The #Idéo2017 platform
Longhi Julien

To cite this version:
Longhi Julien. The #Idéo2017 platform. Journée d’étude CORLI: Traitements et standardisation des corpus multimodaux et web 2.0., May 2018, Paris, France. hal-01873815

HAL Id: hal-01873815
https://hal.archives-ouvertes.fr/hal-01873815
Submitted on 13 Sep 2018

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.
The #Idéo2017 platform
Julien Longhi
Université de Cergy-Pontoise – Centre de Recherche AGORA (julien.longhi@u-cergy.fr)
Members of the project: Hassine N., Marinica C., Borzic B., Alkhoul A.

General context: the CoMeRe project
- Aims to build a kernel corpus of computer-mediated communication (CMC) genres in French
- Mono and multimodal interactions stemming from networks including the Internet and telecommunications that may be synchronous or asynchronous
- Members had previously collected and structured different types of CMC corpora within their local teams (in a variety of formats with disparities in corpus compilation choices)
- Corpora are structured and referred to in a uniform way in order that they may form part of the forthcoming French National Reference Corpus

First project: the Polititweet corpora
- Development of the Interaction Space (IS) model to model CMC interaction (Charrier & Jin, 2013).
- Includes descriptions of time, set of participants, online location(s) defined by the properties of the sets of environments used by participants.
- Description of the IS within the TEI header and messages and turns encoded in the TEI body using a common <post> element

Openness
- Corpora released as open data – paved way for scientific examination, replication and cumulative analyses
- Released on ORTOLANG (French equivalent of DARIAH the European infrastructure for Humanities)
- Bibliographic reference created for each corpus and given in <title>SMT of TEI header: e.g.

Openness
- Corpora released as open data – paved way for scientific examination, replication and cumulative analyses
- Released on ORTOLANG (French equivalent of DARIAH the European infrastructure for Humanities)
- Bibliographic reference created for each corpus and given in <title>SMT of TEI header: e.g.

Interests and limits
- The analysis of political tweets during the election campaigns, or specific events, is increasing and can be seen as a specific type of political discourse (Longhi, 2013).
- Extensive literature on the analysis of political tweets: but these works are difficult to gather because they come either from the computer sciences, either from the humanities and social sciences.
- Despite the unquestionable interest in outlining political facts, citizens do not have access to these results.

Challenges of #Idéo2017
- #Idéo2017 is a web platform available online allowing the analysis of the messages, posted on Twitter, related to political news.
- Its objective is to make available on the web for average citizens a set of statistical analyses and data visualization tools applied to the Twitter messages. #Idéo2017 allowed French electors to analyse the discourse of the candidates by means of their tweets.
- Citizens can make their own queries (based on linguistic and textometric criteria, more precisely, the most commonly used words by political personalities, analyses of similarities, ALCESTE algorithm, etc.) and obtain comprehensible results.

Tools and analysis

Analysis
The #Idéo2017 platform proposes:
The Analysis 1 Analyze the Tweets that Contain the Word [Word]
The Analysis 2 Analyze the Tweets of [Candidate]
The Search Engine

Tool Development
Extraction of the set of candidate's tweets
Extracting tweets
Removing tweets
Dividing tweets
Search Engine
Elaborate the analysis of tweets

Conclusions
- #Idéo2017 combines different technologies and inputs, which give citizens the opportunity to grasp a part of the discursive issues of the election.
- This development, which can be enriched, allows citizen to easily use a set of features usually accessible by software requiring different transformations of the data.
- For the period from September 1st 2016 to May 7th 2017, 42290 tweets were extracted for the 11 candidates. These tweets were gathered in a collection published in a TEI corpus in the standards of the Ortolang platform, thanks to the Coril Consortium.