

Guest Editorial

Special issue: Foundations of the language of argumentation

Patrick Saint-Dizier^{a,*} and Manfred Stede^{b,**}

^a *Institut de Recherches en Informatique de Toulouse, IRIT – CNRS, 118, route de Narbonne 31062, Toulouse Cedex, France*

^b *EB Kognitionswissenschaften, University of Potsdam, Karl-Liebknecht-Str. 24-25, D-14476 Golm, Germany*

Argumentation mining, the automatic search for arguments in texts (in favour or against a particular standpoint) has become increasingly popular during the last few years. One reason is that interesting applications, including some of commercial relevance, are waiting to be explored here. For instance, there is an obvious connection to sentiment analysis or opinion mining: While this discipline seeks to find out how people feel about certain products or events or people, argumentation mining aims at discovering what people call for, why they do that, and how they support their claims.

In mainstream Computational Linguistics, the usual practice is to devise schemes for manual annotation, produce corpora annotated along those lines, and then employ machine learning regimes of various kinds so that unseen data can be analysed automatically. Large repositories of annotated data are now available for many purposes and tasks. For successful, robust applications, the features underlying the automatic classifiers tend to be relatively close to the linguistic surface, so that established tools can produce them reliably. For several years, this line of work has now been applied to argumentation mining, too.

However, full-fledged argumentation mining, given a standpoint toward some critical question, is a very difficult problem, and the approaches are still in a relatively early development stage. The research area is much more complex than, for example, computing textual entailment, factoid question answering or even ‘why’-question answering. To go beyond the current state of the art in argument mining, we feel some methodological guidelines should be taken into account with much care: (1) Proceed step by step: given a standpoint, gradually identify and annotate arguments in isolation. (2) Assume that arguments are rational and are expressed with expressions of a certain regularity, with moderate rhetorical effects. (3) Analyze their linguistic and possibly some simple pragmatic characteristics and generalize them. (4) Identify recurring features to be annotated; test and evaluate them; and define guidelines, which may be revised if needed.

*Corresponding author. E-mail: stdizier@irit.fr.

**Corresponding author. E-mail: stede@uni-potsdam.de.

At the moment, argumentation mining is based essentially on surface linguistic (or even typographic) cues. Most applications consider only specific types of arguments or domains, such as consumer evaluations or citizen opinion on a political decision. The ultimate goals often are

- to construct a graph structure representing the components of the argumentation and their relations,
- to identify underlying value systems that motivate the interpretation of certain arguments, and
- to identify the recurrent argument schemes and to analyze their validity, perception and persuasion impact.

Finally, the counterpart of argumentation mining, argument synthesis is only beginning to be explored now. This step is crucial to be able to present or summarize the mined arguments in a way that is accurate and informative for users.

The workshop from which this special issue is derived focused not on the applications or on the general task of argumentation mining, but looked at a specific side of the problem. Corpus observation shows that argumentation is a complex activity that comes in many linguistic flavours, and in order to more fully understand it with computational methods, it is necessary to pay attention to linguistic detail: How can authors' claims be distinguished from their premises – or from "neutral" text that is not part of the argument proper? More generally, how can the units of a complex argument be delimited, and the relations among them be identified? What is the underlying reasoning pattern or argument scheme? As soon as we move beyond relatively simple text, this becomes a very hard task, which requires attention to detail, both at the linguistic surface and underneath it. – This is the theme underlying the papers in this volume, which all address different aspects of argumentation, but they have a common interest in the linguistic underpinnings.

This issue is a follow-up activity of the workshop "Foundations of the Language of Argumentation", which was a satellite event of the conference "Computational Models of Argument" (COMMA) 2016 in Potsdam/Germany. As a result of a post-workshop resubmission and review procedure, we selected six papers that address specific linguistic issues which play important roles for an accurate automatic analysis of arguments.

The paper entitled "Semantic clause types and modality as features for argument analysis" by M. Becker et al. develops an in-depth analysis of the role of semantic clause types and modality in argumentation. For that purpose, micro-texts of about 5 sentences are annotated according to two main features: situation entity classes and modal verbs with their modal senses. These features are established linguistic categories, but so far they have not been systematically exploited for argument annotation. Correspondences between these two linguistic features are investigated according to several main dimensions of argumentation (such as premises and conclusions) and their functions (such as support and rebuttal). Useful correlations are established and some forms of evidence is found concerning the role played by modal senses of verbs for identifying premises and conclusions. These results are of much interest for an automatic analysis of arguments to construct an argumentation graph.

Then, "Finding enthymemes in real-world texts: A feasibility study" by O. Razuvayevskaya and S. Teufel deals with the complex problem of enthymeme reconstruction. This is a crucial point in the construction of an argumentation graph, since enthymemes, qua their absence, can be reconstructed by listeners in various manners. This is clearly a very challenging task that requires the analysis of a diversity of linguistic and pragmatic factors. A feasibility study is proposed in this contribution dedicated to finding and expanding enthymemes involving a fortiori arguments in real world texts. The authors show that given a sufficiently strict reformulation of the human annotation task, substantial agreement can be

achieved. In particular, to limit complexity, the English “let alone” construction is explored together with its pragmatic effects.

The next article is entitled “Discourse relations: Genre-specific degrees of overtness in argumentative and narrative discourse” by C. Hofmockel et al. This contribution explores the impact of the notion of linguistic genre on the realizations of discourse relations, paying particular attention to the argumentative text type and its preferential realizations of Contrast, Continuation, Elaboration and Explanation/Result. The investigation is realized on two very different types of genres: editorials and personal narratives. This article shows significant and systematic differences in the realization of discourse relations. Genre is then analyzed as a kind a frame that constrains the production and the interpretation of discourse relations. The study of genre-specific variation of overt vs. implicit discourse relations realization is of much interest for argument mining to predict elements of argumentation structure from discourse structures.

The article “Rhetorical strategies in German argumentative dialogs” by A. Hautli-Janisz and M. El-Assady investigates specific aspects of verbal and non-verbal rhetoric. For that purpose, the authors employ an interdisciplinary approach that combines linguistic analysis with argument identification and information visualization. A classification of rhetorical information triggered by a comprehensive set of 27 discourse particles in German is proposed and validated. This classification can support the computational analysis of the framing of argumentation in multi-party natural language. Concerning the original feature of visualization, the challenge lies in representing the complexity of the interactions between the different levels of argumentation.

Next, “Evidently epistential adverbs are argumentative Indicators: A corpus-based study” by E. Musi and A. Rocci focusses on specific adverbs in English and Italian that express evidentiality, namely “evidently” and “evidentemente” as used in comparable corpora of newspaper articles. These texts are taken as good candidates for argumentative discourse and studying argument structures. Data analysis shows that these indicators operate both at the structural and at the inferential level: besides pointing to the presence of premise-conclusion relations, they recurrently pattern with causal argument schemes from the effect to the cause. The Italian adverb seems to be less polysemous and less frequent than its English counterpart.

Finally, “Knowledge-driven argument mining based on the qualia structure” by P. Saint-Dizier develops an analysis of the impact of knowledge on argument mining. Given a standpoint and a set of texts, the problem is to find relevant arguments for or against this standpoint. Via a corpus analysis and specific annotations, the author investigates various forms of knowledge that are typically used in arguments. Basically, concepts related to parts, functions, purposes and goals of the concepts in the standpoint are the main types of knowledge that contribute to identifying arguments. The author provides interesting figures on the crucial impact of knowledge to mine arguments.

We wish to thank the members of the reviewing board for their help in choosing and shaping the papers: Miriam Butt, Nancy Green, Iryna Gurevych, Graeme Hirst, Ralf Klabunde, Robert Mercer, and Maite Taboada.

Patrick Saint-Dizier, CNRS – IRIT
Manfred Stede, University of Potsdam
Guest editors.