results from the correlation between rhetorical relations and argument schemes: *Presentational* RST relations such as REASON and ANTITHESIS select SUPPORT and REBUT relations, respectively, but do not point to specific argument schemes, while the relation EVIDENCE correlates with MEREOLOGICAL and AUTHORITY argument schemes. *Subject-matter* rhetorical relations such as CAUSE seems to constrain the argument scheme at issue.

7. Bibliographical References

Becker, M., Palmer, A., and Frank, A. (2016). Argumentative texts and clause types. In *Proceedings of the Third Workshop on Argumentation Mining*, Berlin. Association for Computational Linguistics.

Feng, V. W. and Hirst, G. (2011). Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 987–996. Association for Computational Linguistics.

Fleiss, J. L. and Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33(3):613–619.

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

Ghosh, D., Muresan, S., Wacholder, N., Aakhus, M., and Mitsui, M. (2014). Analyzing argumentative discourse units in online interactions. In *ArgMining@ ACL*, pages 39–48.

Green, N. (2017). Manual identification of arguments with implicit conclusions using semantic rules for argument mining. In *Proceedings of the 4th Workshop on Argument Mining*, pages 73–78.

Habernal, I. and Gurevych, I. (2017). Argumentation mining in user-generated web discourse. *Computational linguistics*, 43(1):125–179.

Kienpointner, M. et al. (1986). Towards a typology of argument schemes. *Proceedings of ISSA 1986*.

Kirschner, C., Eckle-Kohler, J., and Gurevych, I. (2015). Linking the thoughts: Analysis of argumentation structures in scientific publications. In *ArgMining@ HLT-NAACL*, pages 1–11.

Lippi, M. and Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Trans. Internet Technol.*, 16(2):10:1–10:25.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.

Musi, E., Ghosh, D., and Muresan, S. (2016). Towards feasible guidelines for the annotation of argument schemes. In *Proceedings of the third workshop on argument mining (ArgMining2016)*, pages 82–93.

Peldszus, A. and Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1):1–31.

Peldszus, A. and Stede, M. (2016). An annotated corpus of argumentative microtexts. In *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon 2015 / Vol. 2*, pages 801–816, London. College Publications.

Reed, C. and Rowe, G. (2004). Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(04):961–979.

Rigotti, E. and Morasso, S. G. (2010). Comparing the argumentum model of topics to other contemporary approaches to argument schemes: the procedural and material components. *Argumentation*, 24(4):489–512.

Schneider, J., Samp, K., Passant, A., and Decker, S. (2013). Arguments about deletion: How experience improves the acceptability of arguments in ad-hoc online task groups. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1069–1080. ACM.

Stab, C. and Gurevych, I. (2014). Annotating argument components and relations in persuasive essays. In *COLING*, pages 1501–1510.

Stede, M. and Taboada, M. (2017). Annotation guidelines for rhetorical structure. *manuscript*.

Stede, M., Afantenos, S. D., Peldszus, A., Asher, N., and Perret, J. (2016). Parallel discourse annotations on a corpus of short texts. In *LREC*.

Van Eemeren, F. H. and Houtlosser, P. (2006). Strategic maneuvering: A synthetic recapitulation. *Argumentation*, 20(4):381–392.

Walker, M., Tree, J. F., Anand, P., Abbott, R., and King, J. (2012). A corpus for research on deliberation and debate. In *Proceedings of the Eighth Conference on International Language Resources and Evaluation (LREC)*, Istanbul/Turkey.

Walton, D., Reed, C., and Macagno, F. (2008a). *Argumentation schemes*. Cambridge University Press, Cambridge.

Walton, D., Reed, C., and Macagno, F. (2008b). *Argumentation schemes*. Cambridge University Press.