Structuring a CMC corpus of political tweets in TEI: corpus features, ethics and workflow
Julien Longhi, Ciara R. Wigham

To cite this version:
Julien Longhi, Ciara R. Wigham. Structuring a CMC corpus of political tweets in TEI: corpus features, ethics and workflow. Corpus Linguistics 2015, Jul 2015, Lancaster, United Kingdom. 2015. halshs-01176061

HAL Id: halshs-01176061
https://halshs.archives-ouvertes.fr/halshs-01176061
Submitted on 14 Jul 2015

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 context of politweets: HumaNumDJ

This work is part of the “Digital Humanities and Data Journalism” transdisciplinary project (funded by the Foundation of the Cergy-Pontoise University, France).

The purpose of this research was to take advantage of discussions produced on social media to leverage the semantics of discourse in relation to social issues. The corpus was built starting from seven French politicians from six different political parties. In order to generate political tweets, a set of lists citing these politicians was generated (7087 lists), and 66% of 4649 tweets were recovered. In order to be sure that we selected politicians’ tweets (and not, for example, those of journalists), only the accounts cited in more than 12 lists were considered; 205 politicians were tweeting. We took the last 205 tweets of each of the 205 accounts on 27 March 2014 (13,273 tweets). This allowed us to recover data that focused on the period between the two rounds of the 2014 municipal elections in France.

Analyses have started to be carried out: some ideas have been launched in Djenli, Longhi et al. (2014) but further analyses must adhere rigorously to methodologies stemming from the natural language processing (NLP) field.

Openness

- Released on ORTOLANG (French equivalent of DARIAH the European infrastructure for Humanities)
- Bibliographic reference created for each corpus and given in «dcllmit» of TEI header.
- Members of the project are collaborating with the CMC
depository to release the data in the future.

Corpus structuring in TEI

- Development of the Interaction Space (IS) model to model CMC interaction (Charlier & 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.

Staged quality control process

- Using the Twitter API to get the last 205 tweets of each of the 205 accounts on 27 March 2014 (13,273 tweets).
- Retweets of each of the 205 politicians are added to the corpus.

The politweets

- The corpus is divided into two parts: the first is a study of the Twitter users and the second is dedicated to a study of the content generated by these users.
- The corpus is defined as a collection of Twitter messages from certain Twitter accounts, and the content of each message is defined as a set of elements.
- The corpus is defined as a collection of Twitter messages from certain Twitter accounts, and the content of each message is defined as a set of elements.
- The corpus is defined as a collection of Twitter messages from certain Twitter accounts, and the content of each message is defined as a set of elements.

Type of material

- Textual material
- Retweets: number of retweets

Ethical issues

- On https://twitter.com/losi4n-en we can read:
  8. Restrictions on Content and Use of the Services

  Please review the Twitter Rules (which are part of these Terms) to better understand what is prohibited on the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the aim of improving the quality of the Service. We reserve the right to alter these rules at any time, with the am