

Investigating Argument Relatedness Based on Linguistic Knowledge

Marie Garnier, Sarah Bourse

Cultures Anglo-Saxonnes
 Université Toulouse 2 Jean Jaurès
 5 allée Antonio Machado
 31058 Toulouse, France

Email: mgarnier@univ-tlse2.fr,
 sarah.bourse@univ-tlse2.fr

Patrick Saint-Dizier

CNRS - IRIT
 118 route de Narbonne
 31062 Toulouse Cedex France
 Email: stdizier@irit.fr

Abstract—In this contribution, we explore the type of linguistic knowledge that is required to establish relatedness between a claim and a justification which may be distant or in different texts, within the framework of argument mining. We propose an original annotation method based on XML-Frames and a linguistic analysis of the main resources which are needed to establish relatedness on a linguistic basis.

Keywords—Argument Mining; Linguistic Knowledge; Relatedness; Annotation.

I. AIMS AND MOTIVATIONS

One of the main challenges of argument mining is to correctly identify the statements that are justifications or that support a given claim across different types of online texts (for example, news articles, blogs, consumer reviews). This issue is called argument relatedness and it is a central point in information retrieval. It is also essential in argument mining [1]-[3]. The main objective of research on argument relatedness is to be able to mine statements which develop the same topic as the given claim and have an argumentative orientation. Broadly speaking, relatedness is a measure of the semantic and topical proximity of two text spans. These segments may differ lexically (via the use of synonyms, hyponyms or hypernyms, to name a few possibilities) or syntactically (for example, with alternations, such as active vs. passive structures). Previous research [4] has shown that establishing relatedness between an argument and a statement requires knowledge in 58% to 88% of situations, depending on the topic of the claim.

Since supports and attacks (elements that oppose a claim) of a claim mainly address the purposes, goals, functions or structure of the main concepts of the claim, previous studies have used the Qualia structure of the Generative Lexicon [5] as a knowledge representation system for relatedness. This approach pairs domain knowledge with lexical semantics in an efficient and principled way. However, this previous work also shows that Qualia structures are difficult to develop and must be defined for each topic. This makes knowledge-based argument mining an approach that, although effective, is difficult to reuse over different domains.

The current research project examines and evaluates the possibility of establishing relatedness solely on the basis of linguistic knowledge and lexical semantics. The development of general-purpose linguistic processes and resources that characterize relatedness would make the implementation of

relatedness much simpler and much more reusable over domains. This contribution explores this hypothesis as well as the linguistic knowledge which is required, in particular lexical semantics.

The rest of the paper is structured as follows. The analysis protocol and the annotation system are presented in Section II, while Section III deals with the future steps that need to be taken to be able to establish relatedness on a linguistic basis.

II. ANALYSIS PROTOCOL

This research introduces two specificities. First, the analysis of relatedness is based on the topical content of the claim, as a claim-driven analysis allows the analysis to focus on the features of the claim and to integrate the new elements found in various statements related to that claim. In addition, the annotation is not based on standard linear text annotations but on the use of frames encoded in XML: the use of an XML-Frame approach is motivated by the fact that the elements found in statements and that are decisive for the analysis of the semantic elements of relatedness may not be adjacent, which makes text annotation, which is linear, almost intractable.

In this framework, relevant statements are extracted from the source text and fed into XML-Frames in which the features are filled in manually by annotators. Each statement found to be related to the claim and with an argumentative orientation originates an instance of the frame. The result is a set of frames which can be organized as a tree, where the root is the frame representing the claim and the children are those statements found in texts and that introduce additional constraints on the topic. These additional constraints on the claim topic characterize relatedness.

The relations of each statement with the claim are described in each frame instance through the use of features indicating the linguistic and conceptual links between the claim and the statement. Our corpus is based on texts about controversial social issues, addressing topics such as affirmative action or the gender pay gap. These are relatively complex issues, which guarantees that the need for linguistic and conceptual knowledge will be apparent. The goal is then to mine statements which are related to this claim in various texts. These statements must have a topic that is subsumed by the claim topic and an argumentative orientation which may support or attack the claim, depending on the content of the statement. The argumentative orientation is given by evaluative expressions

such as scalar adjectives, possibly modified by an adverb of intensity. Those statements are also frequently associated with discourse structures which further develop them.

The frame template we have defined for the study of relatedness is very detailed. Inside the general frame <statement>, four sub-frames are embedded (<topic>, <evaluative>, <discourse>, <arg_scheme>), with the first two subframes also including two additional subframes which allow for the identification of a main topic and a field of application for this topic.

As an illustration, let us consider the following claim: *affirmative action in education is good for the economy*. This claim is composed of a topic: *affirmative action in education* and an evaluative expression: *is good for the economy*. The topic is itself composed of a main concept, *affirmative action* and a field of application for this concept, *in education*. The evaluative expression is analyzed in the same way (*is good and for the economy*). The annotation scheme is presented below in more detail:

```
<statement> <topic> <top_main markers= ,
  link= , concept_op= ,
  restrictions= ,
  annotator_confidence= >,
  <top_field markers= , link= ,
  concept_op= ,
  restrictions= ,
  annotator_confidence= > <\topic>
<evaluative> <ev_main markers= ,
  polarity= , strength= ,
  restrictions= ,
  annotator_confidence= >,
  <ev_field markers= , link= ,
  concept_op= ,
  restrictions= ,
  annotator_confidence= >
  <\evaluative>
<discourse> <text= , type= > </discourse>
<arg_scheme
  type= , annotator_confidence= >
<\statement>
```

To say it briefly, this frame allows the description of most features that characterize relatedness. The 'link' and 'concept_op' features respectively specify the linguistic link (exact words, derivation, semantic field, etc.) and the conceptual operation taking place between the words of the claim and the words of the text (reformulation, summarization, definition, etc.). The same description is made for the evaluative expression with, in addition, the orientation and strength of the evaluation.

The <discourse> subframe describes elements such as elaborations, illustrations, comparisons, conditions or circumstances that are not directly argumentative but can be seen as being part of the argument. Finally, the annotator is invited to specify the kind of argument scheme(s) that has been used, from a standard list of arguments [6] [7].

III. TOWARD A LINGUISTIC CATEGORIZATION OF RELATEDNESS

The 'link' and 'concept_op' features are specifically designed to allow for the linguistic categorization of relatedness. To describe the linguistic and conceptual links with the claim,

the annotators can use predefined categories or natural language, until a stable list of categories can emerge through the collective observation and analysis of the corpus. Then, a categorization of the main linguistic operations can be carried out, and the associated resources can be developed and structured from existing resources. The aim of this categorization is to characterize the linguistic operations behind relatedness and to evaluate the efficiency and scope of a linguistic approach, i.e. how much of relatedness analysis can be resolved via linguistic processes.

The parameters which are under investigation, categorization and evaluation are as follows:

- the paradigmatic lexico-semantic transformations developed from the topic of the claim and its restrictions, in particular, forms of synonymy, partial reformulations (*the lower representation of women in paid work*), paraphrases, restrictions, opposites, forms of inchoativity (terms describing the result instead of the process) or vice-versa (for example: *gender pay parity* → *gender pay gap*).
- the functional transformations which are related to the nature of the topic, and may induce some domain dependent lexical data (for example: *providing a better balance of job opportunities for all*). These functional terms are derived from linguistic resources that develop the goals or purposes of entities. These may be found, for the simplest kinds, in WordNet (examples and data can be found in [13]).
- the local syntactic transformations on the claim topic, (*the gap in salary between genders*),
- forms of discourse transformations such as: summarization (when the topic is long), illustration or instantiation, expression of consequence,
- the lexical data which is necessary, its structure according to lexical semantics principles [8], and its availability. A number of resources are already present in our <TextCoop> platform that realizes discourse analysis in English and French with high accuracy (about 90% accuracy in the case of the domains considered here). The version 5.1 of this platform is available at [14] while system foundations and examples can be found in [15].

ACKNOWLEDGMENTS

This research project is partly funded by a grant from the Maison des Sciences de l'Humain et de la Société, Université de Toulouse, France.

REFERENCES

- [1] R. Mochales Palau and M.F. Moens, "Argumentation mining: the detection, classification and structure of arguments in text." In Proceedings of the 12th International Conference on Artificial Intelligence and Law, pp. 98-107, 2009.
- [2] A. Peldszus and M. Stede, "From argument diagrams to argumentation mining in texts: a survey." In International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), vol. 7, no 1, pp. 1-31, 2013.

- [3] R. Swanson, B. Ecker and M. Walker, "Argument mining: extracting arguments from online dialogue," In Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 217-226, 2015.
- [4] P. Saint-Dizier, "Argument Mining: the bottleneck of knowledge and lexical resources," In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), pp. 977-983, 2016.
- [5] J. Pustejovsky, *The Generative Lexicon*, MIT Press, 1995.
- [6] D. Walton, C. Reed, and F. Macagno, *Argumentation Schemes*, Cambridge University Press, 2008.
- [7] V. W. Feng and G. Hirst, "Classifying arguments by scheme." In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 987-996, 2011.
- [8] A. Cruse, *Lexical Semantics*, Cambridge University Press, 1986.
- [9] A. Fiedler and H. Horacek, "Argumentation within deductive reasoning." *International Journal of Intelligent Systems*, 22(1):49-70, 2007.
- [10] H. Nguyen and D. Litman, "Extracting argument and domain words for identifying argument components in texts." In Proceedings of the 2nd Workshop on Argumentation Mining, pp. 22-28, 2015.
- [11] M. P. Villalba and P. Saint-Dizier, "Some facets of argument mining for opinion analysis," *COMMA*, vol. 245, pp. 23-34, 2012.
- [12] M. Walker, P. Anand, J.E. Fox Tree, R. Abbott and J. King, "A corpus for research on deliberation and debate," In LREC, pp. 812-817, 2012.
- [13] WordNet presentation and database access, last version available at: <https://wordnet.princeton.edu/>, Princeton University, 2018.
- [14] TextCoop archive, V5.1, freeware, available at: <https://www.irit.fr/~tilde'Patrick.Saint-Dizier/#projets..> CNRS, IRIT, Toulouse, 2012.
- [15] P. Saint-Dizier, *Challenges of Discourse Processing: the case of technical documents*, Cambridge Scholars Publishing, 2014.