Transdisciplinary Analysis of a Corpus of French Newsreels: The ANTRACT Project

Jean Carrive, Abdelkrim Beloued, Pascale Goetschel, Serge Heiden, Antoine Laurent, Pasquale Lisena, Franck Mazuet, Sylvain Meignier, Bénédicte Pincemin, Géraldine Poels, et al.

Transdisciplinary Analysis of a Corpus of French Newsreels: The ANTRACT Project

JEAN CARRIVE[1], ABDELKRIM BELOUED[1], PASCALE GOETSCHEL[2], SERGE HEIDEN[3], ANTOINE LAURENT[4], PASQUALE LISENA[5], FRANCK MAZUET[2], SYLVAIN MEIGNIER[4], BÉNÉDICTE PINCEMIN[3], GÉRALDINE POELS[1], RAPHAËL TRONCY[5]

[1] INA – Institut national de l'audiovisuel
4 avenue de l'Europe, 94366 Bry-sur-Marne Cedex, France
[2] Centre d'histoire sociale des mondes contemporains, UMR 8058 (Université Paris 1/CNRS)
Campus Condorcet, bâtiment recherche sud, 5 cours des Humanités, 93300 Aubervilliers, France
[3] Univ. Lyon, IHRIM, Institut d'histoire des représentations et des idées dans les modernités, UMR 5317
ENS de Lyon, 15 parvis René Descartes, BP 7000 69342 Lyon Cedex 07, France
[4] LIUM, Laboratoire d'Informatique de l'Université du Mans
Avenue Olivier Messiaen F-72085 – Le Mans Cedex 9, France
[5] EURECOM, 450 Route des Chappes, 06410 Biot, France

[1]jcarrive *at* ina.fr, abeloued *at* ina.fr, gpoels *at* ina.fr
[2]fmazuet *at* free.fr, pascale.goetschel *at* univ-paris1.fr
[3]slh *at* ens-lyon.fr, benedicte.pincemin *at* ens-lyon.fr
[4]sylvain.meignier *at* univ-lemans.fr, antoine.laurent *at* univ-lemans.fr
[5]pasquale.lisena *at* eurecom.fr, raphael.troncy *at* eurecom.fr

Abstract

The ANTRACT project is a cross-disciplinary apparatus dedicated to the analysis of the French newsreel company *Les Actualités Françaises* (1945-1969) and its film productions. Founded during the liberation of France, this state-owned company filmed more than 20,000 news reports shown in French cinemas and throughout the world over its 24 years of activity. The project brings together research organizations with a dual historical and technological perspective. ANTRACT's goal is to study the production process, the film content, the way historical events are represented and the audience reception of *Les Actualités Françaises* newsreels using innovative AI-based data processing tools developed by partners specialized in image, audio, and text analysis.

This article focuses on the data processing apparatus and tools of the project. Automatic content analysis is used to select data, to segment video units and typescript images, and to align them with their archival description. Automatic speech recognition provides a textual representation and natural language processing can extract named entities from the voice-over recording; automatic visual analysis is applied to detect and recognize faces of well-known characters in videos. These multifaceted data can then be queried and explored with the TXM text-mining platform.

The results of these automatic analysis processes are feeding the Okapi platform, a client-server software that integrates documentation, information retrieval, and hypermedia capabilities within a single environment based on the Semantic Web standards. The complete corpus of *Les Actualités Françaises*, enriched with data and metadata, will be made available to the scientific community by the end of the project.

1. Implementing a Transdisciplinary Research Apparatus on a Film Archive Collection: Opportunities and Challenges

The ANTRACT[1] project brings together research organizations with a dual historical and technological perspective, hence the reference to the transdisciplinary in the project's name. It applies to a collection of 1262 newsreels (mostly black and white footage) shown in French movie theaters between 1945 and 1969. These programs were produced by *Les Actualités Françaises* newsreel company during the French *Trente Glorieuses* era. The project develops automated tools well suited to analyze these documents: automatic speech recognition, image classification, facial recognition, natural language processing, and text mining. These software are used to produce metadata and to help organize media files and documentation resources (i.e. titles, summaries, keywords, participants, etc.) into a manageable and coherent corpus usable within a dedicated online platform.

Working together on these newsreels divided into 20,232 news reports, ANTRACT historians and computer scientists collaborate to optimize the research on large audiovisual corpora through the following questions:

- What is the best technological approach to the systematic and exhaustive study of a multimedia archive collection?
- What instruments can compile, analyze and crosscheck the data extracted from such documents?
- Can these extracted data be combined and integrated into versatile user interfaces?
- Can they provide new opportunities to humanities research projects through their assistance in the processing of numerous multi-format sources?

In order to implement a strong cooperation between AI experts and history scholars (Deegan and McCarty, 2012), the key objective of the project is to provide scholars and media professionals working on extensive collections of film archives with an innovative research methodology fit to address the technological and historical questions raised by this particular corpus.

From a technological perspective, the goal is to adapt automatic analysis tools to the specificity of the *Actualités Françaises* corpus, i.e. its historical context, vocabulary, image type. Adapting the language models used by the automatic transcription tools with the help of the typescripts of voice overs underlines this orientation. As a film collection including footage, sound and text produced more than half a century ago, *Les Actualités Françaises* corpus presents an unprecedented challenge to instruments specialized in audiovisual content extraction and identification. Far from separately considering a social and cultural history of cinema on the one hand, and the use of automatic analysis tools on the other hand, the project aims to link the two. Thus, a good understanding of the technical conditions for recording the audio leads to improved audio recognition. Shot in black and white with limited equipment and often under difficult filming conditions, these old newsreels do not meet the quality standards set by the high definition video and audio recordings feeding today's image and speech recognition algorithms. Moreover, several film reels of the collection digitized under high compression formats show pixelated images that cannot be processed by analysis programs and some of the commentary typescripts display printing defects caused by the typewriters used for their production.

Along with these material obstacles comes the problem raised by the transfiguration of film content over time. This is the case for leading figures regularly filmed by the company's cameramen throughout its 24 years of activity. It is also the case for the recurring topographical data caught on

their film. The automatic identification of these ever-changing elements recorded on monochromatic footage requires a considerable amount of resources. As part of this process, ANTRACT historians have selected a sample of the most distinctive representations of notable characters present in *Les Actualités Françaises* newsreels in order to build a series of extraction models.

From an historical perspective, ANTRACT aims to approach topics beyond the notion of newsreels as a wartime media subjected to state censorship and political ambitions (Atkinson, 2011; Bartels, 2004; Pozner, 2008; Veray, 1995). In the wake of existing studies, one of its primary objectives is to extend the historical scope of the cinematographic press to question its role as a vector of social, political and cultural history shaping the opinion of the public during the second half of the 20th century (Fein, 2004, 2008; Althaus et al., 2018; Chambers et al., 2018; Imesch et al., 2016; Lindeperg 2000, 2008). This series of cinematographic documents is not the only legacy left by a newsreel company which witnessed world history from the liberation of France to the late 1960's. The dope sheets filled out by its cameramen, the written commentaries of its journalists and the records left by its management give us rare insight into the content of a film collection as well as its production process. Despite its historical value, *Les Actualités Françaises* corpus has eluded a thorough examination of its entire content. Scattered across different inventories, the numerous films, audio records and typescripts produced by the newsreel company have forestalled such a project. In this regard, the challenge presented by an exhaustive study of *Les Actualités Françaises* is similar to those of other abundant multi-format collections and inspires a recurring question regarding their approach: how can one identify and index thousands of hours of film archives associated with hundreds of text files produced over an extended period of time? The tools developed by the consortium partners working on the project are intended to cast a new light on the French company newsreels through the combined treatment of data extracted from its whole collection and correlatively studied on the Okapi and TXM platforms. This apparatus should open new semantic fields previously overlooked by the fragmentary research conducted on specific inventories of the company records. Focused on film content, the project is also committed to scrutinize the production process and the different trades involved in the making of *Les Actualités Françaises* newsreels emphasizing the political and economic background of a company controlled by a democratic state. Underlining the notion that media participate in events (Goetschel and Granger, 2011), this dual analysis - both technological and historical - will be completed with the study of the public reception of these weekly journals in light of its request patterns, i.e. audience expectation for sensational and exotic news and its interest in the daily life of renowned figures (Maitland, 2015).

Through audio and video analysis tools dedicated to corpus building and enrichment (section 2) and platforms for historical interactive analysis (section 3), this article presents the results from the first phase of the project, which sets the focus on the technological side of the research, specifically its data processing apparatus and tools. Nevertheless, historians are involved in most of these computational preliminary steps, by contributing to the implementation and testing tasks. At the same time, we explore temporary results of historical investigations, while the full potential for historical studies will be developed in the forthcoming second phase of the project that will be addressed in a follow-up article.

2. Corpus building and enrichment

2.1 Organization of the video corpus with automatic content analysis technologies

Automatic content analysis technologies are used to obtain the most consistent, complete and homogeneous corpus as possible, allowing historians to easily search and navigate through the documents (digitized films, documentation notes and typescripts). When considering that the whole archive would not be relevant, a preliminary step was to realize that for some tasks, we had to define how our corpus would be composed and structured. One cannot just input the data into the computer and see what happens. For instance, textometric analysis would be hindered if all the available videos were kept, because of numerous duplicates which would artificially inflate word frequencies. Duplicates could be due to either multiple copies of a single news report, or to the use of the same report in several regional editions. As a collaborative decision involving newsreel experts, corpus analysis researchers, and historians, ANTRACT's main corpus was restricted to the collection of all national issues of *Les Actualités Françaises* newsreels, each issue being composed of topical report sub-units. Then, the next goal was:

1) to get a corpus made of exactly one digital video file by edition (which was a requisite condition for TXM data import, see Section 3.1),
2) to get archival descriptions of the reports temporally linked to these files, as an edition is made of a succession of reports.

This led us to take the following actions:

1) physically segment video files initially coming from the digitization of film reels, so that each file contains exactly one edition, starting at timecode 0.
2) keep only archival descriptions linked to either one edition or one report included in one of the editions, namely "summary" and "report" archival descriptions. Thus, archival descriptions corresponding to other content, such as rushes or unused material, called "isolated" archival descriptions were discarded. Around 10,700 archival descriptions have thus been kept in this first version of the corpus.

The remaining of this section explains how automatic analysis has been used to temporally synchronize archival descriptions with digital video files.

Segmentation of reports. Each one of the 1,200 editions of the newsreels corresponds to more than one digital video file, either because several digitized copies of one given edition exist in the collection, or because the film has been digitized several times, for quality reasons for instance. When they exist, timecodes of archival descriptions may refer to one or the other digital video file. One objective is to get all archival descriptions of one edition referring to the same video file, with timecodes. About 9,500 out of 10,700 archival descriptions have timecodes referring to the video file of the whole edition, which left around 1,200 archival descriptions to manage. A report with timecode is called "*segmented*". One important step is to segment each edition into its constitutive reports, by detecting report boundaries. In most cases, reports are separated by black images, easily detected by simple image analysis methods (the *ffmpeg* video library offers an efficient "blackdetect" option for instance). Reports may also be separated by sequences of a few frames to a few seconds of a motion blur shot by a camera, used as a syntactic punctuation. In some cases, when these sequences are long enough, they can be detected as a simple threshold on the horizontal dimension of the optical flow,

computed with existing algorithms such as OpenCV (Bradski, 2000). A more robust detection method is still under development using machine learning algorithms.

Transfer of timecodes. When timecodes refer to a video file different from the main video file, timecodes on the main file may be computed using copy detection techniques. The principle is illustrated by Figure 1. In the figure, reports on "Rugby" and "Kennedy's visit" (from the edition of May 31st, 1961) refer to two video files, both distincts from the video file corresponding to the whole edition. To identify the location of the reports within the main video file, we used the audio and video copy detection method based on fingerprinting methods developed at INA (Chenot and Daigneault, 2014), eventually allowing the transfer of timecodes for more than 800 reports.

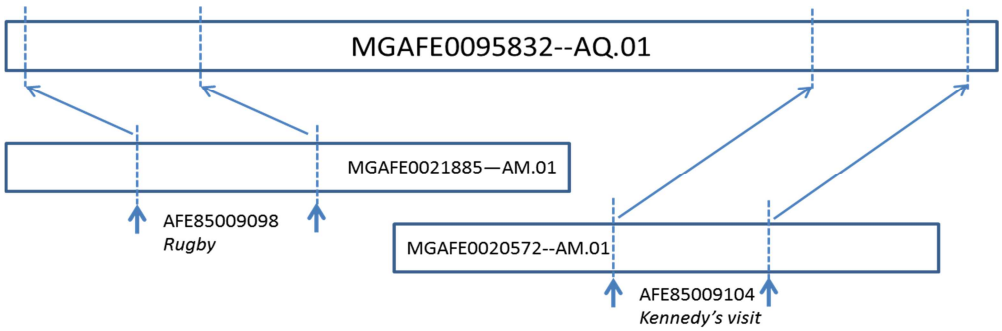


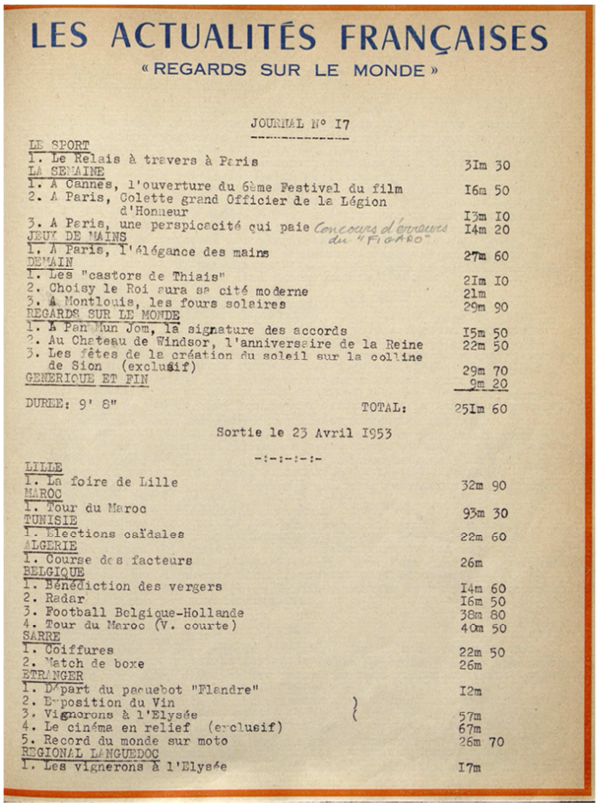
Figure 1. Transfer of archival description timecodes

Timecoding reports using transcripts. We tried to identify the temporal boundaries of the remaining 400 *unsegmented* remaining reports by comparing the text coming from corresponding archival descriptions (title + summary + keywords for instance), with the automatic speech transcription (ASR) of segments of the video file not already corresponding to one report (see Section 2.3). Simple similarity text measures such as the Jaccard distance, or *ratio* metrics in the Fuzzywuzzy Python package give encouraging but not entirely satisfying results. We plan to use a corpus-specific *TF-IDF* measure, or embedding methods such as word2vec or BERT in the future.

2.2 Typescripts: from page scans to structured textual data

Typescripts of the voice-overs have been linked to books and typescripts of each edition and separated with pages giving the summary of the edition (see Figure 2). This represents around 9,000 pages. At the beginning of the project, these documents were scanned in a good quality format (TIFF, color, 400 dpi). An optical character recognition (OCR) tool has thus been applied (Google Vision API in "Document" mode), giving spatially-located digital texts.

Once digitized, typescripts have to be separated from summaries. In order to achieve that, an automatic classifier has been trained by specializing the state-of-the-art Inception V3 classifier (Szegedy et al., 2016) with a few manually chosen examples. This gave about 2,600 pages of summaries and 6,400 pages of voice-overs.



LES ACTUALITÉS FRANÇAISES

« REGARDS SUR LE MONDE »

JOURNAL N° I7

LE SPORT
1. Le Relais à travers à Paris 3Im 30
LA SEMAINE
1. A Cannes, l'ouverture du 6ème Festivel du film I6m 50
2. A Paris, Colette grand Officier de la Légion d'Honneur I3m I0
3. A Paris, une perspicacité qui paie Concours d'erreurs du "FIGARO" I4m 20
JEUX DE MAINS
1. A Paris, l'élégance des mains 27m 60
DEMAIN
1. Les "castors de Thiais" 2Im I0
2. Choisy le Roi aura sa cité moderne 2Im
3. A Montlouis, les fours solaires 29m 90
REGARDS SUR LE MONDE
1. A Pan Mun Jom, la signature des accords I5m 50
2. Au Chateau de Windsor, l'anniversaire de la Reine 22m 50
3. Les fêtes de la création du soleil sur la colline de Sion (exclusif) 29m 70
GENERIQUE ET FIN 9m 20

DUREE: 9' 8" TOTAL: 25Im 60

Sortie le 23 Avril I953

-:-:-:-:-

LILLE
1. La foire de Lille 32m 90
MAROC
1. Tour du Maroc 93m 30
TUNISIE
1. Elections caïdales 22m 60
ALGERIE
1. Course des facteurs 26m
BELGIQUE
1. Bénédiction des vergers I4m 60
2. Radar I6m 50
3. Football Belgique-Hollande 38m 80
4. Tour du Maroc (V. courte) 40m 50
SARRE
1. Coiffures 22m 50
2. Match de boxe 26m
ETRANGER
1. Départ du paquebot "Flandre" I2m
2. Exposition du Vin
3. Vignerons à l'Elysée 57m
4. Le cinéma en relief (exclusif) 67m
5. Record du monde sur moto 26m 70
REGIONAL LANGUEDOC
1. Les vignerons à l'Elysée I7m

REGARDS SUR LE MONDE Journal I7/53 2

1. A Pan Mun Jom: la signature des accords 53.166

A Pan Mun Jom, les pourparlers qui viennent de reprendre sont entrés enfin, dans la voie des réalisations. Le Général Lee Sang-Cho qui conduit la délégation sino-coréenne, et l'Amiral Daniel, chef des délégués alliés, se sont rencontrés dans la "baraque de la paix" pour parapher l'accord sur l'échange des prisonniers malades et blessés. Et quelques minutes plus tard, l'Amiral Daniel pouvait agiter joyeusement le document signé, qui marque un premier pas vers la paix.

2. A Windsor, l'anniversaire de la Reine 53.167

... Au Chateau de Windsor, berceau des souverains britanniques, la Reine Elizabeth, refaisant un geste inauguré par Charles II, est venue à l'occasion de son anniversaire, remettre leur étendard aux grenadiers de la garde et recevoir leurs voeux dans les trois "houras" traditionnels.

3. Sur la colline de Sion, les fêtes de la création du soleil 53.168

Perpétuant une ancienne tradition, les fidèles ont attendu, sur les toits de la synagogue du Mont Sion, le lever du soleil. La croyance affirme en effet, que le soleil reprend tous les 28 ans, la place qu'il occupa au jour de sa création. Perçant les nuées, annoncé au son des trompes de David, l'astre, tel qu'il sortit des mains divines apparait au peuple élu qui, manifeste sa joie en embrassant les rouleaux de la loi. De grandes réjouissances commencent alors, pour célébrer cet évènement auquel peu de juifs peuvent se vanter d'avoir assisté trois fois dans leur vie.

DEMAIN

1. les "Castors de Thiais" 53.163

Demain, trouverons nous à nous loger? Cette question si souvent répétée, les "Castors de Thiais" ont décidé de la résoudre par leurs propres moyens, en construisant eux-mêmes leur pavillon. Et, chaque soir, en sortant de leur travail, et durant leur congé hebdomadaire, des hommes venus de tous les métiers, se retrouvent sur leur chantier, aux portes de Paris, pour fabriquer des parpaings et participer à l'édification de la petite cité qu'ils habiteront demain. Lentement, mais sûrement, sous la conduite de l'ingénieur qui assure la direction du groupe, les murs s'élèvent et les maisons prennent forme. Pour les "Castors de Thiais", demain ne pose plus de problèmes du logement.

2. Choisy le Roi aura sa cité moderne 53.164

Demain, également, Choisy le Roi sera dotée d'un magnifique ensemble de constructions modernes. Dernier vestige du passé, la cheminée d'une usine de faïencerie a été solennellement démolie. En présence d'un nombreux public, le Maire a mis le feu au brasier qui provoquait peu après, l'effondrement spectaculaire des 55 mètres de maçonnerie. Les démolisseurs ont terminé leur travail. Sur les 35.000 m2 qu'ils ont déblayés celui des bâtisseurs va commencer. Demain, une cité moderne blanche et aérée, qui comprendra 500 logements remplacera l'ancienne usine. Et, 500 familles auront enfin trouvé un foyer.

3. Les fours solaires de Mont-Louis 53.165

Dans les Pyrénées, à 1600 mètres d'altitude, fonctionne à l'intérieur de la citadelle de Montlouis, le plus grand four solaire du monde. Cet appareil comprend un miroir orientable, composé de 500 glaces, commandé par une cellule photo-électrique, et un miroir parabolique

Figure 2. Typescripts of voice-overs and summaries

Spatial and temporal alignment of transcripts. The objective of this alignment is to associate each report with the corresponding section of the typescripts. The available metadata allows processing this alignment year by year. This operation is done in two stages, by using on the one hand the result of the automatic speech transcription of the voice-over from the video files, and by using on the other hand the result of the OCR of the typescripts. The first step is done by minimizing a comparison measure between strings in order to find for each subject the corresponding typescripts page. The *partial ratio* method of the Fuzzywuzzy Python package allows looking for a partial inclusion of the speech-to-text into the OCR. Since topics and pages are approximately chronological, exhaustive searching is not required. The second step consists in spatially locating the text of the voice-over in the corresponding typescript page. For that, we use the alignment given by the Dynamic Time Warping algorithm (DTW), slightly modified to overcome the anchoring at the ends of the found path. The typescript area thus identified in the output of the OCR makes it possible to obtain the spatial coordinates of the commentary in the typewritten page. However, the method used does not allow locating transcripts overlapping over two pages. Additional treatment should be considered, for instance in order to get aligned text units for textometric analysis (see section 3.1).

2.3 Automatic audio analysis

The work on the audio part consists in detecting the speakers, transcribing speech into words (ASR) and detecting named entities (NE) using the systems we have developed for contemporary radio and television news.

Audio analysis of an old data set is an interesting challenge for automatic analysis systems. The recording devices used between 1945 and 1969 are very different from today's analog or digital devices. 35-mm films, which contain both sound and image, deteriorated before being digitized in the

2000s. Moreover, the acoustic and language models are generally trained on data produced between 1998 and 2012. This 50-year time gap has consequences on the system's performance.

Technically, acoustic models for ASR and speakers were trained on about 300 hours drawn from several sources of French TV and radiophonic broadcast news[2] with manual transcripts. The ASR language models were trained on these manual transcripts, French newspapers, news websites, Google news and the French GigaWord corpus, for a total of 1.6 billion words. The vocabulary of the language model contains the 160k most frequent words. The NE models were trained only on a subset of manual transcripts[3].

Prior to the transcription process, the signal is cut into homogeneous speech segments and grouped by speakers. We refer to this process as the Speaker Diarization task. Speaker Diarization is first applied at the edition level, where each video record is separately processed. Then, the process is applied at the collection level, over all the 1,200 editions, in order to link the recurrent speakers. The system is based on the LIUM S4D toolkit (Broux et al., 2018), which has been developed to provide homogeneous speech segments and accurate segment boundaries. Purity and coverage of the speaker clusters are also one of the main objectives. The system is composed of acoustic metric-based segmentation and clustering followed by an i-vector-based clustering applied to both edition and collection levels.

The ASR system is developed using the Kaldi Speech Recognition Toolkit (Povey et al., 2011). Acoustic models are trained using a Deep Neural Network which can effectively deal with long temporal contexts with training times comparable to standard feed-forward DNNs (chain-TDNN (Povey et al., 2016)). Generic 3 and 4-gram language models, which allow users to compute the probability of emitting one word knowing a history of 2 or 3 words, were also trained and used during decoding. To help the reading, two sequence labeling systems (Conditional Random Field models) have been trained over manual transcripts to add punctuation and upper-case letters respectively.

The NE system, based on the NeuroNLP[4] toolkit, helps the text analysis. The manually annotated transcripts are used to train a text-to-text sequence labeling system. The system detects eight main entity types: amount, event, function, location, organization, person, product and time.

ASR was performed on the full collection of 1,200 national editions in order to feed Okapi and TXM platforms for historians' analyses (see Section 3): about 300 hours of video, resulting in more than 1.5 million words. A subset of 12 editions from 1945 to 1969 were manually transcribed to evaluate the audio analysis systems. Due to the 50-year time gap, human annotators had some difficulties with the spelling of NE, especially regarding people and foreign NE. Thanks to Wikipedia and INA thesaurus, most of NEs have been checked. However, speakers are very hard to identify. Most of them are male voice-overs. Their faces are never seen and their names are rarely spoken, nor displayed on the images. Only journalists performing interviews and well-known people, such as politicians, athletes and celebrities, can be accurately identified and named.

The quality of an ASR system is evaluated using the so-called Word Error Rate (WER). This metric consists of counting the number of insertions, deletions and substitutions of words between the transcripts automatically generated by the ASR and the human transcripts considered as an oracle. The WER is 24.27% on ANTRACT data using the generic ASR system trained on modern data. The same system evaluated on 2010 data[5] achieves 13.46%. It is known that ASR systems are sensitive to acoustic and language variations between train corpus and test corpus. Here, the WER is almost double. It is generally difficult to exploit transcripts in a robust way when WER is above 30%. Most of the errors come from unknown words (which are not listed in the 160k vocabulary). These out of vocabulary words (OOV) are confused with acoustically close words, which have a negative impact on neighboring words. The system always selects the most likely word sequence containing the word replacing the OOV.

Additional contemporary data, such as archival descriptions and typescripts, would be useful to adapt the language model. Therefore, abstracts, titles and descriptions have been extracted from the archival descriptions. OCR sentences (see Section 2.2) have been kept when at least 95% of the words belong to the ASR vocabulary. A "in domain" training corpus composed of 1.3 million words from archival descriptions and 4.7 million words from typescripts was built. The 4,000 most frequent words were selected to train the new ANTRACT language model, which reduces the error rate by half: from 24.27% to 12.06% WER. Figure 3 shows a sample of automatic transcription of the July 14, 1955 edition. The gain is significant thanks to the typescripts which are very similar to manual transcriptions. This "in domain" training corpus is contrary to the rules usually set during the well-known ASR system evaluations: a test data set should never be used to build a training corpus. However, in our case, the main goal is to provide the best transcripts to historians.

Future work will focus on ASR acoustic models improvement. We plan to use an alignment of typescripts with the editions, as well as historian users' feedback providing some manually revised transcriptions. The objective is to select zones of confidence to be added to the learning data. Evaluation of the Named Entities is the next step in the roadmap. The speaker evaluation will be more difficult because of their identities, which are not available. We plan to evaluate both the detection of voice-overs and interviewers. Furthermore, some famous persons, selected in collaboration with historians for their relevance in historical analyses, will also be identified, with the possible help of crossing results with image analysis as described in Section 2.4.

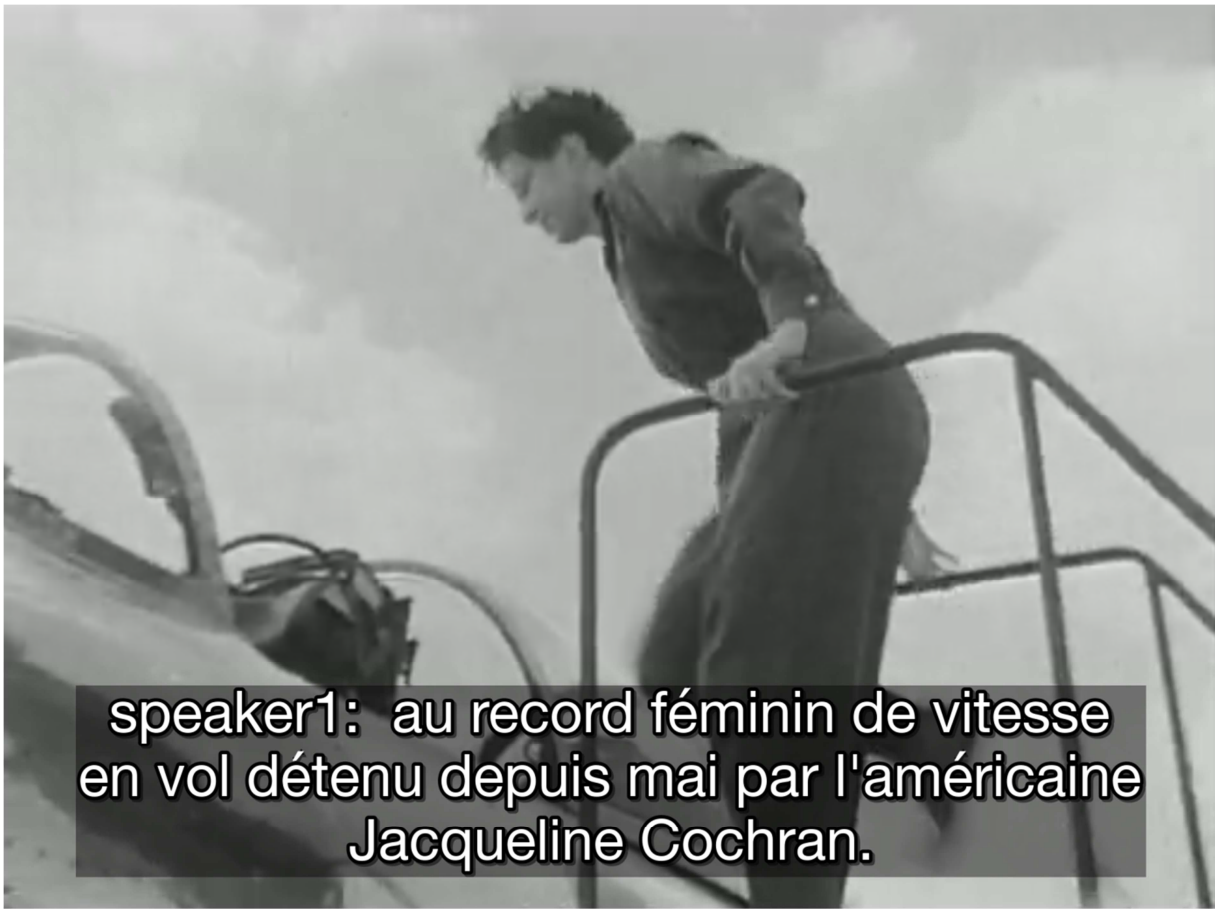


Figure 3. Sample “Actualité Francaise July 14, 1955 from 6:06 to 6:49”. Subtitle is an ASR file with in domain language model, automatic punctuation and upper case. @INA

2.4 Automatic visual analysis

Identifying the people appearing in a video is undoubtedly an important cue for its understanding. Knowing who appears in a video, when and where, can also lead to learning interesting patterns of relationships among characters for historical research. Such person-related annotations could provide ground for value added content. An historical archive such as the *Actualités Françaises* corpus contains numerous examples of celebrities appearing in the same news segment as De Gaulle and Adenauer (see Figure 4). However, the annotations produced manually by archivists do not always identify with precision those individuals in the videos. On the other side, the web offers an important amount of pictures of those persons, easily accessible through Search Engines using their full name as search terms. In ANTRACT, we aim to leverage these pictures for identifying faces of celebrities in video archives.

Figure 4. De Gaulle and Adenauer together in a video from 1959. @INA

There has been much progress in the last decade regarding the process of automatic recognition of people. It generally includes two steps: first, the faces need to be detected (i.e. which region of the frame may contain a person face) and then recognised (i.e. to which person this face belongs to).

The Viola-Jones algorithm (Viola, 2004) for face detection and Local Binary Pattern (LBP) features (Ahonen, 2006) for the clustering and recognition of faces were the most famous techniques used until the advent of deep learning and convolutional neural networks (CNN). Nowadays, two main approaches are in use to detect faces in video and both are using CNNs. The Dlib library (King, 2009) provides good performance for frontal images but it requires an additional alignment step (which can also be performed using the Dlib library) before face recognition can be performed. The recent Multi-task Cascaded Convolutional Networks (MTCNN) approach provides even better performance using an image-pyramid approach and integrates the detection of face landmarks in order to re-align detected faces to the frontal position (Zhang, 2016).

Having located the position and orientation of the faces in the video images, the recognition process can be performed in good conditions. Several strategies have been detailed in the literature to achieve recognition. Currently, the most practical approach is to perform face comparison using a transformation space in which similar faces are close together, and to use this representation to identify the right person. Such embeddings, computed on a large collection of faces, are often available to the research community (Schroff, 2015).

Within ANTRACT, we developed an open source Face Celebrity Recognition system. This application is made of the following modules:

- A web crawler which, given a person's name, automatically downloads from Google a set of k photos that will be used for training a particular face model. In our experiments, we generally use k = 50. Among the results, the images not containing any face or containing more than one face are discarded. In addition, end users (e.g. domain experts) can manually exclude wrong results, for example, corresponding to pictures that do not represent the searched person.
- A training module where the retrieved photos can be converted to black-and-white, cropped and resized in order to obtain images only containing a face, using the MTCNN algorithm (Zhang, 2016). A pre-trained Facenet (Schroff, 2015) model with Inception ResNet v1

architecture trained on VGGFace2dataset (Cao, 2018) is applied in order to extract visual features of the faces. The embeddings are used to train a SVM classifier.

- A recognition module where a newsreel video is received as input and from which all frames are extracted at a given skipping distance *d* (in our experiments, we generally set *d* = 25, namely 1 sample frame per second). For each frame, the faces are detected (using the MTCNN algorithm) and the embeddings computed (Facenet). The SVM classifier decides if the face matches the ones among the training images.
- Simple Online and Realtime Tracking (SORT) is an object tracking algorithm, which can track multiple objects in real-time (Bewley, 2016). Its implementation is inspired by the suggestion code from Linzaer[6]. The algorithm uses the MTCNN bounding box detection and tracks it across frames. We introduced this module to increase the robustness of the library. By introducing this module, while making the assumption that faces do not swap coordinates across consecutive frames, we aim to get a more consistent prediction.
- Finally, the last module groups together the results coming from the classifier and the tracking modules. We observe that even though the face to recognize remains the same over consecutive frames, the face prediction sometimes changes. For this reason, we select for each tracking the most frequently occurring prediction, taking also into account the confidence score given by the classifier. In this way, the system provides a common prediction for all the frames involved in a tracking, together with an aggregated confidence score. A threshold *t* can be applied to this score in order to discard the low-confidence prediction. According to our experiments, t = 0.6 gives a good balance between precision and recall.

In order to make the software available as a service, we wrapped it into a RESTful web API, available at http://facerec.eurecom.fr/. The service receives as input the URI of a video resource, as it appears in Okapi, from which it retrieves the media object encoded in MPEG-4. Two output formats are supported: a custom JSON format and a serialization format in RDF using the Turtle syntax and the Media Fragment URI syntax (Troncy et al., 2012), with normal play time (*npt*) expressed in seconds to identify temporal fragments and *xywh* coordinates to identify the bounding box rectangle encompassing the face in the frame. A third format, again following the Turtle syntax, will be soon implemented so that the results can be directly integrated in the Okapi Knowledge Graph. A light cache system is also provided in order to enable serving pre-computed results, unless the no cache parameter is set which is triggering a new analysis process.

We run experiments using the face model of Dwight D. Eisenhower on a selection of video segments extracted from Okapi, among the ones that have been annotated with the presence of the American president using the *ina:imageContient* and *ina:aPourParticipant* properties in the knowledge graph. In the absence of a ground truth, we performed a qualitative analysis of our system on three videos. For each detected person, we manually assessed whether the correct person was found or not. Out of the 90 selected segments, the system correctly identified Eisenhower in 33 of them. However, we are not sure that Eisenhower is effectively visually present in all 90 segments. We are currently working on extracting from the ANTRACT corpus a set of annotated segments to be used as ground truth so that it is possible to measure the precision and recall of the system.

In addition, we made the following observations:

- The library generally fails in detecting people when they are in the background, or when the face is occluded.
- When faces are perfectly aligned, they are easier to detect. Improvements on the alignment algorithm are foreseen as future work.

- When setting a high confidence threshold, we do not encounter cases where we confuse one celebrity by another one. Most errors are about confusing an unknown face with a celebrity in the dataset.

In order to easily visualize the results and to facilitate history scholars' feedback, we developed a web application that shows the results directly on the video, leveraging on HTML5 features. The application also provides a summary of the different predictions, enabling the user to directly jump to the relative part of the video where the celebrity appears. A slider allows changing the confidence threshold value, in order to better investigate the low-confidence results.
The application is publicly available at http://facerec.eurecom.fr/visualizer/?project=antract (see Figure 5).

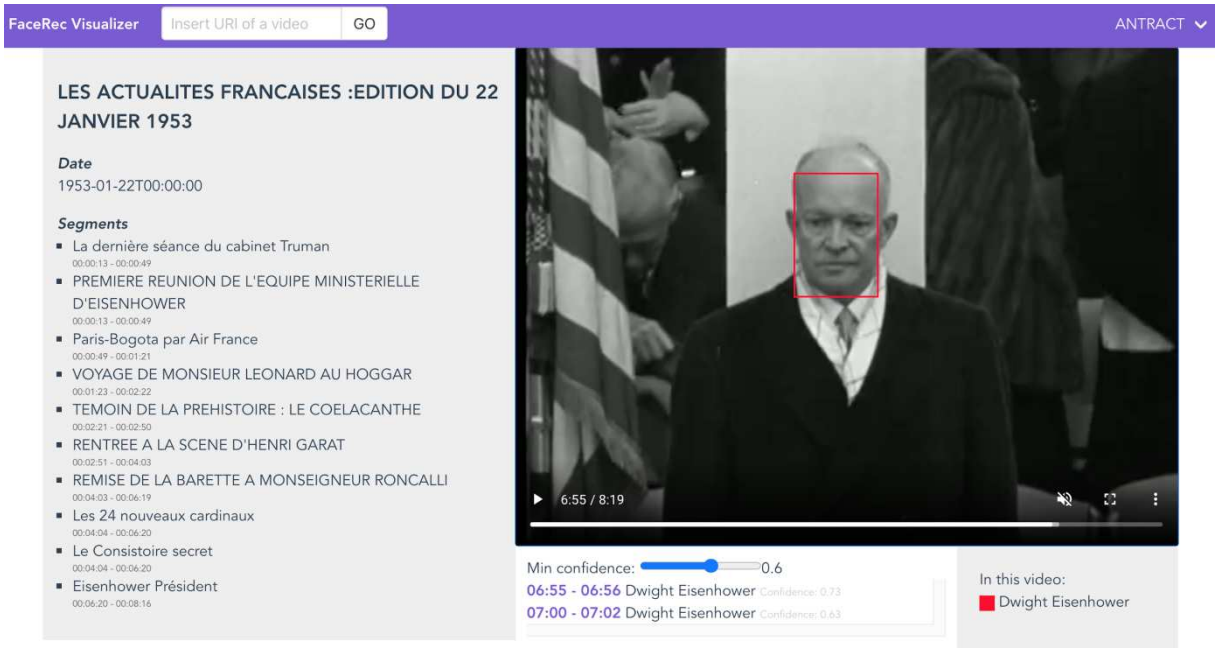


Figure 5: The visualizer of the Celebrity Face Recognition System

3. Platforms for historians' exploration and analysis of the corpus

The corpus built with automatic tools in section 2 is explored interactively by historians using two platforms:

- the TXM platform for analysis of text corpora based on quantitative and qualitative exploration tools, and augmented during the ANTRACT project to facilitate the link between textual data and audio and video data;
- the Okapi knowledge-driven platform for the management and annotation of video corpora using semantic technologies.

3.1 The TXM platform for interactive textometric analysis

Text analysis is achieved through a textometric approach (Lebart et al., 1998). Textometry combines both quantitative statistical tools and qualitative text searching, reading and annotating. On the one hand, statistical functionalities include keyword analysis, collocations, clustering and correspondence analysis. This makes a significant analytical power addition in comparison with usual annotation and search & count features in audiovisual transcription software such as CLAN (MacWhinney, 2000) or

ELAN (ELAN, 2018). On the other hand, yet again in the textometric approach, qualitative analysis is carried out by advanced KWIC concordancing, by placing an emphasis on easy-access to high quality of layout rendering of source documents and by providing annotation tools. Such a qualitative side is marginal if not absent in conventional text mining applications (Hotho et al., 2005; Feinerer et al., 2008; Weiss et al., 2015): most of them process plain text, getting rid of text body markup, if any, and aim at synthetic visualization displacing close text reading.

Textometry is implemented by the TXM software platform (Heiden, 2010). TXM is produced as an open-source software, which integrates several specialized components: R (R Core Team, 2014) for statistical modeling, CQP for full text search engine (Christ, 1994), TreeTagger (Schmid, 1994) for Natural Language Processing (morphosyntactic tagging and lemmatization). TXM is committed to data and software standardization and sharing efforts, and has notably be designed to manage richly-encoded corpora, such as XML data and TEI[7] encoded texts ; for ANTRACT textual data, TXM imports tabulated data (Excel format export of tables from INA documentary databases) and files in the Transcriber XML format provided by speech-to-text software (see Section 2.3). TXM is dedicated to text analysis, but also helps to manage multimedia representations associated with the texts, whether it is scanned images of source material, audio or video recordings: actually, these representations participate in the interpretation of TXM common tools results in their full semiotic context.

In 2018, we began to build the AFNOTICES TXM corpus by importing the INA archival descriptions: each news report is represented by several textual fields (title, abstract, sequence description) and several lexical fields (keyword lists of different types such as topics, people, or places, and credits with names of people shown or cameramen) and labeled by a dozen metadata (identifier, broadcast date, film producer, film genre, etc.) which are useful to contextualize or categorize reports.

In 2019, we began the production of the AFVOIXOFFV02 TXM corpus which makes the voice-over transcripts (see Section 2.3) searchable and available for statistical analysis, synchronized at the word level for video playback and labeled by INA documentary fields.

These corpora may still be augmented by aligning new textual modalities: texts from narration typescripts (OCR text and corresponding regions in the page images) (see Section 2.2), annotations on videos (manual annotations added by historians through the Okapi platform (see Section 3.2), as well as automatic annotations generated by image recognition software (see Section 2.4), named entities, etc.

One of the technical innovations achieved for the project has been the consolidation of TXM back-to-media component (Pincemin et al., 2020), so that any word or text passage found in the result of a textometric tool can be played with its original video; we have also implemented authenticated streamed access to video content from the Okapi media server, which happened to be a key development for video access given the total physical size and the security constraints of such film archive data.

The following screenshots illustrate typical textometric analysis moments of current studies within the ANTRACT project.

In Figure 6 and Figure 7, we study the context of use for the word "*foule*" (crowd), through a KWIC concordance. A double-click on a concordance line opens up a new window (on the right-hand side) which displays the complete transcript in which the word occurs. Then, we click on the music note symbol at the beginning of the paragraph to play the corresponding video. A dialog box prompts for credentials before accessing the video on the Okapi online server. This opportunity to confront textual analysis with the audiovisual source is all the more important here because textual data were generated by the speech-to-text automatic component, whose output could not be fully revised. Moreover, the video may add significant context that is not rendered in plain text transcription.

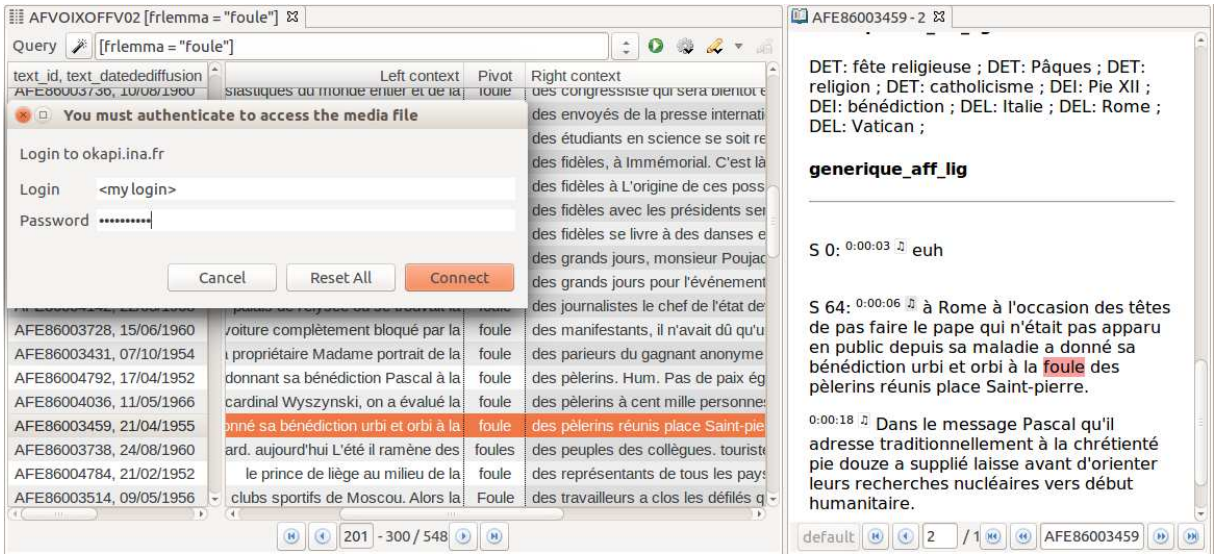


Figure 6. CONCORDANCE of the word "*foule*" (crowd) in the voice-over corpus (left window), voice-over transcript EDITION corresponding to the selected concordance line (right window), and the authentication dialog box to access the Okapi video server to play the video at 0:00:06 (top left window).

Figure 7. Hyperlinked windows managing results associated with the word "*foule*" (crowd): CONCORDANCE (left window), transcript EDITION (middle window) and synchronized video playback (right window)

Our second example is about the place of agriculture and farmers in the *Actualités françaises*, and how the topic is presented. It shows how one can investigate if a given word has the same meaning in documentation and in commentary, or if different words are used when dealing with the same subject. We first get (Figure 8) a comparative overview of the quantitative evolution of occurrences from two word families, derived from the stems of "*paysan*" and "*agricole*"/ "*agriculture*" (see detailed list of words in Figure 9, left hand side window). We complete the analysis with contextual analysis through KWIC concordance views (see Figure 8, lower window) and cooccurrences computing (see Figure 9). We notice that "*paysan*" becomes less used from 1952 onwards, and that it is preferred to "*agriculteur*" when speaking of the individuals present in the newsreels extracts; conversely, "*agricole*"/"*agriculture*" are used in a more abstract way, to deal with new farm equipment and socio-economic transformation of this line of business.

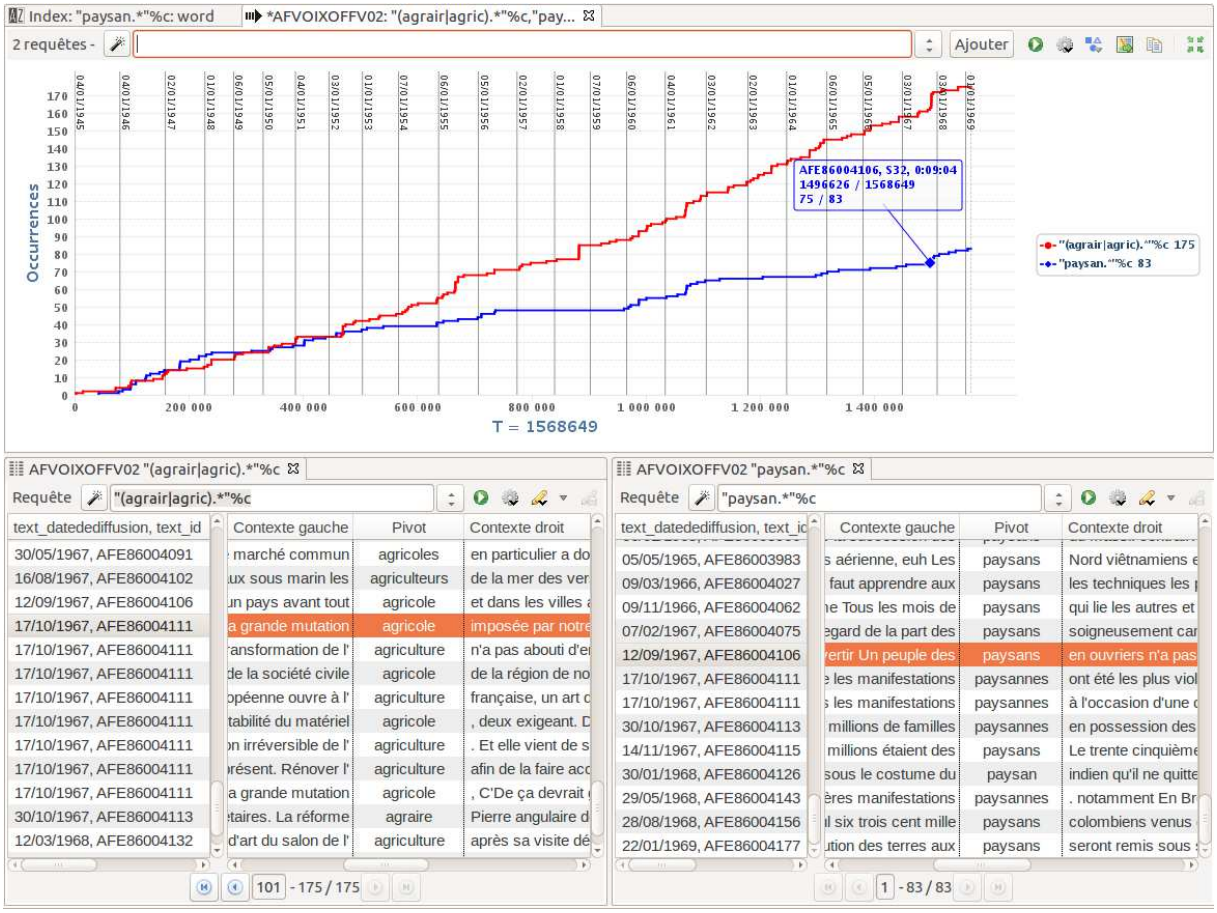


Figure 8. PROGRESSION chart (upper window), and hyperlinked KWIC CONCORDANCES (lower window), to compare two word families related to farming

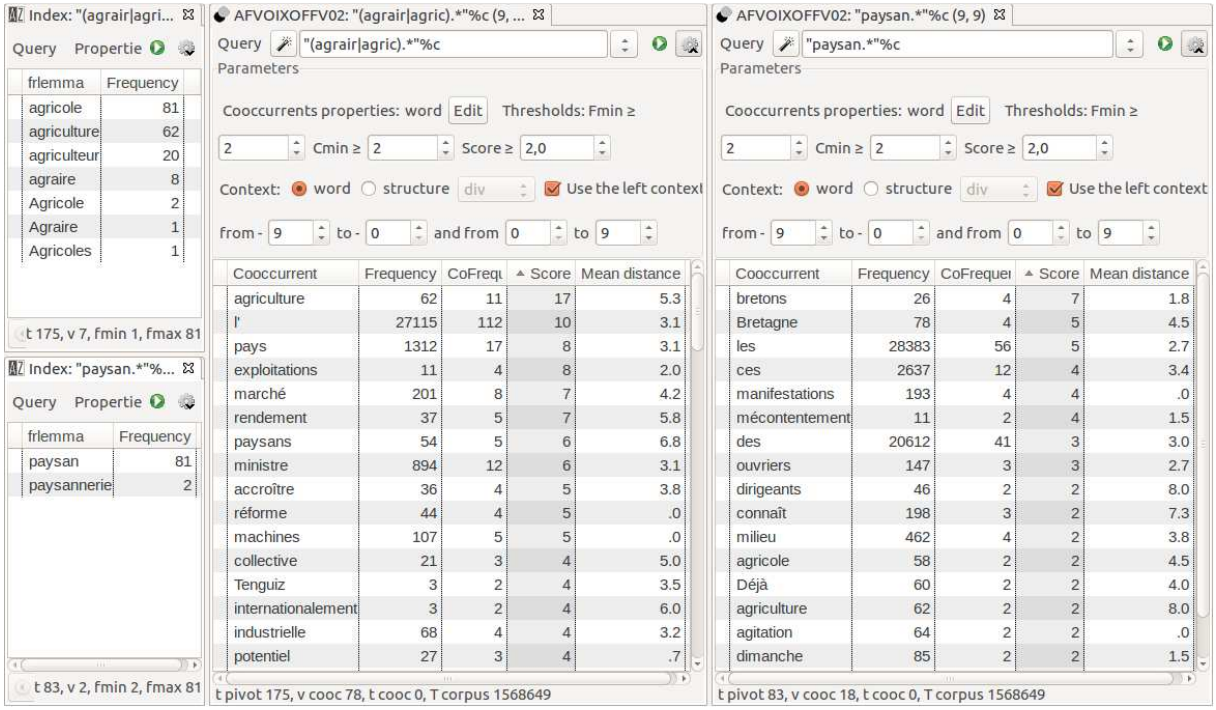


Figure 9. INDEX results detailing the content of two word families (left margin), and COOCCURRENCES statistical analysis to characterize their contexts

Combining word lists (INDEX) and morphosyntactic information is very effective to summarize phrasal contexts. For instance, in Figure 8, we can compare which adjectives qualify "*foule*" in the archival descriptions, and which ones qualify "*foule*" in the voice-over speeches. For a given phrase ("*foule immense*", huge crowd) in the voice-over, we compute its cooccurrences in order to identify in which kind of circumstances the phrase is preferred (funerals, religious meetings). In TXM, full-text search is powered by the extensive CQP search engine (Christ, 1994), which allows very fine-tuned and contextualized queries.

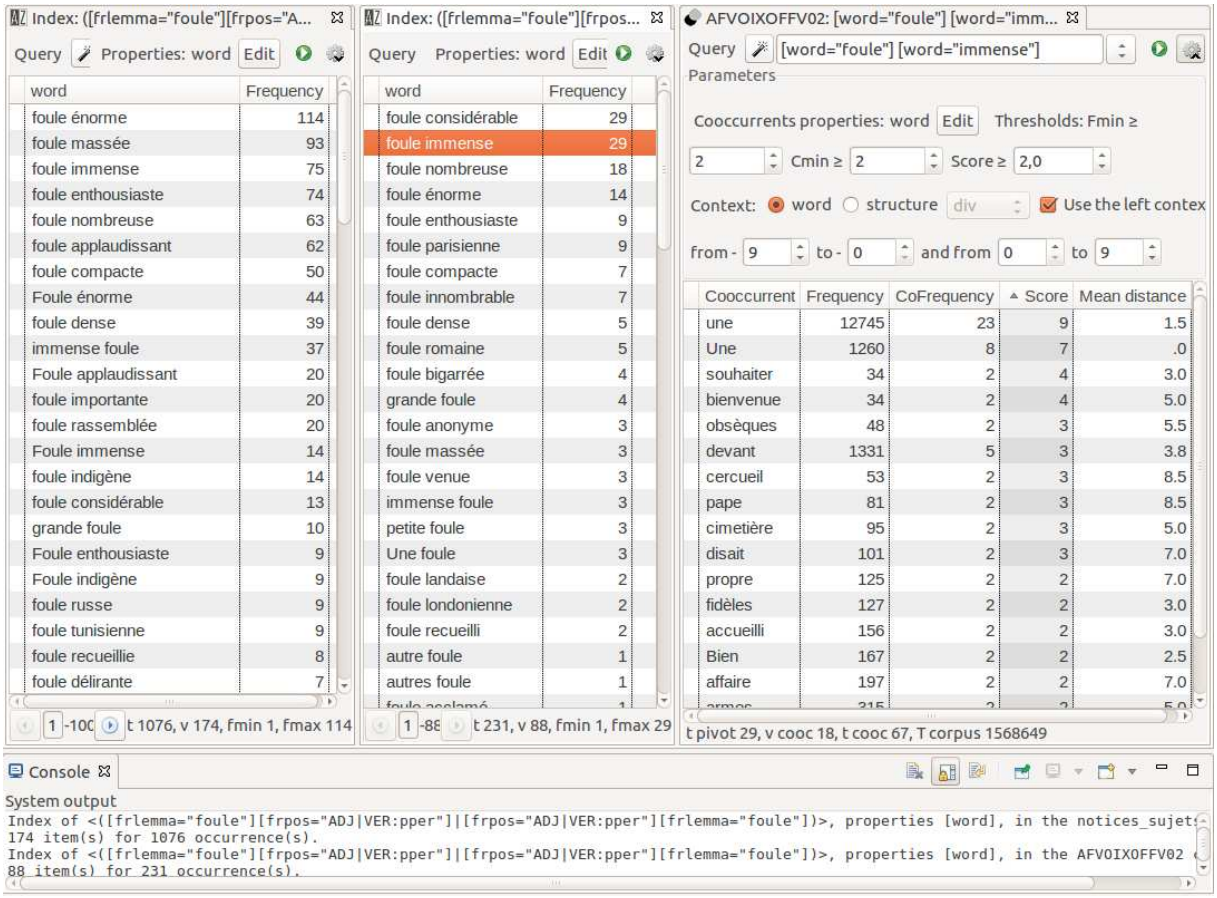


Figure 10. INDEX of "*foule*" (crowd) preceded or followed by an adjective, in archival descriptions (left window) or in voice-over transcripts (middle window). COOCCURRENCES for "*foule immense*" (huge crowd) in voice-over transcripts (right window)

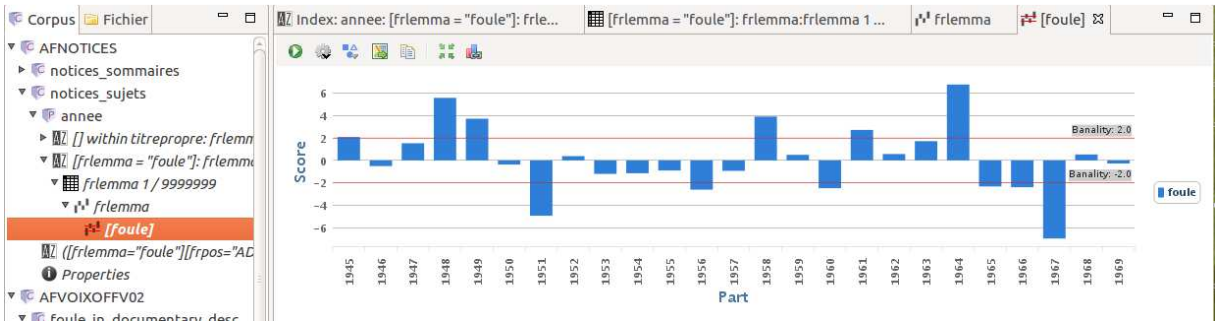
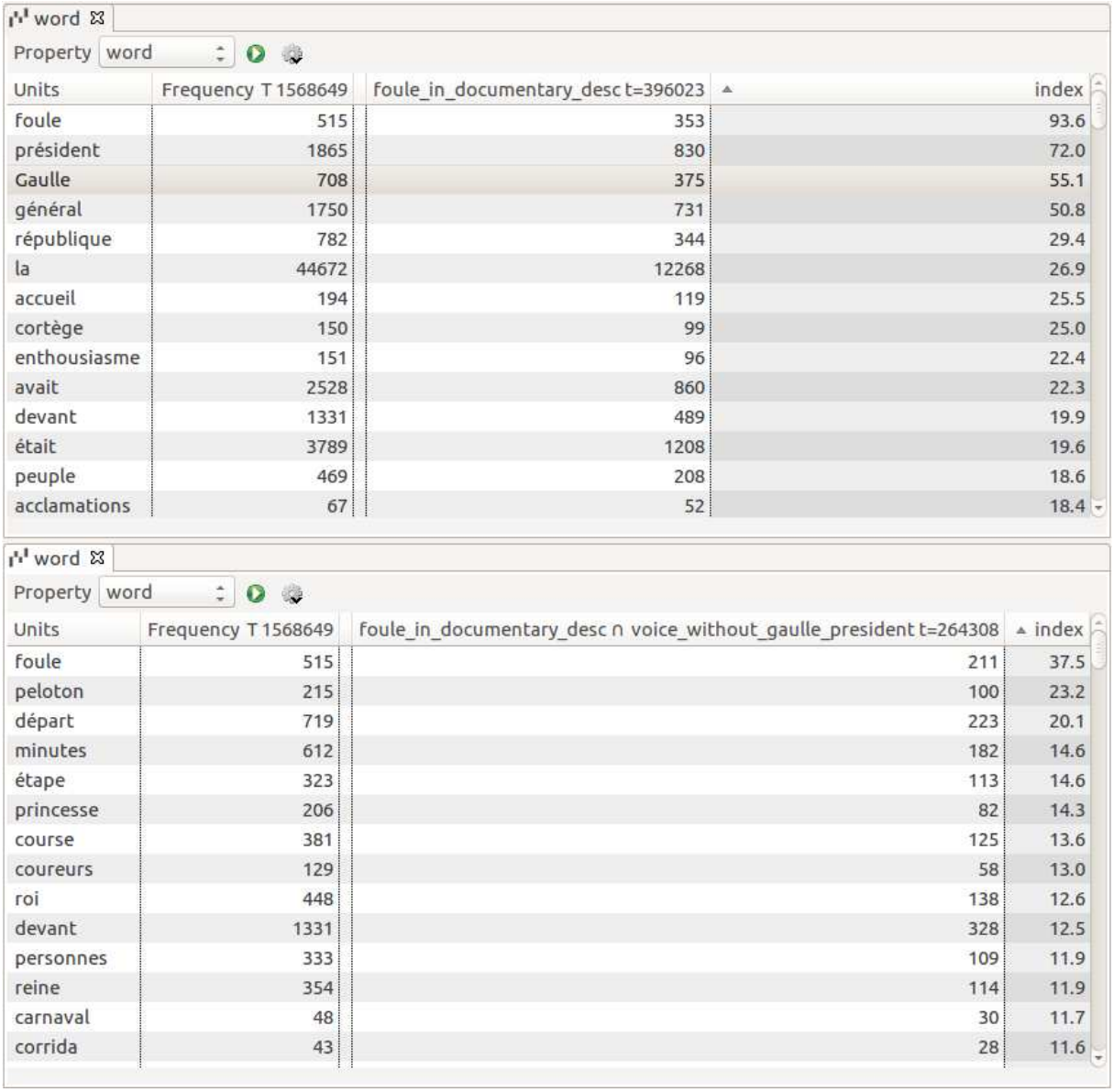


Figure 11. Statistical SPECIFICITY chart for "*foule*" (crowd) over the years

For chronological investigations, we can divide the corpus into time periods in a very flexible way, such as years or groups of years. Any encoded information may be used to build corpus subdivisions. Then the SPECIFICITY command —that implements a Fisher's Exact Test, known as one of the best

calculations to find keywords (McEnery and Hardie, 2012)— statistically measures the steady use, or the singular overuse or underuse of any word. The function can also be used to bring to light the specific terms for a given period, or for any given part of the corpus. For example, (Figure 9) focuses on the word "foule"over the years. Peak years reveal important political events (e.g. the liberation of France after WW2, the advent of the Fifth Republic), which match the high exposure of Général de Gaulle. However, the most frequent occurrences do not necessarily correspond to political upheavals.



word

Property word

Units	Frequency T 1568649	foule_in_documentary_desc t=396023	index
foule	515	353	93.6
président	1865	830	72.0
Gaulle	708	375	55.1
général	1750	731	50.8
république	782	344	29.4
la	44672	12268	26.9
accueil	194	119	25.5
cortège	150	99	25.0
enthousiasme	151	96	22.4
avait	2528	860	22.3
devant	1331	489	19.9
était	3789	1208	19.6
peuple	469	208	18.6
acclamations	67	52	18.4

word

Property word

Units	Frequency T 1568649	foule_in_documentary_desc ∩ voice_without_gaulle_president t=264308	index
foule	515	211	37.5
peloton	215	100	23.2
départ	719	223	20.1
minutes	612	182	14.6
étape	323	113	14.6
princesse	206	82	14.3
course	381	125	13.6
coureurs	129	58	13.0
roi	448	138	12.6
devant	1331	328	12.5
personnes	333	109	11.9
reine	354	114	11.9
carnaval	48	30	11.7
corrida	43	28	11.6

Figure 12. Example of resonance analysis (Salem, 2004): SPECIFIC terms in voice-over comments for reports showing a crowd (according to archival description) (upper window) ; then, SPECIFIC terms in voice-over comments for reports showing a crowd and having no mention of De Gaulle or "président" (president) (lower window)

With Figure 12, we apply a statistical resonance analysis (Salem, 2004). When a crowd is shown (as indicated by the archival description), what are the most characteristic words said by the voice-over? "Président" and "[le général De] Gaulle" represent the main context (Figure 12, upper window). In a second step, we remove all the reports containing one of these two words and focus on the remaining reports to bring out new kinds of contexts associated with the view of a crowd (Figure 12, lower window), such as sports, commemorative events, demonstrations, festive events, etc. The recurring

term "*foule*" (crowd) in the voice-overs promotes a sense of belonging to a community of fate. From a methodological perspective, this kind of cross-querying combined with statistical comparison between textual newsreel archival descriptions and commentary transcripts helps investigate correlations or discrepancies between what is shown in the newsreels and what is said in their commentaries. Such a combination of statistics across media is rarely provided by applications.

Figure 13 provides a first insight of a correspondence analysis output: we computed a 2D-map of the names of people who are present in more than 20 reports, in relation with the years in which they are mentioned. We thus get a synthetic view of the relationship between people and time in the *Actualités françaises* reports. In terms of calculation, as textometry often deals with frequency tables crossing words and corpus parts (here we crossed people's names and year divisions), it then opts for correspondence analysis, because this type of multidimensional analysis is best suited to such contingency tables (Lebart et al., 1998).

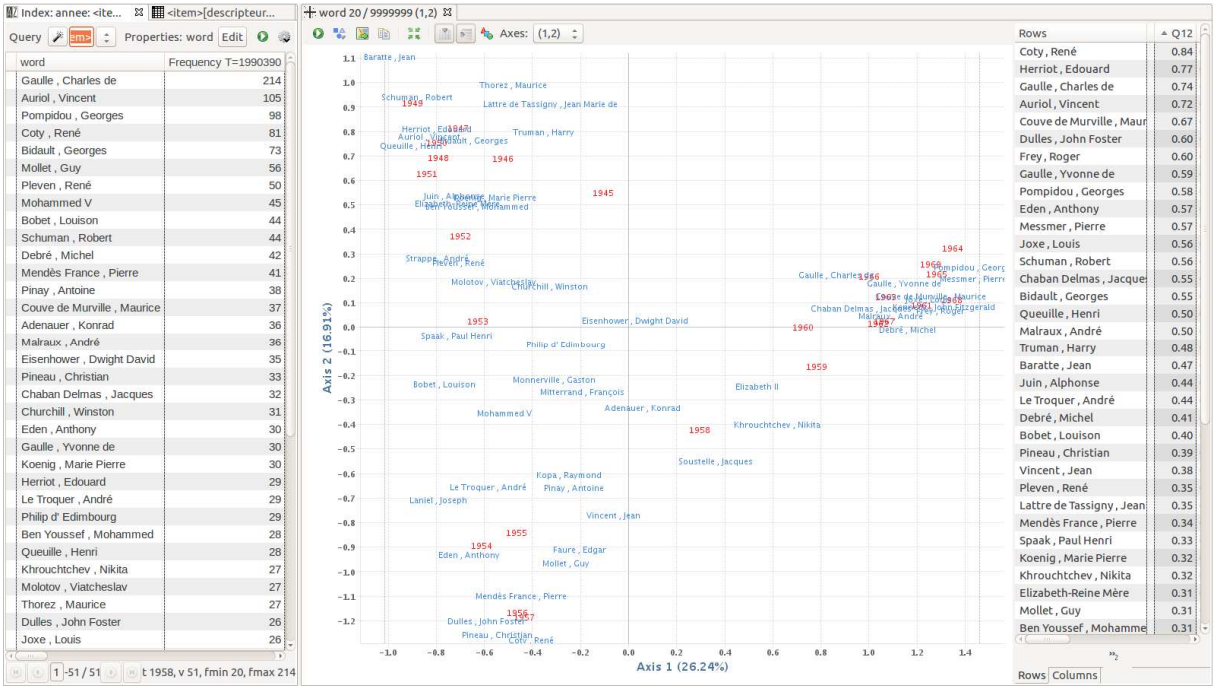


Figure 13. CORRESPONDENCE ANALYSIS (first plane) of the frequency table crossing the years and the names of 51 people that are present in at least 20 reports

3.2 Okapi platform for interactive semantic analysis

Okapi (Open Knowledge Annotation and Publication Interface) (Beloued et al., 2017) is a knowledge-based online platform for semantic management of content. It is at the intersection of three scientific domains: Indexing and description of multimedia content, knowledge management systems and Web content management systems. It takes full advantage of semantic web languages and standards (RDF, RDFS, OWL (Motik et al., 2012)) to represent content as graphs of knowledge; it applies semantic inferences on these graphs and transforms them to generate new hypermedia content like web portals.

Okapi provides a set of tools for analyzing multimedia content (video, image, sound) and managing corpora of annotated video and sound excerpts as well as image sections. Analysis tools allow the semantic indexing and description of content using domain ontology. The corpus management tools provide services for the constitution and visualization of thematic corpora as well as their annotation and enrichment in order to generate mini-portals or thematic publications of their contents.

The Okapi's knowledge management system stores knowledge as graphs of named entities and provides services to retrieve, share and present them as linked open data. These entities can be aligned with other entities in existing knowledge bases like dbpedia and wikidata and so makes Okapi interoperable with the Linked Open Data ecosystem (Bizer et al., 2009). The named entities can be of different types and categories and vary according to the studied domain. For instance, for audiovisual archives, entities may concern persons, geographical places and concepts.

Finally, the Okapi's Content Management System (Okapi's CMS) considers the characteristics of the studied domain and user preferences to generate web interfaces and tools for Okapi as well as content portals adapted to the domain. This publishing framework allows also authors to focus on their authoring work and to create thematic portals without any technical skills. The author can specify his thematic publication as a set of interconnected multimedia elements (video, image, sound, editorial texts). The framework applies thereafter a set of publishing rules on these elements and generates a web site.

The Okapi platform is used by historians to constitute thematic corpora and to publish their portals as explained in the above paragraphs. Okapi can also be used by researchers in computer science and data scientists to show and improve the results of their automatic algorithms (face detection and recognition, automatic speech recognition, etc.). The following sections show some examples of how the Okapi platform can be used on the collection "*Les Actualités Françaises*" (AF) in the context of the ANTRACT project.

The media analysis can be carried out manually by annotators or automatically by algorithms on several axes as shown in Figure 14. In this example, thematic analysis (the layer entitled "*strate sujets*") of the AF program "*Journal Les Actualités Françaises : émission du 10 juillet 1968*" consists in identifying the topics addressed in this program, their temporal scope and a detailed description of the topic in terms of the subject we are talking about, the places where it happens and the persons who are involved in this subject.

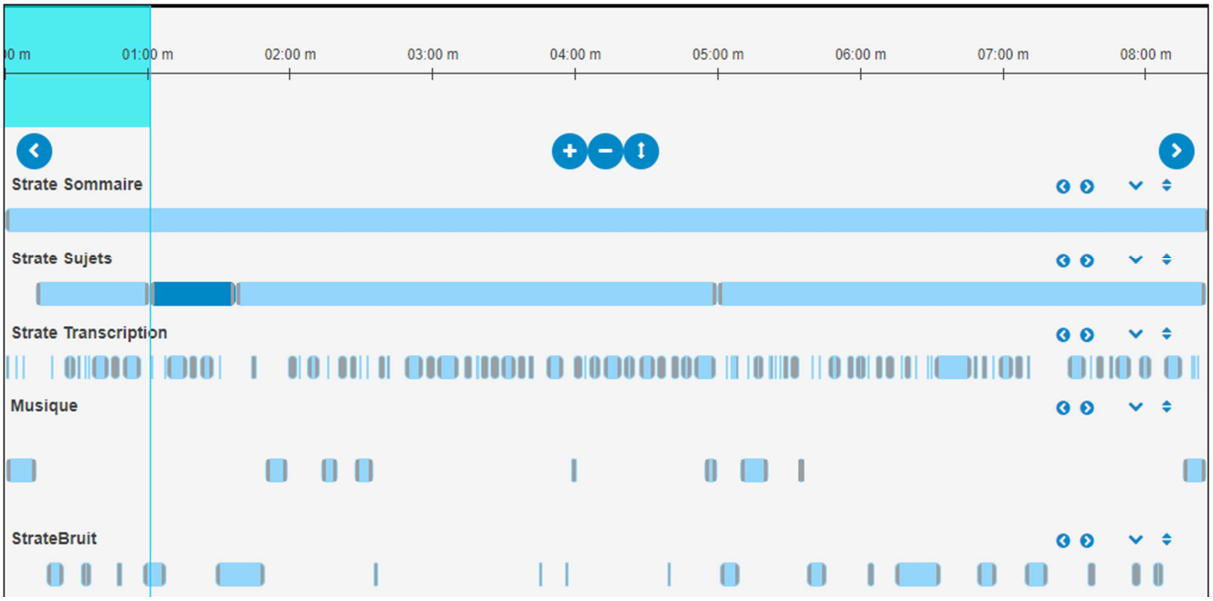


Figure 14. Timeline for Media Analysis

The user can create and remove analysis layers and their segments as well as the description of each segment and its timecodes. Considering the second segment in the example where we are talking about the **water sports** (**concept**) in **England** (**Place**), especially the adventures of the solo sailor **Alec Rose** (**Person**) as indicated in the following form (Figure 15): The user can edit this form to change and

create new description values of the selected segment. These concepts, places and persons are a subset of named entities that are managed and suggested by Okapi to complete the description of the segment.

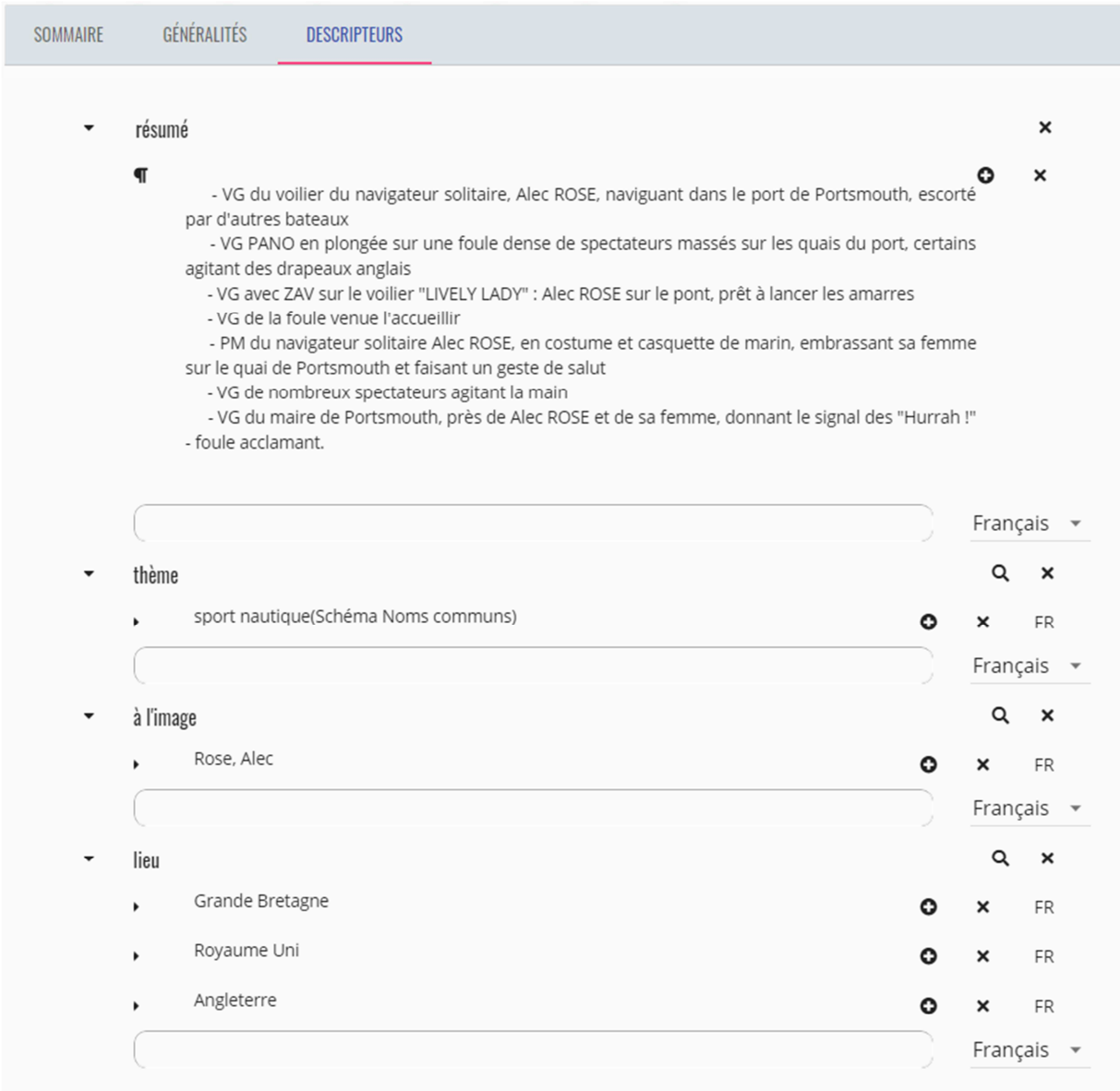


Figure 15. Segment Metadata Form

The other analysis layers (transcription, music detection, etc.) are provided by automatic algorithms. The metadata provided by these algorithms can enrich the ones created manually by users and can be used by the Okapi platform to generate a rich portal that brings value to the content and provides several access and navigation possibilities in the content as shown in Figure 16.

Figure 16. Okapi portal page of the AF news "Journal Les Actualités Françaises: émission du 10 Juillet 1968".

The generated metadata are also used as advanced criteria for looking for video excerpts and so allow users to constitute their thematic corpora focused on some topics. Figure 15 shows an example of an advanced search of segments which talk about "**Water sports**" in "**England**". Like all Okapi's objects, a query is represented as a knowledge graph and then transformed into a SPARQL query. The results of this query, illustrated by Figure 17 and 18, can be used to create a corpus.

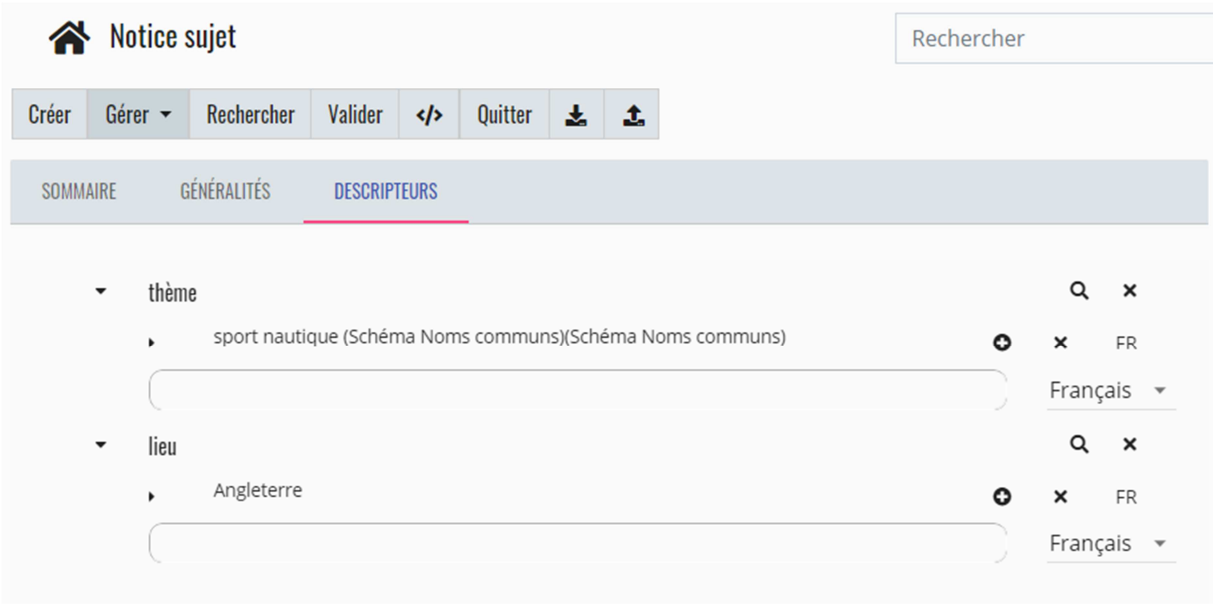


Figure 17. Example of an Okapi Query

Figure 18. Example of query results

The corpus itself is an object to be annotated, i.e, the user can add new metadata on the corpus itself or on its elements (video excerpts) and put rhetorical relations between them. Figure 19 shows a corpus of three excerpts, retrieved from the query presented in the previous paragraph. It displays also a rhetorical relationship between the two segments: "*Robert Manry, 48 ans: Traversée solitaire de l'océan*" which illustrates the other segment "*Alec Rose, après 354 jours sur un bateau: "la terre est ronde"*". All these metadata will be used to create a thematic portal focused on the content of the corpus or integrated into a story through the inclusion of editorial content and preferred reading paths.

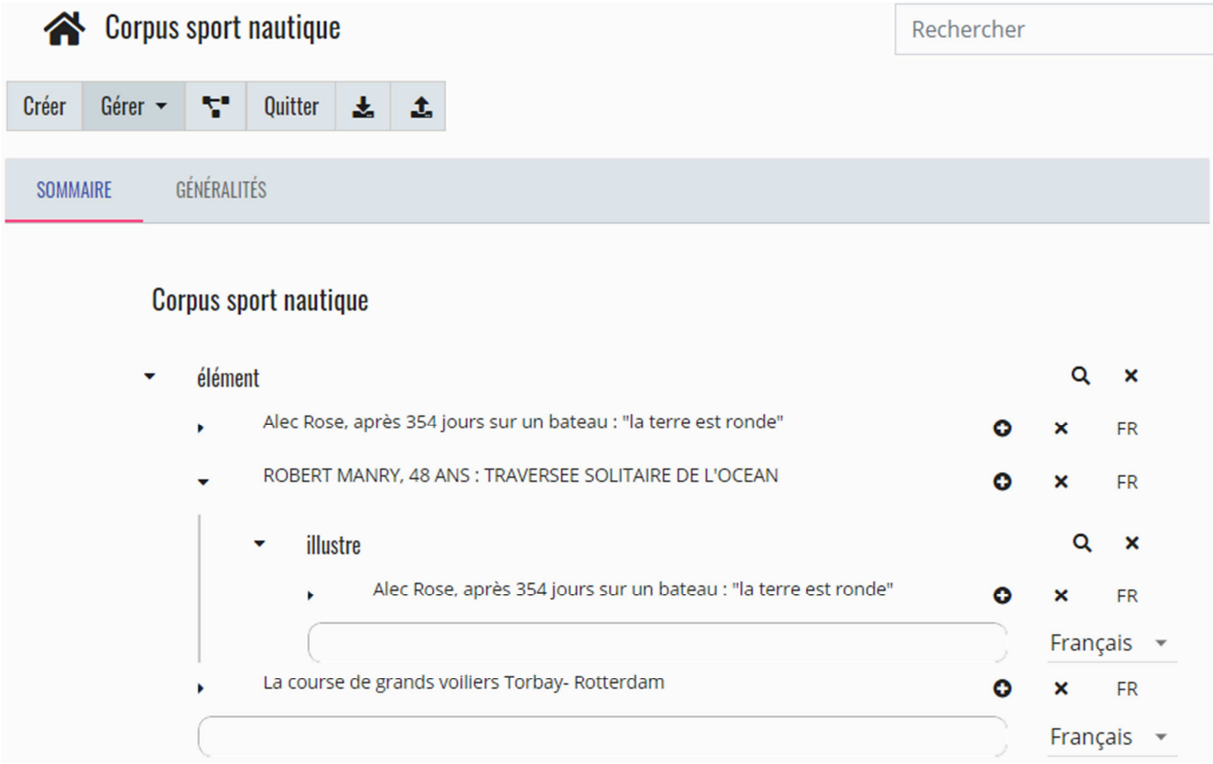


Figure 19. Thematic Corpus “Water Sports”

The Okapi platform exposes a secure SPARQL endpoint and API which allows other ANTRACT tools, especially the TXM platform, to query the knowledge base and to update the stored metadata. For instance, TXM tools could retrieve metadata through the Okapi’s endpoint in order to constitute a corpus. This corpus will then be stored in the knowledge base through the API and used by Okapi to provide thematic publications. Additional semantic descriptors produced by TXM could also be integrated into the Okapi knowledge base.

4. Conclusion

Presented throughout this article, the ANTRACT project’s challenge is to familiarize scholars with the automated research of large audiovisual corpora. Gathering instruments specialized in image, audio and text analysis into a single multimodal apparatus designed to correlate their results, the project intends to develop a transdisciplinary research model suitable to open new perspectives in the study of single or multi-format sources.

At this point of the project, most of the work is dedicated to the development and tuning of the automatic content analysis tools as well as the application of their results to the organization and improvement of the corpus data in connection with research provided by ANTRACT historians (Goetschel, 2019). Their case studies were explored using the TXM textometry platform and the Okapi annotation and publication platform that allows its users to exploit all the data produced by the instruments developed for the project.

From a technological perspective, ANTRACT’s goal is now to further adapt automatic content analysis tools to the specificity of the corpus such as its historical context, its vocabulary, its image format and quality, as it has been done, for instance, by improving the language models used by the automatic transcription tools with the help of the typescripts of voice-overs. Interactive analysis

platforms should also benefit from history scholars feedback in order to improve their user interface and to develop new analytical paths.

At the end of the project, a comprehensive *Les Actualités Françaises* corpus completed with its metadata as well as the results of the research supported by automatic content analysis tools and manual annotations will be made available to the scientific community via the online Okapi platform. To this end, Okapi tutorials will be provided to the public and TXM will continue to be available as an open source software to help the analysis of corpora used in new case studies. Okapi source code will be turned to open source so that other developers can contribute to its enhancement.

Regarding humanities, ANTRACT tools and methodology can be adapted to various types of corpora providing historians as well as specialists from other disciplines such as sociology, anthropology and political science a renewed access to their documents supported by an exhaustive examination of their content.

Acknowledgment

This work has been supported by the French National Research Agency (ANR) within the ANTRACT Project, under grant number ANR-17-CE38-0010, and by the European Union's Horizon 2020 research and innovation program within the MeMAD project (grant agreement No. 780069).

References

[Ahonen 2006] Ahonen, T., Hadid, A., and Pietikainen, M. "Face description with local binary patterns: Application to face recognition", IEEE Transactions on Pattern Analysis & Machine Intelligence, 28.12 (2006): 2037–2041.

[Althaus 2018] Althaus, S., Usry, K., Richards, S., Van Thuyle, B., Aron, I., Huang, L., Leetaru, K., Muehlfeld, M., Snouffer, K., Weber, S., Zhang, Y., and Phalen, P. "Global News Broadcasting in the Pre-Television Era: A Cross-National Comparative Analysis of World War II Newsreel Coverage", Journal of Broadcasting and Electronic Media, 62.1 (2018): 147-167.

[Atkinson 2011] Atkinson, N. S., "Newsreels as Domestic Propaganda: Visual Rhetoric at the Dawn of the Cold War", Rhetoric & Public Affairs, 14.1 (2011): 69-100.

[Bartels 2004] Bartels, U. Die Wochenschau im Dritten Reich. Entwicklung und Funktion eines Massenmediums unter besonderer Berücksichtigung völkisch-nationaler Inhalte. Peter Lang, Frankfurt am Main (2004).

[Beloued 2017] Beloued, A., Stockinger, P., and Lalande, S. "Studio Campus AAR: A Semantic Platform for Analyzing and Publishing Audiovisual Corpuses." In Collective Intelligence and Digital Archives, John Wiley & Sons Inc., Hoboken, NJ (2017): 85-133.

[Bewley 2016] Bewley, A., Ge, Z., Ott, L., Ramos, F. and Upcroft, B. "Simple online and realtime tracking". In IEEE International Conference on Image Processing (ICIP) (2016): 3464–3468.

[Bizer 2009] Bizer, C., Heath, T., and Berners-Lee, T. "Linked data - the story so far", International Journal on Semantic Web and Information Systems, 5 (2009): 1-22.

[Bradski 2000] Bradski, G. "The OpenCV Library", Dr. Dobb's Journal of Software Tools (2000).

[Broux 2018] Broux, P.-A., Desnous, F., Larcher, A., Petitrenaud, S., Carrive, J., and Meignier, S. "S4D: Speaker Diarization Toolkit in Python" Interspeech, Hyderabad, India (2018).

[Cao 2018] Cao, Q., Shen, L., Xie, W., Parkhi, O. M. and Zisserman, A. “Vggface2: A dataset for recognising faces across pose and age”. In 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG) (2018): 67–74.

[Chambers 2018] Chambers, C., Jönsson, M., and Vande Winkel R. (eds.) Researching Newsreels. Local, National and Transnational Case Studies. Global Cinema, Palgrave Macmillan, London (2018).

[Chenot 2014] Chenot, J.-H., and Daigneault, G. “A large-scale audio and video fingerprints-generated database of TV repeated contents” In 12th International Workshop on Content-Based Multimedia Indexing (CBMI), Klagenfurt, Austria (2014).

[Christ 1994] Christ, O. “A modular and flexible architecture for an integrated corpus query system” In Ferenc Kiefer et al. (eds.), In 3rd International Conference on Computational Lexicography, Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest (1994): 23-32.

[Deegan 2012] Deegan, M., and McCarty, W. Collaborative Research in the Digital Humanities. Ashgate, Farnham, Burlington (2012).

[ELAN, 2018] ELAN (Version 5.2) [Computer software]. Max Planck Institute for Psycholinguistics, Nijmegen (2018). Retrieved from https://tla.mpi.nl/tools/tla-tools/elan

[Fein 2004] Fein, S. “New Empire into Old: Making Mexican Newsreels the Cold War Way”, Diplomatic History, 28.5 (2004): 703-748.

[Fein 2008] Fein, S. “Producing the Cold War in Mexico: The Public Limits of Covert Communications” In G. M. Joseph and D. Spenser (eds.), In from the Cold: Latin America’s New Encounter with the Cold War, Duke University Press, Durham (2008): 171-213.

[Feinerer 2008] Feinerer, I., Hornik, K., and Meyer, D. “Text Mining Infrastructure in R”, Journal of Statistical Software, 25.5 (2008): 1-54.

[Goetschel 2011] Goetschel, P., Granger, C. (dir.) “Faire l’événement, un enjeu des sociétés contemporaines”, Sociétés & Représentations, 32 (2011): 7-23.

[Goetschel 2019] Goetschel, P. “Les Actualités Françaises (1945-1969) : le mouvement d’une époque”, #1257, 1 (2019): 34-39.

[King 2009] King, D. E. “Dlib-ml: A machine learning toolkit”. Journal of Machine Learning Research, 10 (2009): 1755–1758.

[Heiden 2010] Heiden, S. “The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme” In R. Otoguro, K. Ishikawa, H. Umemoto, K. Yoshimoto, Y. Harada (eds.), 24th Pacific Asia Conference on Language, Information and Computation, Institute for Digital Enhancement of Cognitive Development, Waseda University (2010).

[Hotho 2005] Hotho, A., Nürnberger, A. and Paaß, G. “A brief survey of text mining”, LDV Forum, 20.1 (2005): 19-62.

[Imesch 2016] Imesch, K., Schade, S., Sieber, S. (eds.) Constructions of cultural identities in newsreel cinema and television after 1945. MediaAnalysis, 17, transcript-Verlag, Bielefeld (2016).

[Lebart 1998] Lebart, L., Salem, A. and Berry, L. Exploring textual data. Text, speech, and language technology, 4, Kluwer Academic, Dordrecht, Boston (1998).

[Lindeperg 2000] Lindeperg, S. Clio de 5 à 7 : les actualités filmées à la Libération, archive du futur. CNRS, Paris (2000).

[Lindeperg 2008] Lindeperg, S. “Spectacles du pouvoir gaullien: le rendez-vous manqué des actualités filmées”. In J.-P. Bertin-Maghit (dir.), Une histoire mondiale des cinémas de propagande, Nouveau Monde Éditions, Paris (2008): 497-511.

[MacWhinney, 2000] McWhinney, B. The CHILDES Project: Tools for Analyzing Talk. L. Erlbaum Associates, Mahwah, N.J. (2000).

[Maitland 2015] Maitland, S. “Culture in translation: The case of British Pathé News” In Culture and news translation, Perspectives: Studies in Translation Theory and Practice, 23.4 (2015): 570-585.

[McEnery 2012] McEnery, T. and Hardie, A. Corpus linguistics: method, theory and practice. Cambridge University Press, Cambridge (2012).

[Motik 2012] Motik, B., Patel-Schneider, P. F., Parsia, B. “OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax (Second Edition)”. W3C Recommendation (2012).

[Pincemin 2020] Pincemin, B., Heiden, S. and Decorde, M. “Textometry on Audiovisual Corpora. Experiments with TXM software”, 15th International Conference on Statistical Analysis of Textual Data (JADT), Toulouse (2020).

[Povey 2011] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P.., Silovsky, J., Stemmer, G. and Vesely, K. “The kaldi speech recognition toolkit” In IEEE 2011 workshop on automatic speech recognition and understanding, IEEE Signal Processing Society, Hilton Waikoloa Village, Big Island, Hawaii, US (2011).

[Povey 2016] Povey, D., Peddinti, V., Galvez, D., Ghahremani, P., Manohar, V., Na, X., Wang, Y., and Khudanpur, S. “Purely sequence-trained neural networks for ASR based on lattice-free MMI” Interspeech, San Francisco (2016): 2751–2755.

[Pozner 2008] Pozner, V. “Les actualités soviétiques durant la Seconde Guerre mondiale : nouvelles sources, nouvelles approches” In J.-P. Bertin-Maghit (dir.), Une histoire mondiale des cinémas de propagande, Nouveau Monde Editions, Paris (2008): 421-444.

[R Core Team 2014] R Core Team., “R: A Language and Environment for Statistical Computing”, R Foundation for Statistical Computing, Vienna, Austria (2014).

[Salem 2004] Salem, A. “Introduction à la résonance textuelle” In G. Purnelle et al. (eds.), 7èmes Journées internationales d’Analyse statistique des Données Textuelles, Presses universitaires de Louvain, Louvain (2004): 986–992.

[Schmid 1994] Schmid, H. “Probabilistic Part-of-Speech Tagging Using Decision Trees” In Proceedings of International Conference on New Methods in Language Processing, Manchester, UK (1994).

[Schroff 2015] Schroff, F., Kalenichenko, D. and Philbin, J. “Facenet: A unified embedding for face recognition and clustering”. In Proceedings of the IEEE conference on computer vision and pattern recognition (2015): 815–823.

[Szegedy 2016] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. “Rethinking the Inception Architecture for Computer Vision” In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas (2016).

[Troncy 2012] Troncy, R., Mannens, E., Pfeiffer, S. and van Deursen, D. “Media Fragments URI 1.0 (basic)”. W3C Recommendation (2012).

[Veray 1995] Veray, L. Les Films d’actualités français de la Grande Guerre. SIRPA/AFRHC, Paris (1995).

[Viola 2004] Viola, P. and Jones, M. J. “Robust real-time face detection”. International Journal of Computer Vision, 57.2 (2004): 137–154.

[Weiss 2015] Weiss, S. M., Indurkhya, N., and Zhang, T. Fundamentals of Predictive Text Mining. Springer-Verlag, London (2015).

[Zhang 2016] Zhang, K., Zhang, Z., Li, Z. and Qiao, Y. “Joint face detection and alignment using multitask cascaded convolutional networks”. IEEE Signal Processing Letters, 23.10 (2016): 1499–1503.

[1] ANTRACT: ANalyse TRansdisciplinaire des ACTualités filmées (1945-1969).

[2] ESTER 1 & 2, EPAC, ETAPE, and REPERE corpus available in ELRA catalogues (http://www.elra.info/).

[3] ETAPE, and QUAERO corpus available in ELRA catalogues (http://www.elra.info/).

[4] https://github.com/XuezheMax/NeuroNLP2

[5] Challenge REPERE, test data.

[6] https://github.com/Linzaer/Face-Track-Detect-Extract

[7] Text Encoding Initiative, https://tei-c.org