Oral History and Linguistic Analysis. A Study in Digital and Contemporary European History

Florentina Armaselu
Luxembourg Centre for Contemporary and Digital History (C^{2}DH)
University of Luxembourg
florentina.armaselu@uni.lu

Elena Danescu
Luxembourg Centre for Contemporary and Digital History (C^{2}DH)
University of Luxembourg
elena.danescu@uni.lu

François Klein
Luxembourg Centre for Contemporary and Digital History (C^{2}DH)
University of Luxembourg
francois.klein@uni.lu

Abstract

The article presents a workflow for combining oral history and language technology, and for evaluating this combination in the context of two use cases in European contemporary history research and teaching. Two experiments have been devised to analyse how interdisciplinary connections between history and linguistics are built and evaluated within a digital framework. The longer-term objective of this type of enquiry is to draw up an "inventory" of strengths and weaknesses and potentially build an online collection of use cases to share reflections and render more transparent the process of applying language technology to research and teaching in different areas of study in the humanities.

1 Introduction

To what extent can the combination of digital linguistic tools and oral history assist research and teaching in contemporary history? How can this combination be evaluated? Is there any added value in using linguistic digital methods and tools in historical research/teaching as compared with traditional means? What are the benefits and limitations of this type of method? The paper will address these questions starting from two experiments based on an oral history collection, XML-TEI annotation and textometric analysis.

In her outline of an oral history "à la française", Descamps (2013: 109-110) talks about a "linguistic age" or a "first age of recorded speech" starting in the 1910s when language scientists began to show an interest in oral sources. With the "invention of oral history, in the 1960s", the use of the spoken word emerged in the historical discipline, subsequently becoming an "indispensable method for contemporary history".[1] Various linguistic aspects have since been considered in the study of spoken corpora. More traditional approaches dealt with this type of data from a number of different perspectives, such as formal and functional narrative analysis of oral versions of personal experiences (Labov and Waletzky, 1967), discursive analysis of the construction of gender identity in life story interviews (Slabakova, 2016), linguistic analysis of metaphor and agency in narrative-biographical interviews (Leonardi, 2018) or close reading by applying discourse analysis and systemic functional linguistics to human rights-related testimonies (Bock, 2007). Digitally oriented research, on the other hand, adopted methods such as topic modelling and sentiment analysis for oral communication data (Choudhury et al., 2018), discourse structure analysis and automatic segmentation of speech corpora transcripts (Zhang and Soergel, 2006), word frequency and co-occurrence computation and qualitative

[1] Fr. "[...] un premier âge de la parole enregistrée, *l'âge linguistique* [...]"; "[...] l'invention de l'histoire orale, dans les années 1960 [...]"; "[...] une « méthode » incontournable de l'histoire contemporaine." (Descamps, 2013: 109).

Florentina Armaselu, Elena Danescu and François Klein 2019. Oral History and Linguistic Analysis. A Study in Digital and Contemporary European History. *Selected papers from the CLARIN Annual Conference 2018*. Linköping Electronic Conference Proceedings 159: 15–27.

analysis for oral history life-course interviews (Hájek and Vann, 2015), and corpus linguistics for dialect speech data or for self-representation in life story interviews (Anderwald and Wagner, 2007; Sealey, 2009).

Since the mid-1970s, the resources and methods of European integration history research have been enhanced with sources from oral history, which are now regularly used alongside both traditional and digital text- and image-based sources (archives, published material, official publications, etc., as well as Web archives and online databases). This "epistemological continuity between written and oral sources" (Bloch, 1999) confirms that oral sources and resources are contributing to the creation and transmission of historical knowledge, while also adding a dimension related to memory and heritage (Ritchie, 2003). Oral history can be seen as a "negotiated history" (Janesick, 2010) or as an "intermediated, influenced history" (Descamps, 2006). In other words, it is "recreated" by the historian in cooperation with the interviewee. It is, therefore, a subdiscipline of the humanities in which critical analysis remains vital. Oral sources are complementary and often prolific, but they should never be viewed in isolation; historians must constantly compare and contextualise them by referring to other sources, especially written sources, which confirm or refute them.

Bridging oral history and linguistics in a digital context has also been the object of dedicated event-oriented initiatives and research, both inside and outside the framework of CLARIN (CLARIN-PLUS OH, 2016; Oral History meets Linguistics, 2015; Georgetown University Round Table on Languages and Linguistics, 2001). Within this context, different tools and perspectives have been adopted, such as language technologies for annotating, exploring and analysing spoken data (Drude, 2016; Van Uytvanck, 2016; Van Hessen, 2016), online platforms for Multimodal Oral Corpus Analysis (Pagenstecher and Pfänder, 2017) or the use of oral histories as "data" for discourse analysts (Schiffrin, 2003).

However, the question of how oral history and linguistics may impact the historian's exploration and interpretation of data seems so far to have been the focus of less research. The theme of digital tool adoption by humanist scholars, and in particular by historians, has already been addressed, either within the scope of tool-building projects and attempts to identify user needs (Gibbs and Owens, 2012; Kemman and Kleppe, 2014) or within the areas of digital tool criticism and digital hermeneutics (Traub and Van Ossenbruggen, 2015; Koolen et al., 2018). Our study is situated in between these approaches: it explores how digital linguistic methods are applied (to answer specific research questions) and perceived by historians (especially as far as added value and innovative potential are concerned). It presents a methodology for preparing and analysing oral history data via tools of corpus linguistics and for observing the "human factor" while dealing with this language technology to accomplish history-related tasks. The proposal aims to contribute to this topic (which in our opinion is of potential interest for the CLARIN community, as it is related to building and evaluating interdisciplinary connections between history, linguistics and digital technologies) and consists of a workflow for: (1) transforming and processing historical spoken data intended for linguistic analysis; (2) evaluating the impact of the use of language technologies in historical research and teaching.

2 Methodology

The growing enthusiasm among the European Union (EU) institutions for oral history on the theme of European integration[2] has led to the systematic use of audiovisual sources for university-based research in this field. Adopting this approach, the Centre virtuel de la connaissance sur l'Europe (CVCE) composed an extensive collection of original historical interviews (more than 160 hours)[3]

[2] Since 1997, the European institutions (Commission, Parliament and Council) have begun gathering a series of oral accounts which have now been compiled into a dedicated collection within the Historical Archives of the European Union. The European Commission was a pioneer in this field, with its "Voices of Europe" programme (1997) (a collection of oral accounts from politicians, diplomats and senior officials who made a significant contribution to the European integration process and its early developments) and "European Commission (1958-1972) – History and memories of an institution" (2002) (a series of oral accounts on the history of the European Commission at the time of the Six, from the creation of the Common Market and Euratom institutions to the eve of the first enlargement). Since 2009, the European Parliament has been building up an oral history collection entitled "Oral history of the European Parliament Presidents" (http://www.europarl.europa.eu/historicalarchives/en/multimedia-gallery/interviews-of-the-presidents.html).

[3] https://www.cvce.eu/histoire-orale. The CVCE is now part of the Luxembourg Centre for Contemporary and Digital History (C²DH) at the University of Luxembourg, https://www.c2dh.uni.lu/.

with key actors and witnesses of the European integration process from Luxembourg and Europe, conducted in French, English, German, Spanish and Portuguese (Klein, 2011-2017).

The present study is based on a selection from this oral history collection, focused particularly on the topic of Economic and Monetary Union (EMU). These interviews represented entirely new sources for the topic under examination and more broadly for the research community as a whole and, given their heritage value, for other sectors of the public. The selection referred to in this paper included 5-10 hours of filmed recordings and transcriptions, in French. The selected transcriptions were converted to a structured format, XML-TEI[4], then imported into the TXM[5] textometry software (Heiden et al., 2010) for linguistic analysis. Two experiments were devised. The first (EUREKA_2017) functioned as a pilot using a smaller corpus and involved a small group of C^{2}DH researchers. The second (MAHEC_2018) was part of a course in Political and Institutional History for Master's students in Contemporary European History at the University of Luxembourg. For each experiment, a set of research questions was prepared, and questionnaires were designed to investigate the role of the language technology in answering these research questions (or in identifying other related research questions).

2.1 Corpus selection and research questions

The "History of European political integration" course, part of the Master's in Contemporary European History, looks at the history of European integration from the early 20th century to the Treaty of Maastricht in 1993 from a political and institutional angle. The learning objectives are not just to provide students with a solid grounding in the political and institutional processes involved in European integration (its origins and development, interconnected structures, mechanisms and players, etc.), including the role played by Luxembourg and its elites, but above all to give them the skills they need to apply critical examination and analysis techniques to the various conceptual and historical perspectives on the building of a united Europe. In terms of methodology, it is hoped that the use of digital primary sources (textual, audio, visual) and methods and tools for digital analysis and visualisation will foster a new historical approach and facilitate access to the complex issues involved in the European integration process.

In light of these goals, we identified the topic of EMU as being of particular interest. EMU not only represents a vital stage in European integration, of which the euro is a tangible result; it is also a valuable object of study in terms of the lessons in economic governance learned following the 2008-2018 economic and financial crisis. Examining the historical processes that gave rise to these events can help shed light on early warning signs pointing to the crisis and avenues for resolution. The corpus that was compiled for the course arose from the "Pierre Werner and Europe" interdisciplinary research project, which was based on a thorough exploration of the Werner family private archives, opened for the first time for research purposes (Danescu, 2013). A series of historical interviews (see Appendix) conducted with key figures from Luxembourg and the international community (more than 55 hours of footage in total) complement the extensive research carried out in these and other archives, offering added value and new resources for the research community.

The corpus developed for the EUREKA and MAHEC experiments is composed of original oral history sources that particularly focus on the plan for the establishment by stages of an economic and monetary union (the Werner Report), the events that subsequently led to EMU, the Luxembourg Compromise, the accession of the United Kingdom, and cooperation between the Benelux countries and the Belgium-Luxembourg Economic Union (BLEU).[6] The number of selected interviewees varied from six (EUREKA) to eight (MAHEC), including figures such as Jean-Claude Juncker, Viviane Reding, Jacques Delors and Étienne Davignon (see Appendix). The selection criteria focused on important milestones in the development of the European Union and the interviews had to be in French for homogeneity purposes. One research question was proposed for the pilot experiment and seven for the second. They were either general queries, e.g. discern the multiple dimensions of the European integration process (EUREKA), or more specialised questions related to the topic of the

[4] http://www.tei-c.org/index.xml.

[5] http://textometrie.ens-lyon.fr/?lang=en.

[6] All these interviews, their transcriptions and translations into English, French and German are published in E. Danescu, The Werner Report of 8 October 1970 in the Light of the Pierre Werner Family Archives (research corpus), Source:. https://www.cvce.eu/project/werner/.

course, e.g. identify the European institutions mentioned in the interviews, their role and interconnections, reconstruct the process of Economic and Monetary Union or determine which of the interviewees is speaking more about Luxembourg's role in European integration, which less, and why (MAHEC).

2.2 Corpus preprocessing

Figure 1 shows the general workflow for preprocessing the corpus before TXM analysis. The filmed recordings were first transcribed[7] into Microsoft Word or Open Office formats. In this project we used the transcriptions as Microsoft Word files that contained markers for identifying the interviewer/respondent and, occasionally, timecodes. As the interviews were structured and included tables of contents and sections, heading styles were added to mark section titles in the documents.

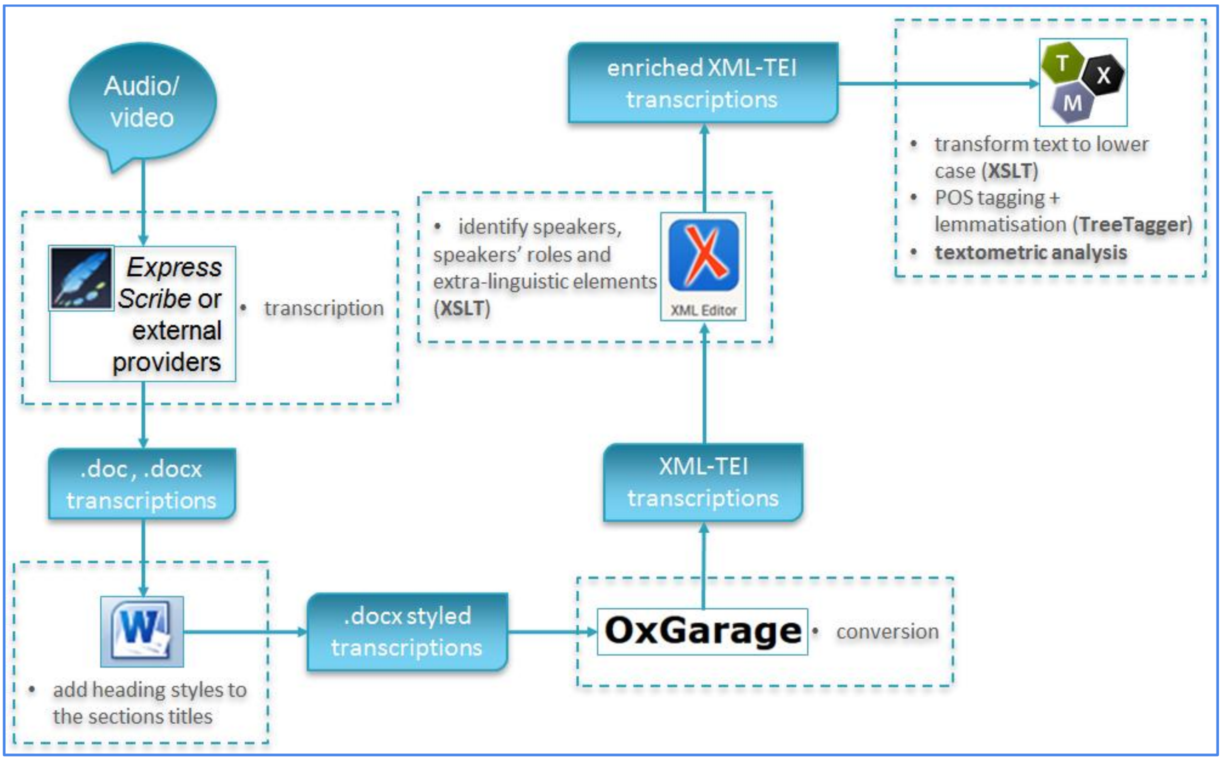
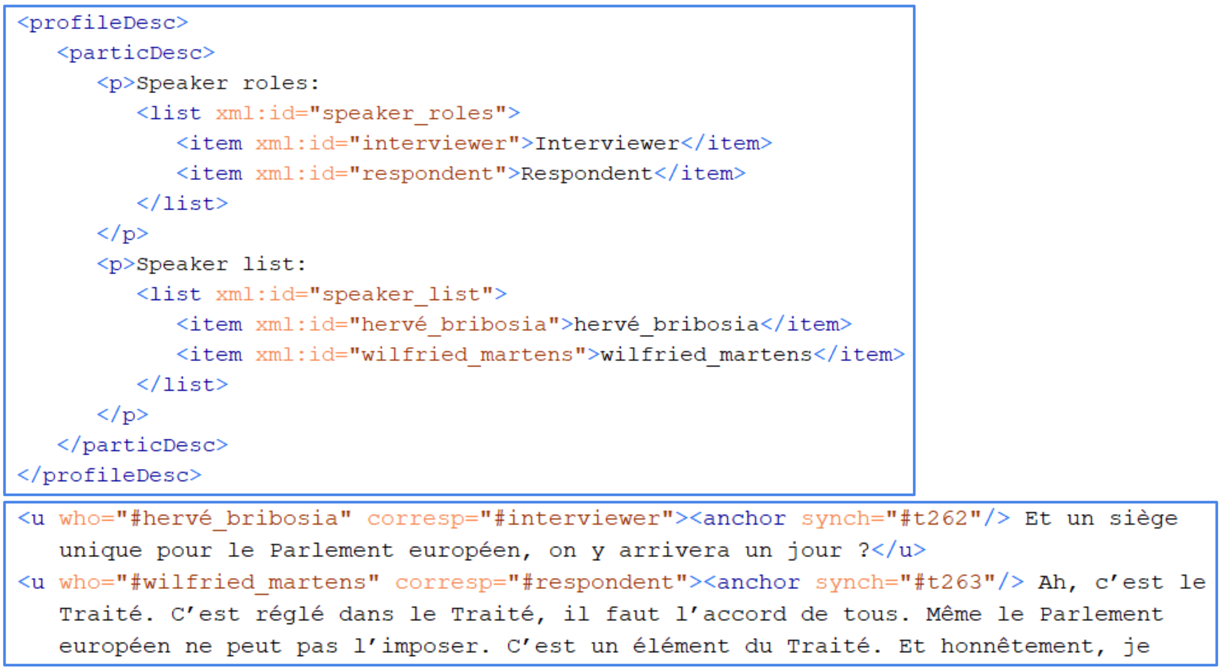

Audio/ video
Express Scribe or external providers
transcription
.doc, .docx transcriptions
add heading styles to the sections titles
.docx styled transcriptions
OxGarage
conversion
XML-TEI transcriptions
identify speakers, speakers' roles and extra-linguistic elements (XSLT)
XML Editor
enriched XML-TEI transcriptions
T X M
transform text to lower case (XSLT)
POS tagging + lemmatisation (TreeTagger)
textometric analysis

Figure 1. Preprocessing workflow for TXM analysis

The transcriptions were first converted from styled Microsoft Word .docx to a raw XML-TEI[8] version that contained only generic encoding for the metadata in the teiHeader and divisions, paragraphs or highlighting marks for the content (body) area. A series of XSLT[9] stylesheets, specially created for this purpose, were then applied to the converted output,[10] in order to transform it into specific TEI encoding for the transcription of speech. Additional information was inserted into the teiHeader, e.g. speaker roles and speaker list in the participant description to serve as a reference that could be pointed to from the body of the text.

The extract in Figure 2 illustrates how the identity (name) and type of speaker (interviewer/respondent) were encoded in the <particDesc> area from the teiHeader and by using the <u> tag (utterance) and the @who and @corresp attributes in the body of the document. Time points (when present) were encoded by <timeline> and <anchor/> elements, in order to mark the text with respect to time. Extra-linguistic aspects, although rare in the selected data, e.g. <pause>, <kinesic>, marking a pause within the utterance or a gesture, were also considered.

[7] From H264 format, using *Express Scribe* (https://www.nch.com.au/scribe/index.html) or by external providers.
[8] Via the OxGarage online service, http://www.tei-c.org/oxgarage/.
[9] https://www.w3.org/TR/xslt/.
[10] Using oXygen XML Editor, https://www.oxygenxml.com/.



```
<profileDesc>
   <particDesc>
      <p>Speaker roles:
         <list xml:id="speaker_roles">
            <item xml:id="interviewer">Interviewer</item>
            <item xml:id="respondent">Respondent</item>
         </list>
      </p>
      <p>Speaker list:
         <list xml:id="speaker_list">
            <item xml:id="hervé_bribosia">hervé_bribosia</item>
            <item xml:id="wilfried_martens">wilfried_martens</item>
         </list>
      </p>
   </particDesc>
</profileDesc>
```

```
<u who="#hervé_bribosia" corresp="#interviewer"><anchor synch="#t262"/> Et un siège
   unique pour le Parlement européen, on y arrivera un jour ?</u>
<u who="#wilfried_martens" corresp="#respondent"><anchor synch="#t263"/> Ah, c'est le
   Traité. C'est réglé dans le Traité, il faut l'accord de tous. Même le Parlement
   européen ne peut pas l'imposer. C'est un élément du Traité. Et honnêtement, je
```

Figure 2. XML-TEI encoding of speakers and utterances – interview with Wilfried Martens

2.3 TXM analysis

TXM is a piece of textometry software based on a methodology allowing quantitative and qualitative analysis of textual corpora by combining developments in lexicometric and statistical research with corpus technologies (Unicode, XML, TEI, NLP, CQP, R) (TXM Manual; TXM Website).

The corpus in XML-TEI format was imported into TXM, lemmatised[11] and parts of speech were tagged. An XSLT stylesheet was also created and applied during the import to convert the text to lower case. The analysed samples contained a total of 38,687 (EUREKA) and 110,563 (MAHEC) word occurrences. Given the encoding, it was possible to build sub-corpora and partitions corresponding to the name and type of the speaker. Separate sub-corpora were created for interviewer and respondent, respectively, and inside them, partitions for the speakers corresponding to each role, by selecting a structural element (<u>) and an appropriate property (attribute @corresp or @who). Taking into account their potential for contrasting and quantitative/qualitative exploration, the following TXM features were recommended to the participants to be used in their tasks of finding answers to the proposed questions or formulating new research questions: specificities[12] (Lafon, 1980), index, concordances and co-occurrences (TXM Manual).

Figure 3 illustrates specificities, that is a comparative view of the vocabularies of the respondents. The tool allows direct computation of specificities, based on a single property (e.g. *word*, *lemma*, *part of speech*) or more complex processing. For instance, particular queries can be entered via the index using single properties or a combination of properties (e.g. different parts of speech). Lexical tables,[13] specificity scores and diagrams may then be built based on query results. The figure shows the results of computing specificities for the combination *noun* + *adjective*. For the top five European institutions most frequently mentioned in the text, an overuse can be observed for *banque centrale*[14] (first vertical bar for each speaker) in the discourse of Yves Mersch and Jean-Claude Juncker (speakers 8 and 5), and an underuse in the speech of Étienne Davignon (speaker 2), with scores over or under a banality threshold of +/- 2.0 marked by horizontal red lines in the figure.

[11] Via TreeTagger.

[12] "The Specificities command calculates a statistic indicating whether in each part of a partition the occurrences of a word or CQL query appear in abundance (or in decline)." (TXM Manual: 94)

[13] "A Lexical Table assembles together the different lexical units of a partition and displays them in table form." (TXM Manual: 111)

[14] Eng. "Central Bank."

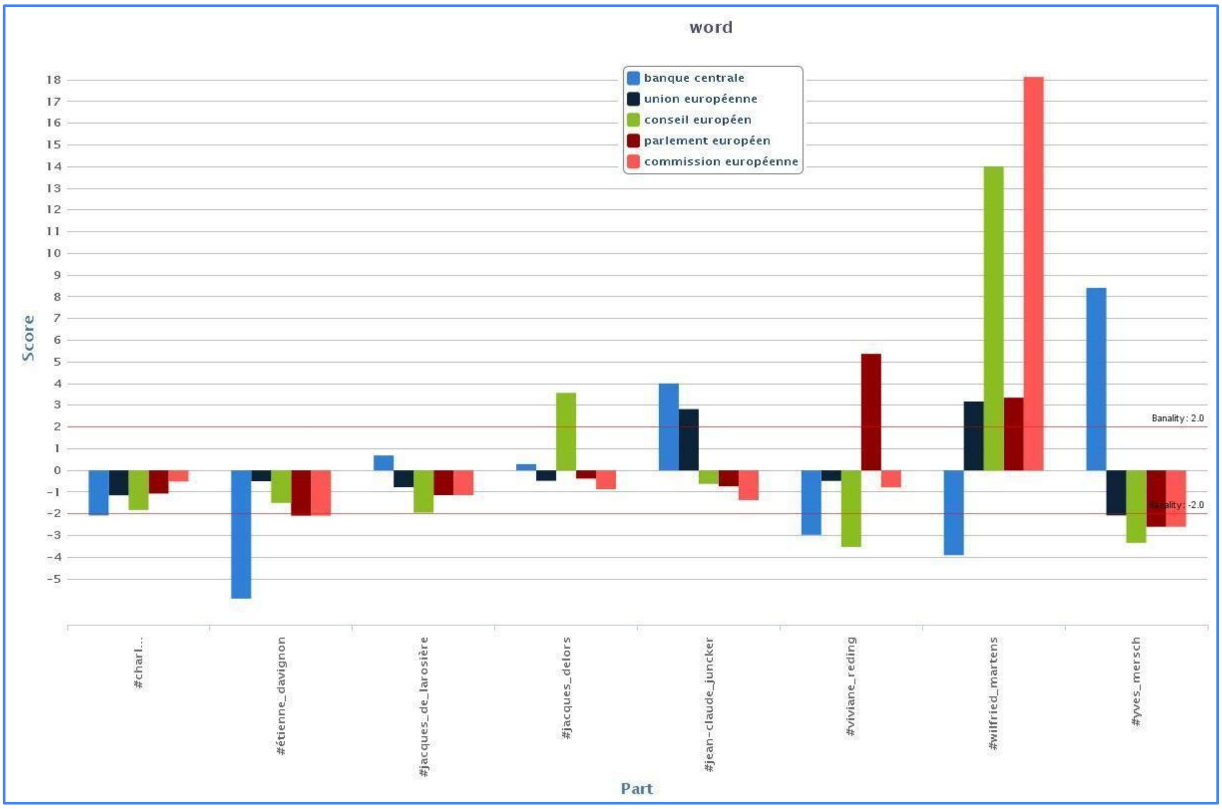
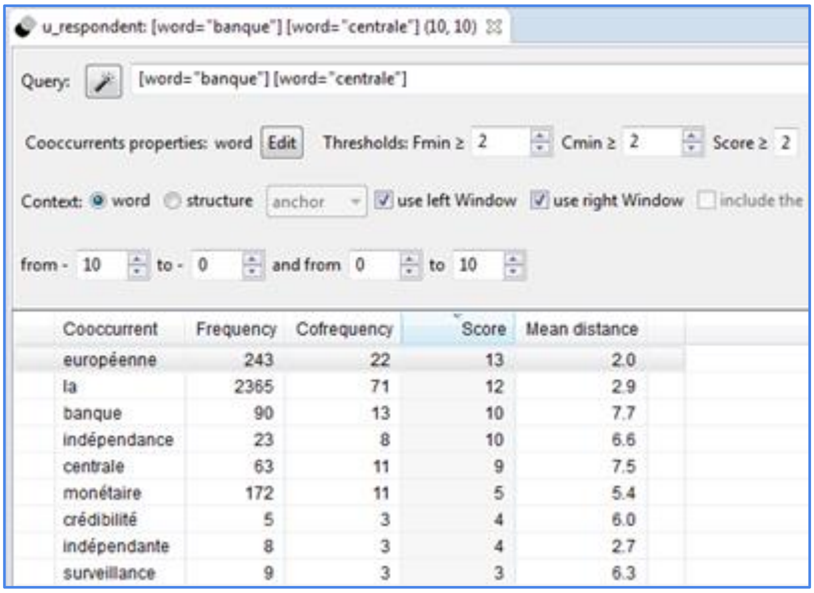


Figure 3. Specificities for European institutions within the respondents' partition (MAHEC_2018)

Other features allowed detection of forms having a tendency to occur together (co-occurrences, e.g. *banque centrale* + *européenne*) or a switch from a synthetic, tabular view to mini-contexts (concordances, e.g. *la banque centrale européenne est en charge de la politique monétaire …*[15]) or document visualisation (Figure 4a, b).



u_respondent: [word="banque"] [word="centrale"] (10, 10)

Query: [word="banque"] [word="centrale"]

Cooccurrents properties: word Edit Thresholds: Fmin ≥ 2 Cmin ≥ 2 Score ≥ 2

Context: ● word ○ structure anchor ☑ use left Window ☑ use right Window □ include the

from - 10 to - 0 and from 0 to 10

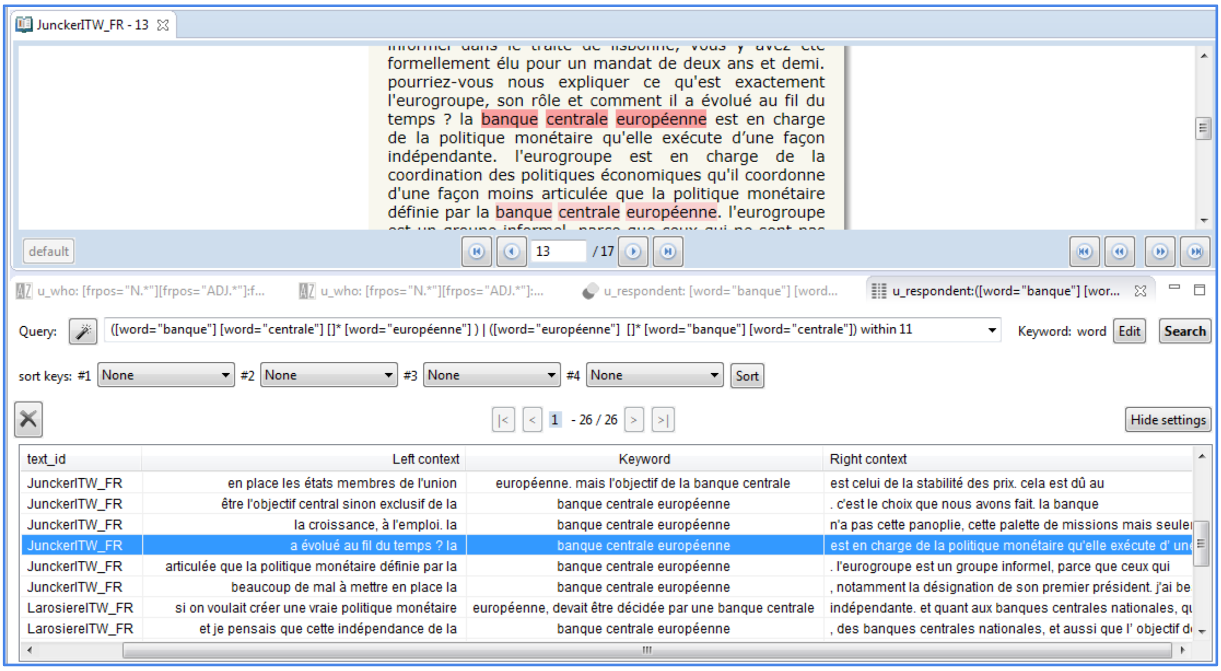
Cooccurrent	Frequency	Cofrequency	Score	Mean distance
européenne	243	22	13	2.0
la	2365	71	12	2.9
banque	90	13	10	7.7
indépendance	23	8	10	6.6
centrale	63	11	9	7.5
monétaire	172	11	5	5.4
crédibilité	5	3	4	6.0
indépendante	8	3	4	2.7
surveillance	9	3	3	6.3

Figure 4a. Co-occurrences (MAHEC_2018)

Our hypothesis was that this type of linguistic analysis, mingling quantitative and qualitative perspectives, may help the participants in their quest for answers to the proposed questions or new questions. For instance, we assumed that different dimensions of European integration (e.g. monetary, economic, political, diplomatic or legal) may be discerned by analysing the specific vocabularies of each of the interviewees as an expression of the particular roles they played in the process (EUREKA). It was supposed that more precise questions may be answered as well. For example, examining the combinations of *pronoun* + *verb* or *noun* + *adjective*, *noun* + *noun,* query by *numerals*, etc. and their specific usage in the respondents' speech may provide insight into nuanced role distinctions such as *actor/witness* in the events

[15] Eng. "the European Central Bank is in charge of monetary policy … ."

discussed, enable identification of important entities (institutions and key figures) and their respective roles or highlight temporal milestones and how concepts have evolved (MAHEC). Given that the degree of familiarity of the participants with the tool was not high and the aim of the experiments was also didactic, suggestions for possible paths of exploration in TXM were made either when assisting with the tasks (EUREKA) or within the assignment instructions themselves (MAHEC). At the same time, the participants were encouraged to look for alternative solutions on their own.


JunckerITW_FR - 13
default
13
/ 17
u_who: [frpos="N.*"][frpos="ADJ.*"]:f...
u_who: [frpos="N.*"][frpos="ADJ.*"]:...
u_respondent: [word="banque"] [word...
u_respondent:([word="banque"] [wor...
Edit
Search
Sort
|<
<
>
>|
Hide settings

formellement élu pour un mandat de deux ans et demi. pourriez-vous nous expliquer ce qu'est exactement l'eurogroupe, son rôle et comment il a évolué au fil du temps ? la banque centrale européenne est en charge de la politique monétaire qu'elle exécute d'une façon indépendante. l'eurogroupe est en charge de la coordination des politiques économiques qu'il coordonne d'une façon moins articulée que la politique monétaire définie par la banque centrale européenne. l'eurogroupe

Query: ([word="banque"] [word="centrale"] []* [word="européenne"]) | ([word="européenne"] []* [word="banque"] [word="centrale"]) within 11

Keyword: word

sort keys: #1 None #2 None #3 None #4 None

1 - 26 / 26

text_id	Left context	Keyword	Right context
JunckerITW_FR	en place les états membres de l'union	européenne. mais l'objectif de la banque centrale	est celui de la stabilité des prix. cela est dû au
JunckerITW_FR	être l'objectif central sinon exclusif de la	banque centrale européenne	. c'est le choix que nous avons fait. la banque
JunckerITW_FR	la croissance, à l'emploi. la	banque centrale européenne	n'a pas cette panoplie, cette palette de missions mais seule
JunckerITW_FR	a évolué au fil du temps ? la	banque centrale européenne	est en charge de la politique monétaire qu'elle exécute d' une
JunckerITW_FR	articulée que la politique monétaire définie par la	banque centrale européenne	. l'eurogroupe est un groupe informel, parce que ceux qui
JunckerITW_FR	beaucoup de mal à mettre en place la	banque centrale européenne	, notamment la désignation de son premier président. j'ai be
LarosiereITW_FR	si on voulait créer une vraie politique monétaire	européenne, devait être décidée par une banque centrale	indépendante. et quant aux banques centrales nationales, qu
LarosiereITW_FR	et je pensais que cette indépendance de la	banque centrale européenne	, des banques centrales nationales, et aussi que l' objectif d

Figure 4b. Concordances and document view (MAHEC_2018)

2.4 Evaluation

The evaluation[16] was intended to confirm/disconfirm the above-mentioned hypothesis and to "measure" the impact of the linguistic technology, its innovative aspects and limitations, when applied to the study of history. Evaluation questionnaires were designed via Google Forms and made available at the end of each test phase or assignment. For anonymisation purposes, identification codes (ID) were assigned to the participants and distributed in sealed envelopes before they answered the questionnaires. The links to the questionnaires were communicated by email (EUREKA) or via the MOODLE page of the course (MAHEC). The language used for the questions/answers was French, as for all the materials (instructions, tutorials, slips indicating ID codes) previously distributed.

Each questionnaire was designed to contain three sections: (1) Participant, including: participant ID code list, age range and gender, main field of expertise, self-evaluation on a scale of 1 to 5 (*Not at all* to *Expert*) in the fields of: *European integration history*, *Multimedia and oral history* (EUREKA) and *Textometry*. Agreement to use the anonymised answers for research/publication was explicitly required by a *Yes/No* question. All the answers from this section were mandatory. (2) Evaluation of the: a) multimedia technology and the oral history collection (EUREKA, first phase); b) textometric analysis (EUREKA, second phase; MAHEC). (3) Evaluation of the proposed experimental scenario.

The overall protocol and questionnaire content were simplified for the second experiment in order to make the students' work more straightforward. However, the general structure, most of the questions and the types of queries were maintained. Sections 2 and 3 above included three types of questions: (1) *Yes/No* questions; (2) Likert-scale queries (with five possible answers from *Don't agree at all* to *Fully agree, Very weak* to *Essential* or *Not at all interesting* to *Very interesting*); (3) open questions.[17]

[16] For legibility purposes, English translations of sample questions/answers are provided in footnotes. When the description applies to only one experiment/phase, this is mentioned in brackets; otherwise, the prose/examples apply to both experiments.

[17] Examples: (1) "Did you find answers to the research questions?"; "Would you like to formulate other research questions

3 The experiments

3.1 Description

The pilot experiment EUREKA_2017[18] took place from 11 to 15 and 18 to 22 September 2017 and involved the study of: (1) online filmed interview sequences (5 hours, 6 interviewees) and transcriptions; (2) XML-TEI transcriptions imported into TXM and ready for analysis (sub-corpora for respondents and interviewers and corresponding speaker partitions were provided). The participants were four C^{2}DH researchers specialised in *European integration*, *Contemporary history* and *History and political science*. While the profile data showed specialisation in European integration with medium knowledge in multimedia and oral history, the self-evaluation of the textometry skills was placed at the lower end of the scale (Table 1, left).

The second experiment, MAHEC_2018, involved five Master's students in Contemporary European History and took place from 16 April to 14 May 2018. The data sample contained transcriptions of interviews (10 hours, 8 interviewees) in XML-TEI format imported into TXM (with sub-corpora and partitions prepared in advance). Links to access the selected video sequences online were also provided but their analysis was not part of the tasks (however, a student reported having consulted them to learn more about the history of European integration). The students' backgrounds varied from *History* and *Contemporary European history* to *Mediaeval history*, with medium and good knowledge of *European integration history* reported. Compared with the previous experiment, the self-evaluation of the textometric analysis skills covered a larger spectrum (Table 1, right).

<table>
<tr><th colspan="11">EUREKA_2017</th><th colspan="11">MAHEC_2018</th></tr>
<tr><th colspan="2">Age range</th><th colspan="2">Gender</th><th colspan="2">Area of expertise</th><th colspan="5">Knowledge</th><th colspan="2">Age range</th><th colspan="2">Gender</th><th colspan="2">Area of expertise</th><th colspan="5">Knowledge</th></tr>
<tr><td rowspan="3"><i>20 – 34</i></td><td rowspan="3"><b>1</b></td><td rowspan="9">F
M</td><td rowspan="9"><b>3</b>
<b>1</b></td><td rowspan="3"><i>European integration</i></td><td rowspan="3"><b>1</b></td><td colspan="5"><i>History of European integration</i></td><td rowspan="9"><i>18 – 34</i></td><td rowspan="9"><b>5</b></td><td rowspan="9"><i>F</i>
<i>M</i></td><td rowspan="9"><b>1</b>
<b>4</b></td><td rowspan="3"><i>History</i></td><td rowspan="3"><b>2</b></td><td colspan="5"><i>History of European integration</i></td></tr>
<tr><td colspan="5"><i>None at all</i> <i>Expert</i></td><td colspan="5"><i>None at all</i> <i>Expert</i></td></tr>
<tr><td></td><td></td><td><b>1</b></td><td><b>3</b></td><td></td><td></td><td></td><td><b>3</b></td><td><b>2</b></td><td></td></tr>
<tr><td rowspan="3"><i>35 – 44</i></td><td rowspan="3"><b>2</b></td><td rowspan="3"><i>Contemporary history</i></td><td rowspan="3"><b>2</b></td><td colspan="5"><i>Multimedia + Oral history</i></td><td rowspan="3"><i>Contemporary history</i></td><td rowspan="3"><b>2</b></td><td colspan="5" rowspan="4"><i>Textometry</i></td></tr>
<tr><td colspan="5"><i>None at all</i> <i>Expert</i></td></tr>
<tr><td></td><td><b>1</b></td><td><b>2</b></td><td><b>1</b></td><td></td></tr>
<tr><td rowspan="3"><i>45 - 54</i></td><td rowspan="3"><b>1</b></td><td rowspan="3"><i>History and political science</i></td><td rowspan="3"><b>1</b></td><td colspan="5"><i>Textometry</i></td><td rowspan="3"><i>Mediaeval history</i></td><td rowspan="3"><b>1</b></td></tr>
<tr><td colspan="5"><i>None at all</i> <i>Expert</i></td><td colspan="5"><i>None at all</i> <i>Expert</i></td></tr>
<tr><td><b>3</b></td><td><b>1</b></td><td></td><td></td><td></td><td><b>1</b></td><td><b>1</b></td><td><b>2</b></td><td><b>1</b></td><td></td></tr>
</table>

Table 1. Profile of the participants in the two experiments

3.2 Discussion of results

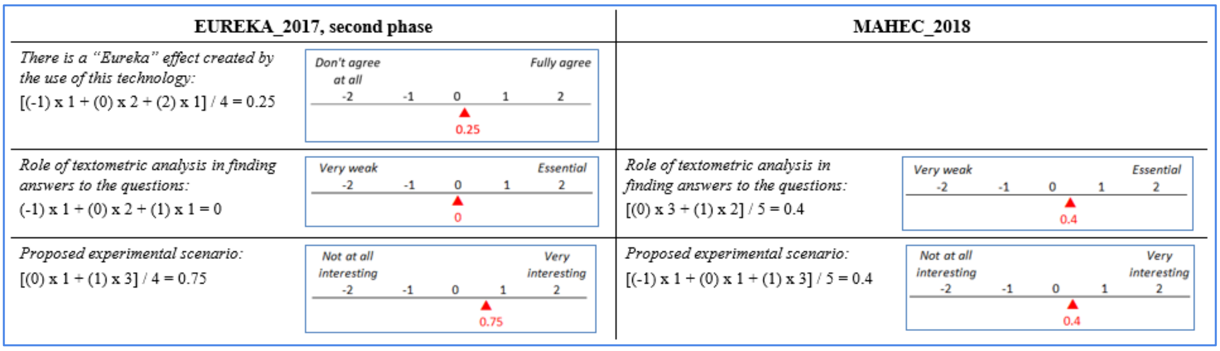
Outcomes from the evaluation of the textometric analysis only (EUREKA, second phase; MAHEC) are presented, since they are more closely related to the topic targeted in the article. In general, a slightly higher percentage of positive responses was observed in the first experiment (75%) than in the second (60%) to the *Yes/No* questions asking (1) whether answers and (2) new questions were found or (3) whether there is any added value in applying textometric techniques as compared to direct exploration of the online collection or to a more traditional non-digital approach: (1) 3 positive/4, (2) 2 positive/4, (3) 4 positive/4 (EUREKA, second phase); (1) 4 positive/5, (2) 1 positive/5, (3) 4 positive/5 (MAHEC). This difference might be explained by the fact that the students seemed to be more reticent than the researchers (20/50% positive) in (2) formulating new questions based on the TXM analysis.

For the Likert-type queries, the results of the first experiment (Figure 5, left) indicate moderate value attributed to the (1) role of textometric analysis in finding the answers to questions, (2) the occurrence of a "Eureka" effect as a result of this technology and (3) the evaluation of the proposed scenario. For analytical and comparative purposes, the five values of the scales were transposed to a numerical range (-2 to +2). Average scores were calculated by considering the numeric values and the distribution of responses by number of participants and answer type, e.g. the role of the textometric analysis was scored as -1 by one, 0 by two and +1 by one participant, with an average value of 0. Slightly higher values were observed for the two other questions. In the second experiment (Figure 5,

related to the proposed ones?" (2) "There is a 'Eureka' effect created by the use of this technology in this study." (EUREKA); "How do you view the role played by textometric analysis in finding answers to the questions?"; "How do you view the proposed experimental scenario?" (3) "Please provide a short description of the 'Eureka' effect, or the absence of this effect, observed during the experiment." (EUREKA); "[...] please describe this 'added value' in a few sentences"; "Other reflections on the innovative character of the considered technology and/or its limitations, bias, etc. for the studied case"; "Please list some strong/weak points of this approach" [proposed scenario].

[18] Presented at Les rendez-vous de l'histoire. Eurêka-inventer, découvrir, innover, Blois, France, 4-8 October 2017.

right), the average value of the role of textometric analysis in finding the answers was slightly higher (0.4/0) but the experimental scenario got less points (0.4/0.75) than in the first case.



EUREKA_2017, second phase		MAHEC_2018	
There is a "Eureka" effect created by the use of this technology: [(-1) x 1 + (0) x 2 + (2) x 1] / 4 = 0.25	*Don't agree at all* … *Fully agree* -2 -1 0 1 2 0.25		
Role of textometric analysis in finding answers to the questions: (-1) x 1 + (0) x 2 + (1) x 1 = 0	*Very weak* … *Essential* -2 -1 0 1 2 0	*Role of textometric analysis in finding answers to the questions:* [(0) x 3 + (1) x 2] / 5 = 0.4	*Very weak* … *Essential* -2 -1 0 1 2 0.4
Proposed experimental scenario: [(0) x 1 + (1) x 3] / 4 = 0.75	*Not at all interesting* … *Very interesting* -2 -1 0 1 2 0.75	*Proposed experimental scenario:* [(-1) x 1 + (0) x 1 + (1) x 3] / 5 = 0.4	*Not at all interesting* … *Very interesting* -2 -1 0 1 2 0.4

Figure 5. Average Likert-based scores for textometric analysis

More insight into the feedback was provided by the answers to the open questions. In terms of the (1) added value of textometric analysis, the participants in the first experiment mentioned: usefulness for analysing textual corpora by quantitative/statistical techniques allowing observation at both local and more general level, rapid identification of the main themes, and graphical representation of results.[19] The responses to the question (2) asking to describe the "Eureka" effect observed (or not) while using this method of analysis reiterated and enforced some of the above reflections, especially concerning the visual transformation of results and the possibility to highlight and de-contextualise/re-contextualise the linguistic units via a quantitative/qualitative perspective shift. [20] As in a previous quote, considering the second phase (textometric analysis) rather as a "refinement" of the first (online exploration of the videos and transcriptions), another participant noted that no new elements were detected; the only difference was the speed at which different topics could be identified.[21] The nature of the data sample and the usability of the tool were evoked as factors preventing the Eureka effect.[22] It was also observed that textometric analysis alone is not sufficient for research.[23] Other comments provided as (3) additional reflections on the innovative character and limitations of the method reiterated concerns about the impact of the data sample selection/size on the analysis results and highlighted the potential but also drawbacks, difficulties and uncertainties in using the method/interface.[24] Other benefits were mentioned and suggestions for alternatives were provided as (4) comments on the proposed experimental scenario.[25] One of the participants also enquired about the amount and type of preprocessing work necessary for this type of analysis.[26]

[19] Participant's code is provided in square brackets. "Possibility for quantitative and technical statistical analysis to explore a text-based corpus, study of occurrences of a linguistic motif, graphical visualisation of results" [EKA_PIL-P03]; "[...] overview, comprehensive view of the discourse of the interviewees without having to view the videos" [EKA_PIL-P04]; "[...] enables identification of the main themes addressed by the interviewees in a few clicks" [EKA_PIL-P02]; "[...] can be used to refine the results found in the first phase and study the views expressed in the discourse" [EKA_PIL-P01].

[20] "[...] the co-occurrences made it possible to contextualise words or groups of words and to stay close to the text. Textometry therefore combines quantitative and qualitative approaches." [EKA_PIL-P02]; "It highlights 'units', the possibility of visually transforming results through graphs and tables. Extracting elements from their original context but also being able to reintegrate them if needed [...]" [EKA_PIL-P01]

[21] "[...] [it] didn't bring out any new elements as compared with the results of the first phase. However, it enabled the different topics to be identified more quickly [...]" [EKA_PIL-P02]

[22] "The sample studied is not representative enough – it is too consensual for a real Eureka effect. Difficulty in getting to grips with the tool." [EKA-PIL_P03]

[23] "There is a Eureka effect but it should be viewed with caution since using textometric analysis alone is insufficient for research. However, textometric analysis can be a good tool for 'mind mapping'". [EKA-PIL_P04]

[24] "This technology has great potential but more time and a much larger sample are needed in order to fully exploit the potential of the tool." [EKA-PIL_P03]; "[...] The scores are not always effective for analysis and the words are not always representative of the discourse [...] The selection of interviews and excerpts is subjective, which may produce bias in the critical analysis of the research question." [EKA-PIL_P04]; "[…] without prior knowledge in linguistics and discourse analysis, I don't see how I can interpret the 'underuse' of a term […]" [EKA-PIL_P01]; "The interface could be more intuitive and the visualisations and graphics more appealing." [EKA-PIL_P02]

[25] "Textometric analysis can certainly be very useful in examining a large research corpus [...]" [EKA-PIL_P02]; "[...] another possible scenario. Define 2 groups. Group 1 works on the analysis of the interviews using traditional methods [...].

In the second experiment, only four of the five students answered the open questions. The (1) added value elements as compared to more "traditional" analysis methods were similar to those mentioned in the first experiment, e.g. enabling analysis of a large corpus of texts, "fast reading", speed and rigour.[27] As (2) additional reflections on the innovative characteristics and limitations of the studied technology, respondents pointed out the possibility to compare different interviews and the lack of features allowing annotation or modification of the texts.[28] Unlike the first experiment, the facility to pass from quantitative to qualitative view didn't seem to be fully grasped, or perhaps what was meant is that the quantitative aspect is more "tempting", which can lead to overlooking the qualitative facet needed in an enquiry of this nature.[29] As with the first experiment, it was observed that the analysis often served to prove something already known, rather than providing new information.[30] As (3) strong points of the experimental scenario, respondents noted the queries based on combined properties and the suitability of textometric analysis for assisting interpretation.[31] (4) Weak points mentioned were the size of the text/results window and the heterogeneity of the questions asked to interviewees.[32]

4 Conclusion and future work

Given time and resource constraints, the experiments had certain limitations. The number of participants was small and their background and familiarity with the proposed topic were not very diverse, since, as specialists or students in the field, the subject of European integration was relatively well known to them. The data samples, although selected to cover a given theme and percentage from the total collection of interviews, were not very large and did not involve a high number/variety of interviewees. The time allocated to TXM training prior to the experiments was limited (no training but a tutorial and assistance for EUREKA, 90 minutes of training and a tutorial and assistance for MAHEC). Taking into account these limitations, it can be hard to draw out generalisations, though various observations can be made.

Although the speed of processing and visualising linguistic features in large numbers of texts was mainly seen as a plus point, and attributes such as "innovative", "audacious" and "avant-garde" were used in the comments, the results showed a certain degree of reservation as to the innovative added value of the analysis tool. This was expressed both by a lower percentage of proposed new research questions and by explicit statements casting doubt on the new information gained as a result of the method. While this type of response can be partly explained by the above-mentioned limitations – which were also referred to by the participants through concerns raised about the data sample and the need for better knowledge of the tool – it can also indicate, to a certain degree, a specific approach to digital tools. That is, they are seen more as a means for proving hypotheses or known information than as "serendipitous" instruments for envisaging new paths of enquiry. However, this is an aspect that needs to be further examined in future experiments.

On the other hand, the results demonstrated awareness, from both the researchers and the students, of the different aspects involved in applying language technology to answering/identifying questions, such as data, methods, interface and general context of use. These aspects were repeatedly evoked in

Group 2 works on the interviews using the textometric tool [...] Comparison of the results [...]" [EKA-PIL_P03]

[26] "What about the manual effort needed to prepare a large corpus for textometric analysis?" [EKA-PIL_P02]

[27] "Textometric analysis enables the study of a large corpus of texts and saves a lot of time for historians. The analysis of the vocabulary used is greatly facilitated in particular." [TXM-HO_P01]; "Possibility of analysing several documents instead of reading them one by one." [TXM-HO_P02]; "Speed, rigorous analysis." [TXM-HO_P06]; "More efficient for 'fast reading' […]" [TXM-HO_P10].

[28] "[...] it is possible to compare the results for the documents, but this requires the interviews to be transcribed so that they can be read using the tool." [TXM-HO_P02]; "What is missing is a function to mark or modify the text [...]" [TXM-HO_P10]

[29] "Another problem is distancing from quality; with the tool it is very appealing to take a large number of documents for analysis [...]" [TXM-HO_P02]

[30] "An issue in textometric analysis is whether there is a real gain of new information. In most cases, textometric analysis proved the position and the known role of a person, but did not really contribute any new information." [TXM-HO_P01]

[31] "I liked the functionalities grouping certain queries, e.g. personal pronouns, nouns, adjectives, verbs, etc. [...]" [TXM-HO_P02]; "The approach is audacious and avant-garde in the field of history. It makes us reflect on different ways of reading sources, as well as on the logic that connects words in a text. [...]" [TXM-HO_P10]

[32] "The window displaying the text and analysis results should be larger [...]." [TXM-HO_P10]; "I would have liked interviews on a specific theme for all the respondents, [...] to compare the answers and see the result [...]." [TXM-HO_P02].

answers referring to the selection of interviews and the questions proposed to the interviewees and in observations about the qualitative/quantitative enquiry allowed by the tool, the more or less useful or easy-to-understand features provided by the interface and the experimental scenario itself. We would argue that this type of reflection is important not only for the "development of tools to be compatible with specific research methods of scholars" (Kemman and Kleppe, 2014) or for building "reflective tools and methods" (Koolen et al. 2018), but also to shed light on the process of research and teaching via digital tools. In this regard, we agree with Traub and Van Ossenbruggen's (2015) suggestion "to collect use cases and to compare evaluations of different tools" but with the intent of going beyond the creation of "checklists and guidelines for both, tool builders and users". Collecting use cases in this way also represents a means of sharing experiences and rendering the research process more transparent, thus improving understanding of emerging shifts in humanities practices brought about by digital technologies.

To sum up, the project combined original sources of oral history and digital linguistic analysis, and evaluated the use of language technology via two use cases of history research and teaching. Two experiments were devised. Although the data samples and the groups of participants were small and not very diverse, and additional experiments are needed for generalisations to be made, the results provided an insight into how researchers and students apply this type of tool and reflect on its use. We would argue that the creation (potentially within the framework of CLARIN) of an interdisciplinary collaborative platform containing an online collection of use cases, evaluation data and workflow descriptions from different areas of study will encourage the pooling of experience and practices with a view to stimulating debate, creativity and the exchange of ideas within the academic community. By sharing reflections and drawing up an "inventory" of strengths and weaknesses, it will thereby facilitate understanding of current and emerging practices in applying language technology to research and teaching in the humanities.

Acknowledgements

We would like to thank our colleagues and students for participating in the experiments, and Sarah Cooper, from the Faculty of Language and Literature, Humanities, Arts and Education at the University of Luxembourg, for English proofreading.

References

[Anderwald and Wagner 2007] Lieselotte Anderwald and Susanne Wagner. 2007. "FRED — The Freiburg English Dialect Corpus: Applying Corpus-Linguistic Research Tools to the Analysis of Dialect Data". In *Creating and Digitizing Language Corpora*. 10.1057/9780230223936_3.

[Bloch, 1999]. Marc Bloch 1999. "Réflexions d'un historien sur les fausses nouvelles de la guerre". Allia, Paris (Re-publication of the article of the same name, originally published in 1921 in the *Revue de synthèse historique*, T.33).

[Bock 2007] Zannie Bock. 2007. *A Discourse Analysis of Selected Truth and Reconciliation Commission Testimonies: Appraisal and Genre*. PhD Thesis, University of the Western Cape, Republic of South Africa, November 2007.

[Choudhury et al. 2018] Prithwiraj (Raj) Choudhury, Natalie A. Carlson, Dan Wang and Tarun Khanna. 2018. "Machine Learning Approaches to Facial and Text Analysis: An Application to CEO Oral Communication". Working Paper 18 – 064, Harvard Business School.

[CLARIN 2016] CLARIN. 2016. CLARIN-PLUS OH workshop: *Exploring Spoken Word Data in Oral History Archives*. University of Oxford, United Kingdom. https://www.clarin.eu/event/2016/clarin-plus-workshop-exploring-spoken-word-data-oral-history-archives. (Accessed March 31, 2019)

[Danescu 2013] Elena Danescu 2013. *A rereading of the Werner Report of 8 October 1970 in the light of the Pierre Werner family archives*. Research corpus. Source: https://www.cvce.eu/en/project/werner/introduction. (Accessed March 31, 2019)

[Dedman 2009] Martin Dedman 2009. *The Origins & Development of the European Union: 1945–2008*. Routledge, London (Second edition).

[Descamps 2013] Florence Descamps. 2013. “Histoire orale et perspectives. Les évolutions de la pratique de l’histoire orale en France”. In F. d’Almeida and D. Maréchal (Ed.), *L’histoire orale en questions*, p. 105–138. INA, Paris.

[Drude 2016] Sebastian Drude. 2016. “ELAN as a tool for oral history”. CLARIN-PLUS OH workshop.

[EU Consilium 2018]. Council of the European Union 2018. *Blue guide to the Archives of Member States’ Foreign Ministries and European Union institutions*. https://www.consilium.europa.eu/media/29595/blueguide_pdf_201404.pdf. (Accessed March 31, 2019)

[Georgetown University 2001] Georgetown University. 2001. *Georgetown University Round Table on Languages and Linguistics (GURT)*, Washington, DC, USA.

[Freiburg Institute for Advanced Studies 2015] Freiburg Institute for Advanced Studies. 2015. Conference *Oral History meets Linguistics*, Freiburg, Germany. https://www.frias.uni-freiburg.de/en/events/frias-conferences/conference-oral-history-and-linguistics. (Accessed March 31, 2019)

[Gibbs and Owens 2012] Fred Gibbs and Trevor Owens. 2012. “Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs”. In *Digital Humanities Quarterly*, 2012, Volume 6, Number 2. http://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html. (Accessed March 31, 2019)

[HAEU 2018]. *Historical Archives of the European Union 2018*. http://www.eui.eu/Research/HistoricalArchivesOfEU/Index.aspx. (Accessed March 31, 2019)

[Hájek and Vann 2015] Martin Hájek and Barbara H. Vann. 2015. “Gendered Biographies: The Czech State-socialist Gender Order in Oral History Interviews”. *Sociologický ústav AV ČR*, v.v.i., Praha.

[Heiden et al. 2010] Serge Heiden, Jean-Philippe Magué and Bénédicte Pincemin. 2010. “TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement”. In Sergio Bolasco, Isabella Chiari, Luca Giuliano (Ed.), *Proc. of 10th International Conference on the Statistical Analysis of Textual Data* - JADT 2010, Vol. 2, p. 1021–1032. Edizioni Universitarie di Lettere Economia Diritto, Roma, Italy. https://halshs.archives-ouvertes.fr/halshs-00549779/fr/. (Accessed March 31, 2019)

[Hix and Høyland 2011] Simon Hix and Bjørn Høyland 2011. *The Political System of the European Union*, Basingstoke, Palgrave MacMillan (Third edition).

[Janesick 2010] Valerie J. Janesick. 2010. *Oral history for the qualitative researcher. Choreographing the story*. Guilford Press, New York.

[Kemman and Kleppe 2014] Max Kemman and Martijn Kleppe. 2014. “User Required? On the Value of User Research in the Digital Humanities”. In CLARIN 2014 Selected Papers; Linköping Electronic *Conference Proceedings # 116*. http://www.ep.liu.se/ecp/116/006/ecp15116006.pdf. (Accessed March 31, 2019)

[Klein 2010-2017]. François Klein 2010–2017. *Oral history of European integration collection*: Source: https://www.cvce.eu/en/oral-history/presentation. (Accessed March 31, 2019)

[Koolen et al. 2018] Marijn Koolen, Jasmijn van Gorp, Jacco van Ossenbruggen. 2018. “A Hands-on Approach to Digital Tool Criticism. Tools for (self-)Reflection”. Conference on *Digital Hermeneutics in History: Theory and Practice*, University of Luxembourg, 25 October 2018. https://docs.google.com/presentation/d/1om6BK4xNNJ0-_hKKwOYArdHrHAn-VRPp0mH6dRo8ViI/edit#slide=id.p. (Accessed March 31, 2019)

[Labov and Waletzky 1967] William Labov and Joshua Waletzky. 1967. “Narrative analysis: Oral versions of personal experience”. In J. Helm (Ed.), *Essays on the verbal and visual arts*. Seattle, WA: University of Washington Press. pp. 12–44.

[Lafon 1980] Pierre Lafon. 1980. “Sur la variabilité de la fréquence des formes dans un corpus”. *Mots*, no. 1, p 127–165. http://www.persee.fr/doc/mots_0243-6450_1980_num_1_1_1008. (Accessed March 31, 2019)

[Leonardi 2018] Simona Leonardi. 2018. “Metaphors in the Life Story of A German- Jewish Immigrant to Palestine/Israel. How Metaphorical Constructions and Remembering Process Interweave”. *Remembrance and Research*, ILOHA, no. 2, January 2018, pp. 51–68.

[Pagenstecher and Pfänder 2017] Cord Pagenstecher and Stefan Pfänder. 2017. “Hidden Dialogues: Towards an Interactional Understanding of Oral History in Interviews”. In *Oral History Meets Linguistics*, edited by Erich Kasten, Katja Roller, and Joshua Wilbur, pp. 185–207. Fürstenberg/Havel: Kulturstiftung Sibirien, Electronic Edition. http://www.siberian-studies.org/publications/PDF/orhili_pagenstecher_pfaender.pdf. (Accessed March 31, 2019)

[Passerini 2006] Luisa Passerini 2006. *Memory and Utopia: The Primacy of Inter-Subjectivity*. Equinox Publishing: London.

[Portelli 2009] Alessandro Portelli 2009 “What Makes Oral History Different”. In Giudice L.D. (Ed.), *Oral History, Oral Culture, and Italian Americans. Italian and Italian American Studies*. Palgrave Macmillan, New York, pp.21–30.

[Radelli and Featherstone 2003] Claudio Radelli and Kein Fetherstone 2003 (Ed.) *The Politics of Europeanization*. Oxford University Press, Oxford.

[Ritchie 2003]. Donald A. Ritchie 2003. *Doing Oral History*. Oxford University Press, New York.

[Schiffrin 2003] Deborah Schiffrin. 2003. “Linguistics and History: Oral History as Discourse”. Georgetown University Round Table on Languages and Linguistics (GURT) 2001: *Linguistics, Language, and the Real World: Discourse and Beyond*, Deborah Tannen and James Alatis (Ed.), pp. 84–113, Georgetown University Press, Washington, D.C. http://faculty.georgetown.edu/schiffrd/index_files/Linguistics_and_oral_history.pdf. (Accessed March 31, 2019)

[Sealey 2009] Alison Sealey. 2009. “Probabilities and surprises: A realist approach to identifying linguistic and social patterns, with reference to an oral history corpus”. In *Applied Linguistics*: 1–21, Oxford University Press, doi:10.1093/applin/amp023.

[Slabakova 2016] Radka Slabakova. 2016. “The Meaning of His Life Was Work: The Construction of Identities in the Oral Narratives of Older Czech Men”. *Gender Studies*. 15. 10.1515/genst-2017-0008.

[Traub and Van Ossenbruggen 2015] Myriam C. Traub and Jacco van Ossenbruggen. 2015. “Workshop on Tool Criticism in the Digital Humanities”. CWI *Techreport*, 1 July 2015, https://pdfs.semanticscholar.org/d337/ce558c2fd1d8be793786c9cfc3fab6512dea.pdf. (Accessed March 31, 2019)

[TXM Manual]. 2018. *TXM User Manual*, Version 0.7 ALPHA, February 2018. http://textometrie.ens-lyon.fr/files/documentation/TXM%20Manual%200.7.pdf. (Accessed March 31, 2019)

[Van Hessen 2016] Arjan van Hessen. 2016. “Increasing the Impact of Oral History Data with Human Language Technologies, How CLARIN is already helping researchers”. CLARIN-PLUS OH workshop.

[Van Uytvanck 2016] Dieter van Uytvanck. 2016. “CLARIN Data, Services and Tools: What language technologies are available that might help process, analyse and explore oral history collections?” CLARIN-PLUS OH workshop.

[Zhang and Soergel 2006] Pengyi Zhang and Dagobert Soergel. 2006. “Knowledge-Based Approaches to the Segmentation of Oral History Interviews”. MALACH Technical Report, University of Maryland. College of Information Studies, May 2006.

Appendix. List of interviewees in the “Pierre Werner and Europe” project

The figures that have been interviewed thus far in connection with the “Pierre Werner and Europe” research project are as follows (in alphabetical order): Michel Camdessus, Honorary Governor of the Banque de France, Managing Director of the IMF (1987–2000); Luc Frieden, Luxembourg Finance Minister (2009–2013); Albert Hansen, Secretary-General of the Luxembourg Government (1979–1998); Edmond Israel (1924–2011), Luxembourg banker, President of the Board of Directors of Cedel International (1970–1999); Jean-Claude Juncker, President of the European Commission since 2014, Prime Minister of Luxembourg (1995–2013), first permanent President of the Eurogroup (2005–2013); Helmut Kohl (1930–2017), Chancellor of the FRG (1982–1998); Philippe Maystadt (1948–2017), Belgian Finance Minister (1988–1998), President of the European Investment Bank (2000–2011); Yves Mersch, Member of the Executive Board of the European Central Bank (since 2012), President of the Banque centrale du Luxembourg (1998–2012); Guy de Muyser, Marshal of the Grand Ducal Court (1971–1981); Charles-Ferdinand Nothomb, President of the Belgian Chamber of Representatives (1979–1980, 1988–1995), Honorary President of the Pierre Werner European Circle; Viviane Reding, Member of the European Commission (1999–2010), Vice-President of the European Commission with responsibility for Justice, Fundamental Rights and Citizenship (2010–2014), Member of the European Parliament (since 2014); Lex Roth, Director of the Information and Press Service of the Luxembourg Government (1988–1993); Charles Ruppert, Chairman of the Luxembourg Bankers’ Association (1992–1995), Chairman of the Pierre Werner Foundation; Fabrizio Saccomanni, Vice President of the European Bank for Reconstruction and Development (EBRD) (2003–2006), Italian Minister for Economic Affairs and Finance (2013–2014); Jacques Santer, Prime Minister of Luxembourg (1984–1995), President of the European Commission (1995–1999); Bernard Snoy et d’Oppuers, International President of the European League for Economic Cooperation, President of Robert Triffin International; Gaston Thorn (1928–2007), Prime Minister of Luxembourg (1974–1979), President of the European Commission (1981–1985); Hans Tietmeyer (1930–2017), President of the Deutsche Bundesbank (1993–1999), Member of the Werner Committee (1970); Niels Thygesen, Member of the Delors Committee (1988–1989), Chairman of the European Fiscal Board (since 2016), Professor at the Institute for New Economic Thinking; Sir Brian Unwin, President of the European Investment Bank (1993–1999), Governor of the European Bank for Reconstruction and Development (1993–1999), Chairman of the Supervisory Board of the European Investment Fund (1994–1999); Henri Werner, son of Pierre Werner; Marie-Anne Werner, daughter of Pierre Werner. Other accounts emerged as a result of the project “Accounts by Luxembourg Ambassadors” (Jean-Jacques Kasel, Adrien Meisch and Jean Mischo), and interviews were also conducted with Étienne Davignon, Member of the European Commission (1977–1981) and Vice-President of the European Commission (1981–1985); Jacques de Larosière, Assistant Director (1967–1974) then Director of the French Treasury (1974–1978), Managing Director of the IMF (1978–1987); Jacques Delors, President of the European Commission (1985–1995); Mark Eyskens, Belgian Finance Minister (1980–1981) and Prime Minister (1981); and Wilfried Martens, Prime Minister of Belgium (1979–1981/1981–1992).