

#ww1. The Great War on Twitter

Frédéric Clavert

frederic@clavert.net

Université de Lausanne, Switzerland

Since 2014, some of the countries that were formerly belligerent of the Great War – most particularly France and UK – have organised a series of commemorations of the First World War, known as the ‘Centenaire’ (France) or the ‘Centenary’ (UK). We can assume that there is a strong link – that cannot let a historian indifferent – between those commemorations, collective memory and historical studies.

Though studies about collective memory are numerous since the famous works of the French sociologist Maurice Halbwachs (Halbwachs, 1950), few of them are examining how collective memories are being expressed – maybe even transformed – on social networks on-line.

In the case of the Centenary of the First World War, a set of questions can be asked: What is the on-line echo of the commemoration of the centenary of the 1st World War? What is the behaviour of Memorial/Heritage Institutions about the 1st World War on Twitter? How do they transmit information about the Centenary? Is there an influence of the English predominance on Twitter about the Centenary on how non-english-speaking twitter accounts are considering the 1st World War? Are there specific subjects that are discussed on-line? Which ‘temporalities’ are present in tweets when Twitter users speak about the Great War on-line?

Though we are not yet able to respond to all those questions, we’ll use our database of tweets in order to answer them at least partially.

Indeed, since the 1st April 2014, around 1.5 millions of tweets containing a hashtag (keyword) linked to the 1st World War were written by over 350 000 Twitter accounts in several languages (mainly English and French). Twitter is a good field to analyse relationships between history and collective memory, memorial institutions and citizens, historians and a wide non-academic audience. We started to explore this database (which is still expanding): we intend to show how a historian can collect, analyse and interpret those tweets, using Digital Humanities methodologies and software in order to answer questions about collective memory of the First World War online.

Tools and Methodologies

We are using 14odev, a PHP open source script within a LAMP environment to collect tweets through the Twitter streaming API. The tweets are then stored in a MySQL database. Diverse information (tweets and their metadata, hashtags, user information, mentions, retweets) about those tweets can easily be extracted through SQL queries.

Those queries can also be used to extract different kind of relations: between tweets, between Twitter users or even between hashtags (*ie* if a Twitter user mentioned or retweeted another twitter user, if two users wrote the same hashtags, etc). Concerning privacy, we respect the Twitter API Terms.

To analyse tweets, we are using mainly two sets of methodologies/software: social network analysis and network visualisations (with Gephi: mention, retweets or hashtags are considered a link); text analysis through the theory of the *mondes lexicaux* (Reinert, 1993) as it is implemented in the IRaMuTeQ software (Ratinaud and Dejean, 2009). The combination of both tools and methodologies has been described by (Smyrnaiois and Ratinaud, 2014). IRaMuTeQ, thanks to time-stamped metadata, can also help us working on temporalities. Indeed, clusters that are defined by this software can be projected in time: we can know, day-by-day, the most used kind of tweets. It helped us, for instance, finding that French fallen soldiers are not described with the same words the 11th of November in comparison to the rest of the year.

The methodologies and tools that remain to be found for this research concern temporalities – even if IRaMuTeQ has helped us answer some question on time. There are several temporalities that are expressed in this corpus: the constant feed of information that is the nature of Twitter; the temporality of each twitter user; the temporality of the Centenary (which is different from one country to the other, and from the Great War temporality); and the temporality of the War itself.

First results

Language

English is overwhelmingly present in this corpus. Around 10% only of the collected tweets are not in English. Among those 10%, French is largely in majority and German almost absent, even though German hashtags are collected. The fact that Twitter is an English-based social network does not explain fully this disequilibrium between English and other languages. The Memorial institutions’ communication policies on Twitter are better factors to explain it.

The decentralized communication policy of British memorial institutions (the BBC and all its Twitter accounts or the Imperial War Museum for instance) is obviously more efficient than the French centralized communication policy of the *Mission du centenaire*. French WW1-related museums do not have Twitter accounts or do have one but do not follow twitter implicit rules such as the use of a general hashtag like #ww1 or the French #pgm.

British and French are not commemorating WW1 the same way

The most striking difference between the French corpus and the English one is the fact that both linguistic areas

do not commemorate the Great War the same way. There are two major differences between both countries:

- French are mainly remembering the soldiers (*Poilus*). British citizens are remembering soldiers, but also battles.
- The French are focusing on the end of the war, the Armistice, on the 11th November. The British are focusing on the way they entered the war.

English public history and French history amateurs

Thanks to the Network visualisations, this corpus also helps understand how public history is present in Britain, in contrary to France where it just begins to appear. The presence of amateurs of history in the French corpus also shows that French historians are not on twitter, in contrary to amateurs who, next to the *Mission du Centenaire*, are structuring discussions about the First World War on Twitter.

Conclusion

Comparing multilingual corpora

To compare our two main corpora (the French one and the English one) that can be extracted from the database, we had to use the two main pieces of software the same way on both corpora and then to 'humanly' compare the results. We could not find any tools able to compare two corpora that are in different languages.

Distant reading / Close reading

This research project shows that, for historians, it is still important to keep a direct link with each single primary source, as some information can be learned from the interpretation of single tweets. Though methods used in this research are dealing with Franco Moretti's notion of *distant reading* (Moretti, 2007), it proved strategic to be able to go back to every single tweet. The software used, if metadata are kept all along their use, allow this.

Twitter and the rest of the web

Why Twitter? The fact that the Twitter API, though sometimes very unstable, is very convenient to use is one of the criteria of this choice. Is it really pertinent in terms of research? Shouldn't we have broader sources? How to extrapolate the project's results to other on-line social networks? Last but not least, the difficulty to anticipate hashtags to be collected might introduce biases in our research.

Future of this research project

The question of 'temporalities' and their imbrications (the temporality of Twitter / the temporality of users / the temporality of the commemorations / the temporality of the First World War itself) should be the next step

of this research. But, as it will require the use of Named Entity Recognition, extending our research to places will be possible as well.

Bibliography

- Halbwachs, M.** (1950). *La mémoire collective*, Paris: Albin Michel.
- Moretti, F.** (2007). *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.
- Ratinaud, P. and Dejean, S.** (2009). IRaMuTeQ: Implémentation de la methode ALCESTE d'analyse de texte dans un logiciel libre [Implementation of the ALCESTE method of text analysis in an open-source software]. *Presentation*. Available at: http://reperer.no-ip.org/Members/pratinaud/mes-documents/articles-et-presentations/presentation_mashs2009.pdf (accessed 4 March 2016).
- Reinert, M.** (1993). Les 'mondes lexicaux' et leur 'logique' à travers l'analyse statistique d'un corpus de récits de cauchemars. *Language et Société*, **66**(1): 5–39. doi:10.3406/lsoc.1993.2632
- Smyrnaio, N. and Ratinaud, P.** (2014). Comment articuler analyse des réseaux et des discours sur Twitter. *tic and société*, **7**(2). doi:10.4000/ticetsociete.1578

Notes

- ¹ <http://14odev.com/> (accessed 4 March 2016).
- ² <http://www.iramuteq.org/> (accessed 4 March 2016) - Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. IRaMuTeQ is a free software based on python and R. It is available in French, English, German and Spanish (interface and analyses).
- ³ IRaMuTeQ works in dividing the corpus in small segments of text (around 40 words). In our case each segment is a tweet and each tweet is also a text.