FLA

Foundations of the Language of Argumentation

COMMA 2016 Workshop
University of Potsdam, September 13, 2016

Proceedings

Edited by:
Patrick Saint-Dizier (IRIT/Univ. Paul Sabatier, Toulouse)
Manfred Stede (University of Potsdam)

Preface

Argumentation mining, the automatic search for arguments in text, 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. 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, 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. For successful, robust application, the features underlying the classifiers tend to be relatively close to the linguistic surface, so that established tools can produce them. This work is now already in full swing for argumentation mining, too.

Our workshop looks at a different side of the problem. We are convinced 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 a lot of 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 the overall task - but have a common interest in the linguistic underpinnings.

Patrick Saint-Dizier
Manfred Stede

Program Committee

Miriam Butt (Univ. Konstanz)
Katarzyna Budzynska (Polish Academy of Sciences and Univ. of Dundee)
Mathilde Janier (Univ. of Dundee)
Ralf Klabunde (Ruhr-Univ. Bochum)
Beishui Liao (Zhejiang University)
Bernardo Magnini (FBK Trento)
Chris Reed (Univ. of Dundee)
Andrea Rocci (USI Lugano)
Patrick Saint-Dizier (CNRS-IRIT Toulouse)
Manfred Stede (Univ. of Potsdam)
Maite Taboada (Simon Fraser University)
Serena Villata (INRIA Sophia Antipolis)

Contents

Clause types and modality in argumentative microtexts

Maria BECKER, Alexis PALMER, Anette FRANK
Leibniz ScienceCampus "Empirical Linguistics and Comp. Language Modeling"
Department of Computational Linguistics, Heidelberg University
Institute of German Language, Mannheim

Abstract. This work investigates the role of semantic clause types and modality in argumentative texts. We annotate argumentative microtexts with Situation Entity (SE) classes and additionately label the segments that contain modal verbs with their modal senses. We analyse both the correlation of SE classes and of modal verbs and senses with components of argument structures (such as premises and conclusions) and their functions (such as support and rebuttal). We find interesting tendencies in the correlations between both argument components and argumentative functions with SE types. We also see interesting distribution differences of modal verbs and senses within argumentative components and functions, as well as evidence that modal senses can be helpful to distinguish conclusions and premises. We conclude that both semantic clause types as well as modal senses can be deployed for automatic recognition and fine-grained classification of argumentative text passages.
NOTE: some of the material presented in this paper overlaps with [1].

Keywords. [semantic clause types, modality, argumentation mining, modal verbs, modal senses, argumentative microtexts]

1. Introduction

The aim of this study is to better understand linguistic features of argument components in argumentative texts. We take an analytical approach, investigating the roles of both modality and semantic clause types in an annotated corpus of argumentative microtexts [2]. The argumentative microtexts are short (usually 5 sentences) written texts elicited in response to a trigger question, such as (1).[1]

(1) Sollte es Supermärkten und Einkaufszentren erlaubt werden an beliebigen Sonn- und Feiertagen zu öffnen?
Should shopping malls generally be allowed to open on holidays and Sundays?

One complete microtext appears in Table 1. The texts were written in German and translated to English; our analysis addresses the original German texts.

The microtexts are dense argumentative texts; each segment contributes to the argument. Each text contains one segment stating the **conclusion**, or the

[1] 23 of the 112 texts were written as teaching tools, thus not associated with trigger questions.

Arg	German/English (+SE-Labels & Modal Senses)
Concl.	Also, ich als Arbeitnehmer fände es sehr praktisch, (S) \\ zumindest an Wochenenden einkaufen zu **können** [**dynamic**]. (GEN) *Well, I as an employee find it very practical to be able to shop at least on weekends.*
Prem. Opp:Reb	Klar **müssen** [**deontic**] dann auch Leute am Wochenende in den Läden arbeiten, (GEN) *Sure, other people have to work in the shops on the weekend,*
Prem. Pro:Und	aber die haben dann eben innerhalb der Woche frei (GEN) \\ und **können** [**dynamic**] in Ruhe Termine wahrnehmen, (S) \\ während ich im Büro sitze. (S) *but they can have days off during the week \\ and run errands at their leisure \\ while I'm stuck in the office.*
Prem. Pro:Supp	Ausserdem **will** [**deontic**] doch der Staat, (S) \\ dass ich mein Geld ausgebe, (EV) *Plus, the state wants me \\ to spend my money,*
Prem. Pro:Add	und wie **soll** [**deontic**] ich dass machen, (Q) \\ wenn die Geschäfte nicht offen sind, (GEN) \\ wenn ich frei habe? (S) *and how am I supposed to do that \\ when the shops aren't open \\ when I'm off work?*

Table 1. Sample microtext (`micro_b015`), German and English versions, with SE labels (EV=Event, S=State, GEN=Generic Sentence, GS=Generalizing Sentence, Q=Question), proponent/opponent status, and argumentative functions (support, undercut, rebuttal, addition). Modal verbs and their senses are marked in boldface; \\ indicates clause segment boundaries.

main claim of the **proponent**. The other segments in the text are **premises** either supporting the conclusion or presenting the view of a potential **opponent**. Premises are related to the conclusion or to other premises via the argumentative functions of (a) **support**; (b) **rebuttal**, in which a premise directly challenges a conclusion or premise; or (c) **undercut**, in which a premise challenges the acceptability of an inference between two other segments (premises or premise and conclusion).

In previous work [1], we have annotated the 112 microtexts with semantic clause types in the form of situation entities.[2] We analyzed the correlations between clause types and various argument components; a portion of that analysis is presented here. In this paper we extend the analysis to investigate the role of modal verbs (Section 3) in argumentative texts, asking the following questions:

- Do particular **clause types, modal verbs or senses** correlate with specific argument components (i.e. conclusion, proponent or opponent premise)?
- Do particular **clause types, modal verbs or senses** correlate with specific argumentative functions (i.e. support, rebuttal, and undercut)?

2. Clause types: background and analysis

To investigate the semantic types of clauses found in **argumentative text passages**, we use the inventory developed by [3] and extended in later work [4,5]. **Situation entity (SE) types** describe the semantic types of situations (including states, events, generics, habituals, questions, and imperatives) evoked in discourse by individual clauses of text. SE types are recognizable (and annotatable) through a combination of linguistic features of the clause and its main verb, and it was found

[2]See Section 2, and for details of the dataset and the annotation process please see [1].

[6,7,1] that the distribution of SE types in text passages correlates to some extent with whether the text passage is (e.g.) narrative, informative, or argumentative. The use of linguistic features for distinguishing text passages is closely related to Argumentative Zoning [8,9], where scientific texts are segmented into types such as Methods or Results. Notions related to SE type have been widely studied in theoretical linguistics [10,11,12,13,14,15] and have seen growing interest in computational linguistics [16,17,18,19,20,21,22].

Situation entity types. From [1], each segment of each microtext is annotated with a SE type label.[3] Only four of the ten SE types described in [5] occur often enough in the microtexts corpus to allow for meaningful analysis.[4]

The inventory of SE types starts with **states** and **events**.

1. State (S): *Armin has brown eyes.*
2. Event (EV): *Bonnie ate three tacos.*

Modality, negation, future tense, and conditionals, when coupled with an Event-type clause, cause a coercion to State. In these constructions reference is made to actual or potential states of the world rather than actual events. Several examples appear below; note that no coercion happens for State clauses.

- EV → S: *Carlo should get the job.* cf. S → S: *Georg could have blue eyes.*
- EV → S: *Darya did not answer.* cf. S → S: *Reza is not short.*
- EV → S: *If he wins the race, ...* cf. S → S: *If it is warm tomorrow, ...*

The other two frequently-occurring SE categories in argumentative texts are **generic sentences** and **generalizing sentences**, the latter sometimes referred to as habituals. While the former predicate over classes or kinds, the latter describe regularly-occurring events, such as habits of individuals.

3. Generic Sentence (GEN): *Birds can fly. / Scientists make arguments.*
4. Generalizing Sentence (GS): *Fei travels to India every year.*

Three of these types can be used to express general knowledge about the world, to varying extents. GENs express the most general knowledge (attributes of classes), States introduce facts held to be generally true, and GSs talk about "what usually happens." Events, on the other hand, typically express episodic information.

Correlations between SE types and argument components. Table 2 shows the distributions over the four SE types for the various types of argument components in the microtext corpus. Generic Sentences are the most frequent type overall, distinguishing argumentative texts from other genres, in which Generic Sentences occur quite infrequently [5]. In the microtexts, conclusions are almost exclusively Generic Sentence or State, while premises, next to a majority of Generic Sentence, include a higher proportion of Generalizing Sentences and Events (cf. Table 2 and Figure 1(a)). This is in line with the role of the conclusion in the argument – it should concisely state the supported position,

[3] As seen in Table 1, an argumentative component can in fact consist of several SE segments. The SE clause segments have been pre-annotated in [1].

[4] Within all 668 annotated segments, six types occurred not at all or infrequently: Question (10 segments), Proposition (2), Resemblance (2), Fact (0), Report (0), and Imperative (0).

	All segments		Premises only		Functions		
	Concl	**Prem-All**	**Prem-Pro**	**Prem-Opp**	**Supp**	**Rebut**	**Undercut**
#	112	464	339	125	263	108	63
GEN	.48 (.48)	.49 (.49)	.52 (.52)	.42 (.42)	.51 (.51)	.42 (.42)	.56 (.56)
GS	.05 (.05)	.12 (.12)	.08 (.08)	.13 (.13)	.12 (.12)	.13 (.13)	.11 (.11)
S	.44 (.17)	.29 (.17)	.29 (.18)	.31 (.18)	.29 (.19)	.32 (.17)	.24 (.08)
EV	.03 (.30)	.10 (.22)	.11 (.22)	.14 (.27)	.08 (.18)	.13 (.28)	.09 (.25)

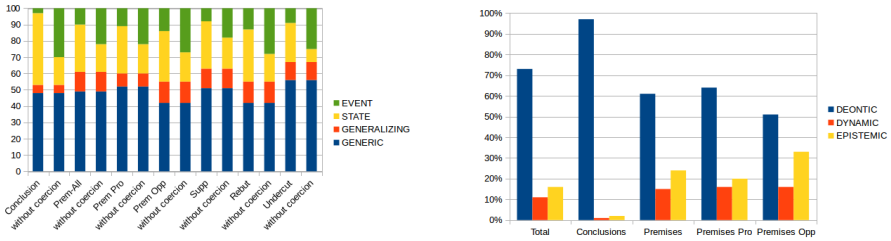
Table 2. Distributions of SE types for argument components and functions, along with absolute number of each type of segment in the microtext corpus (in brackets: subset w/o coerced cases).

which is more likely to be a statement about a class of things (e.g. *Not everyone should be obliged to pay TV licence*) or a condition or attribute of the world (e.g. *Raising the retirement age on the basis of physical fitness of the average citizen is fair*) than reference to an episodic event.

Looking at premises only, GENERIC SENTs are extremely frequent for proponent premises (52%) and less so for opponent premises (42%) (cf. Figure 1(a)).

Correlations between SE types and argumentative functions. Turning to correlations between SE types and argumentative functions of premises, we focus on the three most frequent functions: support, rebuttal, and undercut (Table 2, rightmost section). Supporting premises have a SE type distribution very similar to that of proponent premises overall. This is not surprising given that most microtexts contain only a single opponent premise. Undercutting premises show an even higher frequency of GENERIC SENTENCES (56%), with the caveat that undercuts occur less frequently than the other types. Finally, rebutting premises show a lower frequency of GENERIC SENTENCES (only 42%) and a fairly high occurrence of STATES (32%). What is notable about rebutting premises is that they show the highest proportion of (overall low-frequency) EVENTS (13%) among all argumentative functions. This points to a tendency to rebut or argue against premises or conclusions by reference to episodic events.

Analysis on the dataset without considering coercions. As mentioned above, linguistic features such as modality, negation, future tense or conditionals cause a coercion from EVENT-type to STATE. We find 84 of those coercions in our dataset, 30 within conclusions and 54 within premises. 59 of the total amount of 84 coercions are due to modal verbs. We reiterated our analyses on a dataset undoing these coercions. The distributions can be found in Table 2 (in brackets) and are visualized in Figure 1(a). Since coercion affects the proportion of STATE- and EVENT-types, we find a smaller number of STATES and a higher number of EVENTS in this analysis. As Figure 1(a) shows, the relation between STATES and EVENTS changes most for conclusions. Instead of only 3% EVENTS (when applying coercion), we find 30% EVENTS (without coercion), i.e., a difference of 27 percentage points (pp) (for premises, the difference is only 12 pp). Moreover, while in the analysis considering coercions we find a higher percentage of STATES compared to EVENTS, in the analysis ignoring coercion the reverse is true. Here we find more EVENTS than STATES among all categories (argument components as well as argumentative functions) – except in supporting premises, where we



(a) (b)

Figure 1. (a) Distribution of SE-types with and without considering coercions; (b) Modal senses within argumentative components.

find 19% States and only 18% Events. This supports the assumption that episodic events are less suitable for supporting a conclusion or a premise, whereas the high proportion of Events within rebutting and undercutting premises (28% and 25%) emphasizes the tendency to attack a conclusion or another premise by referring to episodic events – while often in modal, conditionalized or negated contexts.

3. Modal verbs: background and analysis

In our annotation we found that modal verbs constitute a high-frequency linguistic phenomenon in microtexts. The overall 188 modal verbs included in the microtexts occur in 31% of all argumentative segments (57% of conclusions and 26% of premises) and are thus at least partly responsible for the high number of States, due to the coercion of Events to States when embedded under a modal verb. In comparison, only 5% of the segments of a comparison set of texts, including different genres such as fiction, report, commentary and TED talks (see [1]), contain modal verbs. Due to this coercion, the contribution of modality indicators is orthogonal to SE type. We thus complement the analysis of SE types with a deeper investigation of modality-indicating expressions, here modal verbs.

Features of the verb such as modal auxiliaries, tense, and mood have been widely used in previous work for classifying argumentative vs. non-argumentative sentences [23,24]. Our analysis looks closely at distributions of both modal verbs and modal senses and is performed on a subset of the segments from the SE type analysis, as we are concerned only with segments containing modal verbs.

Modal verbs and modal senses. Modal verbs convey extra-propositional meaning of clauses, encoding information about possibility, obligation, necessity, permission, wishes, and requests [25]. Classifiers that automatically distinguish modal senses in context have been developed by [26] and improved in subsequent work [27,28,29]. We annotate all modal verb occurrences in the microtexts with their modal sense, following the inventory of modal senses and annotation guidelines from [27].[5] They distinguish three senses: **epistemic (ep)**, indicating possibility, **de-**

[5]Agreement between the two annotators for modal senses is high (K=0.75). Disagreements were resolved by an expert annotator, and we use the adjudicated annotations for the analysis. If both annotators chose the option "can't decide", the adjudicator in general kept this label.

verb	occ.	translation	senses	verb	occ.	translation	senses
sollen	81	shall/should	ep, de	dürfen	9	may/might	ep, de
können	45	can/could	ep, de, dy	mögen	7	may/want to	ep, de
müssen	36	must/have to	ep, de	wollen	10	want to	de

Table 3. German modal verbs with English translations and available modal senses.

ontic/bouletic (de), indicating permission/request/wish, and **dynamic (dy)**, indicating ability. Table 3 shows the German modal verbs and their available senses.[6]

Correlations between modal verbs/senses and argument components. Overall, the most frequent modal verb (MV) in the microtexts is *sollen* (81 occurrences, 41% of MVs), and the most frequent sense across all MVs is deontic (133 occurrences, 75%). Of the **conclusions** that contain a modal verb, an overwhelming 97% (62/64) are deontic; compare this to MV-containing **premises**, for which only 61% (71/117) are deontic (Fig. 1(b)).[7]

The other nearly 40% of premises are *epistemic* and *dynamic* MVs, which occur in conclusions with a combined frequency of only 3%. The prevalence of these two senses is reflected in the high frequency of *können*, which frequently occurs as dynamic or epistemic.[8] The tendencies can be seen clearly in the upper part of Table 4. A similar observation concerns **proponent and opponent** premises (Fig. 1(b)). Here, 64% (54/84) of proponent components with a MV are *deontic*, while this is true only for 53% (18/34) of opponent components with MVs. We find a relatively high proportion of *epistemic* senses (32%) for opponent premises, and a similarly high proportion for *epistemic* sense is found with **rebutting premises**.

Correlations between modal verbs/senses and argumentative functions. Epistemic modals refer to alternative, often counterfactual worlds, and as such are well suited for rebutting preceding or hypothesized arguments or premises.[9] Turning to *dynamic sense*, we find it more strongly associated with *supporting function*. We typically find this sense realized with *können*, bringing forth abilities or capabilities of entities, presented as options and chances, in support of a premise.[10] Finally, even though *deontic* is the most frequent modal sense within all of the microtexts, it is especially frequent within **undercuts** (76%, 13/17).[11] In **rebuttals** with MVs (30 occurrences) we have only 50% deontic and 39% epistemic. This

[6] *wollen* is not a modal verb, but was included because *mögen* (may/want to) is ambiguous, and thus we included the synonymous *wollen* and tagged this sense as *deontic* (wish).

[7] The prevalence of *sollen* within conclusions (86%) warrants a caveat: the trigger questions all contain some form of *sollen*, and it's possible this influences the form of the conclusions. However, the distributions of MVs within the subset of microtexts not associated with trigger questions are very similar to the overall distributions (among conclusions: sollen=77%, können=8%, müssen=15%). This suggests that the observed distributions are not overly influenced by the trigger questions.

[8] As is generally the case, *müssen* in epistemic sense is rather rare.

[9] For example: *Although IBM's numbers haven't been staggering recently, and the mood in the stock market could (können) be better on the whole, you should (sollen) buy IBM shares.*

[10] For example: *... Supermarkets and large shopping centres dominate the market today already.* **Thus smaller shops could (können) benefit from the additional freedom,** *...*

[11] As an example, we find: *But surely these costs could (müssten) be covered by a minimal increase in rent over the course of the entire lease.*

	All segments		Premises only		Functions		
	Concl	**Prem-All**	**Prem-Pro**	**Prem-Opp**	**Supp**	**Rebut**	**Undercut**
# occ.	68	120	85	35	60	30	17
deontic	.97	.61	.64	.53	.62	.50	.76
dynamic	.015	.15	.16	.15	.18	.11	.12
epistemic	.015	.24	.20	.32	.20	.39	.12
sollen	**.86**	.19	.21	.14	.18	.10	.40
können	.03	**.36**	.38	.31	.42	.30	.24
müssen	.07	**.25**	.22	.31	.25	.33	.24
dürfen	.02	.07	.08	.04	.06	.03	.06
mögen	-	.06	.03	.14	.02	.17	.06
wollen	.02	.07	.08	.06	.07	.07	-

Table 4. Distributions of modal verbs and modal senses for various argument components and functions, plus absolute number of modal-verb-containing segments in the microtext corpus.

echoes the finding in the SE analysis that rebuttals pattern somewhat differently from the other argumentative functions.

4. Conclusion

This study extends our understanding of linguistic features of argumentative texts by investigating SE types and modal verb senses for modeling argumentative text regions. Although the corpus is small and the observed phenomena can be interpreted only as tendencies, our analyses show that both semantic clause types and modal senses can be useful features for automated argumentation analysis – not as stand-alone features, but as part of a larger system.

Our analyses revealed some clear-cut distinctions as well as weaker tendencies. First, SE types, especially Generic Sentences, allow for a clear distinction between argumentative and non-argumentative text passages.[12] The high frequency of Generic Sentences and the rareness of Events within conclusions compared to premises can be helpful to identify conclusions and premises within argumentative texts. Proponent premises were found to be characterized by a high proportion of Generic Sentences and very few Events, whereas a higher amount of Events are found within rebutting premises when controlling for coercions. Overall, this shows that episodic events are less frequently used in conclusions or supporting premises, while they seem more appropriate in rebuttals.

Modal verbs and senses can be used to distinguish conclusions from premises since conclusions show a strong tendency to be deontic, with different lexicalizations. Proponent and opponent premises differ from each other, as the former show a higher proportion of deontic sense. Furthermore, we found rebutting premises to contain more epistemic modal verbs than supporting or undercutting premises.

[12]Similar to SE types, a high proportion of modal verbs was found to be a characteristic feature of argumentative texts as opposed to other text genres.

References

[1] Maria Becker, Alexis Palmer, and Anette Frank. Clause types and argumentative texts. In *Proceedings of the 3rd Workshop on Argument Mining*, 2016.

[2] Andreas Peldszus and Manfred Stede. An annotated corpus of argumentative microtexts. In *Proceedings of the First European Conference on Argumentation: Argumentation and Reasoned Action*, 2015.

[3] Carlota S Smith. *Modes of discourse: The local structure of texts*, volume 103. Cambridge University Press, 2003.

[4] Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. A sequencing model for situation entity classification. In *Proceedings of ACL*, 2007.

[5] Annemarie Friedrich and Alexis Palmer. Situation entity annotation. In *Proceedings of the Linguistic Annotation Workshop VIII*, 2014.

[6] Alexis Palmer and Annemarie Friedrich. Genre distinctions and discourse modes: Text types differ in their situation type distributions. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and NLP*, 2014.

[7] Kleio-Isidora Mavridou, Annemarie Friedrich, Melissa Peate Sorensen, Alexis Palmer, and Manfred Pinkal. Linking discourse modes and situation entities in a cross-linguistic corpus study. In *Proceedings of the EMNLP Workshop LSDSem 2015: Linking Models of Lexical, Sentential and Discourse-level Semantics*, 2015.

[8] Simone Teufel. *Argumentative zoning: Information extraction from scientific text*. PhD thesis, University of Edinburgh, 2000.

[9] Diarmuid O'Seaghdha and Simone Teufel. Unsupervised learning of rhetorical structure with un-topic models. In *Proceedings of COLING*, 2014.

[10] Zeno Vendler. *Linguistics in Philosophy*, chapter Verbs and Times, pages 97–121. Cornell University Press, Ithaca, New York, 1957.

[11] H.J. Verkuyl. *On the Compositional Nature of the Aspects*. Reidel, 1972.

[12] David Dowty. *Word Meaning and Montague Grammar*. Reidel, 1979.

[13] Carlota S. Smith. *The Parameter of Aspect*. Kluwer, 1991.

[14] Nicholas Asher. *Reference to Abstract objects in Discourse*. Kluwer Academic Publishers, 1993.

[15] Gregory N. Carlson and Francis Jeffry Pelletier, editors. *The Generic Book*. University of Chicago Press, 1995.

[16] Eric V. Siegel and Kathleen R. McKeown. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–628, 2000.

[17] Alessandra Zarcone and Alessandro Lenci. Computational models of event type classification in context. In *Proceedings of LREC2008*, 2008.

[18] Aurelie Herbelot and Ann Copestake. Annotating genericity: How do humans decide? (A case study in ontology extraction). *Studies in Generative Grammar 101*, page 103, 2009.

[19] Nils Reiter and Anette Frank. Identifying Generic Noun Phrases. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 40–49, Uppsala, Sweden, July 2010. Association for Computational Linguistics.

[20] Francisco Costa and António Branco. Aspectual type and temporal relation classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 266–275. Association for Computational Linguistics, 2012.

[21] Anna Nedoluzhko. Generic noun phrases and annotation of coreference and bridging relations in the Prague Dependency Treebank. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 103–111, 2013.

[22] Annemarie Friedrich and Manfred Pinkal. Discourse-sensitive Automatic Identification of Generic Expressions. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*, Beijing, China, July 2015.

[23] Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau, and Chris Reed. Automatic detection of arguments in legal texts. In *Proceedings of ICAIL 2007*, 2007.

[24] Eirini Florou, Stasinos Konstantopoulos, Antonis Kukurikos, and Pythagoras Karampiperis. Argument extraction for supporting public policy formulation. In

Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, 2013.

[25] Angelika Kratzer. The Notional Category of Modality. In H. J. Eikmeyer and H. Rieser, editors, *Words, worlds, and contexts: New approaches in word semantics*, pages 38–74. de Gruyter, Berlin, 1981.

[26] Josef Ruppenhofer and Ines Rehbein. Yes we can!? Annotating the senses of English modal verbs. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 1538–1545, 2012.

[27] Mengfei Zhou, Anette Frank, Annemarie Friedrich, and Alexis Palmer. Semantically Enriched Models for Modal Sense Classification. In *Proceedings of the EMNLP 2015 Workshop LSDSem: Linking Models of Lexical, Sentential and Discourse-level Semantics*, Lisbon, Portugal, 2015.

[28] Ana Marasović, Mengfei Zhou, Alexis Palmer, and Anette Frank. Modal Sense Classification At Large: Paraphrase-Driven Sense Projection, Semantically Enriched Classification Models and Cross-Genre Evaluations. *Linguistic Issues in Language Technology, Special issue on "Modality in Natural Language Understanding"*, 2016. to appear.

[29] Ana Marasović and Anette Frank. Multilingual Modal Sense Classification using a Convolutional Neural Network. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, Berlin, Germany, August 2016.

On the role of discourse particles for mining arguments in German dialogs

Annette HAUTLI-JANISZ [1], Miriam BUTT
Department of Linguistics, University of Konstanz, Germany

Abstract. Argument mining in dialogs or multilogs neccessarily must take into account the pragmatic relations that hold between dialog participants, their arguments and the ongoing discourse. This paper analyzes the role of German discourse particles and the illocutionary force contributed by the particles. We investigate a set of highly frequent discourse particles in German and propose a categorization that complements those levels of analysis that are pursued in opinion mining and dialog act annotation. Incorporating the subtle pragmatic information encoded by the discourse particles into Argument Mining offers a new way of pragmatically underpinning the propositional content of arguments in German dialog data.

Keywords. discourse particles, argument mining, dialogical data

1. Introduction

Argument Mining of dialogical data involves the analysis of different levels of information: On the one hand there is the propositional, i.e. the at-issue content of the discourse; on the other hand, there is a level of pragmatic information which enriches the propositional content. This pragmatic level provides information on how participants steer the discourse, how they relate to it and why some information is important to the other participants at a particular stage. This information is highly relevant to interpreting arguments in dialogical data and becomes crucial when analyzing the exchange of arguments and the kinds of conversational moves that trigger succeeding moves.

In German natural speech, pragmatic information is conveyed, among other means, by discourse or modal particles, a linguistic category that is frequently used in spoken language, but is not confined to it. In this paper we show that discourse particles in German are so frequent in natural speech arguments that they are indeed worth exploring in more detail. We then put forward a first proposal to categorize the pragmatic information that a set of highly frequent German discourse particles contributes to the propositional content of the dialog. We further discuss how this information can enrich the interpretation of argumentative structures in this type of data.

The paper proceeds as follows: We first lay out the necessary background concerning the linguistic information on discourse particles and argumentation mining in dialogical

[1]The research carried out in this paper was supported by the Bundesministerium für Bildung und Forschung (BMBF) under grant no. 01461246 (eHumanities VisArgue project) and the Equal Opportunity Office at the University of Konstanz.

data (Section 2) and then provide quantitative evidence as to the relevance of discourse particles in German natural speech arguments (Section 3). We then go on to provide an initial classification of pragmatic information that is contributed by German discourse particles (Section 4). Section 5 discusses how this information can enrich argumentation mining in dialogical data and concludes the paper.

2. Discourse particles and their pragmatic contribution

In linguistic theory, German modal or discourse particles have been a long-standing issue [1,2,3,4,5]. Overall, they are considered to not contribute to the propositional, i.e. at-issue, content of an utterance, but to its expressive content ([6], [7], [8], among many others). Due to their elusive pragmatic nature, there is no generally agreed-upon treatment of discourse particles. Analyses range from considering them as contributing conventional implicatures [9], adding felicity conditions [6], being modifiers of illocutionary operators [3,10] or being a modifier of sentence types [8]. Despite the breadth of analyses, discourse particles are generally considered as conveying a speaker's stance towards an utterance and situating the utterance in the web of information that comprises the discourse, including the Common Ground held by the discourse participants [11,12]. It is this property that we claim can be capitalized on for mining arguments in dialogical data.

Discourse particles are found across languages, as shown in (1) for English. Here, the particle *like* does not change the descriptive content of the utterance, i.e. it does not change the fact that the process took twenty minutes. Instead, *like* indicates that the speaker is not quite sure of how to say what they mean [13]. In contrast, (2) illustrates the verbal usage of *like* where it contributes to the proposition of the utterance.

(1) They're, *like*, representatives of their clan.

(2) I *like* chocolate very much.

For German, consider (3), where the conclusion phrase ('the tunnels are crucially important') precedes the premise phrase ('because they drastically reduce travel times'). Without the particles triggering the conventional implicatures in the premise, the argument itself does not contain information on how it relates to the previous discourse. However, taking this information into account enriches the analysis of the argument: The discourse particle *ja* 'yes' either establishes or reconfirms the reason for tunnels as part of the common ground between the discussion partners. In contrast, *halt* 'stop' expresses a (resigned) acceptance of an immutable constraint, i.e. the speaker accepts the reason that travel times are reduced, although he or she might not like the fact.

(3) Die Tunnel sind existentiell wichtig,
the tunnels are existential important
weil sie (∅ / *ja* / *halt*) die Fahrzeiten drastisch verkürzen.
because they (∅ / *yes* / *stop*) the travel times drastically shorten

'The tunnels are crucially important, because they drastically reduce travel times, (∅/ *as you know/ regrettably, but indisputably so*).'

For some particles, multiple interpretations exist depending on the intonation. In (4), the stressed *doch* 'indeed' signals the rejection of the common ground, i.e. the speaker rejects an opinion which has been uttered in the previous discourse that implies that the tunnels do not reduce travel times. The unstressed *doch* 'indeed' reminds the hearer of a fact that was uttered in the previous discourse, thereby activating the common ground between the discussion partners.

(4) Die Tunnel sind existentiell wichtig,
the tunnels are existential important
weil sie (*DOCH* / *doch*) die Fahrzeiten drastisch verkürzen.
because they (*indeed* / *indeed*) the travel times drastically shorten

'The tunnels are crucially important, because they drastically reduce travel times, (*in contrast to what was said before/if I may remind you*).'

The examples illustrate that discourse particles make subtle pragmatic contributions that structure the discourse, relate participants to it or help the hearer to understand why some information is important at a particular stage. This information is highly relevant for argument mining in dialogical data and while a number of previous approaches take into account pragmatic information [14,15,16,17,18,19,20], these particles offer additional information by signaling propositional boundaries and inter-propositional relationships: two key aspects of argumentation mining. By incorporating them in argument analysis, we cannot only tie arguments to the context of a dialog, we can also more closely analyze the relation between speakers and the discourse.

In the following we briefly show that discourse particles in German are highly frequent in natural language arguments and are therefore worth exploiting for argumentation mining purposes.

3. Quantitative investigation

For the investigation we employ three different corpora: We first use the transcripts of Stuttgart 21 (S21), a public arbitration in the German city of Stuttgart, where a new railway and urban development plan caused a massive public conflict in 2010 (10.000 utterances, $\sim$ 500.000 tokens).[2] We further use the transcripts of experimentaly controlled discussions on whether or not to allow fracking in Germany (2.000 utterances, $\sim$ 282.000 tokens) and on establishing a hypothetical African government (3670 utterances, $\sim$ 363.000 tokens).[3]

For the investigation we take the 20 most frequent particles (and their combinations) in the three corpora (among them *ja* 'yes', *halt* 'halt', *doch* 'indeed', *eben* 'even', *wohl* 'probably') and calculate their relative frequency in discourse relations like premises, conclusions, contrasts and concessions that are part of argumentative structures and are

[2]Until October 2014 the transcripts were publicly available for download at http://stuttgart21.wikiwam.-de/Schlichtungsprotokolle. A new, edited version of the minutes can be found here: http://www.schlichtung-s21.de/dokumente.html

[3]These experimentally controlled discussions were done as part of our eHumanities VisArgue project, which represented a collaboration between computer science, linguistics and political science at the University of Konstanz. The experiments were conducted by our political science partners.

marked by explicit discourse connectives. These discourse relations span sentences but are not related to each other and therefore do not form chains of arguments.

To extract the information from the dialogs, we use the VisArgue pipeline [14,21], a linguistically motivated, high-quality parsing pipeline that automatically identifies, disambiguates and annotates discourse information in transcribed natural speech, among them the spans triggered by discourse connectives. The number of discourse relations that contain one or more particles are shown in Table 1.[4]

	Premise	**Conclusion**	**Contrast**	**Concession**	**Condition**
Stuttgart21	0.28	0.32	0.20	0.08	0.23
Fracking	0.39	0.46	0.30	0.10	0.34
Africa	0.40	0.43	0.23	0.15	0.29

Table 1. Relative frequencies of explicit argument relations containing discourse particles

Overall, Table 1 shows that particles are frequently used across corpora and across those discourse relations that signal argumentative structures: Conclusion relations rate highest across corpora, with particles contained in 32% to 46% of cases, followed by premise relations where particles occur in 28 to 40% of cases. Contrast and condition relations range in the middle, concessions contain the least number of particles (8% to 15% of cases).

The results indicate that particles are worth taking into account when mining arguments in German dialogs: They are frequently employed in spoken natural language arguments and carry pragmatic information that conveys important information on how arguments and speakers relate to the discourse. In the following we offer a first categorization of the pragmatics contributed by individual discourse particles and their combinations.

4. Categorization

While the long-term goal is to create a language-independent scheme of pragmatic facets that are relevant in relating arguments to natural language discourse, this paper puts a particular focus on the contribution of highly frequent discourse particles in German. One challenge is to make the chosen categories general enough to allow for a meaningful comparison, but also fine-grained enough to allow for detailed interpretation. The second challenge resides in the nature of the discourse particles: The pragmatic information they carry is subtle and therefore hard to pin down. Moreover, the majority of them can be used as "regular" lexical items. The particles *halt* 'stop', *ja* 'yes' and *doch* 'indeed', for instance are also used as verbs and/or adverbs. In the following we briefly discuss each of the dimensions of pragmatic information, summarized in Table 2.

4.1. Dimensions

Our initial classification proposes three major dimensions of pragmatic information that are triggered by 20 of the most frequent discourse particles in the three German corpora discussed in Section 3. Table 2 provides an overview of the classification.

[4]The total number of discourse relations ranges from 107 concession relations in the the Fracking corpus to 3075 conclusion relations in the S21 corpus (across corpora, mean: 1746, median: 1798).

Dimension	Subdimesion	Example
Common ground	Refer to common ground (CG)	*ja* 'yes'
	Activate common ground (ACG)	*doch mal* 'lit. indeed sometime'
	Reject common ground (RCG)	*doch wohl* 'lit. indeed probably'
Constraint	Immutable constraint (IC)	*halt* 'stop'
	External constraint (EC)	*mal* 'sometime'
Consensus	Consensus (C)	*ja* 'yes'
	Minimal consensus (MC)	*immerhin* 'at least'
	Consensus-willing (CW)	*nicht wahr* 'lit. not true'= 'right'

Table 2. Categorization of pragmatic information

Common Ground The first dimension and one in which particles have lately been analyzed as triggering conventional implicatures (CIs) is concerned with the Common Ground, i.e. the knowledge shared between discussion partners. With respect to the discourse particle *ja* 'yes', [22,23,8,24,25, inter alia] show that it is used to *refer to the Common Ground*, i.e. the speaker wants the hearer to know that the utterance refers to content that was previously under discussion. Other discourse particles, for instance *doch* 'indeed', trigger a CI that indicates an *activation of the Common Ground*. Here, the speaker deliberately puts information into the Common Ground of the hearers and expects it to be present in the continuing dialog, thereby extending the knowledge shared in the discussion. Thirdly, the *Common Ground can be rejected*, i.e. information that is shared between the hearers is objected to by the speaker. This CI is for instance triggered by the particle combination *doch wohl* 'lit. indeed probably', illustrated in (5).[5]

(5) ... weil das ist *doch wohl* der Konflikt, um den es hier geht.
... because this is *indeed probably* the conflict about the it here go
'... because *(in contrast to what you are claiming)* this is the conflict that we're dealing with here.'

Constraint This dimension subsumes pragmatic information that either conveys an immutable or an external constraint. Concerning the latter, for instance triggered by *mal* 'sometime', the speaker signals that they are subjected to an external constraint, either imposed by a hearer or by a fact under discussion. With respect to the former, particles such as *eben* 'even' and *halt* 'stop' imply that the speaker resignedly accepts a fact that is immutable, be it out of their own accord or imposed from outside, shown in (6).

(6) Wir machen das, weil das *eben* unsere Arbeit ist.
we make this because this *even* our job is
'We do this, because it's our job (*that is how it is*).'

Consensus While full consensus is generally signaled by adverbials such as *ja* or *ja-wohl* 'yes', the contribution of some particles can best be described as signaling minimal consensus. Using for instance *immerhin* or *zumindest* 'at least', which also have concessive meaning, the speaker signals that they partly accept information that was previously under discussion, but are still concerned about some aspects of it. With respect to

[5]The particle *doch* 'indeed' by itself is actually ambiguous between rejection and activation of Common Ground, see [25] for more discussion.

consensus-willing, the multiword particle *nicht wahr* ‘lit. not true’ was used extensively in one of the corpora as a tag question to signal the overall aim of the speaker to reach a consensus, calling on the hearers to follow suit. An example is shown in (7).

(7) Bei dem ICE geht es wahrscheinlich schneller, *nicht wahr*?
with the ICE go it probably faster *not true*
It’s probably faster with the ICE, *(don’t you agree?)*

This classification of the individual particles and the associated pragmatic information is a necessary first step in operationalizing their content for argumentation mining. The following evaluation shows that out classification is indeed valid and that the discourse particles can be reliably annotated with the proposed dimensions.

4.2. Evaluation

In order to evaluate our classification scheme, we conducted an experiment on the inter-annotator agreement, using five particles that are highly frequent and that belong to one of the chosen categories. For the dimension ‘Common Ground’, we select *ja* ‘yes’, for ‘Constraint’ we select *halt* ‘stop’ and for ‘Consensus’ we use *immerhin* ‘at least’. We also include *doch* ‘yes’ as its pragmatic contribution varies with intonation and is expected to create disambiguation problems for the annotators, as well as *mal* ‘sometime’ whose particle meaning can be hard to differentiate from its literal usages depending on the context. The task for the annotators is to differentiate the particle meaning from the literal meaning, and, in the case of *ja* ‘yes’ and *doch* ‘indeed’, pick the correct pragmatic contribution.

The annotation data consists of a total of 100 sentences which are randomly chosen from the S21 corpus, each containing one of the five particles. The task for each of the four annotators (undergraduates of linguistics) was to assign one dimension to each of the sentences. Across all particles and all annotators, kappa is $\kappa = 0.66$ (“substantial agreement”). Filtering out the least accurate annotator increases κ to $\kappa = 0.85$ (“almost perfect agreement”). A closer investigation of the results shows that *doch* ‘indeed’ is the most problematic particle due to the effects of the intonation on the pragmatic contribution. Here, the agreement between annotators is significantly lower. In all other cases, the annotators were able to assign the right annotations to the data, showing that the proposed classification scheme is systematic and well-defined.

5. Discussion and conclusion

This paper presents an initial survey of German discourse particles and their pragmatic contribution, showing that this linguistic category offers potential for interpreting arguments in dialogical data. Although we see our classification scheme as an initial proposal which may yet be subject to revision, given our results so far, we claim that it is imperative to take discourse particles in account in argument mining. This particularly pertains to a key issue in argument mining, namely the detection of propositional boundaries. Using particles and the scope they have in the surrounding sentence, we believe that this support the differentiation of argumentative text units from those that are non-

argumentative. This is particularly true for the case where only a portion of a sentence coincides with an argumentative unit.

Particles also offer information on the if and how of relations between propositions and more abstract notions: For instance, they allow us to tie arguments to the Common Ground of the dialog participants, i.e. the knowledge shared between discussion partners. Using the pragmatic information contributed by the particle, we cannot only connect individual propositions (or arguments), we can also relate them to a larger set of propositions that constitute the Common Ground, as well as determine the type of the relation. For instance, if a premise contains a reference to the Common Ground, the speaker indicates that they base their conclusion on information that they believe is known among the debate participants. Conversely, if the conclusion implies a rejection of the Common Ground, the speaker signals that the inference rejects information that is known among the participants. For argument mining, this information can yield more detailed insights into how dialog settings differ in the way participants relate their arguments and themselves to the shared knowledge of the discussion – enriching the relations between arguments and the propositions that constitute them. It also opens up the possibility of devising a new set of argumentation schemes based on the notion of the Common Ground, complementing those previously proposed by [26,27] and many others.

Regarding our dimensions of constraint and consensus, the discourse particles that instantiate these dimensions rather offer information on the intention of the speaker and the stance/attitude of the speaker towards the current utterance and the arguments put forward previously. For instance if an immutable constraint particle is used in a premise, the speaker implies that the conclusion results from a constraint that cannot be avoided. Speakers can indicate that they are willing to make concessions or are seeking consensus, they can also use them to strengthen the force of their argument by invoking the idea of an immutable constraint that is governing their argumentation.

Besides discourse particles, natural language contains a variety of further strategies for structuring the discourse and for expressing pragmatic information that contributes to the overall illocutionary force [28]. A prominent example in English are tag questions. The operationalization of the pragmatic contribution of discourse particles presented here needs to be extended to other relevant natural language phenomena.

Overall, the proposed categorization of pragmatic information is a first step in analyzing the breadth of pragmatic information that structures and guides dialogical arguments. This operationalization has the potential to complement other levels of pragmatic information. Such an analysis can for instance pave the way for establishing a scheme of linguistically-motivated types of arguments, i.e. causal arguments where a conclusion is drawn based on an immutable constraint in one of the premises. By not only analyzing the pragmatic role of discourse particles, but also taking into account other rhetorical means such as tag questions and rhetorical questions, we can arrive at a much richer interpretation of the discourse and exploit information that is inherent in natural speech argumentation.

References

[1] G. v. d. Gabelentz, *Die Sprachwissenschaft, ihre Aufgaben und Methoden*. Tübingen: Narr, 1891.

[2] J. Jacobs, *Fokus und Skalen: Zur Syntax und Semantik der Gradpartikeln im Deutschen*. Tübingen: Niemeyer, 1983.

[3] J. Jacobs, "On the semantics of modal particles," in *Discourse Particles* (W. Abraham, ed.), pp. 141–162, Amsterdam: Benjamins, 1991.
[4] E. König, "Zur Bedeutung von Modalpartikeln im Deutschen: Ein Neuansatz im Rahmen der Relevanztheorie," *Germanistische Linguistik*, vol. 136, pp. 57–75, 1997.
[5] M. Coniglio, *Die Syntax der deutschen Modalpartikeln: Ihre Distribution und Lizenzierung in Haupt- und Nebensätzen*. Berlin: Akademie Verlag, 2011.
[6] A. Kratzer, "Beyond "oops" and "ouch": how descriptive and expressive meaning interact," in *Cornell Conference on Theories of Context Dependency*, vol. 26, March 1999.
[7] E. Karagjosova, *The Meaning and Function of German Modal Particles*. PhD thesis, Universität des Saarlandes, 2004.
[8] M. Zimmermann, "Discourse Particles," in *Semantics. (= Handbücher zur Sprach- und Kommunikationswissenschaft HSK 33.2)* (K. v. Heusinger, C. Maienborn, and P. Portner, eds.), vol. 2, pp. 2011–2038, Berlin: De Gruyter, 2011.
[9] M. Doherty, *Epistemische Bedeutung*. Studia grammatica 23, Berlin: Akademie Verlag, 1985.
[10] K. Lindner, "'Wir sind ja doch alte Bekannte' — the use of German *ja* and *doch* as modal particles," in *Discourse Particles: Descriptive and theoretical investigations on the logical, syntactic and pragmatic properties of discourse particles in German* (W. Abraham, ed.), vol. 12, pp. 163–201, Amsterdam: John Benjamins, 1991.
[11] C. Gunglogson, "Declarative questions," in *SALT XII*, (Ithaca, NY), pp. 124–143, Cornell University, 2002.
[12] R. Stalnaker, "Common ground," *Linguistics and Philosophy*, vol. 25, no. 5, pp. 701–721, 2002.
[13] L. Schourup, *Common Discourse Particles*. New York: Garland, 1991.
[14] T. Bögel, A. Hautli-Janisz, S. Sulger, and M. Butt, "Automatic detection of causal relations in German multilogs," in *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, pp. 20–27, April 2014.
[15] K. Budzynska and C. Reed, "When inference?," tech. rep., University of Dundee, 2011.
[16] M. A. Walker, J. E. Fox Tree, P. Anand, R. Abbott, and J. King, "A corpus for research on deliberation and debate," in *LREC*, pp. 812–817, 2012.
[17] K. Budzynska, M. Janier, C. Reed, and P. Saint-Dizier, "Towards extraction of dialogical arguments," in *Proceedings of Computational Models of Natural Argument (CMNA13)*, 2013.
[18] K. Budzynska, M. Janier, J. Kang, C. Reed, P. Saint-Dizier, M. Stede, and O. Yaskorska, "Towards argument mining from dialogue," in *Frontiers in Artificial Intelligence and Applications. Proc. of 5th International Conference on Computational Models of Argument COMMA 2014* (S. Parsons, N. Oren, C. Reed, and F. Cerutti, eds.), vol. 266, pp. 185–196, 2014.
[19] M. Janier and C. Reed, "Towards a theory of close analysis for dispute mediation discourse," *Argumentation*, vol. DOI: 10.1007/s10503-015-9386-y, 2015.
[20] K. Budzynska, M. Janier, C. Reed, and P. Saint-Dizier, "Theoretical foundations for illocutionary structure parsing," *Argument and Computation*, 2016.
[21] V. Gold, M. El-Assady, T. Bögel, C. Rohrdantz, M. Butt, K. Holzinger, and D. Keim, "Visual linguistic analysis of political discussions: Measuring deliberative quality," *Digital Scholarship in the Humanities*, vol. DOI: http://dx.doi.org/10.1093/llc/fqv033, 2015.
[22] H. Zeevat, "Particles: Presupposition triggers, contet markers or speech act markers," in *Optimality Theory and Pragmatics* (R. Blutner and H. Zeevat, eds.), Basingstoke: Palgrave Macmillan, 2004.
[23] A. Kratzer and L. Matthewson, "Anatomy of two discourse particles." Handout, SULA 5, May 2009.
[24] Y. Viesel, "Discourse structure and syntactic embedding: The German discourse particle 'ja'," in *Proceedings of the 20th Amsterdam Colloquium* (F. R. T. Brochhagen and N. Theiler, eds.), pp. 418–428, 2015.
[25] M.-M. Zymla, M. Müller, and M. Butt, "Modeling the Common Ground for Discourse Particles," in *Proceedings of LFG15, CSLI On-line publications* (M. Butt and T. H. King, eds.), 2015.
[26] M. Kienpointner, *Alltagslogik: Struktur und Funktion von Argumentationsmustern*. Stuttgart: Fromman-Holzboog, 1992.
[27] D. Walton, *Argumentation Schemes for Presumptive Reasoning*. Mahwah, N. J., Erlbaum, 1996.
[28] J. Searle and D. Vanderveken, *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press, 1985.

Identifying Problem Statements in Scientific Text

Kevin HEFFERNAN, Simone TEUFEL
University of Cambridge
Computer Laboratory, 15 JJ Thomson Avenue, Cambridge CB3 0FD
forename.surname@cl.cam.ac.uk

Abstract. In this work, we focus on the automatic identification of fine-grained problem-solution structure in scientific argumentation. We operationalise the task of finding problem formulations within scientific text in a supervised setting, using a newly-created hand-curated corpus from the domain of computational linguistics. In terms of linguistic features for their detection, we distinguish features from within the statement, and features representing the surrounding context. Results from a classification task on our corpus show that the task of identifying problem statements is tractable using a mixture of features, whereby features modelling the rhetorical context are particularly successful. Overall, our experiment shows promise for future work in identifying scientific problem-solution structure in a more global way.

Keywords. problem-solving, argumentation, argumentative zoning

1. Introduction

Argumentation is a human activity that can take many shapes and forms. Almost every aspect of our life is governed by communicative needs to persuade somebody of something. The cognitive tasks associated with this have therefore left their traces in almost any extant written or transcribed textual material we as computational linguists might choose for automatic analysis and interpretation.

However, academic study of argumentation has been heavily biased towards areas of human argumentation that are associated with professional activity rather than private ones, and in particular those areas that are seen to be more "objective", such as political speech, legal contracts, and scientific articles. Following Aristotle, there has been a tradition to consider mainly arguments that are logically truth-conditional. More recently, computational linguists' attention has turned to defeasible arguments, i.e. those that people actually use in everyday argumentation, whether they are logically sound or not. Seen this way, interpreting arguments has more to do with assessing plausibility than with formal proof.

In this paper, we will present evidence for one particular facet of argumentation in science – problem-solution structure. The view of science as a problem-solving activity is a common assumption amongst many researchers [1,2,3,4,5]. [6] identified three basic types of scientific article: the "controlled experiment", the "hypothesis testing" and the

"technique description". Each type has its own structure, but according to [7] they can be reduced, either by degradation or by amelioration, to a problem-solution structure.

In earlier work, one of us presented a theory of argumentation moves in science (Argumentative Zoning; [8]), which can be operationalised as a supervised machine learning task that assigns a small number of rhetorical labels to individual sentences. Problem-solution structure is "hard-wired" into the labels and features a way of defining the task. For instance, the fact that an author declares a particular state of the world as "problematic" in a sentence might well lead to that sentence being classified as a research gap (the declared motivation for the knowledge claim that constitutes the paper). At the same time, linguistic features that might indicate problem-status (such as lexical items, the use of the verb "need", negative-polarity adjectives or negated verbs expressing solutionhood) will be detected in such a sentence if they are expressed in an explicit enough manner; this will eventually serve to classify the sentence as "CTR" (the label associated with a research gap).

However, Argumentative Zoning (AZ) treats several aspects of scientific argumentation simultaneously and collates all these phenomena into only 7 (or 12, in follow-on work [9]) classification labels. In contrast, we are interested in detecting descriptions of problems as a separate task here. The task we are setting ourselves in the current paper also differs by its formal definition. Rather than classifying an entire sentence, we will classify shorter linguistic strings extracted from the sentence, which might or might not describe a problem.

In this paper, we present (in section 2.1) the development of a small hand-curated training and testing corpus for a binary problem classification of real-world strings from articles in computational linguistics. We use explicit cue phrases to create this corpus, but hope to be able to apply the classifier to *any* problem description in scientific articles, whether explicit or not. Section 4 will present the results of a supervised machine learning experiment to replicate this classification. We split our features into those internal to the string and those using context around the candidate string. One of the core claims in AZ was that rhetorical labels of certain statements influence each others' rhetorical status; our experiments allows us to quantify this effect, as opposed to the effect of the semantics of the potential problem description itself.

1.1. Linguistic Correlates of Problem-hood

Let us now look at what a description of a problem might look like. A priori, we would expect any description of the body of scientific knowledge or the state of the world in general which is seen as negative. We count as problems descriptions of impracticality, lack of knowledge or of a failure of an existing attempt to rectify such a situation, i.e., an unsuccessful attempt of solving a problem. In this category, we also include statements where a solution unearthed follow-on problems. We also include all task descriptions as problems, i.e., all statements of tasks the authors are setting out to do in the current paper. The phrases we consider can syntactically be noun phrases, verb phrases, propositional phrases, and any other syntactic constituents determined by our parser, as long as they pass a human quality test (cf. section 2.1).

We will now discuss possible linguistic correlates of problem-solutionhood structure. Since descriptions of problems have a strong correlation with negative sentiment, identifying the polarity status of the head of each candidate phrase should intuitively help

in resolving a candidate's problem-hood. For example, in the phrase: "a complication", the head noun here ("complication") clearly identifies this statement as problematic. The syntactic characteristics of a candidate phrase should also help in classifying their status. Since problems are often posed as questions, this observation might be captured with WH- POS tags. Additionally, descriptions of problems often have a large proportion of adjectives or adverbs to qualify their badness (e.g. "the negatively skewed distribution"). Tense, negation and modality also play a role in determining sentiment. Making use of tense is an important aspect to consider when modelling an author's viewpoint. For example, previous work will be cited because it motivated something in the paper. However, it may be cited for use as a method (praise) or as a motivation (dismissal). Negation has been a popular technique shown to improve sentiment classification [10] where the intuition is that any word following a negation (e.g. "not") should be given a negative weight. Modality can also identify the mood of a statement [11] or hedging [12] and so we also took this into consideration. Since many words in our statements may not have a known polarity status, instead of using a null value, the semantic similarity of nouns or verbs in the candidate phrase to those with a known polarity status should help increase our success. Lastly, knowledge of the rhetorical context surrounding a candidate phrase should aid in determining its problem-hood.

We will model each of these linguistic features in section 3. We will now explain our experimental setup (how the corpus was created, and how the experiment was designed).

2. Method

2.1. Corpus

Our new corpus is a subset of the latest version of the ACL anthology released in March, 2016[1] which contains 22,878 articles in the form of PDFs and OCRed text. The 2016 version was also parsed using ParsCit [13]. ParsCit recognises not only document structure, but also bibliography lists as well as references within running text. A random subset of 2,500 papers was collected covering the entire ACL timeline. In order to disregard non-article publications such as introductions to conference proceedings or letters to the editor, only documents containing abstracts were considered. We preprocessed the corpus using tokenisation, sentence splitting and syntactic parsing with the Stanford Parser [14].

In order to define an indisputable ground truth for problem strings, we use textual templates such as "problem is X". These were executed using tregex and tsurgeon [15], a set of tools for structural search in trees and tree manipulation. An example of one of our templates is shown in Figure 1. To increase our recall of different-worded problem descriptions, we additionally use target words which are semantically close to the noun "problem". Semantic similarity was defined by training a deep learning distributional model using Word2Vec [16] on 18,753,472 sentences from a biomedical corpus based on all full-text Pubmed articles [17]. From the 200 semantically closest words to "problem" (decided by cosine similarity with our Word2Vec model), we manually selected 28 clear and unambiguous synonyms for use in the templates. Of the sentences matching the templates, 600 were randomly selected, and the syntactic phrase corresponding to X was

[1] http://acl-arc.comp.nus.edu.sg/

excised from the sentence. Both the template match and the problem phrase X itself were then plausibility-checked by two annotators without communication between them (the two authors of this paper).

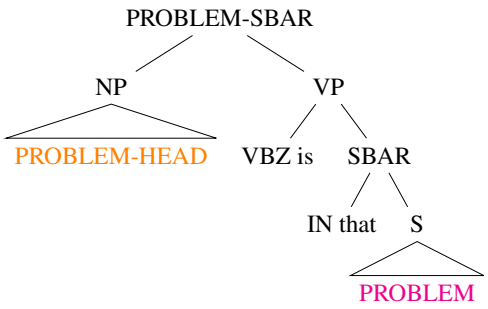


Figure 1. Template for PROBLEM-SBAR. PROBLEM-HEAD indicates the head noun of the NP must be one of our chosen problem words. Example: "The problem is that we do not achieve a significant result."

We also wanted to find similarly shaped negative examples, i.e., guaranteed non-problem strings. We sampled a population of phrases to mimic our 600 problem strings as closely as possible while making sure they really are negative examples. We started from sentences *not* containing any problem words (i.e. those used in problem templates). From each, we at random selected one syntactic subtree; from those we selected 600 that satisfy the following conditions: first, the distribution of the head POS tags of the non-problem strings perfectly matches the head POS tags[2] of the problem strings perfectly. Secondly, the distribution of the lengths of the non-problem strings must not be significantly different to that of the problem strings, using the Pearson's chi-squared goodness of fit test [19].

A human quality-test was then performed on problem and non-problem statements separately. Given a candidate problem statement within a sentence, the candidate was marked as positive if the string represented one of the following:

1. an unexplained phenomenon or a problematic state in science; or
2. a research question or a description of a task; or
3. an artifact that does not fulAfill its stated specification.

Additionally, the lexical material inside the candidate string must not explicitly mark its status as a problem (e.g. "problem" or "difficult" must not appear *inside* in the string). We made this decision as such explicit signals would detract from the real task, that of judging the semantics of the string itself as problematic, without requiring explicit signals.

For each candidate non-problem statement, the candidate was marked positive if it conformed to both of the following rules:

1. The string is neither a phenomenon, a problematic state, a research question or a nonfunctioning artefact.
2. If the string expressed a research task, without explicit statement that there was anything problematic about it, we allowed for it to be defined as a non-problem.

[2]The head POS tags were found using the Collins' head finder [18].

Additionally, there must not be a different other description of a problem in the rest of the sentence (i.e. in the lexical items around the candidate). Non-grammatical/syntactic sentences were excluded (these could appear in our corpus as a result of its source being OCRed text). If the annotator found that the sentence had been slightly misparsed, but did contain a non-problem or problem, they were allowed to move the boundaries for the candidate string. This resulted in cleaner text, e.g., in the frequent case of coordination, when non-relevant constituents could be removed. This quality-test was conducted by both authors independently. From the set of sentences where both annotators agreed, 500 problem and 500 non-problem statements were randomly chosen.

The scientific documents containing statements resulting from the quality-test were converted to SciXML [8]. An AZ [20] model was trained on 80 computational linguistics papers (mutually exclusive to our quality-test document collection) which then predicted AZ [20] zones for each document.

2.2. Feature Extraction

To construct our feature sets, we began with a bag of words baseline using only the words within candidate phrases. This will tell us about the disambiguation ability of the problem description's semantics alone. Polarity of known words was then taken into account by first finding the head of each candidate phrase and then performing word sense disambiguation of each head using the Lesk algorithm [21]. We then looked up the polarity of the resulting synset using SentiWordNet [22]. Tense, negation, and modality were then added. To model negation, we specified a small set of negative words (e.g. "not") and for each word following a negation, appended "_not" until a phrase marking (e.g. ".,?"). Syntactic features were then added by including the POS tag distribution. We were careful not to base our model only on the head POS tag and the length of each candidate phrase, as these were features used for generating the non-problem candidate set. Since some phrasal heads may have been left without a sense by the Lesk algorithm (and thus with value NONE), we decided to use distributional semantic similarity between all nouns and verbs in each candidate phrase to words with a known polarity. We chose the words "poor" and "excellent" as these have been shown to be good indicators of polarity status in previous studies [23,24]. Semantic distance was calculated as before (cf. Section 2.1) using cosine similarity with our Word2Vec model. To take the rhetoric context into account, we used the AZ zones of the four sentences prior to each candidate phrase. However, when calculating the context of candidate sentences in the main body we never included the abstract, as these two sections fulfill different rhetorical functions and should not affect each other. For all features we decided not to use the additional textual material (other than the candidate phrase) contained in the sentence itself. This is done in order not to distort the task's difficulty.

3. Results

As shown in Figure 2, the bag of words baseline we chose performs better than random. Adding in the polarity of known synsets provides a small improvement, as does tense, negation, and modality. However, making use of the syntax within each candidate phrase provides a significant increase in performance. This may be due our obser-

Feature Sets	Classification Accuracy	
	NB	LR
$\text{Baseline}_{\text{bow}}$	57.1	56.7
+Polarity	55.6	56.9
+Tense, Neg, Mod	57.1	58.8
+Syntax	61.7*	65.6*
+Word2Vec	81.0*	84.5*
+AZ	81.4	84.7

Figure 2. Performance statistics for our classification task using Naïve Bayes (NB) and Logistic Regression (LR). Each consecutive feature set is cumulative. 10-fold cross-validation was used across all experiments. * denotes significance with respect to the previous feature set.

IG	Feature
0.7199	Word2Vec:poor
0.2437	Word2Vec:excellent
0.0258	pos:VB
0.0184	pos:.
0.0147	pos:JJ
0.0119	pos:DT
0.0112	pos:IN
0.0109	pos:TO
0.0104	pos:NNS
0.0103	pos:PRP
0.0102	pos:CD
0.0089	pos:WDT

Figure 3. Information gain (IG) in bits of top features from the best performing model (AZ).

vation that problems often take the form of questions, giving rise to a high concentration of WH- POS tags. Another significant performance increase was caused by using the Word2Vec model. This improvement is likely due to the effect of smoothing mentioned earlier: instead of receiving a null score for unknown words using the synset polarities in SentiWordNet, we are given a distance measurement. The marked improvements from Word2Vec are reflected in Figure 3, where Word2Vec attributes have the greatest information gain.

However, providing knowledge of the rhetoric context using the AZ zones leading up to each candidate statement provides the best performance for both classifiers used. This result supports one of the core claims of AZ: that rhetorical labels of certain statements influence each others' rhetorical status. Therefore, knowledge of the rhetorical context of a problem or non-problem is an important attribute for automatically classifying problem-solving structure within scientific argumentation.

4. Conclusions and Future Work

In this work, we have introduced a new hand-curated corpus of problem and non-problem statements, and shown that identifying and automatically classifying these statements is a tractable task. Our best system beat the baseline by a large margin, with the best performing feature set taking advantage of the statement's rhetorical context using Argumentative Zoning.

In future work, we intend to split the candidate statements into *tasks*, *problems* and *non-problems*. The fact that descriptions of tasks could be both a problem and non-problem in the rubric for hand-crafting our data set, is likely to provide a large degree of noise. Therefore, making this distinction may show a substantial increase in performance. We also plan to explore additional contextual features such as citations, and test the domain specificity of identifying problems and non-problems against corpora from other fields such as chemistry and genetics.

References

[1] Michael Hoey. Signalling in discourse. 1979.

[2] Michael P Jordan. The rhetoric of everyday english texts. 1984.

[3] James P Zappen. A rhetoric for research in sciences and technologies. *New essays in technical and scientific communication*, pages 123–138, 1983.

[4] Bogdan Trawiński. A methodology for writing problem structured abstracts. *Information processing & management*, 25(6):693–702, 1989.

[5] VI Solovev. Functional characteristics of the authors abstract of a dissertation and the specifics of writing it. *Scientific and Technical Information Processing*, 3:80–88, 1981.

[6] Myrna Gopnik. *Linguistic structures in scientific texts*, volume 129. Mouton, 1972.

[7] John Hutchins. On the structure of scientific texts. *UEA Papers in Linguistics*, 5(3):18–39, 1977.

[8] Simone Teufel. The structure of scientific articles: Applications to citation indexing and summarization (center for the study of language and information-lecture notes). 2010.

[9] Simone Teufel, Advaith Siddharthan, and Colin Batchelor. Towards discipline-independent argumentative zoning: evidence from chemistry and computational linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, pages 1493–1502. Association for Computational Linguistics, 2009.

[10] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10, pages 1320–1326, 2010.

[11] Yang Liu, Xiaohui Yu, Zhongshuai Chen, and Bing Liu. Sentiment analysis of sentences with modalities. In *Proceedings of the 2013 international workshop on Mining unstructured big data using natural language processing*, pages 39–44. ACM, 2013.

[12] Ben Medlock and Ted Briscoe. Weakly supervised learning for hedge classification in scientific literature. In *ACL*, volume 2007, pages 992–999. Citeseer, 2007.

[13] Isaac G Councill, C Lee Giles, and Min-Yen Kan. Parscit: an open-source crf reference string parsing package. In *LREC*, 2008.

[14] Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454, 2006.

[15] Roger Levy and Galen Andrew. Tregex and tsurgeon: tools for querying and manipulating tree data structures. In *Proceedings of the fifth international conference on Language Resources and Evaluation*, pages 2231–2234. Citeseer, 2006.

[16] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[17] Kathy McKeown, Hal Daume, Snigdha Chaturvedi, John Paparrizos, Kapil Thadani, Pablo Barrio, Or Biran, Suvarna Bothe, Michael Collins, Kenneth R Fleischmann, et al. Predicting the impact of scientific concepts using full-text features. *Journal of the Association for Information Science and Technology*, 2016.

[18] Michael Collins. Head-driven statistical models for natural language parsing. *Computational linguistics*, 29(4):589–637, 2003.

[19] Alan Agresti and Maria Kateri. *Categorical data analysis*. Springer, 2011.

[20] Simone Teufel et al. *Argumentative zoning: Information extraction from scientific text*. PhD thesis, Citeseer, 2000.

[21] Michael Lesk. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26. ACM, 1986.

[22] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.

[23] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

[24] Tony Mullen and Nigel Collier. Sentiment analysis using support vector machines with diverse information sources. In *EMNLP*, volume 4, pages 412–418, 2004.

Discourse relations: genre-specific degrees of overtness in argumentative and narrative discourse

Carolin HOFMOCKEL[1a], Anita FETZER[a], and Robert M. MAIER[a]
[a] *University of Augsburg*

Abstract. This paper presents a contrastive analysis of the linguistic realization of discourse relations in the argumentative discourse genre of commentary and the narrative genre of short personal narrative, comparing and contrasting co-constructed texts from English editing-based tasks with single-authored ones. The methodological framework integrates Segmented Discourse Representation Theory with Functional Grammar and psycholinguistic models of discourse processing. Genre is understood as a blue-print which constrains the linguistic realization and interpretation of discourse relations. The focus of analysis is on the overt signaling of discourse relations through discourse connectives and non-canonical word order, and on their adjacent and non-adjacent positioning. Across single-authored and co-constructed texts, the narrative genre shows a higher degree of overtness guiding the reader in their inference processes while the argumentative genre displays a lower degree, with single-authored texts generally showing a higher degree of overtness than co-constructed texts.

Keywords. discourse genre, discourse relation, discursive glue, overtness

1. Introduction

Argumentative discourse contains internal and external relationships between premises and conclusions. The internal perspective to argumentation has been introduced by Ducrot (1984) and Anscombre and Ducrot (1983). They claim that every piece of discourse contains explicit and implicit dialogues. Argumentation is expressed through argumentative operators and argumentative connectors, such as *but*, *since, because, although* and *thus,* and *almost, hardly, only, still, few, little, by the way* and *moreover*, through the sequential organization of arguments and through argumentative principles, such as claim, warrant or backing.

Argumentative discourse has also been compared to narrative/expository discourse with respect to genre-dependent structural organization through cohesive means (e.g. Berzlánovich et al. 2012; Berzlánovich & Redeker 2012, Redeker & Egg 2007; Fetzer & Speyer under review) and the types of discourse relations it contains (e.g. Sanders 1997, Wolf & Gibson 2005). With regard to lexical cohesion, for instance, it has been shown to feature a lower "cohesion density" (Berzlánovich et al. 2012: 150) than narrative discourse; as concerns discourse (or: coherence) relations, studies have identified a

1 Corresponding Author.

higher proportion of pragmatic (vs. semantic) discourse relations in argumentative discourse (e.g. Sanders 1997), in which the illocutionary act, rather than its propositional content, serves as the source of coherence.

Viewing lexical coherence, coherence strands, discourse connectives and discourse relations as carriers of discursive glue that contribute to the construal of discourse coherence (Maier, Hofmockel and Fetzer 2016), the present paper continues this line of research, focusing on discourse relations as one particular type of discursive glue. Discourse relations are conceived of as sociocognitive entities that may be realized overtly through discourse connectives and non-canonical word order, or implicitly through coherence strands (Givón 1993) and lexical coherence. As the linguistic realization of discourse relations is assumed to be constrained by discourse genre, genre-dependent preferences regarding their degree of overtness are to be expected.

The goal of this paper is to examine the linguistic realization of discourse relations in the argumentative discourse genre of commentary and in the narrative genre of short personal narrative, comparing and contrasting 19 single-authored texts with 18 co-constructed texts from English editing-based tasks.

1.1. Discourse relations and overtness

In line with Segmented Discourse Representation Theory (SDRT) (Asher & Lascarides 2003), we conceive of discourse relations (DRs) as making manifest logical connections between two propositions in a discourse and thus as an important type of discursive-glue carrier. The classification of DRs is based on the semantics of their connectedness, referred to as their defining condition, and systematized in table 1. The propositions connected through discourse relations are realized through spans of discourse referred to as discursive units (DUs), which typically take the form of clauses, but may also be realized as clause-like units (Marcu 2000) such as complex noun phrases. DUs entering a DR may be adjacently or non-adjacently positioned.

Table 1. DRs and their defining conditions.

Discourse Relation		**Defining conditions**
Coordinating	Continuation	Common topic
	Narration	Common topic Temporal sequentiality
	Contrast	Semantic dissimilarity between ¶[2] and ¶[1]
	Background	¶[2] forms the background of ¶[1] Common topic
Subordinating / Superordinating	Result	¶[1] gives reason for (parts of) events in ¶[2] Connecting two sub-events Temporal precedence of cause
	Comment	¶[2] selects ¶[1] as topic; or: ¶[1] selects ¶[2] as topic
	Elaboration	Topic of ¶[2] specifies topic of ¶[1] mereologically
	Explanation	¶[2] gives reason for (parts of) events in ¶[1] Temporal consequence

Analyzing the linguistic realization of DRs in context requires the explicit accommodation of linguistic cues as carriers of their defining conditions. Carriers of the defining conditions of DRs are proposition-anchored

- coherence strands
    - topic continuity, i.e. topic identity or specification
    - tense and aspectual coherence (including modality) (cf. Givón 1993)
- lexical coherence in particular in lexical relations (antonymy etc.)

and non-propositional

- discourse connectives
- metacomments
- pragmatic word order.

While topic continuity, tense/aspectual coherence and lexical coherence index the defining conditions of a DR implicitly, non-canonical word order (e.g. non-congruently configurated theme zones) (Halliday 1994; Hannay 1994) and discourse connectives like *but*, *however* and *so* cue defining conditions overtly. Thus, depending on whether the defining conditions of a DR are indexed through reference to topic-, tense/aspect- and lexis-based coherence strands, or through discourse connectives and/or non-canonical word order, we differentiate between their implicit and overt realization (cf. Maier, Hofmockel & Fetzer 2016). In (1), for instance, Contrast is realized overtly, with its defining condition being cued trough topic and referential continuity ('London' – 'it'), temporal discontinuity ('was' – 'is') and lexical antonymy ('dowdy'; 'stale' – 'exciting'), with the contrastive discourse connective 'but' and pragmatic word order (fronted temporal adjuncts 'in the 1950s' and 'today') providing additional discursive glue. In (2), Explanation is realized implicitly through the cues of topic continuity, the lexical index 'causing' and temporal-aspectual overlap.

(1) #1/2 [*In the 1950s*,] London *was* a DOWDY place of tea-houses and STALE rock cakes.
#1/3 [*But today*,] it'*s* MUCH MORE EXCITING.

(2) #1/12a English has gradually become the LINGUA FRANCA,
#1/12b **causing** interest in FOREIGN LANGUAGE to wane.

Overt vs. implicit DR realization is assumed to depend on (1) the semantics of the DR, (2) its sequential status in the discourse, viz. in adjacently or non-adjacently positioned DUs realizing the DR, and the contextual constraints of genre, yielding genre-specific degrees of overtness.

1.2. Genre as blueprint

Thibault's definition of genre (2003: 44) differentiates between type and token with genre being a type – or a blueprint – which may be realized with different tokens, regarding both quantity and quality:

> genres are types. But they are types in a rather peculiar way. Genres do not specify the lexicogrammatical resources of word, phrase, clause, and so on. Instead, they specify the typical [original emphasis] ways in which these are combined and deployed so as to enact the typical semiotic action formations of a given community.

Connected intrinsically with 'typical ways' of doing things with words in a discourse genre – or in an activity type, in Levinson's terms (1979: 370) – are inferential schemata:

> … there is another important and related fact, in many ways the mirror image of the constraints on contributions, namely the fact that for each and every clearly demarcated activity there is a set of inferential schemata [original emphasis]. These schemata are tied to (derived from, if one likes) the structural properties of the activity in question

Adopting Thibault's approach to genre supplemented with Levinson's pragmatic conceptualization of activity type we assume that the argumentative genre of commentary does things differently with respect to the overt realization of DRs than other genres, for instance narratives.

2. Data and Method

In order to identify genre-dependent preferences for the overt vs. implicit realization of DRs, we analyzed two datasets: one of single-authored texts comprising 9 argumentative editorials from the British quality newspaper *The Guardian* and 10 short personal narratives[2] from British university students (N = 9.377), and another of co-constructed texts comprising 9 argumentative and 9 narrative texts (N = 3.702). The co-constructed argumentative and narrative texts are from an editing task in which dyads were asked to collaboratively flesh out a 'bare' text which was adapted from a genre-specific original. The skeleton text had been stripped of all adjuncts, disjuncts and conjuncts, and thus reduced to a minimum of propositional content; but it still contained its original sequential organization and default configuration of events. Dyads were requested to turn the bare texts into a well-formed coherent whole; they were allowed to add and delete linguistic material, but they had to maintain the order of the given units.

Based on the model outlined in section 1, all texts were segmented into DUs, coded for DRs and analyzed with respect to their implicit and overt realization of DRs by two inter-raters negotiating deviating analyses. Subsequently, we compared and contrasted the realization and (non-)adjacent positioning of the coordinating DRs of Continuation, Narration and Contrast, and the subordinating DRs of Elaboration, Explanation and Comment across narrative and argumentative texts, and across single-authored and co-constructed ones. The results of the contrastive analysis are presented below.

3. Results

In our data, narrative texts (single-authored and co-constructed) show an overall higher degree of overtness than argumentative texts (cf. table 2). The two genres differ significantly with respect to the realization of Continuation, Explanation and Comment. For Contrast and Elaboration, there are no significant differences, with contrastive DRs showing strong tendencies towards overt realization across genres.

2 The narratives have been collected by N.M. Fronhofer.

Table 2. Degrees of overtness in single-authored and co-constructed commentaries and narratives.[3]

DR	Commentaries (single-authored)		Narratives (single-authored)		Commentaries (co-constructed)		Narratives (co-constructed)	
	%	(total N)	%	(total N)	%	(total N)	%	(total N)
Contrast	100	(40)	100	(38)	100	(15)	100	(1)
Continuation	26.9	(158)	52.2	(155)	57.1	(35)	58.5	(65)
Narration	n/a	(0)	71.1	(38)	n/a	(0)	83.3	(18)
Explanation	37.5	(40)	100	(25)	50	(22)	100	(19)
Elaboration	76.9	(325)	72.1	(298)	57.8	(57)	46.7	(45)
Comment	14.8	(27)	75	(20)	7.7	(13)	0	(6)
overall	59.4	(590)	69.9	(574)	49.3	(142)	60.4	(154)

Genre-dependent differences in the degree of overtness are significant both in single-authored and co-constructed texts, with commentaries being significantly less overt than narratives (χ =13.948, $p < .01$ for single-authored texts; $\chi = 3.674$, $p < .10$ for co-constructed texts). The realization of DRs in co-constructed texts parallels that of the single-authored texts as regards Contrast, Narration and Explanation, but shows differences for Continuation, Elaboration and Comment.

Table 3. Degrees of overtness in single-authored and co-constructed commentaries and narratives.[4]

		Commentaries (single-authored)	Narratives (single-authored)	Commentaries (co-constructed)	Narratives (co-constructed)
adjacent	$\sum_{overt}$	316	338	62	80
	$\sum_{total}$	*467*	*462*	*121*	*127*
	rate (overt)	**67.7 %**	**73.2 %**	**51.2 %**	**66.1 %**
non-adj.	$\sum_{overt}$	34	75	8	13
	$\sum_{total}$	*123*	*136*	*21*	*27*
	rate (overt)	**27.6 %**	**55.1 %**	**38.1 %**	**48.1 %**
χ		64.629	15.957	1.2369	2.051
p		0	< 0.0001	(> 0.1)	(> 0.1)

The data also show genre-dependent differences for the degree of overtness in adjacently/non-adjacently positioned DUs realizing DRs (cf. table 3). While the overall proportion of adjacently and non-adjacently positioned DUs realizing DRs is relatively homogeneous across all texts examined, with argumentative texts having a slightly higher proportion of adjacently positioned DUs realizing DRs (80.3% vs. 78.2%), we find a higher rate of overtness in both adjacently and non-adjacently positioned DU realizing DRs in the narrative texts. Degrees of overtness vary significantly across adjacently and non-adjacently positioned DUs realizing DRs in single-authored texts with less variation in the co-constructed ones.

3 Percentages indicate the percentage of overt DR realizations. Round brackets ('( )') indicate the total numbers of DRs (overtly and implicitly realized) in raw frequencies.

4 For single-authored narratives, table 3 considers all of the DRs realized in the data, and yields, for this reason, a higher N of DRs than table 2.

4. Discussion

Variation in the overt vs. implicit realization of DRs shows that their realization is constrained by their semantics. Across our data, Contrast is always realized overtly through non-propositional material, i.e. discourse connectives or pragmatic word order, indicating that its realization is overt by default, whereas all other DRs have more variable patterns of realization. As reflected in the results of our contrastive study, these patterns are significantly influenced by discourse genre, supporting our view that discourse genre functions as blueprint for the production and perception of DRs. This is corroborated in our comparison of single-authored texts with co-constructed texts, which shows that the joint productions tend to adhere to the overall degrees of overtness of genre blueprints specifying typical ways of coding DRs, despite some variation for Continuation, Elaboration and Comment. That co-constructed texts show a lower degree of overtness across both narrative and argumentative genres is very likely due to their joint construction and their explicitly negotiated wellformedness.

In our data, DR realizations in narrative texts are significantly more overt than in argumentative texts, reflecting differences in the communicative needs the genres fulfil, guiding the interlocutors' production and interpretation processes. Derived from the oral discourse activity of storytelling, narrative genres tend to be closer to spoken modes of language. They relate a sequence of events from a particular perspective and are thus more subjective or, in Chafe's (1982) terms, more involved, than argumentative ones. As the interlocutors by default consider the sequence of DUs to represent temporally unfolding events linked through the DR Narration, we assume that DRs that indicate deviations from this narrative sequence, e.g. Continuation, Comment and Explanation, require overt marking, thus leading to significantly higher degrees of overtness for these DRs in narrative texts as compared to argumentative texts across our data. The close association of narrative discourse with spoken language, which is more redundant to ensure felicitous communication, may be another catalyst for overt DR realization. One means to provide such redundancy is to supplement the linguistic realization of DRs with discourse connectives to trigger generalized conversational implicatures as to their nature of connectedness (cf. Ariel 2008), thus ensuring the activation of relevant defining conditions to guide recipients in their interpretation of the DR as intended by the producer (Maier, Hofmockel & Fetzer 2016).

A huge part of the activity type of argumentation consists of the negotiation of validity claims and thus of their justification and refutation (cf. Fetzer 2007). Prototypical DRs utilized for backing and warranting these claims are Contrast, Explanation and Elaboration. Due to the genre-specific constraints of argumentation, interlocutors are prone to producing/interpreting sequences of DUs as being linked causally in argumentative discourse and therefore not in need of overt signaling. For this reason, it seems redundant to mark justifying DRs such as Explanation or Comment overtly in argumentative discourse, but not in narrative discourse; moreover, argumentative discourse may tend to refrain from using polysemous connectives that may be interpreted as causal if there is no causal interpretation intended. This appears to be reflected, for instance, in the comparatively low incidence of *and* as an overt marker in the realization of Continuation in our argumentative data. Despite their lower degree of overtness, the argumentative data only reveal few cases in which defining conditions of DRs have not been indexed in a

clear-cut manner (5.7%),[5] with lexical cueing as a preferred mode to activate defining conditions, counterbalancing the lack of overt cues. From a processing perspective, the lower degree of overtness in the argumentative genre supports the view that argumentative texts differ from narrative ones in their processing load (cf, e.g., Sanders 1997 and Berzlánovich & Redeker 2012).

Genre-specific differences are also reflected in the structuring of discourse as regards the level of adjacency. There is a significant difference in the overt marking of non-adjacently positioned DU realizing DRs (χ =30.343, p < .000001), with narrative texts signaling DRs overtly across non-adjacent DUs more frequently than argumentative ones. We assume that for the narratives, non-adjacently positioned DU realizing DRs are more overt because deviations from the default chronological story line need to be accounted for.

References

Anscombre, Jean-Claude & Oswald Ducrot, *L'Argumentation dans la Langue*, Mardaga, Brussels, 1983.

Asher, Nicholas & Alex Lascarides, *Logics of conversation*, Cambridge, Cambridge University Press, 2003.

Berzlánovich, Ildikó, Markus Egg & Gisela Redeker, Coherence structure and lexical cohesion in expository and persuasive texts, in A. Benz, M. Stede & P. Kühnlein, *Constraints in Discourse 3. Representing and inferring discourse structure*, Amsterdam, Benjamins, 137-164, 2012.

Berzlánovich, Ildikó & Gisela Redeker, Genre-dependent interaction of coherence and lexical cohesion in written discourse, *Corpus Linguistics and Linguistic Theory* **8**/1 (2012), 183-208.

Chafe, Wallace L., Integration and involvement in speaking, writing and oral literature, in D. Tannen (ed.), *Spoken and written language: Exploring orality and literacy,* Norwood, Ablex, 35-53, 1982.

Dik, Simon, *The Theory of Functional Grammar* (2 vols), Ed. K. Hengeveld, Amsterdam, Benjamins, 1997.

Ducrot, Oswald, *Le Dire et le Dit*, Minuit, Paris, 1984.

Fetzer, Anita. "Well if that had been true that would have been perfectly reasonable": appeals to reasonableness in political interviews, *Journal of Pragmatics* **39**/8 (2007), 1342-1359.

Fetzer, Anita & Augustin Speyer, Discourse Relations across Genres and Contexts: A Contrastive Analysis of English and German Discourse, under review.

Givón, Talmy, *English Grammar. A function-based introduction* (2 vols), Amsterdam, Benjamins, 1993.

Halliday, M.A.K., *Introduction to Functional Grammar*, 2nd ed., London, Arnold, 1994.

Hannay, Mike, The Theme zone, in R. Boogart & J. Noordegraaf (eds.), *Nauwe Betrekkingen*, Amsterdam, Neerlandistiek and Münster, Nodus Publikationen, 107-117, 1994.

Levinson, Stephen C., Activity types and language, *Linguistics* **17**/5-6 (1979), 365-400.

Maier, Robert M., Carolin Hofmockel & Anita Fetzer, The negotiation of discourse relations in context: Co-constructing degrees of overtness, *Intercultural Pragmatics* **13**/1 (2016), 71-105.

Marcu, Daniel, The rhetorical parsing of unrestricted texts: a surface-based approach, Computational Linguistics 26/3 (2000), 395-448.

Redeker, Gisela & Markus, On the interaction of relational coherence and lexical cohesion in expository and persuasive text genres, Paper presented at *10th International Pragmatics Conference*, Göteborg, 8-13 July, 2007.

Sanders, Ted, Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context, *Discourse Processes* **24**/1 (1997), 119-147,

Sanders, Ted & Leo Noordman, The Role of Coherence Relations and Their Linguistic Markers in Text Processing, *Discourse Processes*, **29**/1 (2000), 37-60.

Speyer, Augustin & Anita Fetzer, The coding of discourse relations in English and German argumentative discourse, in H. Gruber & G. Redeker (eds.), *The Pragmatics of Discourse Coherence. Theories and Applications*, Amsterdam, John Benjamins, 87-119, 2014.

Thibault, Paul, Contextualization and social meaning-making practices, in S.L. Eerdmans, C.L. Prevignano, & P.J. Thibault (eds.), *Language and interaction: Discussions with John J. Gumperz*, Amsterdam, John Benjamins, 41–62, 2003.

Wolf, Florian & Edward Gibson, Representing discourse coherence: A corpus-based study*, Computational Linguistics* **31**/2, 2005.

5 This leads to DRs being underspecified and thus to a small number of overlaps, especially among Background and Elaboration, and Continuation and Result.

Rephrase in Argument Structure

Barbara KONAT [a], Katarzyna BUDZYNSKA [b,a], and Patrick SAINT-DIZIER [c]
[a] *Centre for Argument Technology, University of Dundee, UK*
[b] *Institute of Philosophy and Sociology, Polish Academy of Sciences, Poland*
[c] *IRIT-CNRS, Toulouse, France*

Abstract. When making arguments during public consultations, disputants tend to rephrase their premises and conclusions. Citizen Dialogue corpus comprises of transcripts of public meetings in which rephrasing contents plays a significant role. The modification of wording and the repetition of pro-&con- arguments might have an important rhetorical effect, it does not, however, increase the degree or strength of support and attack for or against one's position. This paper makes the first step to set the foundations for argument mining and argument analytics systems to be applied in such types of discourse. We introduce the first corpus for rephrases in argumentative context and we show a preliminary study of the linguistic characteristics for different categories of rephrase relation.

1. Introduction

In this study we propose the analysis of the material from the Citizen Dialogue (CD) corpus as the laboratory for modelling the rephrase relation. Government agencies struggle with structuring and making sense of the feedback collected from citizens during public consultations. Argument mining and argument analytics systems [1] can support deeper understanding of debates by creating structured summaries and automatically extracting claims which divide citizens [2]. Rephrases, however, pose a particular challenge, as these systems fail in measuring the amount of support (or attack) for (or against) a given position, when the premises or conclusions are repeated with different linguistic surface. To our best knowledge, this problem – which is common in argument mining – has not been yet addressed in a systematic manner. The aim of this paper is to propose a linguistic model of rephrase in argument structure using the corpus analysis and to set up the foundation for future implementations enriching argument and debate technologies.

2. Related work

2.1. Argument Mining

Argument mining (see e.g. [3,4,5] for an overview) is a rapidly growing area of computational models of argument and computational linguistics aimed at the automatic detection and extraction of reasons behind opinions, decisions, actions which people express in different situations, media or domains. The task of recognising identical and different statements is of high relevance to argument mining [6]. The argument mining community has begun to address this task by distinguishing between claims, restatements and premises in the annotation schemes [7] and by identifying identical claims for summaris-

ing argumentative conversations [8]. The studies on the applicability of discourse relations to argument mining [9] have resulted in various attempts to develop a system for the recognition of restatement of speaker's claims in argumentative discourse [10,11].

2.2. Paraphrase Mining

Two text spans with similar discourse functions can be treated as paraphrases. The notion of similarity of text spans is what makes paraphrases and rephrases closely related. Various definitions of paraphrase have been proposed in the literature, focusing either on full, close or loose semantic equivalence. Hirst [12] defines paraphrase as "talking about the same situation in a different way" with changes in the wording or syntactic structure. He argues that paraphrases are not fully synonymous: they exhibit pragmatic differences of evaluation, connotation and viewpoint. Melcuk [13], in his analysis of lexical functions in human communication, proposes the concept of approximate synonymy: "An approximate synonymy of sentences is considered as sufficient for them to be produced from the same Semantic Structure". Dras [14] defines paraphrase pair as "a pair of units of text deemed to be interchangeable". The term "quasi paraphrases" is proposed by Bhagat and Hovy [15] and defined as "sentences or phrases that convey approximately the same meaning using different words", as an attempt at providing exhaustive list of linguistic transformations involved in paraphrase identification.

Such understood similarity has limited applicability for argument mining and argument analysis tasks, because two text units, despite being semantically similar, can serve different argumentative functions. Thus, we propose the concept of rephrase defined solely with regards to its function within an argument structure. Notice that a paraphrase can be analysed as a pair of text spans isolated from its context (see e.g. Microsoft Research Paraphrase corpus [16]), while a rephrase requires to consider its argumentative context (function).

3. Rephrase and argument

Rephrase is a relation which holds between two text spans which have the same argumentative function but different linguistic surface and therefore should not be considered as two arguments pro- or two arguments con- by argument analytics and argument mining systems. In other words, two text units are in the relation of rephrase when substitution of one unit for another preserves the argument structure. By argument structure, we understand here the structure built by inference (arguments pro-) and conflict (argument con-) relations.

Rhetorical and dialogical functions of the two rephrased text spans can be very different. Speakers are rephrasing their positions, premises, conclusions for multiple reasons such as a rhetorical effect (repeating a premise can pretend that the argument is stronger than it really is) or for the reduction of cognitive limitations (stating the same point twice can support communication and memory). In Example (1) taken from Citizen Dialogue (CD) corpus, Greg Dyer, a representative of Tennessee Department of Transportation (TDOT) is responding to the citizen's comment about the future road being built very close to her property.

(1) CITIZEN: *What impact is that [the road] going to have when it is so close to a house?*
MR. GREG DYER: *Yes, ma'am. So the right-of-way limits for the corridor is 250 feet, correct, but*

our actual road is going to be, you know, centered within that right-of-way. *So the road is not going to be at the edge of the prop – of the property that we're purchasing,* ***it's going to be in the middle of that right-of-way***. *So we're not going to – you know, maybe the land that we're purchasing is within ten feet of your property, but* ***our road is going to be, you know, within that right-of-way***.

Greg Dyer is creating the argument with the conclusion that the road will not be at the edge of the citizen's property. He provides three premises (text in bold in Example (1)). Figure 1 presents the same text spans analysed with the use of the rephrase relation capturing the fact that there is only one pro-argument, albeit rephrased by using words "centered within", "in the middle" and "within". Each of the text spans serves the same argumentative function and substituting one for another would result in identical argument structure. When assessing the amount of support provided for the conclusion, an argument analytics system would correctly interpret this argument as having only one premise, albeit rephrased in three different statements.

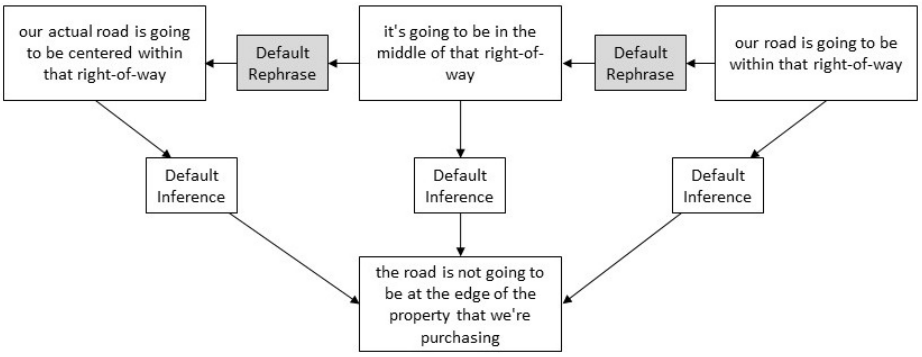


Figure 1. Rephrased premise (argument map CD# 10029)

In Example (2), another representative of DOT (Heather Jensen) formulates two arguments with very similar conclusions (see text in bold at the beginning and end of the turn).

(2) MS. HEATHER JENSEN: *Again,* ***this project got delayed*** *because there were so much tracts in this and we had to figure out a way that – a bite-size chunk that we could do this project in. So that took a little bit of time. And then we had environmental studies, geotechnical studies, all those things that come into play because this was a very large project. So that is part of the reason why* ***this project hasn't gone as quickly as maybe it could have*** *or maybe another comparable project.*

Proposed interpretation of the argument structure includes the relation of rephrase between the conclusions (see Figure (2)), emphasising that the two conclusions are in fact one statement, and premises provided for one of them could be used to support another.

4. Citizen Dialogue corpus

The material for the Citizen Dialogue corpus comprises of three transcripts of public meetings organised by the Tennessee Department of Transportation (TDOT).[1] In the

[1]The transcripts are prepared by a court reporter, were obtained from the official TDOT website: `https://www.tn.gov/tdot/section/public-hearings-meetings`

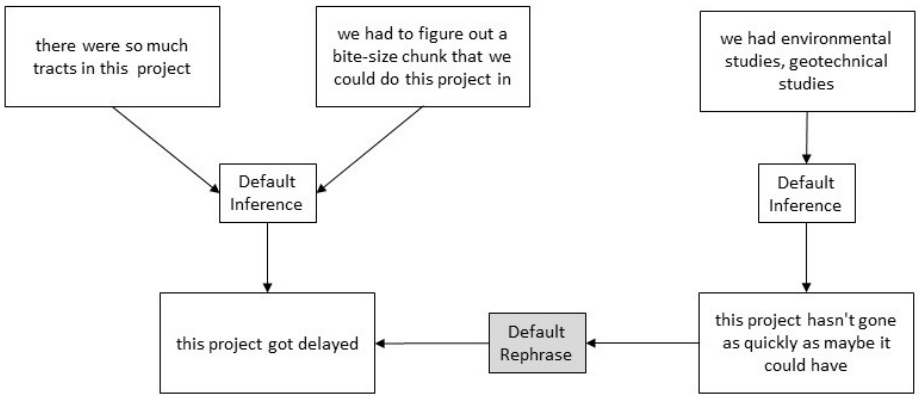


Figure 2. Rephrased conclusion (argument map CD# 10030)

transportation rulemaking system, a public meeting is a part of the decision making process, where a state branch of Department of Transportation is informing and consulting citizens with regards to new regulations and development plans.[2]

The annotation of the transcripts was performed using the OVA+ annotation tool[3] [17] marking three types of relations between propositional contents of utterances: pro-arguments (Default Inference); con-arguments (Default Conflict); and the relation of rephrase (Default Rephrase):

Default Inference holds between two propositions when one proposition provides a reason to accept another proposition. In other words, for a given claim p, a supporting claim q can potentially be used to reply to the question "Why p?" ("Because q").

Default Conflict holds between two propositions which cannot be both true at the same time. Speakers use conflicting propositions to attack another speaker's claims by means of providing counter-claims.

Default Rephrase holds between two non-overlapping text spans which occupy the same position in the argument structure (in other words, which preserve the argument structure).

Resulting Citizen Dialogue corpus is freely available[4] and uses open Argument Interchange Format (AIF) [18] standard for argument representation, constituting a part of the AIFdb database[5] [19]. For the inter-annotator agreement, a systematic sample of 10% of the corpus was extracted and annotated by the second annotator, resulting in Cohen's $\kappa = 0.83$ for all relations (Default Inference, Default Conflict and Default Rephrase).

The frequency of the rephrase relation in Citizen Dialogue is significantly higher in comparison with two other corpora of argument structures available in AIFdb: the Moral Maze 2012 (MM2012) corpus [20] comprising of the transcript of the BBC4 radio

[2] See U.S.Department of Transportation "Public Involvement Techniques for Transportation Decisionmaking" `http://www.fhwa.dot.gov/planning/public_involvement/publications/pi_techniques/fhwahep15044.pdf`

[3] Online Argument Visualisation (OVA+) tool available at: `http://ova.arg-tech.org`

[4] Citizen Dialogue corpus: `http://arg.tech/cd`

[5] Argument Interchange Format database (AIFdb) available at `http://aifdb.org`

Table 1. Summary of Citizen Dialogue corpus

Citizen Dialogue	Argument Maps	Units (propositions)	Words
	31	634	8,667

programme, and the eRulemaking (eR) corpus [2] comprising of an on-line discussion forum material. As presented in Table 2, rephrases constitute large proportion, i.e. 23.8%, of all relations annotated in the CD corpus compared to much lower proportion of 0.9% in the MM2012 corpus and 1.8% in the eR corpus. Although the samples are relatively small, these percentages demonstrate the strong trend in the data confirming that the CD corpus is a good laboratory for studying the relation of rephrase and also surficing the significance of this relation for structuring and analysing the feedback from citizens. [6]

Table 2. Frequency of rephrase relations compared to the frequency of three relations in the corpora: Moral Maze 2012 (MM2012), eRulemaking (eR) and Citizen Dialogue (CD).

Relation/Corpus	MM2012	eR	CD
Default Inference	870	671	108
Default Conflict	215	97	17
Default Rephrase	10	14	39
TOTAL	1095	782	163
Rephrase %	**0.9%**	**1.8%**	**23.8%**

5. A categorization for rephrases

The notion of rephrase is of much relevance for the Citizen Dialogue corpus, due to the specific communicative setting of the meetings: (i) the meetings start with the presentation of a road construction project which is followed by questions, clarification requests, various expressions of doubt or oppositions to the project, (ii) the communication is led by the community leaders, imposing specific form of a communicative hierarchy, and (iii) there are mostly oppositions or supports to the projects at stake: there are seldom opinion confrontations. These interactions are frequently associated with various forms of rephrases to guarantee an optimal mutual understanding and to outline the main points of the project.

In this section, we investigate the different categories of rephrases from a linguistic and argumentative perspective and the role they play in argumentation. Two text units can be in the relation of rephrase, even if (i) their linguistic realisations is substantially different and (ii) they have the same argumentative function but they serve different communication purposes. The analysis of the notion of rephrase presented below involves rephrase pairs from our corpus which are linguistically challenging but do not entail complex domain knowledge and inferences. These constitute about 60% of the total of the rephrase situations that have been manually annotated. The other 40% require a number of contextual considerations (domain and discussion context) which are out of the scope of this initial investigation.

[6] Note that rephrase relations are also present in mononological texts, however here we focus only on the dialogical context.

The categories presented here are essentially based on corpus analysis. They have been developed also in part based upon Rhetorical Structure Theory [21] categories: rephrases are discourse structures, playing a specific role in argumentation. These categories are under elaboration, therefore, they are probably not yet comprehensive and optimal, they may overlap and their scope and definition need further investigations. We propose an initial categorisation of the rephrase relation types, according to their argumentative function and communicative intentions of the speaker:

Reformulation with close terms occurs in pair (3) where the second unit is almost a repetition of the first one:
(3) a. *I'm not sure about any other projects in this area necessarily*
b. *I'm not aware of any other projects in the immediate area*
The second unit is slightly more affirmative than the first one: "not sure" is replaced by "aware"; only the modals differ. This variation slightly increases the strength of the affirmation, and therefore the strength of the argument.

Reformulation using semantically related terms and structure variations from unit:
(4) a. *when we normally start a project, we always begin with utility relocation.*
the following rephrase is uttered:
b. *we will start utility relocation, and then we will work on construction*.
Besides the event re-ordering in their natural sequence, which makes the process clearer, the term "project" is made more specific via the use of the term "construction". Similarly to a simple reformulation, this rephrase should have little impact on the argumentation. The rephrase makes the statement more fluid.

These two first categories are close to paraphrases. More complex rephrases, which are not paraphrases, are related to specific communicative intentions of the speaker:

Confirmation is an important feature of argumentative dialogues, it is frequent in Citizen Dialogue corpus because in this type of exchange it is crucial to make sure that participants clearly understand what is said. The confirmation is often stronger than the initial utterance (CD: #6818):
(5) a. *typically, in other projects those phases overlap* is rephrased as follows with the goal of being more clear and to confirm the general statement that was initially uttered:
b. *those phases will overlap*.

Summarisation In this example (CD: #6822), the summary is also used as a conclusion of a discussion segment, that rephrases the main ideas. It is a frequent strategy in argumentation. The absence of the justification given in (6a) makes it more definitive: (6) a. *there has been a study whether or not there is going to be more truck traffic on this road than what is now coming through Portland since it is a direct connection road* is summarised as: b. *it is going to increase on the traffic*.

Clarification of a complex issue: When speakers have difficulties to express themselves, or when they have a very long request, it is often appropriate to clarify it, for example via summarisation or abstraction, as in the following pair (CD: #6822):
(7) a. *there has been a study whether or not there is going to be more truck traffic on this road than what is now coming through Portland since it is a direct connection road*.
with the clarification request: b. *You're asking about truck traffic ?*.
Other forms of clarifications include the resolution of references, in particular, pronominal, spatial or temporal.

Instantiation of a general statement: (8) a. *some things are out of control* is rephrased via instantiation by: b. *Finance and planning are out of control*. The specification of more precise elements makes the argument in (8) (CD: #6766) rhetorically stronger, similarly to the use of illustrations to support an argument.

More complex types of rephrases: A number of rephrases include several forms of argumentation and communicative intentions. Example (9) (CD: #6755) shows a pair of statements which are not paraphrases, as there is no semantic equivalence between them. Yet they provide one instance of support for the conclusion, therefore can be treated as rephrases: (9) a. *this project got delayed* with the justification that serves as a form of argumentative conceding: b. *but it could have been even more delayed if we hadn't got the help of an extra company*.
Example (10) clearly that rephrase is a notion which is substantially different from paraphrase, linguistically and conceptually, even if a rephrase serves the same argumentative function than the segments it rephrases. The following pair is a rephrase via elaboration (CD: #6822): (10) a. *There will be some truck traffic*
b. *we know you wish to keep you area calm and safe: we made every effort to limit truck traffic*. Elaborations cover a large diversity of phenomena, behaving as a proto-relation.

Automatically recognising the rephrase relation is a very challenging task: it not simply mining for utterances which have a high rate of similarities in terms of words, including direct synonyms, possibly with limited syntactic variations, because there are very few such situations (our corpus has three of such situations). Rephrases often add some meaningful information and convey specific communicative intentions. The development of learning procedures from annotated texts could be a useful approach to automatically recognise and characterise rephrases, but a large number of rephrase pairs must be annotated to guarantee a certain accuracy, considering the diversity of categories, and within each pair, additional tags must be developed to better circumvent the nature of the rephrase and the means which are used (e.g. more generic nouns, modals, sentence reorganisation). A more linguistically-based approach where the deep issues behind the notion of rephrase are investigated seems an alternative, which would have some explanatory and predictive power of interest.

6. Conclusions

This paper reports a preliminary corpus-based study of rephrase which influence the analysis and evaluation of argument structure. Our first results show that rephrase mining is a very complex task which involves the taking into account contextual information. The Citizen Dialogue corpus is currently in development and future work includes building and implementing the linguistic model for automated extraction of rephrases. This work opens new challenges and perspectives to argument analytics and argument mining.

7. Acknowledgements

We would like to acknowledge that the work reported in this paper has been supported in part by the Innovate UK under grant 101777. We also thank Rory Duthie for providing analysis as a second annotator.

References

[1] J. Lawrence, R. Duthie, K. Budzysnka, and C. Reed, "Argument analytics," in *Proceedings of the Sixth International Conference on Computational Models of Argument (COMMA 2016)* (P. Baroni, M. Stede, and T. Gordon, eds.), (Berlin), IOS Press, 2016.

[2] B. Konat, J. Lawrence, J. Park, K. Budzynska, and C. Reed, "A corpus of argument networks: Using graph properties to analyse divisive issues," in *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC 2016)*, 2016.

[3] M.-F. Moens, "Argumentation mining: Where are we now, where do we want to be and how do we get there?," in *FIRE '13 Proceedings of the 5th 2013 Forum on Information Retrieval Evaluation*, 2013.

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

[5] M. Lippi and P. Torroni, "Argumentation mining: State of the art and emerging trends," *ACM Transactions on Internet Technology*, vol. 16, no. 2, 2015.

[6] B. Stein, "Paraphrasing," in *Report of Dagstuhl Seminar Debating Technologies* (I. Gurevych, E. H. Hovy, N. Slonim, and B. Stein, eds.), Dagstuhl Publishing, 2016.

[7] I. Habernal, J. Eckle-Kohler, and I. Gurevych, "Argumentation mining on the web from information seeking perspective," in *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing (ArgNLP)*, 2014.

[8] E. Barker and R. Gaizauskas, "Summarizing multi-party argumentative conversations in reader comment on news," in *Proceedings of the 3rd Workshop on Argument Mining*, Association for Computational Linguistics, 2016.

[9] E. Cabrio, S. Tonelli, and S. Villata, "From discourse analysis to argumentation schemes and back: Relations and differences," in *International Workshop on Computational Logic in Multi-Agent Systems*, 2013.

[10] N. Madnani, M. Heilman, J. Tetreault, and M. Chodorow, "Identifying high-level organizational elements in argumentative discourse," in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, 2012.

[11] A. T. Rutherford and N. Xue, "Robust non-explicit neural discourse parser in English and Chinese," in *Proceedings of ACL 2016*, Association for Computational Linguistics, 2016.

[12] G. Hirst, "Paraphrasing paraphrased," in *Invited talk at the ACL International Workshop on Paraphrase, Sapporo*, 2003.

[13] I. Melcuk, "Lexical functions: a tool for the description of lexical relations in a lexicon," *Lexical functions in lexicography and natural language processing*, vol. 31, pp. 37–102, 1996.

[14] M. Dras, *Tree adjoining grammar and the reluctant paraphrasing of text*. PhD thesis, Macquarie University NSW 2109 Australia, 1999.

[15] R. Bhagat and E. Hovy, "What is a paraphrase?," *Computational Linguistics*, vol. 39, no. 3, pp. 463–472, 2013.

[16] B. Dolan, C. Brockett, and C. Quirk, "Microsoft research paraphrase corpus," *Retrieved March 2005*, 2005.

[17] M. Janier, J. Lawrence, and C. Reed, "OVA+: an argument analysis interface," in *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014)*, pp. 463–464, 2014.

[18] I. Rahwan, F. Zablith, and C. Reed, "Laying the foundations for a worldwide argument web," *Artificial Intelligence*, vol. 171(10-15), pp. 897–921, 2007.

[19] J. Lawrence, F. Bex, C. Reed, and M. Snaith, "AIFdb: Infrastructure for the argument web.," in *Proceedings of the 4th International Conference on Computational Models of Argument COMMA*, pp. 515–516, 2012.

[20] K. Budzynska, M. Janier, J. Kang, B. Konat, C. Reed, P. Saint-Dizier, and O. Yaskorska, "Automatically identifying transitions between locutions in dialogue," in *Proceedings of 1st European Conference on Argumentation: Argumentation and Reasoned Action (ECA 2015)*, 2016.

[21] W. C. Mann and S. A. Thompson, "Rhetorical structure theory: Toward a functional theory of text organization," *Text-Interdisciplinary Journal for the Study of Discourse*, vol. 8, no. 3, pp. 243–281, 1988.

Towards a Model for Ethotic Structures in Dialogical Context

Marcin KOSZOWY [a,b] and Katarzyna BUDZYNSKA [b,c]
[a] *University of Bialystok, Poland*
[b] *Institute of Philosophy and Sociology, Polish Academy of Sciences, Poland*
[c] *Centre for Argument Technology, University of Dundee, UK*

Abstract. The paper aims to make a first key step in developing a model of communicative structures involved in dialogicial manoeuvring which targets speakers' ethos (character or credibility). We present case studies from the real-life rhetorical practice in the UK Parliament for which we propose the representation of dialogical ethotic context. In this preliminary work, we select a specific type of ethotic attacks in which the opponent is accused to be inconsistent. We show a challenge and a solution for dealing with sparse data such as the references to ethos are, and we build the model with the perspective of future applications in ethos mining.

1. Introduction

Aristotle distinguished three modes of communication: logos which is argumentation or propositional content of message; pathos which is emotions of audience; and ethos which is the character (or credibility) of speakers [1]. Although the first one method is most noble and valuable, the latter two are extremely powerful in influencing others.

The communication structures of ethos have been relatively well studied in argumentation theory [14–16], however, their representation is limited to its inferential aspect. In other words, only references to ethos which are dependent on logos, i.e. ethotic argumentation [2, 15, 16, 18, 19], are considered. One of the very first attempts [3] to relax this limitation in order to allow for the representation of dialogical ethotic structures used the framework of Inference Anchoring Theory [5] which captures argumentative (inferential) and dialogical structures and the connections between them. This paper follows this approach in the attempt at exploring the structure of the variety of ethotic manoeuvring which are typical for real-life communication.

In this preliminary study, we select a specific type of referring to ethos in which the opponent is accused to be inconsistent. Its representation is built upon circumstantial *ad hominem*, CAH ("you don't practice what you teach") which in argumentation theory is modelled in terms of argumentation scheme [16, 17]:

CIRCUMSTANTIAL AD HOMINEM ARGUMENTATION SCHEME
Premise$_1$: *a* advocates argument *A*.
Premise$_2$: *a* has carried out an action, or set of actions, that imply that a is personally committed to the opposite of *A*.
(Intermediate) Conclusion: *a* is a bad person.
(Final) Conclusion: *a*'s argument *A* should not be accepted.

Although this scheme allows us to capture the initial ethotic structure, its representation is limited to inferential aspect of communication (it focuses on premise-conclusion structure). The aim of this paper is to extend this approach by considering broader dialogical context (Sect. 3). We also show a method for collecting cases for corpus analysis given the challenge of sparse data typical for references to ethos (Sect. 2) with the perspective of the future applications in argument mining (*cf.* [10, 11]) and ethos mining [7].

2. Method for studying ethotic structures

The method for studying ethotic structures consists of two concurring tasks: (i) elaborating a formal abstract model of ethotic structures, and (ii) applying this model to corpus analysis in order to describe more complex ethotic structures that are typical for the real-life practice. In case of references to ethos, the development of corpus faces the challenge of working on large sparse datasets.

Although ethotic aspects of dialogue are crucially important for establishing and strengthening the credibility relation between the speaker and the audience (or weaken the relation between the speaker's opponents and their audience), ethotic structures have no typical linguistic surface that would be common for all types of ethos supports and attacks. It is rather the case that ethotic structures are context-dependent and thus it is difficult to identify their common linguistic patterns. Despite this difficulty, the ethos mining method (as applied to large datasets such as transcripts of political debates) should be capable of defining such patterns for particular ethotic structures. However, the sparseness problem related to dialogical ethos may be dealt with by generalising the very procedure of dealing with sparse ethotic data in dialogue. This paper is a first step in this direction. To address this challenge we propose to search data repositories using basic, most intuitive keywords for a given ethotic support or attack. In this paper, we work with the Hansard repository which contains transcripts of the UK parliamentary debates structured as sessions.[1] The selection of the genre is motivated by the fact that political debates contain a relatively high number of references to other speakers' (politicians') ethos [7].

In studying attacks on ethos that consist of accusing the opponent to be inconsistent, we create a list of simple and intuitive lexical cues which signals references to inconsistency, such as the word 'inconsistency' (see Table 1), and search for the parliamentary sessions in Hansard which contain such a keyword. The result of the search allows us to reduce the searching space for ethotic structures, e.g. a search for keywords 'inconsistency' and 'inconsistent' returns 37,286 sessions in which these cues were used as opposed to the total number of 747,949 sessions available in Hansard which would have to be search through otherwise. This means that the set of pre-processed texts reduces the further manual search up to 4.98% of whole Hansard with chances to find 6 instances of ethotic structures in every 10 sessions returned by the initial keyword search (we do not assume that every appeal to inconsistency necessarily involves an attack on ethos). The last column shows this proportion for just first lexical cue and not for the other ones as this value is to indicate the general trend in the data. This initial search allows us also to find less obvious, context-dependent keywords that are likely to generate interesting results. For example, a cue 'back benchers' is a good candidate for an extended list of

[1] The complete records of Hansard sessions from 1803 till 2005 are available at: `http://hansard.millbanksystems.com/`.

Table 1. Lexical cues for searching for inconsistency-related ethotic structures (out of the total number of 747,949 sessions in Hansard)

Lexical Cue	**Number of sessions**	**Proportion of search space reduction**	**Proportion of ethotic interactions**
inconsistent (inconsistency)	37,286	4.98 %	60 %
not consistent	14,596	1.95 %	-
inconsequent	373	0.05 %	-
consistent (consistency)	163,313	21.8 %	-
consequent	260,136	34.77 %	-

keywords, as we noticed that in the UK parliamentary debates politicians tend to use this word most typically in the attempt to attack ethos of others. For example, ethos of the party may be attacked by ascribing inconsistency between what front benchers say and what back benchers actually do. For the analysis of excerpts containing the keyword 'back bencher' see sections 3.2 and 3.3.

3. Dialogical ethotic structures involving inconsistency

In this section, we discuss three case studies of ethotic structures involving circumstantial *ad hominem* (CAH). The first case (see section 3.1) was found by searching the hansard repository with one of the intuitive linguistic cues ('inconsistency'). The next two examples (see sections 3.1 and 3.3) were found using the context-dependent cue 'back bencher'. Our examples show that in addition to ethotic argumentation represented as inferential scheme, they contain a number of interesting dialogical ethotic structures that cannot be captured by the argumentation scheme approach alone. We build upon a theoretical framework proposed by Inference Anchoring Theory (IAT; [5]) and annotate data using the Online Visualisation of Arguments (OVA+) tool [8][2] following the Argument Interchange Format standard [6, 12] for argument representation.

Inference Anchoring Theory allows us for linking dialogue structure (right hand side of an IAT diagram; see e.g. Figure 1) with argument structure (left hand side of the IAT diagram).[3] More specifically, on the right hand side we have two types of nodes: locutions which are consecutive moves (utterances) performed in the dialogue; and transitions which represent relations between locutions, i.e. they provide the information about which previous locution a given locution is the response for. For instance, the first *Default Transition* node from the top in Figure 1 shows that the third locution is the response to first and second locutions. Next, on the left hand side we have contents of dialogical structures: propositions are contents of locutions (or occasionally – transitions); and relations between propositions are contents of transitions. We have two types of argument relations between propositions: pro-argument structures which capture giving reasons in support of claims – see *Default Inference* in Figure 1; and con-argument structures which represent giving counterclaims as attacks – see *Default Conflict*.

The actual link between these two types of structures is represented by illocutionary connections (such as asserting or arguing; see e.g. the nodes in the middle of the IAT di-

[2] Available at: `http://ova.arg-tech.org`.

[3] For the same methodology see e.g. the analysis for mediation [9] or moral debate [4, 20].

agram in Figure 1).[4] We say that arguments are anchored through illocutionary connections in dialogue structures. In the annotation scheme used in this study, propositions are linked with locutions via illocutions of asserting, questioning (pure, assertive and rhetorical), challenging (pure, assertive and rhetorical), and popular conceding; while propositional relations are anchored in transitions via illocutions of arguing, agreeing and disagreeing. The rationale behind anchoring an illocutionary connection in a transition between locutions rather than the locution itself is that a type of connection and/or a content can be reconstructed only if we know what the locution is the response to (i.e. we need to know the full transition relation for the reconstruction). For instance, if the proposition "It is not inconsistent" in example (1) is uttered in the isolation, then it can be treated only as an assertion; but if we know that this is the response (related via transition) to something stated earlier, then we understand that it is in conflict and disagreement with something uttered before.[5]

Finally, we represent ethotic structures which express the character of the speaker (e.g. the node "The Attorney-General has ethos" in Figure 1) and allow us to capture support of (*Default Inference*, not present in the examples in this paper) or attacks on (*Default Conflict*) opponent's ethos.

3.1. Circumstantial AH with direct responses

The following example illustrates simple circumstantial *ad hominem* of the type "what you say today is inconsistent with what said yesterday":

(1) a. Captain Crookshank: *The Government are bound, he (the Attorney-General) says, to have this Clause (...) But, when the Attorney-General spoke to this Clause before, he started off by saying that it deals with a form of tax avoidance. I find what he says today hard to reconcile with the normal meaning of tax avoidance, as used in debates on this Bill.*

b. The Attorney-General: *It is not* ***inconsistent****.*

c. Captain Crookshank: *I should have thought it was.*

This example was found using the keyword 'inconsistent' (in bold). In the move (1-a), Crookshrank points out that there is an inconsistency between what the Attorney-General says now and what he said earlier in the House. In (1-b), the Attorney-General directly defends himself by saying that those views are not inconsistent. The move (1-c) is in turn Crookshrank's direct attack on Attorney-General's defence.

The top grey box in Figure 1 represents the CAH technique. The first move of Captain Crookshrank introduces the content of the first premise "The Attorney-General says that the Government are bound to have this Clause" and the second premise stating "The Attorney-General said that it deals with a form of tax avoidance". These two premises are inferentially linked (by means of CAH) with the conclusion stating "I find what he says today hard to reconcile with the normal meaning of tax avoidance".

The second tactics is represented in the bottom grey box Figure 1. The Attorney-General defends himself from the charge of inconsistency by saying "It is not inconsistent". Note that his defence also constitutes an attack on Crookshrank's conclusion of the argument represented in the top grey box. By saying "I should have thought it was", Crookshrank defends his conclusion by at the same time retrying to attack Attorney-General's consistency. To sum up, either of these two moves plays two mutual roles: attack and defense.

[4] The illocutionary connections build upon the concept of illocutionary force introduced in speech act theory, *cf.* [13].

[5] In other words, one can not argue for, agree with, disagree with, if nothing else was uttered beforehand.

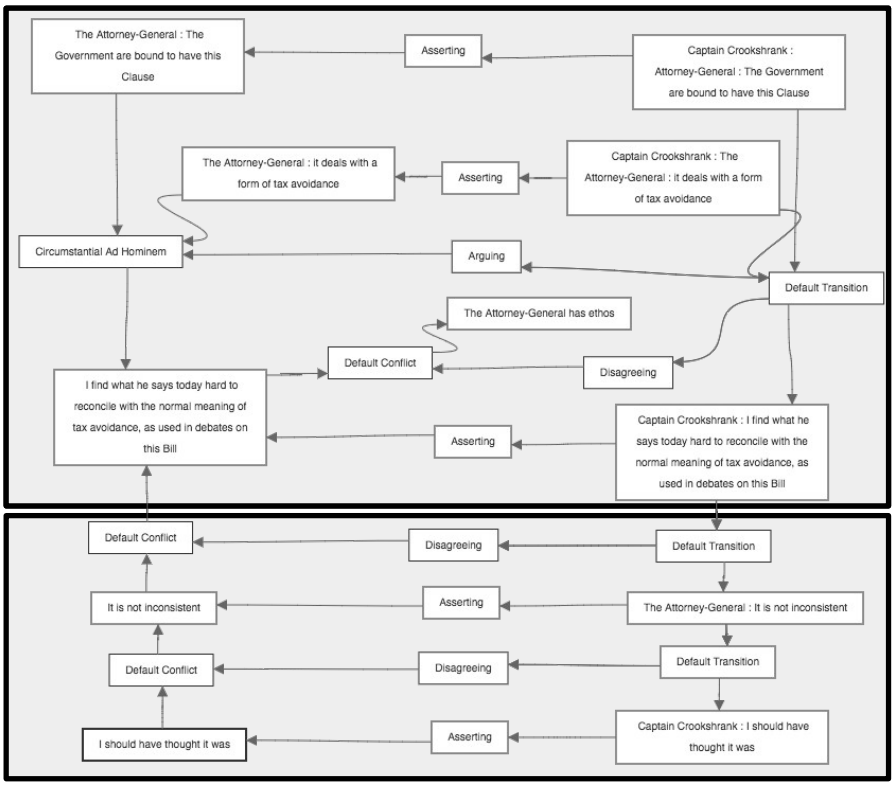


Figure 1. Circumstantial AH (top grey box) with two direct responses (bottom grey box)

3.2. Using Back Benchers to indicate party's inconsistency

The next Hansard excerpt shows more complex ethotic interactions:

(2) a. Mr. Hogg: *(...) I understood that Labour was pretending to be a European party, but Labour* ***Back Benchers*** *are showing that they have an extreme dislike of the European Community. It would be interesting to know which represents the proper view in the Labour party.*

This example is a result of searching the Hansard repository with the use of the keyword 'back bencher'.[6] It illustrates the technique of using back benchers against their own party's ethos.

As shown in Figure 2, Mr. Hogg introduces the content of two premises: "Labour was pretending to be a European party" and "Labour Back Benchers are showing that they have an extreme dislike of the European Community". These premises support the conclusion the content of which is "Labour Party's views are inconsistent". This attack on Labour Party's consistency has a form of CAH as the conclusion of the argument which points to Labour Party's inconsistency. This attack is represented by the *Default Conflict* box which links the content of the conclusion with the ethotic box "Labour Party has ethos".

[6]The term 'back bencher' denotes a Member of Parliament (MP) or a legislator who holds no governmental office and is not a front bench spokesperson in the opposition.

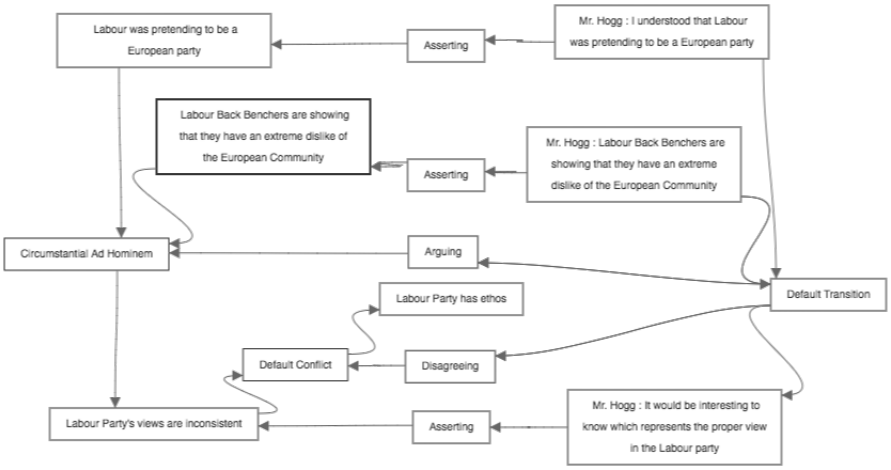


Figure 2. Using Back Benchers against their own party's ethos

3.3. Using a party member to indicate party's inconsistency

Another ethotic interaction is present in the following example:

(3) a. Mr. Darling: *Given that the Tories want to cut 21 billion of investment, it is not surprising that, to get around the gap that that leaves in the health service, they have to emphasise the importance that they attach to increasing private health provision. It is another example of how they are drifting further from the mainstream into the right wing of politics. I do not think that the hon. Member for Grantham and Stamford (Mr. Davies), to whom I am about to give way, believes in that nonsense. He spent 18 years in the wilderness on the Tory* ***Back Benches*** *because he does not believe that. (...)*

b. Mr. Davies: *The right hon. Gentleman affords me rather more seniority than I can claim. (...) Of course we are not against spending money efficiently on health and education. (...)*

Again, the keyword 'back bench' was used to find this example. Figure represents using a party member against his own party's ethos and defence.

The first box in solid line and two next boxes in dotted line illustrate how a single member of the party is used in order to attack the whole party's ethos. In this part of the diagram, there are two different *ad hominem* techniques employed by Mr. Darling in order to attack the conservative party (one of them is circumstantial). The node "Tories have ethos" is firstly attacked using the basic, direct AH technique ("they are drifting further from the mainstream into the right wing of politics") and secondly – by means of CAH ("(Mr Davis) does not believe in that nonsense" being inconsistent with Tories' beliefs).

In response, represented in Figure 3 with the bottom dotted line box, Mr. Davies repels both of the attacks by, firstly, undercutting CAH ("The right hon. Gentleman affords me rather more seniority than I can claim" is undercutting the inference which supports the premise of CAH) and then undercutting direct AH ("we are not against spending money efficiently on health and education" is undercutting the premise of AH).

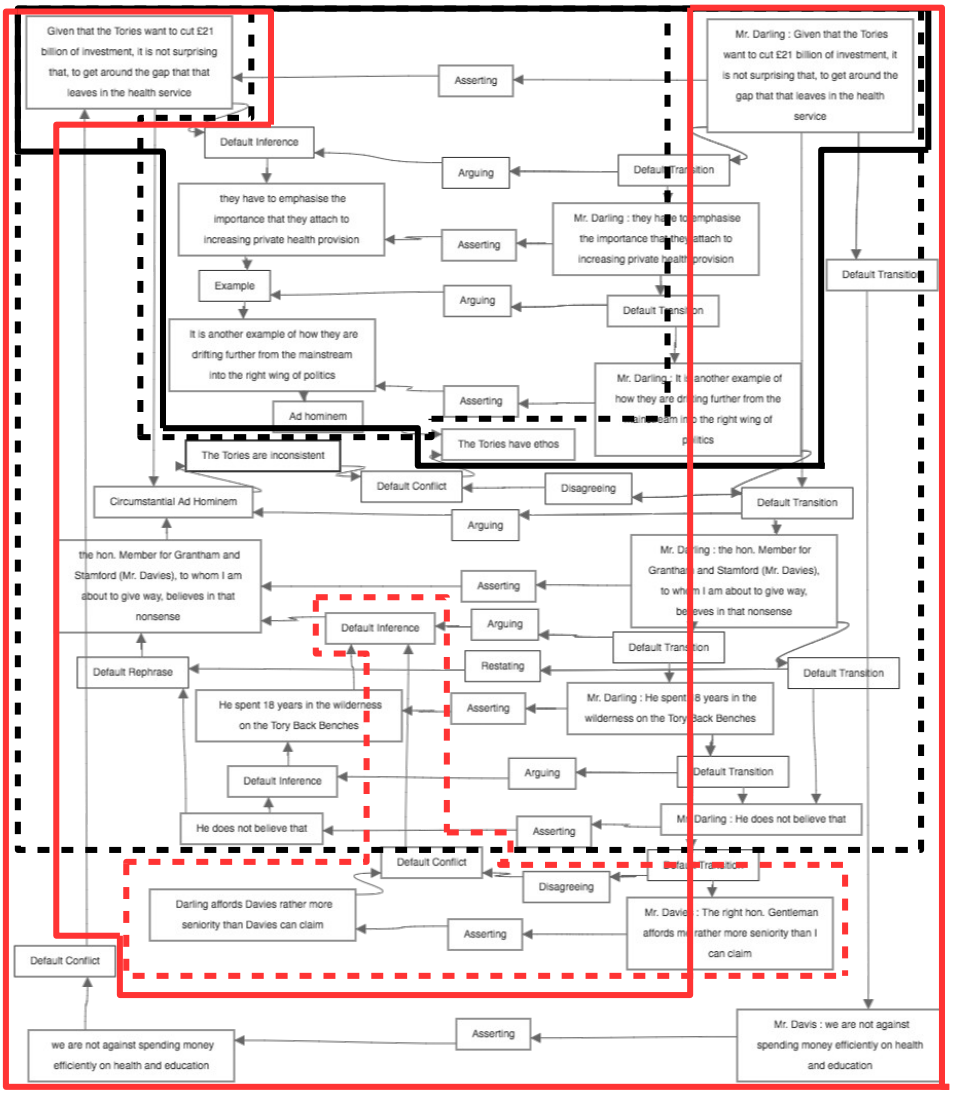


Figure 3. Using a party member against his own party's ethos (first box in solid line and two next boxes in dotted line) and defence (the bottom dotted line box and solid line box)

4. Conclusions and future work

In this paper, we propose: the method for the development of corpora which cope with sparseness of ethotic data; and the case studies of dialogical ethotic structures which show: mutual exchange of shallow attacks on opponents' consistency, attacking party's ethos by pointing to inconsistency between frontbenchers and back benchers, and attacking party's ethos by pointing to inconsistency between the party and its member. In the future we plan to: (i) analyse other types of ethotic arguments (such as appeals to ignorance or appeals to authority) and distinguish particular types of ethos within a given type of argument; (ii) propose the taxonomy of ethotic structures; (iii) extend the list of lexical cues that are markers of various forms of ethos; and specifically (iv) elaborate the

list of linguistic cues for moral ethos (i.e. references to speaker's ethical dispositions) and evaluate whether these cues are efficient in detecting ethos supports and attacks. This work would provide the foundations for future systematic corpus analysis and corpus development, and for future application in ethos mining [7].

Acknowledgements

The work reported in this paper was supported in part by the Polish National Science Centre under grant 2015/18/M/HS1/00620.

References

[1] Aristotle. *On Rhetoric (G. A. Kennedy, Trans.)*. New York: Oxford University Press, 1991.

[2] A. Brinton. Ethotic argument. *History of Philosophy Quarterly*, 3:245–257, 1986.

[3] K. Budzynska. Circularity in ethotic structures. *Synthese*, 190:3185–3207, 2012.

[4] K. Budzynska, M. Janier, C. Reed, and P. Saint-Dizier. Theoretical foundations for illocutionary structure parsing. *Argument and Computation*, 2016.

[5] Katarzyna Budzynska and Chris Reed. Whence inference. Technical report, University of Dundee, 2011.

[6] C. Chesnevar, J. McGinnis, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, and S. Willmott. Towards an argument interchange format. *The Knowledge Engineering Review*, 21(4):293–316, 2006.

[7] R. Duthie, K. Budzynska, and C. Reed. Mining ethos in political debate. In *Proceedings of the Sixth International Conference on Computational Models of Argument (COMMA 2016)*, 2016, accepted for publication.

[8] M. Janier, J. Lawrence, and C. Reed. OVA+: an argument analysis interface. In *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014)*, pages 463–464, 2014.

[9] M. Janier and C. Reed. Towards a theory of close analysis for dispute mediation discourse. *Argumentation*, 0.1007/s10503-015-9386-y, 2015.

[10] M.-F. Moens. Argumentation mining: Where are we now, where do we want to be and how do we get there? In *FIRE '13 Proceedings of the 5th 2013 Forum on Information Retrieval Evaluation*, 2013.

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

[12] I. Rahwan, F. Zablith, and C. Reed. Laying the foundations for a world wide argument web. *Artificial Intelligence*, 171(10-15):897–921, 2007.

[13] J. Searle. *Speech Acts: An essay in the philosophy of language*. Cambridge, 1969.

[14] F.H. van Eemeren, B. Garssen, E.C.W. Krabbe, F.A. Snoeck Henkemans, B. Verheij, and J.H.M. Wagemans. *Handbook of Argumentation Theory*. Springer, 2014.

[15] D. Walton. *Appeal to Expert Opinion : Arguments from Authority*. University Park, Pa., Penn State Press, 1997.

[16] D. Walton. *Ad Hominem Arguments*. Tuscaloosa: University of Alabama Press., 1998.

[17] D. Walton. Case study of the use of a circumstantial ad hominem in political argumentation. *Philosophy & Rhetoric*, 33:101–115, 2000.

[18] D. Walton. Argumentation schemes and historical origins of the circumstantial ad hominem argument. *Argumentation*, 18:359–368, 2004.

[19] D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge: Cambridge University Press., 2008.

[20] O. Yaskorska and M. Janier. Applying Inference Anchoring Theory for the analysis of dialogue structure in debate. In *European Conference on Argumentation (ECA)*, 2015.

Obviously Epistentials are Argumentative Indicators: Evidence from an Italian and English Corpus of Newspaper Articles

Elena MUSI [a,1], Andrea ROCCI [b]

[a] *Columbia University, Center for Computational Learning Systems, New York*
[b] *Università della Svizzera italiana, Istituto di Argomentazione, Linguistica e Semiotica, Lugano*

Abstract. Recent studies have shown that epistentials are good candidates to work as argumentative indicators, useful tool for argumentation mining. In this paper we investigate the argumentative properties of the Italian adverbs *evidentemente* and *ovviamente* and their English counterparts *evidently* and *obviously* in comparable corpora of opinion articles through a multilayer annotation. The results show that *evidentemente* and *evidently* presuppose the presence of premises-conclusion relations where premises are textually adjacent, while *ovviamente* and *obviously* indicate (inter)subjectivity.

Keywords. epistentials, argumentative indicators, newspaper articles

1. Introduction

Epistentials are linguistic constructions which present symultaneously modal and evidential functions. Modals, which express the capacity of thinking that things might be otherwise, and evidentials, which point to the presence of justifications to support assertions [1], refer to moves which are central to argumentative discourse.

Recent studies at the semantics-argumentative interface [19,13,12] have shown that a set of Italian epistentials work as argumentative indicators [25] since they contrain discourse moves both at the level of argumentation structure (e.g. indicators of standpoint, of premises of different types) and argument schemes, the inferential relations linking premises to standpoints. For instance, epistential verbs *dovere* ("must") and *sembrare* ("seem") function as argumentative indicators only in m-performative constructions [15]. and impose different constraints on argument schemes on the basis of their lexical semantics [20,12]. Knowledge about linguistic indicators is highly relevant for the development of computational techonologies for the automatic detection of arguments.

[1]Corresponding Author: Elena Musi, Columbia University, Center for Computational Learning Systems, 475 Riverside Drive 10115, New York, USA; E-mail: em3202@columbia.edu

In this paper we present the results of a semantic-argumentative corpus based analysis of the epistential Italian adverbs *evidentemente* and *ovviamente* and their English counterparts *evidently* and *obviously*. [2]

In sentiment analysis these adverbs have been considered as indicators of degrees of commitment and belief [2,23], while their evidential function has been neglected.

In argumentation theory they have been treated either as indicators of the degree of support that the arguments offer to the conclusion [5] or as indicators of the degree of the speaker's epistemic commitment to the standpoint [24], in the context of the act of advancing a standpoint [25]. In both frameworks, different expressions pertaining to epistemic certainty have been considered roughly equivalent. As Freeman [7, p.12] puts it "adverbial expressions such as 'obviously', 'evidently', 'surely', 'apparently' all have force" but "their cognitive meaning is sufficiently vague as to render their force the only aspect of their meaning worthy of interest. Where their cognitive meaning is clear it is synonymous with *necessarily*". In the same vein, Tseronis [24, p. 54] argues that "there is no difference in the discourse effect that using a modal or an evidential adverb has when qualifying a standpoint".

Our analysis challenges this *negative hypothesis* showing that the two adverbs exhibit both in English and Italian semantic differences. Their properties constrain in different ways the argumentative function of the propositional content they modalize as well as the way they relate to the beliefs and inferences of the arguer and of the critic.

2. Data and Methods

2.1. Data: Corpus

The proposed analysis is corpus based. The use of constructed examples has been limited to grasp the semantic difference between the two adverbs and to test contextual incompatibilities.

We have chosen the text genre of opinion articles: opinion articles express the journalist's argued *prise de position* about a specific issue of public interest [8]. Moreover, evidentials have proved to be particularly frequent in this text genre where sources of information play a key role in the negotiation of stance [3]. As to Italian, our data consist of opinion articles taken from the daily newspapers *La Stampa* and *La Repubblica* (1053 articles, 580.352 words) and relative contributions to the online comment space (49.043 posts, 605.496 words), collected within the Project *From Perception to inference: evidential, argumentative and textual aspects of perception predicates in Italian* (Swiss National Science Foundation grant n. 141350). In this corpus the adverb *evidentemente* occurs 36 times, while the adverb *ovviamente* is more frequent for an overall number of 99 occurrences.

With the aim of obtaining a comparable sample for the analysis of English *evidently* and *obviously* we have randomly selected an equivalent number of occurrences from the subsection of the *SiBol/Port* corpus including articles from *The Guardian* and the *New York Times*, available through the corpus tool *Sketch Engine* [9].

[2] Although the whole paper has been the result of a continuous process of interaction between the two authors, Elena Musi is the main responsible of sections 1-3, while Andrea Rocci of sections 4-5.

2.2. Methods: Annotation Process

The annotation has been conducted using *UAM Corpus Tool* (ver. 3.3e). It has been carried out only by the two authors. The aim of the annotation process was not, in fact, that of developing analytical guidelines which could guarantee a high interannotator agreement, but that of offering a significant sample of analysis to investigate the semantic and argumentative properties of the two adverbs. The empirically unraveled role of the adverbs as argumentative indicators is meant to be *per se* relevant in view of computational applications. The considered layers of analysis are the following:

1. the semantic type of proposition in which the two adverbs occurr, according to Freeman's taxonomy [6]: broadly logically necessary statements, whose truth/falsity is logically determinate (e.g. "One physical object cannot be in two places at the same time") are opposed to indeterminate ot contingent statements. The modal nature of the adverbs excludes their compatibility with broadly logical necessary statements. Among contingent statements *evaluations*, propositions containing an axiological predicate (e.g. "Non mi è piaciuto tutto questo, ovviamente", "I did not like all this, obviously", *La Repubblica*) are distinguished from non evaluative, natural statements. In turn, natural statements include *interpretations* and *descriptions*. The former are "intensional statements which contain a meaningful or nomic connection to the actual world" [6, p.155]. In other words, they assert or presuppose causal or other explanatory relations (e.g. "If you need one of those [power reserve indicators] you are obviously not trying hard enough", source *New York Times*). The latter are extensional statements which encode potentially observable facts in the actual word through perception (e.g. "[...]You notice a shiny object on the ground. It is obviously some sort of metal instrument", source *The Guardian*) or in internal states through introspection (e.g. "What distinguishes these evidently painful personal events is the story they warrant told", source *The Guardian*).
2. whether the proposition is presented as the result of an inference of the author, instead of testimony and direct experience: in the sentence "I sondaggi dicono che i tifosi rossoneri sono in larga maggioranza-non nella totalità, ovviamente, di destra e di centrodestra" ("Surveys say that the Rossoneri fans are to a large majority - not all, of course, right wing and center right wing", source *La Repubblica*) for instance, the proposition cannot be inferred by the speaker since he has direct access to the numerical results of the surveys.
3. whether there are textually recoverable premises and how they are textually manifested (intersententially, intrasententially, intraclausally, multiple manifestations);
4. whether the proposition is presented as inferrable by the addressee: this happens when the premises are textually expressed or when they are easily recoverable from the commonground. In such cases the premises constitute non-subjective evidence and express intersubjectivity as intended by Nuyts [16]. According to the scholar, the dimension of subjectivity does not rely on the quality of the speaker's evidence (mathematically measurable or vague) in support of an epistemic evaluation [10], but rather on the interactive terms of the question whether the evidence is available only to the speaker or more widely known. In particular, "subjectivity involves the speaker's indication that (s)he alone knows (or has ac-

cess to) the evidence and draws conclusions from it" while intersubjectivity 'involves his/her indication that the evidence is known to (or accessible by) a larger group of people who share the same conclusion based on it" [16, p. 393]. In an argumentative perspective the epistential adverbs realize the act of effecting an inference in the addressee [17], who is invited to follow the same inferential path constructed by the speaker;

5. what argument scheme is activated in the inferential step from the premises to the modalized proposition: the adopted taxonomy of argument schemes is that proposed in the *Argumentum Model of Topics* (AMT) framework [18,21]. It is a hierarchical taxonomy based on frame semantic principles where a primary distinction is made among:

   - Intrinsic argument schemes: the state of affairs expressed by the premise and that expressed by the claim belong to a unitary frame, either within the same possible world (e.g. parts and whole) or in causally accessible possible worlds (e.g. material cause).
   - Extrinsic argument schemes: the state of affairs expressed by the premise and that expressed by the claim belong to separate, independent or alternative, frames that are compared or contrasted.
   - Complex argument schemes: they have a pragmatic rather than a semantic-ontological connection with the standpoint. (e.g. argument from authority).

This typology offers several advantages and allows to build feasible guidelines for the annotation of argument schemes [14].

3. Analysis of Annotation Results: *Evidentemente* and *Evidently*

Both the Italian adverb *evidentemente* and the English *evidently* have in our sample mainly scope over propositions of the interpretative type (respectively 24/36 and 21/36 occurrences). The compatibility of the two adverbs with descriptive propositions could allow thinking that they express direct evidentiality [26], the most suitable evidential mode to present state of affairs as factual. However, neither of the two adverbs can have scope on propositions encoding a state of affairs that the speaker has personally witnessed or experienced:

Context: speaking fo someone well known to be a miser...
"*Evidentemente mi ha chiesto se potevo pagare il conto"
"*Evidently he asked me if I could pay the bill"

The two adverbs can have scope over descriptions since they present the embedded proposition as factual "according to the information that we have" [22, p. 161] at the moment of utterance.

Evidentemente functions as a strictly inferential evidential strategy: in every occurrence the proposition is presented as inferred by the speaker. Most of the times (28) there is also inferrability by the addressee, except a few cases from the online readers' comments (8) in which the inferences are about the addressee, who therefore knows from direct experience if what the speaker has attributed to him corresponds to reality or not:

"Evidentemente vivi all'estero, perché io vivo dove hanno inventato la brugola e so che qui producono tutto fuorché roba che i cinesi sono in grado di copiare" (source *La Repubblica*)
"Evidently you live in abroad, because I live where they invented the Allen wrench, and, fortunately, I know that here they produce anything but stuff that the Chinese are capable of copying"

Evidently differs from its Italian counterpart in that it can also express hearsay, though marginally (2/36 occurrences). When having an inferential value, *evidently*, like *evidentemente*, is accompanied by clearly recoverable textual premises. These premises can be positioned inter or intra sententially and before or after the proposition over which the adverbs have scope; when located in a different sentence, the premises are always attested in adjacent position. This tendency distinguishes *evidently* and *evidentemente* from other inferential evidential strategies which do not seem to impose any constrain on the premises position.

Both for *evidentemente* and *evidently* it has always been possible to single out specific argument schemes, distributed as follows:

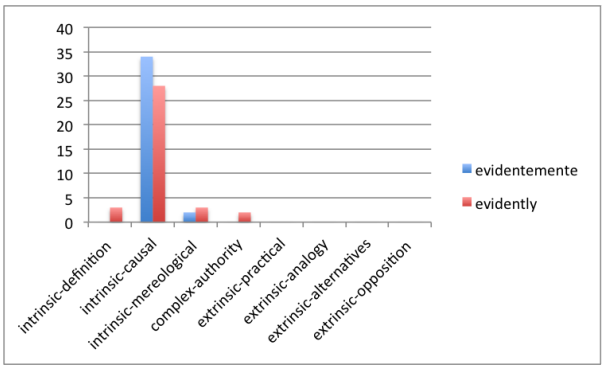


Figure 1. Distribution of argument schemes with *evidentemente* and *evidently*

As shown in Figure 1, with the exclusion of complex argumentation schemes expressed by *evidently* when showing a hearsay value, both adverbs select intrinsic argument schemes for the majority of the causal type. Further zooming into the causal argument schemes it has emerged a clear prevalence of the efficient cause, and, within it, of the direction from effect to cause:

"Better to focus on the picture, many have evidently concluded, hence the torrent of pictures of Mrs Clinton's new style." (source *The Guardian*)

More specifically, it seems that the two adverbs are routinely used to signal the inference from human observable behaviors to not observable intentions, thoughts or values, as in the preceding example.

4. Analysis of Annotation Results: *Obviously* and *Ovviamente*

As a premise, it has to be pointed out that the English adverb *obviously* semantically differs from its morphological equivalent *ovviamente* since it can also work as a manner adverb with the meaning "in a clear perceptible way" [22]:

> "Forse sei la prima persona arrivata lì. Mentre ci pensi noti un oggetto luminoso per terra. Si tratta *ovviamente/chiaramente di una sorta di strumento metallico"
> "Perhaps you are the first person there. While thinking about it you notice a shiny object on the ground. It is obviously some sort of metal instrument" (source *The Guardian*).

This meaning of *obviously* is accessible in the occurrences in which the adverb has scope restricted on an adjective or it is modified by a degree word or by another manner verb [22]. Our analysis has concentrated on the occurrences in which *obviously* functions as a sentence adverb. Differently from *evidentemente* and *evidently*, both *ovviamente* and *obviously* have preferably scope over propositions of the descriptive type (respectively 43/99 and 55/99). A possible explanation for this bevahior relies in the peculiar function of these adverbs at the level of speaker-hearer orientation. As underlined by Barbaresi [11, p. 21-22], *obviously* has the function of presenting the embedded proposition as inferable by the hearer. The same applies to the Italian counterpart *ovviamente* which would be odd in contexts where an inference on the part of the hearer is hardly conceivable:

> **Context** Boy comes home at an unlikely hour, walks into the kitchen and opens the fridge. Mom says:
> "?Obviously, you think this house is an hotel" "?Ovviamente pensi che questa casa sia un albergo"

In the preceding example *ovviamente* would suggest that what the boy thinks is not something that the mother inferred from his behavior but rather something that they both inferred from what is usually to be expected in such a situation. Corpus analysis has, in fact, revealed that in every occurrence of *ovviamente* and *obviously* the modalized proposition is presented as inferrable by the addressee, and frequently not inferred by the speaker (who knows it from direct experience). In other words, the premises from which the inference is drawn are always intersubjective, namely accessible to the hearer [15]. As an effect, *ovviamente* works as a strategy of objectification [4] promoting the hearer's recognition of the inferred proposition as factual.

The annotation results of the layer concerning premises suggest that the evidential value of the adverb is less prominent than in the couple *evidentemente* and *evidently*: the premises are in circa half of the occurrences not textually recoverable; when textually expressed they tend not to be textually adjacent. Turning to the analysis of argument schemes where premises are textually-contextually recoverable, it has emerged that *obviously* and *ovviamente* are compatible with both intrinsic and extrinsic argument schemes, though having a preference for the former.

5. Conclusion

We have investigated the semantics and argumentative properties of the Italian epistential adverbs *evidentemente/evidently* and their English counterparts *obviously/ovviamente* through a systematic multi-layer annotation in comparable corpora of newspaper articles. The main aim was that of unraveling their role as argumentative indicators in view of computational applications.

As a preliminary result, it has emerged that the Italian adverbs have more specific evidential functions compared to their English counterparts: differently from *evidentemente*, *evidently* can signal not only inferential evidentiality, but also hearsay (although rarely); *obviously* can be used as a manner adverb, while *ovviamente* has always scope on a proposition.

Focusing on the inferential occurrences, the analysis has shown that *evidentemente* and *evidently* function as indicators of premises-conclusion relations imposing constraints on the position of premises – which are intrasentential or adjacent – and the argument scheme linking them-which is of the intrinsic type and generally causal. *Ovviamente* and *obviously* show a preference for intrinsic argument schemes too, but do not point to specific premises-conclusion configurations, while they systematically signal the presence of (inter)subjectivity. The two couples of adverbs seem to play quite distinct rhetorical roles in editorials: *ovviamente /obviously* are used by the knowledgeable journalist to manage the expectations of a reader that is presented as intelligent and aligned with the writer; *evidentemente* and *evidently* are used by the journalist to point to the hidden but necessarily true causes of observable facts. A typical pattern is the one where the intentions or beliefs behind an action of a public figure or group are uncovered.

The conducted annotation on the corpus only concerns micro-argumentation. The informal examination of the broader argumentative structures suggests that the propositions on which the adverbs have scope never appear to be modalizing the main standpoint of the opinion article. In particular, *ovviamente* and *obviously* often modalize propositions that are fairly peripheral in the overall argument, while the hidden causes uncovered in propositions modalized by *evidentemente* and *evidently* function as premises to support a higher-level evaluative standpoint.

As to future work, we plan to extend this kind of analysis to the whole class of evidentials in order to shed light on the argumentative patterns they signal at the level of argumentation structure, argument schemes and rhetorical strategies. One of the hypothesis we would like to test is the tendency of epistentials which indicate high degree of certitude and are lexically related to perception to work as indicators of intrinsic argument schemes [14].

References

[1] Lloyd B Anderson. Evidentials, paths of change, and mental maps: typologically regular asymmetries. 1986.

[2] Farah Benamara, Carmine Cesarano, Antonio Picariello, Diego Reforgiato Recupero, and Venkatramana S Subrahmanian. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *ICWSM*. Citeseer, 2007.

[3] Caroline Clark. Evidence of evidentiality in the quality press 1993 and 2005. *Corpora*, 5(2):139–160, 2010.

[4] Musi Elena. Strategies of objectifications in opinion articles: the case of evidentials. In *Proceeding of the OSSA Conference, 18-21 May 2016*, to appear.

[5] James B Freeman. *Dialectics and the macrostructure of arguments: A theory of argument structure*, volume 10. Walter de Gruyter, 1991.

[6] James B Freeman. What types of statements are there? *Argumentation*, 14(2):135–157, 2000.

[7] James B Freeman. *Argument Structure: Representation and Theory*, volume 18. Springer Science & Business Media, 2011.

[8] Gilles GAUTHIER. La structure et les fondements de l'argumentation éditoriale. *Les Cahiers du journalisme*, 17:322–342, 2007.

[9] Adam Kilgarriff, Pavel Rychly, Pavel Smrz, and David Tugwell. Itri-04-08 the sketch engine. *Information Technology*, 105:116, 2004.

[10] John Lyons. Semantics.(2 vols.) cambridge, 1977.

[11] Lavinia Merlini Barbaresi. "obviously" and "certainly": two different functions in argumentative discourse. *Folia linguistica*, 21(1):3–24, 1987.

[12] Johanna Miecznikowski and Elena Musi. Verbs of appearance and argument schemes: Italian sembrare as an argumentative indicator. In *Reflections on Theoretical Issues in Argumentation Theory*, pages 259–278. Springer, 2015.

[13] Elena Musi. Evidential modals at the semantic-argumentative interface: appearance verbs as indicators of defeasible argumentation. *Informal Logic*, 34(4):417–442, 2014.

[14] Elena Musi, Ghosh Debanjan, and Smaranda Muresan. Towards feasible guidelines for the annotation of argument schemes. In *Proceedings of the 3rd Workshop on Argumentation Mining*, to appear.

[15] Jan Nuyts. *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective*, volume 5. John Benjamins Publishing, 2001.

[16] Jan Nuyts. Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of pragmatics*, 33(3):383–400, 2001.

[17] Robert Pinto. *Argument, inference and dialectic: Collected papers on informal logic*, volume 4. Springer Science & Business Media, 2001.

[18] Eddo Rigotti and Sara Greco Morasso. Comparing the argumentum model of topics to other contemporary approaches to argument schemes: the procedural and material components. *Argumentation*, 24(4):489–512, 2010.

[19] Andrea Rocci. Modality and argumentative discourse relations: a study of the italian necessity modal dovere. *Journal of Pragmatics*, 44(15):2129–2149, 2012.

[20] Andrea Rocci. Modal conversational backgrounds and evidential bases in predictions: the view from the italian modals. *Time: Language, Cognition and Reality*, pages 128–152, 2013.

[21] Andrea Rocci. *Modality in Argumentation*. Springer, to appear.

[22] Anne-Marie Simon-Vandenbergen and Karin Aijmer. *The semantic field of modal certainty: A corpus-based study of English adverbs*, volume 56. Walter de Gruyter, 2007.

[23] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

[24] Assimakis Tseronis et al. *Qualifying standpoints. Stance adverbs as a presentational device for managing the burden of proof.* LOT, Netherlands Graduate School of Linguistics, Utrecht, 2009.

[25] Frans H Van Eemeren, Peter Houtlosser, and AF Snoeck Henkemans. *Argumentative indicators in discourse: A pragma-dialectical study*, volume 12. Springer Science & Business Media, 2007.

[26] Thomas Willett. A cross-linguistic survey of the grammaticization of evidentiality. *Studies in language*, 12(1):51–97, 1988.

Recognising enthymemes in real-world texts: a feasibility study

Olesya RAZUVAYEVSKAYA and Simone TEUFEL
Computer Laboratory,
University of Cambridge,
15 JJ Thomson Avenue, Cambridge, UK
e-mail: *or264@cam.ac.uk, sht25@cam.ac.uk*

Abstract. We present a feasibility study for the task of finding and expanding enthymemes, i.e, arguments with missing propositions, in real-world texts. We split the task into three subtasks: 1. finding the beginning and the end of the text span describing a minimal argument block, 2. deciding whether that span really represents an *enthymematic* minimal argument block, and 3. describing the missing premise in natural language. We argue that an objective ground truth for these tasks must be found before we can address automatic annotation. In our opinion, such a ground truth can only come from agreement on independent human annotation. In other words, we require that two or more annotators agree on the existence and textual span of an enthymematic minimal argument block, and on the actual paraphrase of the missing premise. We present a case study using the two authors of this paper as annotators, where we test three cue phrases for their suitability to this task, *because, therefore* and *let alone*. We find that minimal argument blocks centred around the cue phrase *let alone* are of a particularly high quality. We also discuss pragmatic effects of *let alone* and how they relate to argumentation theory.

Keywords. argumentation, enthymemes, annotation experiment, pragmatics

1. Introduction

In this paper, we investigate the feasibility of finding arguments with missing premises or conclusions, also referred to as *enthymemes* [1], in unrestricted texts.

Many of the arguments expressed in a form of natural language are incomplete [6]. The majority of implicit propositions represent trivial facts that the speaker tends to avoid to say in order not to bore the listener [7]. This complies with Grice's Maxim of Quantity - *Do not make your contribution more informative than is required* [8]. Sometimes, the speakers also reduce the amount of required information in order to avoid potentially fallacious propositions [3]. Jackson and Jacobs [9] notice that enthymemes also play an important role in maximizing a listener's agreement, because additional information always increases the possibility of disagreement.

Automatic reconstruction of enthymemes is an interesting task from a text understanding perspective. As enthymemes are known to establish textual entailment relations between two propositions [10], the ability to reconstruct missing premises would provide a clear demonstration that the argument was understood. For the systems that aim at verifying the validity of arguments, the task of enthymeme reconstruction is important from a practical viewpoint, because such systems need to have access to the omitted parts as well as the explicitly stated ones. Moreover, universally quantified premises express general truths about our world, and could therefore be theoretically used as a source of information that is objective enough to be put into general knowledge databases.

All types of arguments can be expressed in a truncated form, but truncated syllogisms are the most studied types of these [2]. Standard-form syllogisms are convenient for the analysis of en-

[1] *Enthymemes* were first defined by Aristotle [1] as a standard-form syllogisms with one missing proposition. However, in the scope of modern usage of this term in argumentation theory [2] [3] [4] [5], it is widely accepted to refer to any type of an argument where one of the principal inferences is missing as enthymematic.

thymemes, because such arguments consist of only three terms and satisfy a set of strict requirements [11]. Therefore, given any two syllogistic propositions, the goal of restoring the missing premise or conclusion can be achieved by following a set of deterministic steps. For instance, consider the enthymeme below:

Pugs are mammals, *[minor premise]*
because they are dogs. *[conclusion]*

As can be inferred from the formal definition of a syllogism, this is a categorical syllogism with the *minor premise* and the *conclusion* expressed. There exists a general procedure [2] that allows us to conclude that the missing proposition is:

All dogs are mammals. [major premise, universally quantified]

Syllogisms allow us a straight-forward and objective definition of the missing proposition. Enthymematic syllogisms already contain all three terms required for the reconstruction, and no external information is necessary.

However, it is almost impossible to find well-formed standard syllogisms in everyday language [2]. Even once spotted, the transformation of such a naturally stated argument into the syllogistic form requires much effort and general knowledge. It therefore does not seem feasible to automate this step to collect a large corpus of well-formed syllogisms in this way. Additionally, many of the naturally occurring arguments cannot be translated into syllogisms because they are not logically valid, but only *defeasible*, i.e. some exceptions to their application exist. It therefore does not seem feasible to use real-world arguments directly as a source of well-formed syllogisms.

The argumentation theory literature is generally pessimistic about the feasibility of insertion of missing premises or conclusions: it has been called a challenging and subjective task even for human experts, often resulting in a wrong interpretation of the arguments [12] [13] [14]. Hitchcock [14] mentions two problems in particular associated with enthymemes, a) the difficulty of distinguishing enthymemes from deductively valid arguments and from the arguments that are to be rejected (”demarcation problem”) and b) the difficulty of evaluating whether the inferred expression was the one that was originally intended. Therefore, successful automation of enthymeme reconstruction is not possible without first establishing the ground truth about the premise insertion goal. We plan to investigate the feasibility of several subtasks associated with this task:

- detection of the beginning and the end of the text span describing a minimal argument block;
- agreement on whether that span really represents an enthymematic minimal argument block;
- agreement on the missing premise explicitly stated in natural language.

Ground truth for these tasks can only be established by means of human annotation, because the interpretation of enthymemes is known to be inherently subjective. If two or more annotators, working independently, can agree on the existence and identity of the missing premise, we can take it as proof that the enthymeme objectively ”exists”. This turns the problem of human subjectivity of enthymeme reconstruction into an empirical question – it can now be answered simply by measuring agreement of annotators.

We believe that the first step towards making the task of premise insertion objective is to detect *enthymematic minimal argumentation blocks* (EMAP for short). We define them to be text pieces containing a minimal enthymematic argument that can be interpreted without additional context, in particular, without general knowledge. This will make the task of inserting missing premises more objective. In what follows, we are developing a corpus of minimal argumentation blocks which are enthymematic, using an input stream of unrestricted text. We will assume that most of these arguments are defeasible. We need to employ an automatic procedure to find the text spans expressing the arguments; in particular, we need to avoid including non-relevant material in these text spans.

Under our scheme, annotators have to decide for each candidate EMAP, whether the given text contains a stand-alone argument. One of the possible questions that can arise is: *What if these stand-alone arguments are not really enthymematic, i.e., what if they already contain all of the required propositions?* We make the assumption, also made by Dijk and Kintcsh [15], that we can always

insert an implicit premise between two propositions, for the pragmatic reasons already sketched out above – humans tend not to spell out links between statements. Therefore, for any EMAP, there should always be the logical ”space” to insert a new statement.

2. Discourse markers for enthymeme detection

The result of the human annotation studies will be a corpus of EMAPs. To conduct these experiments, we first need to build an initial corpus of candidate EMAPs for annotation. The use of discourse markers for automated recognition of arguments is standard practice [16] [17] [18] [19]. By using argumentative discourse markers, i.e., explicit signals used by the speaker, we can assume that the argumentative moves are indeed present.

However, not all of the discourse markers are equally suitable for our goal of building a corpus of EMAPs. Plausible discourse markers should have a high likelihood of indicating an EMAP in their context. We will refer to this property as the *locality* of a discourse marker. The relevant context around the discourse marker which forms part of the candidate EMAP should ideally be as small as possible.

We define three acceptable locality ranges. Here, we refer to the sentence where the discourse marker is localized as a *target sentence*.

- *1-span locality (complete locality)* - at least two propositions, the premise and the conclusion, are usually stored in the target sentence itself
- *2-span locality*
  - *backward-looking locality* - the conclusion and its preceding premise are stored in the following span:
    [1 sentence before the target sentence]+[target sentence]
  - *forward-looking locality* - the premise and its preceding conclusion are stored in the following span:
    [target sentence]+[1 sentence after the target sentence]
- *3-span locality* - the premise can be stored both before and after the conclusion, therefore, we need a larger span to make sure that both propositions are captured:
  [1 sentence before the target sentence]+[target sentence]+[1 sentence after the target sentence]

Our expectation is that smaller locality ranges should make objective enthymeme detection and reconstruction easier. However, we also predict that successful reconstruction of an enthymeme depends not only on the locality. Consider the following two arguments:

A must be true, because A has never observed to be false
A must be true, because X has told so

While the same *because* discourse marker is used in both cases, these enthymematic arguments are based on entirely different reasoning types. The first argument is based on the lack of evidence about any controversial case, whereas the second argument is based on the reliability of the claim's source. These arguments also correspond to different argument schemes, *argument from expert opinion* and *argument from ignorance*; therefore, different reconstruction techniques must be applied to these enthymemes. In order to apply these techniques, discourse markers must uniquely identify their associated argument schemes. We further refer to this characteristic of a discourse marker as *scheme specificity*. With respect to scheme specificity, *because* can be used with almost any argument scheme; therefore, people are likely to disagree on the missing inference type even more than with a scheme-specific discourse marker.

3. Methodology and Results

We report an experimental human annotation study that supports the assumption about the importance of both the locality and the scheme specificity of discourse markers for reaching good agree-

ment between annotators on the inserted inference. We performed a case study using two annotators, the authors of this paper, where we test three discourse markers for their suitability to this task: *because* and *therefore*, both of which are scheme-non-specific, and a scheme-specific *let alone*.

3.1. Annotation data

100 random sentences containing each discourse marker (300 in total) of the average length of 25 tokens were extracted from the British National Corpus (BNC).

Let alone constructions were limited to the target sentence only (The assumption is that both the premise and the conclusion are usually located inside a 1-span text). For each *therefore* cue phrase, we extracted the target sentence along with the immediately preceding one, because this discourse marker is of a 2-span type (The assumption is that the preceding sentence may contain the premise for the conclusion). For each *because* marker, target sentences were extracted with one directly preceding and one directly following sentence (The assumption is that both the preceding and following sentences may contain the conclusion for the premise).

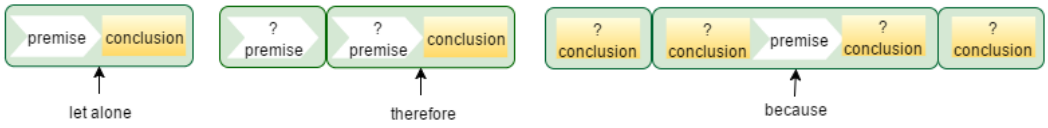


Figure 1. Spans occupied by *let alone, therefore,* and *because* EMAPs

3.1.1. Boundary detection task

Two annotators independently decided whether there was enough context to detect an EMAP. For the cases where they could detect an EMAP, they were also asked to indicate the EMAP-irrelevant information. This would allow us to study the locality of the discourse marker.

Table 1 shows the distribution of the total number of positive choices, i.e., the cases where annotators agreed on the presence of an EMAP, per marker. As can be seen, the number of positive choices was high for *let alone* (183/200), in contrast to *therefore* (67/200) and *because* (83/200).

The inter-rater agreement was calculated by means of Cohen's *Kappa* statistics [20].

As far as the locality property is concerned, for 2-span *therefore*, the annotators only specified EMAP-irrelevant information (17/67 cases) to the left of a discourse marker. This demonstrates that our assumption that preceding sentence often contains a premise for the conclusion was correct. For 3-span *because*, the amount of EMAP- irrelevant information was higher (25/83 to the left and 30/83 to the right of a discourse marker). This shows that the conclusion for the premise is often located either to the left or to the right of the target sentence. For 1-span *let alone*, there were only 2/183 cases where either of annotators specified EMAP-irrelevant information to the left of a discourse marker.

Table 1. Inter-annotator agreement results for the boundary detection task

Marker	**Number of positive choices**	**Kappa-statistics**	**Number of times left boundary was moved**	**Number of times right boundary was moved**
Therefore	67	K=0.451 (k=2, n=2, N=100)	17	0
Because	83	K=0.577 (k=2, n=2, N=100)	25	30
Let alone	183	K=0.729 (k=2, n=2, N=100)	2	0

3.2. Experimental Study on Proposition Insertion Tasks

Looking at the desired properties of discourse markers for the task of EMAP detection, we can see, based on the enthymeme detection task (Section 3.1.1), that *let alone* is always concentrated in a single sentence, i.e., its locality range is small. We will further demonstrate that it also has the other desired property, scheme-specificity.

3.2.1. Linguistic Background on let alone

Let alone represents an unusual linguistic phenomenon. According to [21], *let alone* can be syntactically treated as a coordinating conjunction, where the interpretation of the second part of the construction depends on information provided in its first part. Consider the following examples:

*This task is difficult for an **adult**, let alone a **child**.*
*The baby cannot **sit** yet, let alone **walk**.*

The phrases linked by *let alone* display parallelism with respect to grammatical functions; e.g. direct object (*adult* and *child*) or verbal complement to auxiliary (*sit* and *walk*). In the terminology of [22], the stressed elements in the first and second part of the sentence are called *correlate* and *remnant* respectively. *let alone* is similar to comparatives in this respect.

Unlike comparatives, *let alone* is frequently treated as a negative polarity construction [21]. However, [23] refined this treatment by distinguishing the following three main cases: explicit negative (a), implicit negative (b) explicit positive (c):

(a) *He cannot solve simple, let alone starred tasks.*
(b) *The task is too difficult to be understood, let alone solved.*
(c) *He could find the exact answer, let alone an approximate solution.*

From a pragmatic point of view, [24] observe that the remnant is usually more relevant to the context, and the correlate is in some respect "more likely". The pragmatic negation of *let alone* results in the negation of the remnant by first negating the more likely correlate.

[22] observes that some form of scaling information is always present in *let alone* sentences, and that a pragmatic entailment relation holds between the remnant and the correlate. The hearer can pragmatically infer, based on common knowledge [8, 25], that these stressed elements are ordered based on some scale. This hidden scalar reasoning is highly relevant to our task of finding arguments with hidden premises.

We argue that *let alone* sentences are by definition enthymemes, and that the missing scalar relation is an important part of the missing proposition.

3.3. Argument Scheme for let alone *Sentences*

The pragmatic effects observed in the *let alone* construction have an obvious similarity with *a fortiori* arguments, which were first mentioned in Aristotle's Rhetoric [1]. The main principle behind *a fortiori* arguments is that if there is a case where a certain quality is more likely to exist than in some other, less likely, but that the quality does not even exist in this case, it certainly cannot exist in the less likely case. The speakers therefore use the more likely case to reject the less likely one. [26] designed the following argumentation scheme for such arguments:

If even X does not have property P and it is a less likely case that Y has property P than that X has property P, then Y does not have P
(Even) X does not have P
(Therefore) Y does not have P

Let alone seems to be one of the signals of such *a fortiori* inferences in natural language. We extend a scheme suggested by Kienpointer to the scheme consisting of four terms. As an additional term, we introduce is the scaling relation holding between X and Y. In the next section, we will present such scaling relations.

3.4. Annotation Scheme for "let alone" scale

We categorized the scaling relations based on about 250 cases of *let alone* sentences observed, resulting in the scheme in Table 2.

The categories are as follows.

1. Smaller than: The scale concerned is the one with standard measurements or cardinalities of the remnant and the correlate.

Table 2. Annotation scheme for *let alone* scales

	Relation type	Example
1	Smaller than	*You wouldn't make to New York, let alone the West Coast*
2	Part of	*This does not apply to Germany, let alone all of Europe.*
3	Precondition for	*Your talent isn't enough to participate, let alone win.*
4	Other lexical entailment	*He doesn't even sleep, let alone snore.*
5	Earlier date	*They didn't have electricity in 1923, let alone 1909.*
6	Additional constraint	*You don't know what a middle-aged person feels like, let alone a middle-aged prince.*
7	Additional referent	*The company does not even insure their employees, let alone their families.*
8	Cumulative/independent	*This Easter-egg packaging does not even protect its contents, let alone have anything to do with Easter.*
9 10 11	More extreme case than; Easier than; Less likely than	*They refused to refer to Kursk, let alone Moscow.* *I could not solve the first, let alone the last tasks.* *I have not even seen Mary, let alone Rose there.*

2. Part of: The relationship between the referents in remnant and correlate is that of superset.

3. Precondition: The action in the remnant necessarily requires an action in the correlate having taken place earlier (In example 3 from Table 2, *participating* is a precondition for *winning*).

4. Other lexical entailment: Covers the larger class of actions where the proposition expressed in the remnant lexically entails the proposition expressed in the correlate. There can be temporal overlap between the two propositions (In example 4 from Table 2, *snoring* entails *sleeping*.)

5. Earlier date: A special case of *Smaller than*: if a state first becomes true at time t_0, it by definition does not hold at any earlier time $t < t_0$. It can only be applied to irreversible events.

6. Additional constraint: In his category, a number of semantic predicates (constraints) applied to the correlate are compared with that same predicates plus additional predicates in the remnant.

7. Additional referents: Covers cases where the set of referents in the remnant logically includes the referent from the correlate.

8. Cumulative/independent: Here, the remnant of *let alone* is not directly comparable to the correlate, but becomes interpretable if we read it as an additive constraint.

The three categories *Easier than*, *Less likely than*, and *More extreme case than* are fallback options, which are only to be used if none of the more specific categories applies. A special rule in the annotation is that annotators should always choose the most specific case.

There is a clear connection between the scale class and the linguistic form of premise that can be constructed from that class:

1 X is smaller than Y
2 X is a part of Y
3 X is a precondition for Y
4 Y lexically entails X
5 Y is earlier than X
6 Y poses additional constraint on X
7 Y in addition contains X
8 X and Y are cumulatively more important than X
9 Y is a more extreme case than X
10 X is easier than Y
11 Y is less likely than X

Annotators were asked to extract the appropriate parts from the given sentence and fill them into the template of their chosen category. For instance, for the first example in Table 2 they were expected to write:

The distance to New York **is smaller than** *the distance to the West coast.*

3.5. Enthymeme Reconstruction Task

In 3.1.1, *let alone* contexts were found to have the highest positive agreement and the only substantial agreement regarding the presence of an EMAP. We next conducted more fine-grained annotation tasks only on those *let alone* sentences that were positively classified by both annotators as enthymematic (90 cases).

In this experiment, the annotators performed two tasks:

- They classified EMAPs according to the annotation scheme from Table 2 in order to describe the scale of *let alone* presented.
- They provided a paraphrase of the missing premise by composing a new sentence using material extracted from the given sentence together with the template given in Section 3.4.

3.5.1. Results of Scale Classification Task

The class distribution per category presented in Table 3 is skewed.

The most frequently chosen category was the "precondition for" relation - 61 (33.9% of cases). The "additional constraint" and "additional referent" classes were also frequent (12.7% and 10% respectively). The fact that the fallback categories were chosen relatively frequently suggests that better formulation of the main categories is needed for future experiments.

Table 3. Class distribution (absolute and relative) per relation category

Relation type	**Number of choices**
Smaller than	9 (5%)
Part of	8 (4.4%)
Precondition for	61 (33.9%)
Other lexical entailment with	11 (6%)
Earlier date	5 (2.7%)
Additional constraint on	23 (12.7%)
Additional referent	18 (10%)
Cumulative/independent	9 (5%)
Less likely than	9 (5%)
More extreme case than	12 (6.7%)
Easier than	15 (8.3%)

The agreement between annotators was K=0.75 (k=2, n=2, N=90). This indicates that annotators were overall well able to distinguish relation categories (scales). The small numbers prevent us from analysing problematic categories systematically, but we noticed that *earlier date* and *precondition for* seem to be relatively harder for annotators to distinguish reliably.

3.5.2. Results of Premise Insertion Task

In order to evaluate the annotators' agreement of inserted premises, we considered the 71 cases where the annotators agreed on the chosen scale category. Of those cases, one of the annotators judged 69 premises (97%) to be identical or near-identical. The first disagreement case was:

> *Few adherents to the new classical macroeconomics trouble even to question it, let alone provide an analytical basis to justify it.*

While both annotators have specified that "questioning" is a precondition for "providing an analytical basis to justify", they specified partially different propositions:

(a) *Questioning something is a precondition for providing an analytical basis to justify something.*
(b) *Questioning classical macroeconomics is a precondition for providing an analytical basis to justify classical macroeconomics.*

The second EMAP on which the annotators disagreed was:

It's not exactly as if Wimbledon are an English force – let alone European.

Here, both of the premises were still nearly the same, but the annotators disagreed on the ellipsis resolution:

(a) *For Wimbledon, to be an English force is easier than to be a European force.*
(b) *To be an English force is easier than to be a European force.*

4. General Discussion and Conclusion

In this work, we presented our first experiments on whether it is possible to automatically find enthymemes in unrestricted texts. The results support the hypothesis that signals, such as discourse markers of particular types, can be used as initial objective indicators of argumentative relations between two minimal units of text. For the case of *let alone*, we have observed both significant inter-annotator agreement and high positive agreements on the decision on whether the text units are argumentative. For the cases where annotators agreed, we have further investigated whether the annotators also agree on the exact type of pragmatic relation that holds between two propositions. The results show that although there exist some types of relations that are not very distinguishable from each other, the annotators are able to achieve significant agreement on this task. The fact that the exact premises specified by the annotators are identical in the overwhelming number of cases, suggests that *let alone* constructions are very good candidates for future collection, investigation and reconstruction of enthymemes. It also means that all the information needed to reconstruct an enthymeme could indeed be located in the local context of the sentence itself. As a result, human annotators can perform this task with apparent ease.

In future work it is essential to perform this task on a larger scale, with independent annotators and a larger dataset. Another aspect worth mentioning is that the amount of surrounding context for *therefore* and *because* EMAPs was larger than for *let alone* ones, which can be plausibly assumed to decrease the agreement probability. In the future, it important to identify more precise argumentative boundaries of EMAPs before they are presented to the annotators. It is also essential to describe all types of relations that have demonstrated low agreement rate in a way that will make them more intuitive for future annotators.

Human annotation is only the first step towards automatic reconstruction of enthymemes. In particular, automatic resolution of gapping is a challenging task. Syntactic information is often not enough for intelligent reconstruction, and the semantic information is not available automatically [27] [28]. But it is important to have the ground truth on this subjective task showing that it is achievable by people.

Another insight that this study suggests is the fact that pragmatic information, which is not traditionally used in argument mining, can be beneficially incorporated as important features.

References

[1] E. M. Cope and J. E. Sandys, eds., *Aristotle: Rhetoric*, vol. 1. Cambridge University Press, 2010. Cambridge Books Online.

[2] I. Copi and C. Cohen, *Introduction to Logic*. Maxwell Macmillan international editions, Macmillan, 1990.

[3] D. Walton and C. A. Reed, "Argumentation schemes and enthymemes," *Synthese*, vol. 145, no. 3, pp. 339–370, 2005.

[4] E. Black and A. Hunter, *Using enthymemes in an inquiry dialogue system*, vol. 1, pp. 437 – 444. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[5] 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 - Volume 1*, HLT '11, (Stroudsburg, PA, USA), pp. 987–996, Association for Computational Linguistics, 2011.

[6] M. Lippi and P. Torroni, "Context-independent claim detection for argument mining," in *Proceedings of the Twenty-Fourth International Conference on Artificial Intelligence*, pp. 185–191, 2015.

[7] P. Hurley, *A Concise Introduction to Logic*. Cengage Learning, 2014.

[8] H. P. Grice, "Logic and conversation," in *Syntax and Semantics: Vol. 3: Speech Acts* (P. Cole and J. L. Morgan, eds.), pp. 41–58, San Diego, CA: Academic Press, 1975.

[9] S. Jackson and S. Jacobs, "Structure of conversational argument: Pragmatic bases for the enthymeme," *Quarterly Journal of Speech*, vol. 66, no. 3, pp. 251–265, 1980.
[10] K. Morrell and A. Hewison, "Rhetoric in policy texts: the role of enthymeme in darzi's review of the nhs.," 2013.
[11] J. Z. Sukkarieh, "Mind your language! controlled language for inference purposes," 2003.
[12] M. B. Burke, "Unstated premises," *Informal Logic*, vol. 7, no. 2, 1985.
[13] J. Gough and C. Tindale, "Hidden' or 'missing' premises," 1985.
[14] D. Hitchcock, "Enthymematic arguments," 1985.
[15] T. A. V. Dijk and W. Kintsch, "Strategies of discourse comprehension," 1983.
[16] S. Somasundaran, J. Wiebe, and J. Ruppenhofer, "Discourse level opinion interpretation," in *Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1*, COLING '08, (Stroudsburg, PA, USA), pp. 801–808, Association for Computational Linguistics, 2008.
[17] A. Tseronis, "From connectives to argumentative markers: A quest for markers of argumentative moves and of related aspects of argumentative discourse," *Argumentation*, vol. 25, no. 4, pp. 427–447, 2011.
[18] E. Florou, S. Konstantopoulos, and A. Kukurikos, "Argument extraction for supporting public policy formulation."
[19] J. Eckle-Kohler, R. Kluge, and I. Gurevych, "On the role of discourse markers for discriminating claims and premises in argumentative discourse," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (Lisbon, Portugal), pp. 2249–2255, Association for Computational Linguistics, Sept. 2015.
[20] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.
[21] C. J. Fillmore, P. Kay, and M. C. O'Connor, "Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone," *Language*, vol. 64, no. 3, pp. 501–538, 1988.
[22] M. Toosarvandani, "Letting negative polarity alone for let alone," In Tova Friedman and Satoshi Ito (eds.) Proceedings from Semantics and Linguistic Theory XVIII, 729- 746. Ithaca, New York: CLC Publications., 2008.
[23] O. Sawada, "Rethinking the let alone construction: what are its construction specific characteristics.," 2003.
[24] B. Cappelle, E. Dugas, and V. Tobin, "An afterthought on let alone," *Journal of Pragmatics*, vol. 80, pp. 70 – 85, 2015.
[25] R. Stalnaker, "Common ground," *Linguistics and Philosophy*, vol. 25, no. 5-6, pp. 701–721, 2002.
[26] M. Kienpointner, *Alltagslogik: Struktur und Funktion von Argumentationsmustern*. Problemata (Stuttgart), Frommann-Holzboog, 1992.
[27] M. McShane and P. Babkin, "Automatic ellipsis resolution: Recovering covert information from text," 2015.
[28] N. Suszczaska, J. Romaniuk, and P. Szmal, "Automatic analysis of elliptic sentences in the thetos system1," 2005.

Knowledge-Driven Argument Mining: what we learn from corpus analysis

Patrick SAINT-DIZIER [a,1],
[a] *IRIT-CNRS, 118 route de Narbonne, 31062 Toulouse Cedex France*

Abstract. Given a controversial issue, argument mining from texts in natural language is extremely challenging: besides linguistic aspects, domain knowledge is often required together with appropriate forms of inferences to identify arguments. Via the the analysis of various corpora, this contribution explores the types of knowledge that are required to develop an efficient argument mining system.

Keywords. Argument mining, Natural language processing, Knowledge representation

1. Aims and Challenges of Argument Mining

One of the main goals of argument mining is, given a controversial issue, to identify, in a set of texts, the arguments for or against that issue. Arguments are difficult to identify, in particular when they are not adjacent to the controversial issue, possibly not in the same text, because their linguistic, conceptual or referential links to that issue are rarely direct and explicit. Arguments are often evaluative natural language statements which become arguments because of the specific relations they have with another evaluative statement.

Except in specific contexts, most statements do not have any specific linguistic mark that would allow to directly identify as arguments. It is difficult to identify whether a statement is a support or an attack of a controversial issue, and what it precisely attacks or supports. Argument mining has to deal with two major problems: (1) **reference**: the argument is thematically related to the issue, and (2) **relatedness**: what facets of the issue are involved, how and how much.

Argument mining is an emerging research area which introduces new challenges both in natural language processing (NLP) and in argumentation. It requires the combination of linguistic analysis and language processing with artificial intelligence technology. Argument mining research applies to written texts, e.g. (Mochales Palau et ali.., 2009), (Kirschner et ali., 2015), for example for opinion analysis, e.g. (Villalba et al., 2012), mediation analysis (Janier et al. 2015) or transcribed dialogue analysis, e.g. (Budzynska et ali., 2014), (Swanson et ali., 2015). The analysis of the NLP techniques relevant for argument mining from annotated structures is analyzed in e.g. (Peldszus et al. 2016). Annotated corpora are now available, e.g. the AIFDB dialog corpora or (Walker et al., 2012). These corpora are very useful to understand how argumentation is realized in

[1] COMMA 2016 Workshop on the Linguistic Dimensions of Argumentation.
Corresponding Author: IRIT-CNRS, 118 route de Narbonne, 31062 Toulouse Cedex France stdizier@irit.fr

texts, e.g. to identify argumentative discourse units (ADUs), linguistic cues (Nguyen et al., 2015), and argumentation strategies, in a concrete way, possibly in association with abstract argumentation schemes, as shown in e.g. (Feng et al., 2011). Finally, reasoning aspects related to argumentation analysis are developed in e.g. (Fiedler et al., 2007) and (Winterstein, 2012) from a formal semantics perspective with a study of concessive and contrastive connectors. In opinion analysis, the benefits of argument mining are not only to identify the customers satisfaction level, but also to characterize why customers are happy or unhappy. Abstracting over arguments allows to construct summaries and to define customer preferences or value systems (e.g. low fares are preferred to localization or quality of welcome for some categories of hotel customers).

This paper focuses on the corpus construction and analysis and identifies and categorizes the needs in terms of knowledge to perform efficient argument mining.

2. Corpus Construction and Analysis

To explore and characterize the forms of knowledge that are required to develop argument mining in texts, we constructed and annotated four corpora based on four independent controversial issues. These corpora are relatively small, they are designed to explore the problem, and to elaborate the main features of a more extended empirical analysis and model, but not to design a comprehensive knowledge-driven argument mining system.

For this first experiment, we considered the four following issues, which involve very different types of arguments, forms of knowledge (concrete or relatively abstract) and language realizations. These issues are: (1) Ebola vaccination is necessary, (2) Women's conditions have improved in India, (3) The development of nuclear plants is necessary, and (4) Organic agriculture is the future.

For each of these issues, the corpus constructed and the different arguments found (eliminating duplicates or closely related arguments) are summarized in Table 1. The text fragments which are investigated are extracts from various sources where these issues are discussed, in particular: newspaper articles and blogs from associations. These are documents accessible to a large public, with no professional consideration. Language is French (glosses are given is paper) or English. For example: issue 1: The Lancet (UK journal), a French journal that develops results from the Howard Hughes Medical Institute, a web site: Ebola vaccines, therapies, and diagnostics Questions and Answers ; issue 2: http://saarthakindia.org/womens_situation_India.html, http://www.importantindia.com/20816/ women-in-india-role-and-status-of-women-in-india/; issue 3: Pour/Contre le nucleaire (les centrales) web site in French, etc.

Issue	Corpus size (text extracts)	nb. of annotated different arguments	overlap rate: average nb of similar arguments
(1)	16 texts, 8300 words	50	4.7
(2)	10 texts, 4800 words	27	4.5
(3)	7 texts, 5800 words	31	3.3
(4)	23 texts, 6200 words	22	3.8
Total	56 texts, 25100 words	130	4.07

Table 1. Corpus typology

This corpus shows that the argument diversity per issue is not very large. A preliminary annotation task includes the annotation of all the arguments that have been found. In a second stage, arguments judged by the annotator to be similar or redundant are bound and count for a single utterance. A high overlap rate has been observed: while there are original arguments, authors tend to borrow quite a lot of material from each other, for example in (4) an average repetition rate of 3.8 has been observed. This overlap rate clearly depends on the annotator (personal analysis and knowledge of the issue domain) and is somewhat subjective. In spite of this subjectivity, this rate gives an interesting rough redundancy level. A more detailed analysis of those repetitions would be of much interest from a rhetorical and sociological perspective.

An argument and its context (the discourse structures that modify it) are tagged between XML <argument> tags with attributes. The argument kernel is tagged <main arg>. Attributes informally characterize the knowledge that is required. Tags encode typical features associated with arguments (text span, polarity, strength) and specific features proper to our investigation; these are:
- the **text span involved**, which ranges from a few words to a short paragraph. Arguments are numbered for referencing needs,
- the **discourse relations** associated with the argument, these are annotated using the tags defined in our TextCoop platform (Saint-Dizier, 2012).
- the **polarity of the argument** w.r.t. the issue which has one of the following values: support, concession (argumentative concession is a weak support), contrast (a weak attack), and attack. The concession and contrast categories have been introduced to account for cases where the support or the attack are weak. These however need to be defined more precisely (formally and in annotation guidelines). Concessions and contrasts are both discourse structures and criteria to evaluate argument polarity.
- the **conceptual relation with the issue**: why it is an attack or a support,
- the **knowledge involved**, when appropriate, to identify the argument: list of the main concepts used. These come preferably from a predefined domain ontology, or from the annotator intuitions, if none is available. This list may be quite informal, it nevertheless contributes to identify the nature of the knowledge involved to identify arguments,
- the **strength of the argument**.

A tagged argument for issue (1) is the following:
<argument nb= 11, polarity= concession ,
relationToIssue= limited proofs of efficiency and safety of vaccination,
conceptsInvolved= efficiency measure, safety measures, test and evaluation methods,
strength= moderate>
<concession> Even if the vaccine seems 100% efficient and without any side effects on the tested population, < /concession>
<main arg> it is necessary to wait for more conclusive data before making large vaccination campaigns < /main arg>
<elaboration> The national authority of Guinea has approved the continuation of the tests on targeted populations.</elaboration> < /argument>.

From our manual analysis, the following argument polarities are observed: attacks: 53 occurrences, supports: 33, argumentative concessions: 21, argumentative contrasts: 19 and undetermined: 4. The corpus shows a tendency to argue against an issue: attacks and contrasts = 55%, supports and concessions = 41%. The need of knowledge to iden-

tify the relation between an issue and an argument is summarized in Table 2.

Issue	need of knowledge nb of cases (rate)	total number of concepts involved (estimate)
(1)	44 (88%)	54
(2)	21 (77%)	24
(3)	18 (58%)	19
(4)	17 (77%)	27
Total	100 (77%)	124

Table 2. Evidence for Knowledge

These figures show that for about 77% of the arguments, some form of knowledge is involved to establish an argumentative relation between a statement and a controversial issue. An important result is that the number of concepts involved is not very large: 124 concepts for 100 arguments over 4 domains. Even if the notion of concept remains vague, these results are nevertheless interesting to develop large argument mining systems. These considerations are exemplified below.

3. Analysis of the types of knowledge involved in argument identification

Our analysis below tends to suggest that the Generative Lexicon (Pustejovsky 1986) (GL), with some improvements, is an adequate representation framework to deal with knowledge based argument mining. The main other lexical semantics approaches such as FrameNet or VerbNet mainly concentrate on verb's predicate argument structure and adjuncts: they characterize the roles that these elements (NPs, PPs a,d S) play in the verb and proposition meaning. According to our observations, this are not central features for knowledge-based argument mining. The GL has a relatively simple argument structure feature, which could be enhanfced with FrameNet data, however, w.r.t. our analysis, the features of much interest are the purposes and goals of an object or action, its origin, and its uses. These are those which are evaluated in arguments. These, to the best of our knowledge, are specific features of the GL Qualia structure, in particular the telic and the agentive roles. A main limitation is that the GL has very little resources available.

The GL is an attempt to structure lexical semantics knowledge in conjunction with domain knowledge via a decompositional view of lexical meaning. The Qualia structure of an entity is composed of four fields called roles:
- the constitutive role describes the various parts of the entity and its physical properties, it may include subfields such as material, parts and components,
- the formal role describes what distinguishes the entity from other objects, i.e. the entity in its environment, in particular the entities which are more generic. It may also be structured into several subroles such as shape, dimensions, position, etc.
- the telic role describes the entity functions, uses, roles and purposes,
- the agentive role describes the origin of the entity, how it was created or produced.

Let us now develop the analysis of each corpus.

Main concepts in arguments related to issues (1) and (4)

For issues (1) and (4), arguments mainly attack or support salient features of the main concepts of the issue and closely related ones by means of various forms of evalu-

ations. Samples of arguments found for issue (1) are:
Supports: *vaccine protection is very good; Ebola is a dangerous disease; high contamination risks; vaccine has limited side-effects, no medical alternative to vaccine, etc.*
Attacks: *limited number of cases and deaths compared to other diseases; limited risks of contamination, ignorance of contamination forms; competent staff and P4 lab difficult to develop; vaccine toxicity and high side-effects,*
Concessions or Contrasts: *some side-effects; high production and development costs; vaccine not yet available; ethical and freedom problems.*

For issue (1), the concept *vaccine* is the root of the system. The facets of this concept and the closely related ones used in arguments can be organized as follows:
(1) the parts of a vaccine: the *adjuvant* and the active principle; (2) its super types: a vaccine is a kind of medicine; and (3) the most central aspects of the concept vaccine w.r.t. argument mining, namely its purposes, goals and consequences, and how it is created, tested and sold. For example, the concepts of *side-effect* and *toxicity* are consequences of using a medicine; the concept of *contamination* is related to one of the purposes of a vaccine, namely to avoid *disease dissemination*; and *production costs* are related to the creation and development of any product, including medicines and vaccines, etc.

Without knowing that a vaccine protects humans from getting a disease, it is not possible, e.g. to say that *prevents high contamination risks* is a support for issue (1). Similarly, without the knowledge that the active principle of a vaccine is diluted into an adjuvant that is also injected, it is not possible to say that *the adjuvant is toxic* is an attack, this statement could be e.g. purely neutral or irrelevant to the issue.

The terms used in this short analysis: purpose, properties, creation and development, etc. are foundational aspects of the structure of a concept, relatively well defined in the GL. Using this knowledge allows to identify arguments that attack or support an issue and how. Positive or negative terms found in statements are not sufficient to determine arguments, as shown and illustrated in (Saint-Dizier 2016).

Main concepts in arguments related to issue (2)

The arguments related to issue (2) mainly involve comparisons with men's living conditions or refer to general principles of human welfare. Some arguments are justified by means of figures while others remain vague, possibly not up-to-date or relative to specific situations: women living and social conditions may indeed evolve over time, and may differ depending on locations. Samples of arguments found for issue (2) are:
Supports: *increased percentage of literacy among women; women are allowed to enter into new professional fields; at the upper primary level, the enrollment increased from 0.5 million girls to 22.7 million girls.*
Attacks: *practices of female infanticide, poor health conditions and lack of education still persisting; home is women's real domain; they are suffering the violence afflicted on them by their own family members; malnutrition is still endemic.*
No concessions or contrasts have been observed: argument polarities are very clearcut, with a very positive or negative tonality, proper to highly controversial and overheated debates.

The concepts used in arguments related to issue (2) concentrate on facets of humans in the society. The identification of these concepts is crucial to characterize argumentative relations between issue (2) and their polarity. For example, improving literacy means higher education, better jobs and therefore more independence and social recognition, which are typical of living condition improvements.

The concepts used in arguments for issue (2) can be classified into two categories:
- those related to the services provided by the society to individuals: education, safety, health, nutrition, human rights, etc.
- those related to the roles or functions humans can play in the society: job and economy development, family development, cultural and social involvement, etc.
The first category is close to the formal role of the GL Qualia, while the latter is close to the telic role. These categories act as types which are characterized by more precise predicates in Qualia roles. For example the family category may include:
`have(X, family), develop(X, equality in family), educate(X, children), feed(X, family), choose(X, husband/wife).`

Arguments pros or cons develop evaluations of these concepts for women. In addition, lexical inferences must be developed to establish a complete link between words found in arguments and the main concepts illustrated above. For example, *malnutrition* is provoked by a 'lack of food' and entails 'health problems'.

Main concepts in arguments related to issue (3)

In the case of issue (3) (nuclear plants are necessary), supports or attacks mainly involve comparisons between various sources of energy. Samples of arguments are:
Supports: *energy independence; creates high technology jobs; risks are over-estimated; wastes are well-managed and controlled by AIEA; preserves the other natural resources,*
Attacks: *there are alternative solutions with less pollution; alternatives create more jobs than nuclear; there are risks of military uses : more dangerous than claimed; nuclear plants have high maintenance costs, etc.*
Concessions or Contrasts: *nuclear plants use dangerous products, but we know how to manage them; difficult to manage nuclear plants, but we have competent persons,*
In terms of language realizations, arguments are essentially constructed on the basis of comparatives and facts related to consequences, purposes or uses of nuclear plants. Utterances are quite developed and refer to knowledge about energy, pollution, health, or various forms of dangers, etc.

In terms of knowledge, nuclear plants have relatively complex agentive (construction, maintenance) and telic (roles, purposes) roles which must be decomposed into facets. For this latter facet, subtypes can be defined which develop its main purposes and secondary purposes and consequences, such as the production of nuclear wastes. The same structures must be developed for other sources of energy (coal, sea, wind, etc.) so that comparisons can be evaluated. Most arguments lack precise comparative data, e.g. wastes are all said to be a nuisance, but the degree is not given. These arguments require knowledge to be evaluated against precise data, otherwise they rather play on the pathos. These are called underspecified arguments. Finally, some arguments are fallacious, for example, the 'energetic independence' advocated as a support for nuclear plants is not acceptable since uranium is bought abroad: the use of knowledge contributes to detect them.

4. Summary: from concepts to knowledge representation for argument mining

The introduction of knowledge in argument mining is a crucial feature which allows to improve the identification of:
(1) the potential relatedness of an evaluative statement with a controversial issue,

(2) which aspect(s) of the issue it deals with, characterized by a specific facet in a Qualia role, identified by one or more predicates,
(3) the argument polarity: attack, support, concession, contrast,
(4) neutral, irrelevant, underspecified (when knowledge is not comprehensive enough to decide on the polarity of an argument) and fallacious arguments.
The fully annotated corpus (French and English) will be shortly made available.

In terms of language realizations, the following three main types of argument realizations are observed:
(1) use of **evaluative expressions** (52% of cases), in attribute-value form, where the attribute is proper to the concepts of the controversial issue or of closely derived concepts: *Vaccine development is very expensive, adjuvant is toxic.* Resources which are required are a set of Qualias representing the concepts, the lexical terms which are associated with each element of the Qualia and their quasi-synonyms,
(2) use of **comparatives** applied to related concepts (e.g. electricty produced by nuclear energy compared to coal or wind, 21% of the cases): Qualia of related concepts are needed with lexical realizations, e.g. *number of sick people much smaller than for Malaria, nuclear wastes are more dangerous than coal wastes*.
(3) **evaluation of uses, consequences or purposes** of the main concept of the issue (27% of the cases) requires Qualia structures and associated lexical entries and, when relevant, lexical inference and presuppositions associated with those lexical entries. Examples are: *vaccine prevents bio-terrorism, women are allowed to enter into new profesional fields.*

Knowledge-based argument mining requires relatively complex sources of knowledge, however, an important result is that for a given issue, only a few conecpts (and their Qualias) are needed. We have breifly shown elements of GL Qualia structures seem to be an appropriate knowledge representation framework, with an adequate conceptual typing (e.g. various forms of telicity) which can be combined with lexical data and lexical inference. In addition, this approach allows to:
- reason about arguments on the basis of a domain knowledge representation,
- develop conceptually relevant forms of argument synthesis or clustering, based on Qualia roles.

An open issue remains the acquisition of the Qualia structure data. Given that relatively few Qualias are involved for a controversial issue, some features can be developed manually. An important investigation direction is the acquisition of additional features, probably based on bootstrapping techniques.

5. Conclusion

In this short paper, we present a preliminary analysis of the different forms of knowledge which are frequently required to relate arguments with a given controversial issue. Our short corpus analysis shows that the type of information proposed by the Qualia structure of the GL is a useful knowledge source: object parts, action or object properties, uses, functions, creation and development are often the main topics for arguments. Obviously, this analysis must be expanded in various directions before any development of a knowledge-driven argument mining system. A more detailed empirical analysis is necessary, with a larger corpus, several annotators, and annotation guidelines before conluding that e.g. the Qualia structure is the main resource that is needed.

References

[1] Budzynska, K., Janier,M., Reed, C., Saint-Dizier, P., Stede, M., Yakorska, O., A model for processing illocutionary structures and argumentation in debates. In proc. LREC, 2014.

[2] Feng, V. W., Hirst, G., Classifying arguments by scheme. In Proceedings of the 49th ACL: Human Language Technologies, Portland, USA, 2011.

[3] Fiedler, A., Horacek, H., Argumentation within deductive reasoning. International Journal of Intelligent Systems, 22(1):49-70, 2007.

[4] Janier,M., Reed, C., Towards a Theory of Close Analysis for Dispute Mediation Discourse, Journal of Argumentation, 2015.

[5] Kirschner, C., Eckle-Kohler, J., Gurevych, I., Linking the Thoughts: Analysis of Argumentation Structures in Scientific Publications. In: Proceedings of the 2nd Workshop on Argumentation Mining, Denver, 2015.

[6] Mochales Palau, R., Moens, M.F., Argumentation mining: the detection, classification and structure of arguments in text. Twelfth international ICAIL'09, Barcelona, 2009.

[7] H., Nguyen, D., Litman, D., Extracting Argument and Domain Words for Identifying Argument Components in Texts. In: Proc of the 2nd Workshop on Argumentation Mining, Denver, 2015.

[8] Peldszus, A., Stede, M., From argument diagrams to argumentation mining in texts: a survey. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 2016.

[9] Pustejovsky, J., The Generative Lexicon, MIT Press, 1995.

[10] Saint-Dizier, P., Processing natural language arguments with the TextCoop platform, journal of Argumentation and Computation, vol 3(1), 2012.

[11] Swanson, R., Ecker, B., Walker, M., Argument Mining: Extracting Arguments from Online Dialogue, in proc. SIGDIAL 2015, USA.

[12] Villalba, M.G., Saint-Dizier, P., Some Facets of Argument Mining for Opinion Analysis, COMMA, Vienna, IOS Publishing, 2012.

[13] Walker, M., Anand, P., Fox Tree, J.E., Abbott,R., King, J. A Corpus for Research on Deliberation and Debate. Proc. of the Language Resources and Evaluation Conference (LREC), Istanbul, 2012.

[14] Winterstein, G., What but-sentences argue for: An argumentative analysis of 'but', in Lingua 122, 2012.

Realizing Argumentative Coherence Relations in German: A Contrastive Study of Newspaper Editorials and Twitter Posts

Tatjana SCHEFFLER [1] and Manfred STEDE
University of Potsdam, Germany

Abstract. Using two German corpora of newspaper editorials and Twitter posts, which have been annotated according to Rhetorical Structure Theory, we extract all instances of coherence relations that we regard as potentially argumentative. We suggest a set of linguistic features indicating "argumentativity" and study their distribution in the two corpora of very different genres. We find that content-based lexical features and complexity differ significantly across the two corpora. Similarly, though connectives are often used to indicate argumentative coherence relations (especially for contrastive relations), different kinds of connectives are found in newspaper text vs. tweets.

1. Introduction

Argumentation is an abstract process that is part of many different domains and genres. In this paper, we address the question to which extent linguistic reflexes of certain kinds of argumentative structures are stable across two different text genres. In particular, we study argumentative text in newspaper articles and in Twitter conversations, and investigate the distribution of various linguistic features. We assume that on the one hand, explicit linguistic markers of argumentative structures are somewhat stable across genres, since they contribute to the semantics. On the other hand, properties of the medium or genre may influence the kinds and frequencies of markers that are employed, for example due to enforced brevity of contributions or lack of context in Twitter conversations. In this paper, we want to start investigating the effects of these opposing constraints.

Viewed very broadly, argumentation consists of individual *claims*, which can be substantiated by *justifications* or attacked by *counterarguments*. In the absence of large argument-annotated corpora (especially for German), we propose to use *coherence relations* as a proxy for 'argumentativeness': From corpora annotated with Rhetorical Structure Theory trees (Mann & Thompson, 1988), we extract those relation instances that correspond to argumentative moves. This is facilitated by RST's differentiating 'subject-matter' from 'presentational' relations, where the latter are defined as influencing the state of mind of the readers, which largely (but not exclusively, cf. the ENABLEMENT

[1]Corresponding Author: Tatjana Scheffler, Dept. of Linguistics, UFS Cognitive Sciences, University of Potsdam. E-mail: tatjana.scheffler@uni-potsdam.de.

relation) corresponds to argumentation. In this way, though we do not study all instances of argumentation in our documents, we restrict ourselves to the same subset of argumentative structures in both text genres, to compare how they are marked.

In the following, we characterize the linguistic realization of both contrastive and causal argumentative relations. We show that content-related lexical features as well as sentence complexity differ significantly between the two corpora. We also identify features (connectives and others) that show distinctive patterns in both types of text that can be used for classification.

2. Coherence Relations and Argumentation

Given the scarcity of data annotated for argumentation, esp. across languages and genres, we propose to use corpora with rhetorical structure annotation as a surrogate, as far as the search for features of argumentative language is concerned. In particular, Rhetorical Structure Theory[2] (Mann & Thompson, 1988) suggests a set of coherence relations that is divided into 'subject-matter' (henceforth: semantic) and 'presentational' (henceforth: pragmatic) relations. For the former, the intended effect is that the readers *recognize* some state of affairs; for the latter, the effect is described in terms of changes to their state of mind. In the following, we list the six RST relations that in our annotation projects[3] have been used as 'pragmatic', with a short summary of their definition, and the translation of an example from our data. In the following, we use 'Nuc' and 'Sat' for the Nucleus and Satellite part of the relation, respectively.

Antithesis. Def: Nuc and Sat are in contrast; one cannot have equal positive regard for both. Writer wants to increase reader's positive regard for Nuc. Example (Twitter): *[It's not nice,]$_{Sat}$ [but that's life.]$_{Nuc}$*

Concession. Def: Nuc and Sat are potentially incompatible, but in the current situation they both hold. Writer wants to increase reader's positive regard for Nuc. Example (Twitter): *[I know that he's not going to reply,]$_{Sat}$ [but still I'm somehow waiting for some message...]$_{Nuc}$*

Evidence. Def: Reader might not believe Nuc, but will believe Sat. Sat is an "objective" statement that can increase Reader's belief in Nuc, which Writer intends. Example (Twitter): *[There's two tough months ahead;]$_{Nuc}$ [2 exams, BarCamp, KPT, LPT, BP.]$_{Sat}$*

Justify. Def: Nuc is a subjective claim that Reader might not believe or appreciate. Sat is a statement of a general, for example moral, attitude of the actor in Nuc. Understanding Sat can increase Reader's acceptance of Writer's right to present of Nuc, which Writer wants to achieve. Example (PCC): *[We pay public TV fees for that,]$_{Nuc}$ [since the public mandate requires a healthy mixture of entertainment and information.]$_{Sat}$*

[2]For reasons of space, we cannot explain the concepts of RST here; the reader is referred to the original Mann/Thompson paper or to (Taboada & Mann, 2006).

[3]Our annotation guidelines largely follow the original Mann/Thompson proposal, but there are some differences in terms of categorizing relations as semantic or pragmatic; also, we have added the Reason relation in order to be able to make more fine-grained distinctions in analyzing argumentative text. One relation that is situated on the borderline here is Evaluation. After inspecting our data, We have decided not to include it here, because much non-argumentative material would be included. But for other corpora, this might have to be reconsidered.

Motivation. Def: Nuc is a potential action with Reader as actor. Sat provides an incentive to perform that action, which Writer wants Reader to do. Example (Twitter): *[Join our post card action!]$_{Nuc}$ [our basic right to privacy is in EU-danger.]$_{Sat}$*

Reason. Def: Nuc is a subjective claim that Reader might not believe or appreciate. Sat is also a subjective statement, and it can increase the belief in Nuc, which Writer wants from Reader. Example (Twitter): *[Am happy to follow you]$_{Nuc}$ [– I'm also a vegan :)]$_{Nuc}$*

3. Data

We briefly describe the two datasets that form the basis for our study. The first is an existing set of annotated newspaper editorials, the second is a new corpus of conversations extracted from Twitter; both have been annotated with a variant of RST.

3.1. Newspaper editorials

The Potsdam Commentary Corpus (PCC) (Stede & Neumann, 2014) is a collection of 175 editorials (29,500 words) from a regional German daily (*Märkische Allgemeine Zeitung*). They were collected in the early 2000s and deal with local, regional, or federal political issues. Their length is on average 13 sentences / 168 words. The text type is primarily argumentative, but the "depth" of argumentation can differ: Some texts clearly advance a specific claim and provide justifications; others more or less just re-tell some piece of news and add some opinion to it.

The PCC is publicly available from our website.[4] In the current version, manual annotations of five different layers are included: Sentence syntax, nominal coreference, rhetorical structure (RST), connectives and their arguments, and information-structural topics. On the RST layer, there are 3211 relations in total. All the annotation guidelines are also available (in German) (Stede, 2016).

3.2. Twitter Conversations

Our annotated Twitter data is yet unpublished but will be made available later. The source of the data is the one-month German Twitter snapshot described by (Scheffler, 2014). For our purposes here, we filtered specifically for "conversations", i.e., tweets that are in a `reply-to` relation. These can be either pairs or larger sets, where multiple replies form a tree structure, with the initiating tweet as the root. The corpus consists of 296 conversations ("threads"), which contain 1437 individual tweets (19,561 words). Annotations were done with a variant of Marcu's RSTTool.[5]

Our hypothesis is that coherence can be modeled in such conversations also by means of rhetorical structure. We devised genre-specific guidelines that account for certain dialog phenomena, but the relation set also contains the 'pragmatic' relations mentioned above. We annotated 3267 relations in total, and the proportion of pragmatic relations is 4%. This number is much lower than in the newspaper corpus, because a large

[4] `http://angcl.ling.uni-potsdam.de/resources/pcc.html`

[5] The tool was extended for handling Twitter data by Uladzimir Sidarenka in our group, and it will be available to interested researchers.

Corpus	All Rels.	ANTITH.	CONC.	EVID.	JUST.	MOTIV.	REAS.	nuc	sat
PCC	3211	123	125	99	4	1	267	17.9	19.8
TWI	3267	26	25	22	3	6	45	6.9	7.8
(*r*-rel.)		5	2				2		

Table 1. Our dataset of newspaper editorials (PCC) and Twitter conversations (TWI). Tweet relations are listed separately for relations within tweets (middle row) and relations across different tweets ("*r*-relations", last row). The last columns give the average length (in words) of nucleus and satellite.

part of the relations are Twitter-specific (e.g., ADDRESS for the relation between an addressed user's name and the text sent to him/her) or dialog-specific (e.g., QUESTION-ANSWER) relations. In the Twitter RST annotations, a distinction is made between relations within a tweet (the regular kind) or across two tweets (as when a second speaker provides justification for a previous speaker's point); the second type of relations are called *r-relations*. Here, we analyze both types together because of the small number of *r*-relations.

For building our data set, we automatically extracted all instances of the six relations, i.e., their nucleus and satellite text spans. While in the tweets, relations can hold within a post or at most between two complete posts (and no more than that), in the PCC texts this extraction yields complications: When relations are located higher up in the RST tree, a segment can stretch across multiple sentences, which for our present purposes generates considerable noise for statistical analysis. One theoretical option is to restrict all segments to the central nuclei, but these often are just parts of sentences, and also we run into various complications with multinuclear relations. We therefore decided to prune from our study those instances of a relation where a segment is longer than 20 words (which is the average length of a sentence in PCC).

Table 1 shows the distribution of the coherence relations in the two datasets.

4. Related work

For gathering a set of linguistic features that could signal the presence of argumentative relations, we turned to earlier computational linguistics research on identifying argumentative sentences in text. Most of this work uses word n-grams (n = 1..3) and sometimes also pairs of non-adjacent words as basic features. Our interest here is in the role of non-lexical features, though. In the early work of (Moens, Boiy, Palau, & Reed, 2007), the authors classified sentences in the multi-genre Araucaria corpus as non-/argumentative and used the presence of modal verbs, measures of sentence and token length, as well as two structural features: the depth of the syntactic parse tree and the number of clauses. In addition they employed a list of 286 keywords, which is being characterized as containing causal and contrastive connectives (and probably more). In their experiments, the lexical features were hard to beat (accuracies in the lower 70%s); adding some of the linguistic features yielded roughly one additional percent.

When (Somasundaran & Wiebe, 2009) introduced their "arguing lexicon" (n-grams extracted from the MPQA corpus), they also added modal verb features, but unfortunately did not measure the contribution of those features in isolation. (Florou, Konstantopoulos, Koukourikos, & Karampiperis, 2013) employed verb tense and mood fea-

tures and found that they can contribute to some extent to the differentiation of non-/argumentative sentences in Greek. Working with student essays in English, (Stab & Gurevych, 2014) sought to classify sentences as (two types of) claims, premises, and non-argumentative. They borrowed many of the features from (Moens et al., 2007) and added various features of surrounding sentences, use of 1st person pronouns, the syntactic production rules of the parse tree, and a list of connectives extracted from the PDTB. They found that structural (syntactic) features were most helpful (in addition to lexical n-grams) especially for identifying non-argumentative sentences.

5. Analysis

Following up on the previous work, we have investigated the presence of several proposed linguistic features in our extracted segments. We describe the operationalization of our features in turn.

Number of verbs: The Twitter segments that participate in the adversative and causal relations are naturally much shorter than the newspaper text segments (see Table 1). In order to get a closer estimate of the complexity of the segments (independent of length), we have counted the number of verbs in each segment. In order to identify verbs, we have used the `pattern.de`[6] Python module by (De Smedt & Daelemans, 2012), which provides a version of the Brill tagger.

Connectives: Argumentative relations can be explicitly marked by connectives, though it has been observed that by far not all such relations are marked. We compiled lists of contrastive and causal connectives from the German DiMLex connective lexicon (Stede & Umbach, 1998) and checked whether a segment contains contrastive connectives (if it is part of an adversative relation) or causal connectives (for causal relations).

Modals: We counted occurrences of the modal verbs *können, sollen, müssen, brauchen, dürfen*.

Negation: A feature that was not discussed in the related work is the presence of sentential negation. We are interested in its distribution across causal versus adversative relations, since polarity mismatch can indicate adversative relations. We count the presence of the negation operators *nicht, nie, niemand*, and *kein* (plus inflected forms).

First person: In certain genres (e.g., student essays; see above), the presence of 1st person pronouns can be indicative of the role of a segment. While we do not expect many in the PCC texts, they might well be present in Twitter posts.

[6] `http://www.clips.ua.ac.be/pages/pattern-de`

	PCC news text				Twitter conversations			
	Adversative		Causal		Adversative		Causal	
Features	**Nuc**	**Sat**	**Nuc**	**Sat**	**Nuc**	**Sat**	**Nuc**	**Sat**
verbs	1.44	1.36	1.70	1.71	0.91	1.16	0.85	0.93
connectives	0.61	0.21	0.14	0.14	0.23	0.51	0.03	0.22
modals	0.18	0.09	0.21	0.15	0.05	0.09	0.12	0.09
negation	0.25	0.31	0.21	0.23	0.14	0.35	0.14	0.19
1st person	0.02	0.03	0.00	0.02	0.35	0.33	0.34	0.30
# segments	185	197	273	225	57	57	73	74

Table 2. Linguistic features found in argumentative moves: numbers indicate rates per segment. (Note that the segment length in the Twitter data is less than half of the PCC segments.)

PCC news text				Twitter conversations			
Adversative		Causal		Adversative		Causal	
Nuc	**Sat**	**Nuc**	**Sat**	**Nuc**	**Sat**	**Nuc**	**Sat**
aber	zwar	deshalb	denn	aber	aber	also	weil
doch	aber	also	weil	zwar	wobei	darum	also
sondern	obwohl	denn	also	trotzdem	doch		denn

Table 3. Top three connectives in each segment type. Nuc = nucleus, Sat = satellite.[7]

6. Discussion

The results of the analysis are presented in Table 2. As shown before, the PCC segments are considerably longer, and the number of verbs reflects this difference by and large proportionally. To some extent, the difference in complexity is reflected in the most frequent connectives for each type of segment (Table 3). Considering the adversatives, we find the subordinating conjunctions *zwar* and *obwohl* ('although') marking satellites in the PCC texts, as is expected from common observations in the literature (syntactic subordination corresponds to RST nuclearity status). These subordinators are not present in the top-3 list in Twitter, though, showing that this genre achieves brevity and conciseness by replacing subordination with coordination. An interesting exception here is the presence of subordinating *zwar* in nuclei, which merits a detailed qualitative analysis (future work).

It is interesting to see that adversative relations are much more frequently marked by a connective than causal ones, in both genres alike. Giving a reason can easily occur just implicitly, while objecting tends to involve connectives. Here we have to bear in mind that our data contains only the pragmatic adversative relations, which are both mononuclear. The multinuclear CONTRAST is more likely to occur without connectives, because the contrast can be evoked just by the lexical predicates.

A clear outlier in the connective frequencies is the small value in the nuclei of causal relations in Twitter. This can be explained by an overwhelming use of the subordinating conjunction *weil* for giving reasons in Twitter (Scheffler, 2014), which marks the satellite of the relation. The predominance of *weil* is also observed in spoken language.

[7] Approximate translations for the connectives: *aber* 'but', *also* 'so', *darum* 'therefore', *denn* 'since', *deshalb* 'therefore', *doch* 'however, but', *obwohl* 'although', *sondern* 'instead, rather', *trotzdem* 'anyway', *weil* 'because', *wobei* 'though', *zwar* 'though'.

As expected, the frequency of 1st person pronouns is negligible in PCC but not rare in Twitter. The distribution is almost even, the one exception being a lower frequency in causal satellites. Our hypothesis is that reasons for actions or beliefs are more often non-personal events; the underlying scheme being "I did X, because Y happened".

The distribution of modals is rather inhomogeneous. A tendency is for them to occur in nuclei rather than satellites, but an exception is the adversative relations in Twitter. For causal relations, the distribution might reflect the scheme of concluding a 'desired' state of affairs (something should happen) from a factual observation.

Negations in PCC are more frequent in adversatives, but this does not hold for the nuclei in Twitter, which also merits a more detailed analysis. A uniform tendency is for them to occur more frequently in satellites than in nuclei. As a next step, a polarity mismatch between nucleus and satellite could be studied.

7. Summary and Outlook

Using two quite different linguistic genres, we provided a comparative analysis of various features that previous research has postulated as being characteristic for argumentative text portions. In the absence of multi-genre German data that has genuine argument annotation, we proposed to regard pragmatic rhetorical relations as a proxy for identifying a subset of argumentative sections in text. Then we carried out the analysis on the RST-annotated Potsdam Commentary Corpus and a collection of Twitter conversations.

The longer-term perspective of this research is twofold. On the one hand, we are interested in identifying argumentation in (German) Twitter data automatically, and the study presented here serves as a first step toward establishing a feature set. On the other hand, we plan further investigations into the relationship between RST analysis and argumentation. We will analyze additional RST corpora (also in other languages) in order to determine the extent to which

- the five relations used here indeed indicate argumentative content, and
- argumentative portions of text correlate with additional relations.

On the first point our hypothesis is that the correlation is indeed quite good, as long as the annotation guidelines of the respective projects follow the basic layout of the Mann/Thompson relation inventory. The second point can be more interesting: Argumentation might very well reach beyond the RST subtrees formed by the relations in question, because argumentative material can be in an ELABORATION, CONTRAST, or other relation to a text portion that we have identified. I.e., on the boundaries of the extracted subtrees, it can be necessary to attach more material. This has to be studied by additional corpus analyses.

Acknowledgments

Part of the work reported here was funded by the Federal Ministry of Education and Research (BMBF) in the project "Discourse Analysis of Social Media". We thank Uladzimir Sidarenka and Matthias Bisping for providing the RST annotations on the Twitter data.

References

De Smedt, T., & Daelemans, W. (2012). Pattern for Python. *Journal of Machine Learning Research*, *13*, 2031—2035.

Florou, E., Konstantopoulos, S., Koukourikos, A., & Karampiperis, P. (2013, August). Argument extraction for supporting public policy formulation. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (pp. 49–54). Sofia, Bulgaria: Association for Computational Linguistics. Retrieved from `http://www.aclweb.org/anthology/W13-2707`

Mann, W., & Thompson, S. (1988). Rhetorical structure theory: Towards a functional theory of text organization. *TEXT*, *8*, 243–281.

Moens, M.-F., Boiy, E., Palau, R., & Reed, C. (2007). Automatic detection of arguments in legal texts. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law* (pp. 225–230). Association for Computing Machinery. doi: 10.1145/1276318.1276362

Scheffler, T. (2014). A German Twitter snapshot. In N. C. et al. (Ed.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association (ELRA).

Somasundaran, S., & Wiebe, J. (2009, August). Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP* (pp. 226–234). Suntec, Singapore: Association for Computational Linguistics. Retrieved from `http://www.aclweb.org/anthology/P/P09/P09-1026`

Stab, C., & Gurevych, I. (2014, August). Annotating argument components and relations in persuasive essays. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)* (pp. 1501–1510). Dublin, Ireland: Dublin City University and Association for Computational Linguistics. Retrieved from `http://www.aclweb.org/anthology/C14-1142`

Stede, M. (Ed.). (2016). *Handbuch Textannotation: Potsdamer Kommentarkorpus 2.0* (Vol. 8). Potsdam: Universitaetsverlag. Retrieved from `http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-82761`

Stede, M., & Neumann, A. (2014, may). Potsdam Commentary Corpus 2.0: Annotation for discourse research. In *Proceedings of the ninth international conference on language resources and evaluation (lrec'14)*. Reykjavik, Iceland: European Language Resources Association (ELRA).

Stede, M., & Umbach, C. (1998). Dimlex: A lexicon of discourse markers for text generation and understanding. In *Proceedings of the 17th International Conference on Computational Linguistics* (pp. 1238–1242).

Taboada, M., & Mann, W. (2006). Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies*, *8*(4), 423-459.