HAL
open science

Climate Negotiation Analysis

Pablo Ruiz, Clément Plancq, Thierry Poibeau

To cite this version:

Pablo Ruiz, Clément Plancq, Thierry Poibeau. Climate Negotiation Analysis. Digital Humanities 2016, Jagiellonian University and Pedagogical University, Jul 2016, Cracovie, Poland. pp.663-666. hal-01423299

HAL Id: hal-01423299

https://hal.science/hal-01423299

Submitted on 29 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Digital Humanities 2016
Kraków, 11–16 July

(http://dh2016.adho.org)

DH Home (http://www.dh2016.adho.org) / Abstracts (/abstracts/) / 81 (/abstracts/81)

Show info | How to cite | XML Version (/static/data/357.xml)

Title: Climate Negotiation Analysis
Authors: Pablo Ruiz Fabo, Clément Plancq, Thierry Poibeau
Category: Paper:Short Paper
Keywords: semantic role labeling, relation extraction, climate negotiations, Earth Negotiations Bulletin

Climate Negotiation Analysis

1. Introduction

Text analysis methods based on word co-occurrence have yielded useful results in humanities and social sciences research. For instance, Venturini et al., (2012) describe the use of concept co-occurrence networks in social sciences. Grimmer and Stewart (2013) survey clustering and topic modeling applied to political science corpora. Whereas these methods provide a useful overview of a corpus, they cannot determine the predicates [1] relating co-occurring elements with each other. For instance, if *France* and the phrase *binding commitments* co-occur within a sentence, how are both elements related? Is France in favour of, or against *binding commitments*?

Different natural language processing (NLP) technologies can identify related elements in text, and the predicates relating them. A recent approach is *open relation extraction* (Mausam et al., 2012, among others), where relations are derived from the corpus in a data-driven manner, without having to pre-specify a vocabulary of predicates or actors. We are developing a workflow to analyze the Earth Negotiations Bulletin (vol. 12) [2], which summarizes international climate negotiations. A sentence in this corpus can contain several verbal or nominal predicates indicating support and opposition (see Table 1). Results were uneven when applying open relation extraction tools to this corpus. To address these challenges, we developed a workflow with a domain model, and analysis rules that exploit annotations for semantic roles and pronominal anaphora, provided by an NLP pipeline.

Our system identifies points supported and opposed by negotiating actors and extracts keyphrases and DBpedia [3] concepts from those points. The results are displayed on an interface, allowing for a comparison of different actors' positions. The system helps address a current need in digital humanities: tools for the quantitative analysis of textual structures beyond word co-occurrence.

The abstract is structured as follows. First, related work and the corpus are presented. Then, our system is described. Finally, evaluation is discussed.

Material supplementing the paper and access information to the system will be available on the project's website. [4]

1 - Multiple verbal predicates				
The EU, with NEW ZEALAND and opposed by CHINA, MALAYSIA and BHUTAN, supported including the promotion of natural regeneration within the definitions of "afforestation" and "reforestation."				
Propositions				**Predicate Type**
	Actor	*Predicate*	*Negotiation Point*	
1	European_Union	supported	including the promotion of natural regeneration within the definitions of "afforestation" and "reforestation."	support
2	New_Zealand			support
3	China	~supported	including the promotion of natural regeneration within the definitions of "afforestation" and "reforestation."	opposition
4	Malaysia			opposition
5	Bhutan			opposition
2 - Nominal predicate				
Much of the discussion was on a proposal by the G-77/China to include research and development in the transport and energy sectors in the priority areas to be financed by the SCCF.				
Propositions				**Predicate Type**
	Actor	*Predicate*	*Negotiation Point*	
1	Group_of_77/China	proposal	to include research and development in the transport and energy sectors in the priority areas to be financed by the SCCF.	support

*Table : **Typical corpus sentences**. Sentence 1 has predicates supported and opposed, with several actors each. Example 2 shows a nominal predicate (proposal). For Sentence 1, five ‹actor, predicate, negotiation point› propositions are extracted by the system, and the opposing actors (China, Malaysia, Bhutan) are assigned a proposition which is a negated version (with ~supported as the predicate) of the proposition for the main verb supported.*

2. Related work

Venturini et al., (2014) created concept co-occurrence networks for the ENB corpus, using Cortext Manager [5], a corpus cartography tool. This analysis does not cover which predicates relate concepts and actors. Salway et al., (2014) used *grammar induction* on ENB to identify recurrent actor/predicate patterns; it could be tested whether results with that approach complement ours.

Some studies have used syntactic and semantic parsing for text-analysis of social sciences and humanities corpora. Diesner (2012, 2014) examines the contribution of NLP to the construction of text-based networks. Van Atteveldt (2015) used dependency parsing to apply co-occurrence based methods within sentence elements related to an actor or a predicate. These studies rely mostly on syntactic dependencies and verbal predicates. We are using semantic role labeling as the basis for relation extraction, and treating nominal predicates besides verbal ones. We also developed an interface to navigate the results.

Finally, a relevant resource for text-mining on climate corpora is *climatetagger* API [6], which links concepts against a domain-specific thesaurus (Bauer et al., 2011). This thesaurus could complement our concept-linking results (based on DBpedia, a general ontology).

3. Corpus

Each ENB issue is a 2000 word summary for one day of negotiations. The issues are written by domain experts, who strive for an objective tone and, to avoid biases, use similar expressions when reporting about all participants' interventions (Venturini et al., 2014). The COP meetings are covered in 255 ENB issues, with ca. 35,000 sentences. The original corpus format is HTML, which we preprocessed into clean text. We dated each issue based on ENB's table of contents.

4. System description

The system helps analyze patterns of support and opposition between negotiating parties, and the issues about which parties agree or disagree. To achieve this, the system extracts propositions of shape *‹actor, predicate, negotiation point›*, [7] based on a domain model containing actors and predicates, and applying analysis rules on the outputs of an NLP toolkit. Keyphrases and DBpedia concepts are also extracted from the negotiation points. All extractions, and the corpus itself, are made navigable on a user interface (UI).

4.1. NLP toolkit

We used the IXA Pipes library [8] (Agerri et al., 2014), with default models for **tokenization** and **part-of-speech tagging**. We resolved some types of **pronominal anaphora** based on *CorefGraph* [9] coreference chains.

Semantic Role Labeling (SRL) (Surdeanu et al., 2008) identifies a predicate's arguments and their semantic functions or roles (e.g. *agent*). SRL was performed with ixa-pipe-srl [10], which tags against the PropBank database (Palmer et al., 2005) for verbal predicates and against NomBank (Meyers et al., 2004) for nominal ones.

Keyphrase Extraction: YaTeA [11] was used (Aubin and Hamon, 2006). This library performs unsupervised term extraction using syntactic and statistical criteria.

Entity Linking (EL): The tool from (Ruiz and Poibeau, 2015) was used. It combines outputs from several public EL services, selecting the best outputs with a weighted vote.

4.2. Domain-specific components

The **domain model** contains actors (negotiating countries and groups) and verbal or nominal predicates. Verbal predicates (from PropBank) can be neutral reporting verbs (e.g. *stated*), or verbs related to support and opposition (*recommended*, *criticized*). The nominal predicates (from NomBank) express similar notions to the verbs (e.g. *proposal*, *objection*). The model also specifies a predicate type: *report*, *support*, or *oppose*.

Analysis rules were implemented to identify propositions based on the semantic roles of predicates' arguments, previously obtained with SRL. Most domain predicates involve an agent and a message expressed by the agent (who agrees with the message, objects to it, or just reports it). Thus, actor mentions in a predicate's A0 argument [12] represent the actor who expresses the message, and the predicate's A1 argument 12 often represents the negotiation point addressed by the actor. The generic rule to identify propositions is in Figure 1.

Rule: Generic proposition

for each predicate *p* :
- resolve negation (see below)
- for each pronoun *he, she, it* in *p*'s A0 argument :
  - apply anaphora resolution (see below)
- for each actor-mention *am* in *p*'s A0 argument :
  - create a proposition *‹am, p, point›*, where *point* is a concatenation of *p*'s A1 arguments

Figure : ***General rule to create a proposition***

Sentences with *opposed by* constructions require a different analysis (e.g. *China, opposed by the EU, recommended…*) In such sentences, a different rule creates, for the opposing actors, propositions where the predicate contradicts the main clause's predicate (see Table 1 for an example). Proposition-creation rules for more specific cases have also been implemented.

The treatment of **negation** relies on finding *AM-NEG* roles (see footnote 12) attached to a predicate, or negative items (*not*, *lack*) in a window of two tokens preceding a predicate.

Pronominal anaphora was treated via custom rules operating on the output of a coreference resolver (see footnote 9). We created custom rules since, in the corpus, *he* and *she* (besides *it*) can refer back to a country (pronoun gender depends on the country's delegate).

To facilitate searches by date-range, propositions are assigned their documents' date.

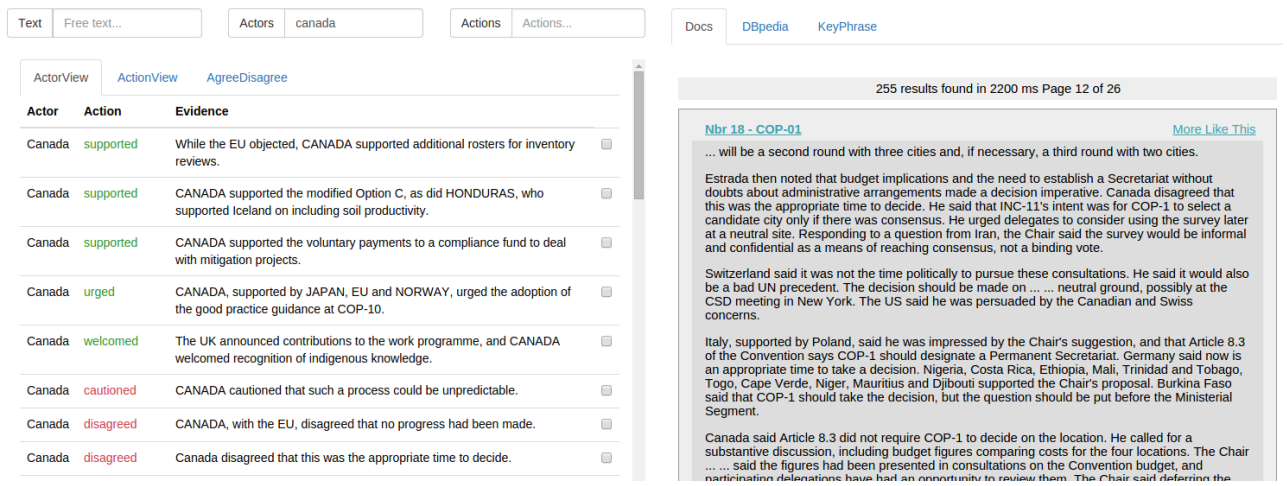

*Figure : **Main view of the interface.** The left panel gives access to the search workflows (Actors, Actions, Points). It also shows propositions for a query (e.g. the actor Canada), and gives access to the AgreeDisagree view. The right panel shows the documents in the Docs tab, as well as aggregated keyphrases and DBpedia concepts for the query or for selected propositions, in the other tabs.*

4.3. User interface

The UI (Figure 2) helps analyze actors' negotiation positions. It allows searching for documents matching a text query (***Text*** search box), and for propositions matching a given actor (***Actors*** box) or a given predicate (***Actions*** box). Propositions matching a query are displayed on the left panel, documents for a query on the right. Aggregated ***keyphrases*** and ***DBpedia*** concepts for the content matching a query (documents or propositions) are displayed in tabs on the right panel. The ***AgreeDisagree*** view provides an overview of keyphrases and concepts from propositions where selected actors agree or disagree. Simultaneous access on the UI to the corpus and the annotations helps researchers validate results.

The implementation framework is Django [13], with Solr search. [14] We're working on allowing the user export results and edit the model's actors and predicates.

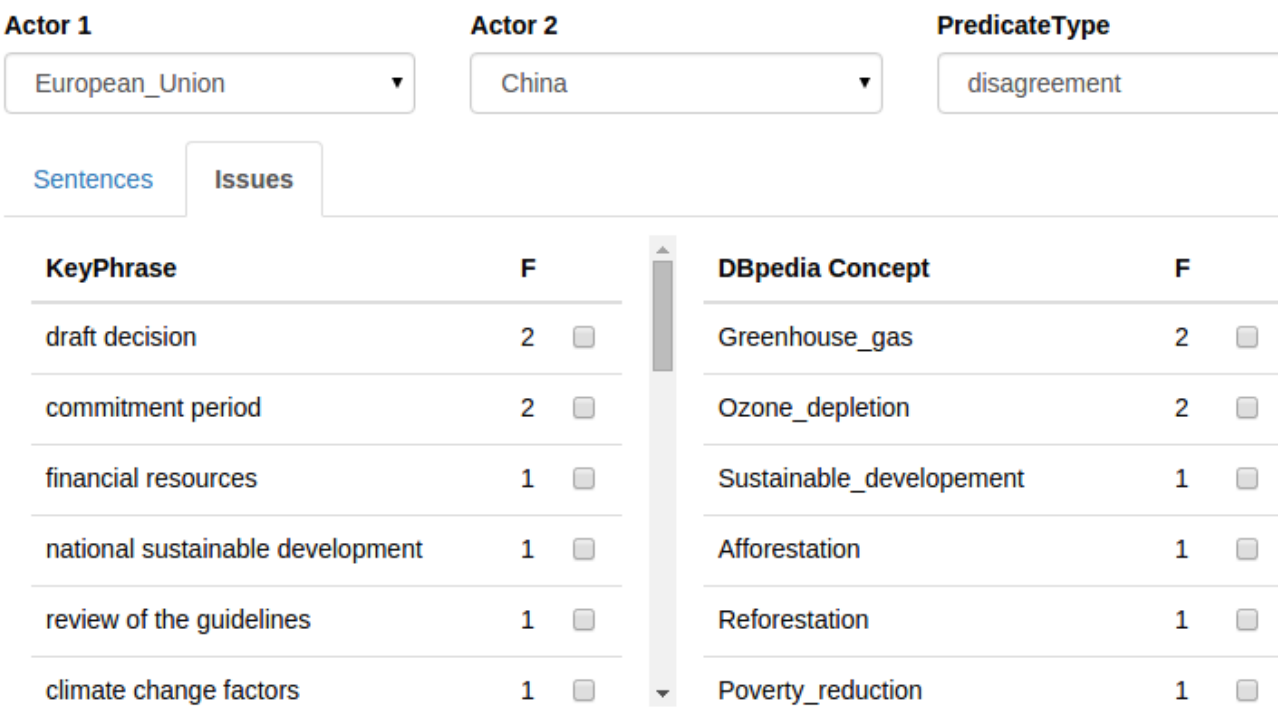

*Figure : **AgreeDisagree** View displays keyphrases and DBpedia concepts from propositions where actors (here the EU and China) agree or (as here) disagree.*

5. Evaluation

It is important to assess whether the system can help domain-experts gain insights they would not have otherwise obtained, e.g. detect previously unnoticed generalizations (see e.g. Berry, 2012). This type of evaluation is ongoing; we are collaborating with political scientists, whose initial feedback on the tool has been positive. User validation of the interface is also ongoing.

The system's NLP components were evaluated in literature cited above. Results are state-of-the-art or competitive, and available on our project's website (sites.google.com/site/nlp4climate).

To evaluate the model and analysis rules that create domain-relevant propositions, we have manually annotated a set of corpus sentences with propositions. Details about the test-set, evaluation metrics and results are on the website. We consider the results satisfactory.

6. Outlook

A useful feature would be an annotation confidence score, that users could employ to establish priorities in manual result revision. A useful application of the propositions extracted would be creating network graphs with different types of edges representing support and opposition among parties, and between parties and issues.

6.1. Acknowledgements

We thank Tommaso Venturini, Audrey Baneyx, Kari de Pryck and Diégo Antolinos-Basso from the Sciences Po médialab in Paris for domain-expert feedback on the system. Pablo Ruiz is supported by a PhD grant from Région Ile-de-France.

Bibliography

1. **Agerri, R., Bermudez, J. and Rigau, G**. (2014). IXA Pipeline: Efficient and ready to use multilingual NLP tools. In *Proceedings of LREC 2014, the 9th Language Resources and Evaluation Conference*. Reykjavik, Iceland.
2. **Aubin, S. and Hamon, T.** (2006). Improving Term Extraction with Terminological Resources. In *Advances in Natural Language Processing: 5th International Conference on NLP, FinTAL 2006*, LNAI 4139. Springer, pp. 380-87.

3. **Auer, S., Bizer, C., Kobilarov, G., Lehman, J., Cyganiak, R., and Ives, Z.** (2007). DBpedia: A nucleus for a web of open data. In *The Semantic Web*, Springer, pp. 722–35.
4. **Bauer, F., Recheis, D. and Kaltenböck, M.** (2011). Data. reegle. info–A new key portal for Open Energy Data. In *Environmental Software Systems. Frameworks of eEnvironment*, Springer Berlin Heidelberg, pp. 189-94.
5. **Berry, D. M.** (2012). *Understanding Digital Humanities,* pp. 1–20. Palgrave Macmillan.
6. **Björkelund, A., Bohnet, B., Hafdell, L. and Nugues, P.** (2010). A high-performance syntactic and semantic dependency parser. In *Coling 2010, 23 rd International Conference on Computational Linguistics: Demonstration Volume*, Beijing, pp. 33–36,
7. **Diesner, J.** (2012). *Uncovering and managing the impact of methodological choices for the computational construction of socio-technical networks from texts*. PhD Thesis. Carnegie Mellon University.
8. **Diesner, J.** (2014). ConText: Software for the Integrated Analysis of Text Data and Network Data. In *Social and Semantic Networks in Communication Research,* at *ICA*, *Conference of International Communication Association*.
9. **Grimmer, J. and Stewart, B. M.** (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, OUP, pp. 1–31.
10. **Mausam, Schmitz, M., Bart, R., Soderland, S. and Etzioni, O.** (2012). Open language learning for information extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 523–34.
11. **Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, V. and Grishman. R.** (2004). The NomBank project: An interim report. In *HLT-NAACL 2004 workshop: Frontiers in corpus annotation*, pp. 24–31.
12. **Palmer, M., Gildea, D. and Kingsbury, P.** (2005). The Proposition Bank: A Corpus Annotated with Semantic Roles. *Computational Linguistics Journal*, **31**: 1.
13. **Ruiz, P. and Poibeau, T.** (2015). Combining Open Source Annotators for Entity Linking through Weighted Voting. In *Proceedings of *SEM. Fourth Joint Conference on Lexical and Computational Semantics,* Denver, U.S., pp. 211–15.
14. **Salway, A., Toulieb, S. and Tvinnereim, E.** (2014). Inducing information structures for data-driven text-analysis. *Proceedings of the ACL Workshop on Language Technologies and Computational Social Science*, pp. 28–32.
15. **Surdeanu, M., Johansson, R., Meyers, A., Màrquez, L., and Nivre, J.** (2008). The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, Association for Computational Linguistics, pp. 159–77.
16. **Van Atteveldt, W., Sheaferm, T., Shenhav, S., and Fogel-Dror, Y.** (2015). Clause analysis: using syntactic information to enrich frequency-based automatic content analysis. In *Symposium New Frontiers of Automated Content Analysis in the Social Sciences*, at the University of Zurich.
17. **Venturini, T. and Guido, D.** 2012. Once upon a text: an ANT tale in Text Analytics. *Sociologica,* **3**: 1–17. Il Mulino, Bologna.
18. **Venturini T., Baya Laffite, N., Cointet, J-P., Gray, I., Zabban, V., and De Pryck, K.** (2014). Three maps and three misunderstandings: A digital mapping of climate diplomacy. *Big Data and Society*,**1**(2): 1–19.

Notes

1.

Predicate in the sense of an expression relating a set of arguments.

2.

http://www.iisd.ca/vol12

3.

wiki.dbpedia.org (Auer et al., 2007)

4.

https://sites.google.com/site/nlp4climate

5.

http://docs.cortext.net

6.

API: http://api.climatetagger.net ; Thesaurus: http://www.climatetagger.net/glossary/

7.

Terminology adopted: *‹Norway, preferred, legally-binding commitments›* is a proposition, with actor *Norway*, predicate *preferred* and *legally-binding commitments* as the negotiation point.

8.

http://ixa2.si.ehu.es/ixa-pipes/

9.

https://bitbucket.org/Josu/corefgraph

10.

https://github.com/newsreader/ixa-pipe-srl ; it provides a wrapper to *mate-tools* (Björkelund et al., 2010)

11.

http://search.cpan.org/~thhamon/Lingua-YaTeA/lib/Lingua/YaTeA.pm

12.

In SRL, *A0* corresponds to a predicate's agent. *A1* is the patient or theme. *AM* roles represent adjuncts (time, location etc.) or negation. See Palmer et al., 2005.

13.

https://www.djangoproject.com/

14.

https://lucene.apache.org/solr/