



Analyse transdisciplinaire d'un corpus d'actualités filmées

Jean Carrive, Abdelkrim Beloued, Pascale Goetschel, Serge Heiden, Steffen Lalande, Pasquale Lisena, Franck Mazuet, Sylvain Meignier, Bénédicte Pincemin, Raphaël Troncy

► To cite this version:

Jean Carrive, Abdelkrim Beloued, Pascale Goetschel, Serge Heiden, Steffen Lalande, et al.. Analyse transdisciplinaire d'un corpus d'actualités filmées. Scopsi, Claire; Roullier, Clothilde; Sin Blima-Barru, Martine; Vasseur, Édouard. Les nouveaux paradigmes de l'archive, Publications des Archives nationales, pp.40-67, 2024, Actes, 978-2-86000-390-2. <10.4000/books.pan.7194>. <hal-04875186>

HAL Id: hal-04875186

<https://hal.science/hal-04875186v1>

Submitted on 8 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Analyse transdisciplinaire d'un corpus d'actualités filmées

L'environnement d'analyse numérique développé par le projet ANTRACT

Jean Carrive, Abdelkrim Beloued, Pascale Goetschel, Serge Heiden,
Steffen Lalande, Pasquale Lisena, Franck Mazuet, Sylvain Meignier,
Bénédicte Pincemin et Raphaël Troncy

Ce travail a été soutenu par l'Agence nationale de la recherche [ANR] dans le cadre du projet ANTRACT (ANR-17-CE38-0010) et par le programme de recherche et d'innovation Horizon 2020 de l'Union européenne dans le cadre du projet MeMAD (accord de subvention n° 780069).

Mise en œuvre d'un dispositif de recherche transdisciplinaire sur une collection d'archives cinématographiques : opportunités et défis

- 1 Le projet ANTRACT¹ réunit des laboratoires de recherche dans une double perspective historique et technologique, ce qui explique le caractère résolument transdisciplinaire du projet. Il s'intéresse à une collection de 1 262 films d'actualités (essentiellement des séquences en noir et blanc) diffusés dans les salles de cinéma françaises entre 1945 et 1969. Ces programmes ont été produits par la société *Les Actualités Françaises* durant une époque qui a pu être qualifiée de Trente Glorieuses. La recherche s'effectue non seulement sur les films proprement dits, mais aussi sur diverses ressources documentaires qui leur sont attachées : tapuscrits des commentaires en voix off contemporains des reportages, scannés et océrisés ; notices documentaires détaillées rédigées par les documentalistes de l'époque et reprises à divers moments par l'INA, avec titres, résumés, mots-clés, description et valeurs des plans, noms des participants, noms de lieux, etc.
- 2 Afin de mener cette recherche collective, des outils automatiques ont été développés et adaptés à l'analyse de cette collection conservée par l'Institut national de l'audiovisuel : reconnaissance automatique de la parole, classification d'images, reconnaissance faciale, traitement du langage naturel et fouille de textes. Ces logiciels sont utilisés

pour permettre à la communauté scientifique de travailler sur un corpus gérable et cohérent, disponible, à terme, à des fins de recherche.

- 3 En travaillant sur ces films d'actualités divisés en quelque 20 232 sujets, les historiens et les informaticiens d'ANTRACT collaborent, en définitive, pour optimiser la recherche sur les grands corpus audiovisuels en posant plusieurs questions clés :
 - Quelle approche technologique peut efficacement épauler l'étude systématique et exhaustive d'un fonds d'archives multimédias ?
 - Quels instruments peuvent compiler, analyser et recouper les données extraites de ces documents ?
 - Ces données extraites peuvent-elles être combinées et intégrées dans des interfaces utilisateurs polyvalentes ?
 - Avec ces traitements automatisés de sources multiformat, quelles nouvelles possibilités sont offertes aux projets de recherche en sciences humaines ?
- 4 Afin de mettre en œuvre une coopération solide entre les experts en informatique et les spécialistes de la discipline historique (Deegan et McCarty, 2012), l'objectif principal du projet est donc de fournir des pistes méthodologiques de recherche innovantes adaptées aux questions technologiques et historiques soulevées par ce corpus particulier.
- 5 **D'un point de vue technologique**, l'objectif est d'adapter les outils d'analyse automatique à la spécificité du corpus des *Actualités Françaises*, c'est-à-dire à son contexte historique, son vocabulaire et son type d'images. L'adaptation des modèles de langue utilisés par les outils de transcription automatique à l'aide de la version tapuscrite des voix off illustre cette orientation. En tant que collection de films comprenant des images, du son et du texte, produits il y a plus d'un demi-siècle, le corpus des *Actualités Françaises* représente un défi sans précédent pour les instruments spécialisés dans l'analyse et l'identification de contenus audiovisuels. Loin de considérer séparément une histoire sociale et culturelle du cinéma, d'une part, et l'utilisation d'outils d'analyse automatique, d'autre part, le projet vise à lier les deux. Ainsi, une bonne compréhension des conditions techniques d'enregistrement du son permet d'améliorer la reconnaissance audio. Tournées en noir et blanc avec un équipement limité et dans des conditions de tournage souvent difficiles, ces anciennes bandes d'actualités ne répondent pas aux normes de qualité fixées par les enregistrements vidéo et audio haute définition qui alimentent les algorithmes de reconnaissance d'images et de parole d'aujourd'hui. De plus, plusieurs bobines de films de la collection, numérisées sous des formats à haute compression, présentent des images pixellisées qui ne peuvent être traitées par les programmes d'analyse et certains des commentaires dactylographiés présentent des défauts d'impression causés par les machines à écrire utilisées pour leur production.
- 6 À ces obstacles matériels s'ajoute le problème posé par la transformation du contenu des films dans le temps. C'est le cas des personnalités régulièrement filmées par les cameramen de la société tout au long de ses 24 années d'activité. C'est aussi le cas des données topographiques récurrentes prises sur leur pellicule. L'identification automatique de ces éléments en constante évolution enregistrés sur des séquences monochromatiques nécessite des ressources considérables. Dans le cadre de ce processus, les historiens d'ANTRACT ont proposé une liste de 121 personnalités présentes dans les *Actualités Françaises* afin de construire une série de modèles d'extraction.

- 7 **D'un point de vue historique**, il convient de noter que le corpus des *Actualités Françaises* n'avait pas fait, jusqu'à présent, l'objet d'une analyse historique systématique de son contenu et de ses modalités de production. Aussi le projet ANTRACT vise-t-il à remédier à cette situation tout en cherchant à proposer des analyses historiques de différentes natures. Elles peuvent, dans le sillage des études déjà existantes sur l'histoire des actualités filmées, renvoyer à des considérations politiques et être, en particulier, centrées sur des études de censures (Atkinson, 2011 ; Bartels, 2004 ; Pozner, 2008 ; Veray, 1995). Elles peuvent aussi questionner le rôle de la presse cinématographique comme vecteur de l'histoire sociale, politique et culturelle façonnant l'opinion publique durant la seconde moitié du xx^e siècle (Fein, 2004, 2008 ; Althaus, 2018 ; Chambers, 2018 ; Imesch, 2016 ; Lindeperg 2000, 2008). De manière plus innovante, les études historiques effectuées dans le cadre du projet proposent d'autres pistes. En effet, au-delà des films eux-mêmes, d'autres sources liées au corpus filmique sont intéressantes. Les fiches d'observation remplies par les cameramen, les commentaires écrits des journalistes et les archives laissées par la direction donnent un aperçu inédit du contenu d'une collection de films ainsi que de son processus de production. Aussi, les études historiques développées dans le cadre d'ANTRACT conduisent à prendre en compte les conditions de production des informations diffusées.
- 8 Au-delà, la mise à disposition des résultats des outils de reconnaissance automatique permet de répondre à une question récurrente pour les historiens lorsqu'il s'agit de travailler sur des collections de grande ampleur : comment, parmi les milliers d'heures d'archives filmiques associées à des centaines de fichiers textes produits sur une longue période, travailler sur des objets d'étude particuliers ? Les outils développés par les partenaires du consortium, par le traitement combiné de données extraites, rendent possible une identification de thématiques de recherche historique plus aisée. Ces outils leur permettent donc de fabriquer des « sous-corpus » liés aux objets d'analyse en dépassant une recherche effectuée sur seulement quelques fragments, par exemple : les foules, les fêtes, les personnages influents, les mutations sociales, les prisonniers de guerre ou les mutilés à l'écran... Très concrètement, au-delà du repérage de « sous-corpus » liés à leurs thématiques, les historiens et historiennes peuvent travailler sur la fréquence, la durée, les moments d'apparition des sujets ou leur traitement médiatique.
- 9 Ainsi, centré sur le contenu des films, le projet s'attache, d'une part, à scruter le processus de production et les différents métiers impliqués dans la réalisation des *Actualités Françaises* en mettant en évidence les ressorts d'une entreprise contrôlée par un État démocratique. D'autre part, ANTRACT nourrit des analyses relatives à l'histoire politique, sociale et culturelle de la France des Trente Glorieuses, prise dans un environnement impérial, mais aussi européen et international. Il alimente les études sur le rôle des médias dans la fabrique des événements (Goetschel et Granger, 2011). Enfin, le projet invite à réfléchir sur le type d'images et de sons offerts aux spectateurs des salles de cinéma : goût pour les nouvelles sensationnelles, extraordinaires et exotiques, mais aussi ordinaires et banales, vie quotidienne de personnalités célèbres (Maitland, 2015), exploits et innovations scientifiques et techniques, mais aussi images évoquant différentes traditions locales, régionales, nationales...
- 10 À travers les outils d'analyse audio et vidéo (et les plateformes d'analyse historique interactive, cet article présente les résultats du projet, en mettant l'accent sur l'aspect

technologique de la recherche et, plus précisément, sur les appareils et les outils de traitement des données, en lien avec plusieurs exemples d'enquêtes historiques.

Analyse audio automatique

- 11 Le travail sur la partie audio consiste à détecter les locuteurs, à transcrire la parole en mots (ASR, pour *Automatic Speech Recognition*) et à détecter les entités nommées (EN, *Named Entities*) en utilisant les systèmes que nous avons développés pour les actualités contemporaines de radio et de télévision.
- 12 L'analyse audio d'un ensemble de données anciennes constitue un défi intéressant pour les systèmes d'analyse automatique. Les appareils d'enregistrement utilisés entre 1945 et 1969 sont très différents des appareils analogiques ou numériques d'aujourd'hui. Les films 35 mm, qui contiennent à la fois le son et l'image, se sont détériorés avant d'être numérisés dans les années 2000. De plus, les modèles acoustiques et linguistiques utilisés par les outils de reconnaissance automatique de la parole sont généralement entraînés sur des données produites entre 1998 et 2012. Ce décalage de 50 ans a des conséquences sur les performances du système.
- 13 Techniquement, pour ANTRACT, les modèles acoustiques pour l'ASR ont été entraînés sur environ 300 heures tirées de plusieurs sources d'actualités télévisées et radiophoniques françaises² associées à des transcriptions manuelles. Les modèles de langage (probabilités de suites de mots) ont été entraînés sur ces transcriptions manuelles, des journaux français, des sites d'information, Google News et le corpus français GigaWord, pour un total de 1,6 milliard de mots. Le vocabulaire du modèle de langage contient les 160 000 mots les plus fréquents. Les modèles EN ont été entraînés uniquement sur un sous-ensemble de transcriptions manuelles³.
- 14 En amont du processus de transcription, le signal est découpé en segments de paroles homogènes et groupés par locuteur. Nous appelons ce processus la tâche de Segmentation et regroupement en locuteur [SRL]. Cette tâche est d'abord appliquée au niveau de l'édition (c'est-à-dire d'un journal entier), où chaque enregistrement vidéo est traité séparément. Ensuite, le processus est appliqué au niveau de la collection, sur l'ensemble des 1 262 éditions, afin de relier les locuteurs récurrents qui sont principalement les voix off. Le système développé par le Laboratoire d'informatique de l'université du Mans [LIUM] (Broux, 2018) est destiné à fournir des segments de parole homogènes contenant la parole d'un seul locuteur et des limites de segment précises marquant un changement de locuteur. Cette étape est essentielle au bon fonctionnement de l'ASR pour éviter les erreurs de transcription des début et fin de phrases. Le regroupement des segments par locuteur au niveau de l'édition ou de la collection n'a pas d'influence sur la qualité de la transcription, mais un regroupement précis des tours de parole d'un locuteur facilite la navigation dans la collection et la compréhension. L'élément clé du système de SRL est la caractérisation du signal au moyen d'un réseau de neurones profond qui extrait toutes les 10 ms un vecteur caractéristique du locuteur. Les algorithmes de segmentation et de regroupement reposent sur les méthodes classiques couramment utilisées en traitement acoustique ou d'image.
- 15 Le système ASR est développé principalement à l'aide de l'outil *open source* Kaldi (Povey, 2011). Il fait intervenir un modèle acoustique pour représenter les unités sonores

élémentaires (les phonèmes), une liste finie de mots potentiels et un modèle de langage qui détermine la probabilité d'une suite de mots. Les modèles acoustiques sont formés à l'aide d'un réseau neuronal profond qui peut traiter efficacement des contextes temporels longs (Povey, 2016). Des modèles de langage calculés avec des contextes de 2 ou 3 mots ont été entraînés sur de vastes corpus audio et textuels. Pour faciliter la lecture, deux systèmes d'étiquetage de séquences ont été entraînés sur des transcriptions manuelles pour ajouter respectivement la ponctuation et les majuscules à la suite de mots générés par l'ASR.

- 16 Le système neuronal d'extraction d'entités nommées complète l'analyse du texte. Le système, entraîné sur des transcriptions manuelles, détecte huit types d'entités principales : les lieux (la ville de Lyon), les organisations (l'ONU), les fonctions (Général), les personnes (Charles de Gaulle), les produits (un avion Caravelle), les événements (la Fête nationale du 14 juillet), les expressions numériques (1 000 francs) et les expressions temporelles (demain). L'annotation des entités nommées a pour but de mettre en avant les éléments de la transcription qui permet de répondre à des questions factuelles : qui, quoi, quand, où et comment.
- 17 L'ASR a été réalisée sur la collection complète des 1 262 éditions nationales afin d'alimenter les plateformes Okapi et TXM pour les analyses des historiens (voir les sections « Analyse textométrique interactive » et « Analyse sémantique interactive ») : environ 300 heures de vidéo, résultant en plus de 1,5 million de mots. Un sous-ensemble de 12 éditions de 1945 à 1969 a été transcrit manuellement pour évaluer les systèmes d'analyse audio. En raison de l'écart de 50 ans, les annotateurs humains ont eu quelques difficultés avec l'orthographe des entités nommées [EN], notamment en ce qui concerne les personnes et les EN étrangères. Grâce à Wikipédia et au thésaurus de l'Ina, la plupart des EN ont été vérifiées. En revanche, les locuteurs sont très difficiles à identifier. La plupart d'entre eux sont des voix off masculines. On ne voit jamais leur visage, leur nom est rarement prononcé et n'est pas affiché sur les images. Seuls les journalistes réalisant des interviews et les personnes connues, comme les politiciens, les athlètes et les célébrités, peuvent être identifiés et nommés avec précision.
- 18 La qualité d'un système ASR est évaluée à l'aide du taux de mots erronés (WER, pour *Word Error Rate*). Cette métrique consiste à compter le nombre d'insertions, de suppressions et de substitutions de mots entre les transcriptions générées automatiquement par le système ASR et les transcriptions humaines prises comme référence. Le WER est d'environ 24 % sur les données ANTRACT en utilisant le système générique ASR entraîné sur des données journalistiques modernes. Le même système évalué sur des données datant de 2010⁴ atteint environ 13 %. Il est connu que les systèmes ASR sont sensibles aux variations acoustiques et linguistiques entre le corpus d'entraînement et le corpus de test. Ici, le WER est presque le double. Il est généralement difficile d'exploiter les transcriptions de manière robuste lorsque le WER est supérieur à 30 %. La plupart des erreurs proviennent de mots inconnus (qui ne sont pas répertoriés dans le vocabulaire de 160 000 mots). Ces mots hors vocabulaire sont confondus avec des mots acoustiquement proches, ce qui a un impact négatif sur les mots voisins. En effet, le système sélectionne toujours la séquence de mots la plus probable contenant le mot qui remplace le mot inconnu du système.
- 19 Des données contemporaines supplémentaires, telles que les notices documentaires et les tapuscrits, s'avèrent utiles pour adapter le modèle linguistique. Par conséquent, les résumés, les titres et les descriptions ont été extraits des notices documentaires. Les

phrases issues des tapuscrits ont été conservées lorsqu'au moins 95 % des mots appartiennent au vocabulaire ASR. Cela a permis de construire un corpus d'entraînement spécifique au domaine, composé de 1,3 million de mots provenant des notices documentaires et de 4,7 millions de mots provenant des tapuscrits. Les 4 000 mots les plus fréquents ont été sélectionnés pour entraîner le nouveau modèle de langage ANTRACT, ce qui a réduit le taux d'erreur de moitié : de 24 % à environ 12 % de WER. La figure 1 montre un exemple de transcription automatique de l'édition du 14 juillet 1955. Le gain est significatif grâce aux transcriptions consignées dans les tapuscrits, qui sont très similaires aux transcriptions manuelles. Ce corpus d'entraînement spécifique va à l'encontre des règles strictes établies pour l'évaluation de systèmes de reconnaissance automatique : les données de test ne doivent jamais être utilisées pour construire un corpus d'entraînement. Cependant, dans notre cas, l'objectif principal est de fournir les meilleures transcriptions possibles aux historiens.

- 20 Cependant, toute erreur a un impact sur les recherches d'information pour les historiens et constitue une limite à nos travaux. Par exemple, transcrire le mot « poule » lorsque le locuteur a prononcé le mot « foule » ne permet pas de visionner l'ensemble des séquences parlant des foules. En fin de projet, une correction manuelle des transcriptions a été réalisée. La correction est effectuée en deux passes. Les corrections orthographiques et grammaticales des mots communs et de la segmentation (essentiellement les débuts et les fins de segments) sont réalisées dans une première phase. Une seconde passe, qui demande plus de temps que la première, se focalise sur la correction des noms propres. Cette dernière nécessite d'interroger des ressources externes aux projets (Wikipédia, dictionnaire, etc.) ou internes (notice, tapuscrit) pour corriger les noms propres souvent inconnus du vocabulaire du système.
- 21 En définitive, au fil du projet, les travaux se sont concentrés sur l'amélioration des modèles acoustiques ASR, grâce, notamment, aux données textuelles issues des tapuscrits des commentaires en voix off et au retour d'information des historiennes et historiens qui ont corrigé manuellement des parties de transcription. L'évaluation des entités nommées n'a pas été réalisée sur les données du projet, faute d'annotations. De même, l'évaluation des locuteurs a été difficile en raison de l'absence d'information sur leur identité. Celle-ci s'est donc concentrée sur la détection des commentateurs (voix off) et des intervieweurs. En outre, certaines personnes célèbres, sélectionnées en collaboration avec des historiens pour leur pertinence dans les analyses historiques, sont identifiées, avec l'aide éventuelle du croisement des résultats avec l'analyse d'image, comme décrit dans la section suivante.

Figure 1. Exemple de l'Édition du 14 juillet 1955 des *Actualités Françaises* (de 6:06 à 6:49).



Le sous-titre est généré automatiquement par l'ASR avec un modèle de langue du domaine, une ponctuation automatique et des majuscules.

© INA

Analyse visuelle automatique

- 22 L'identification des personnes apparaissant dans une vidéo est indéniablement un élément clé pour sa compréhension. Savoir qui figure dans une vidéo, quand et où, peut également permettre de découvrir des modèles intéressants de relations entre les personnes pour la recherche historique. De telles annotations liées aux personnes pourraient permettre de créer un contenu à valeur ajoutée. Une archive historique telle que le corpus des *Actualités Françaises* contient de nombreux exemples de célébrités figurant dans le même segment d'actualités, par exemple Charles de Gaulle et Konrad Adenauer (voir Figure 2). Cependant, les annotations produites manuellement par les documentalistes ne permettent pas toujours d'identifier avec précision les individus présents dans les vidéos. D'autre part, le Web offre une quantité importante de photographies de ces personnes, facilement accessibles par les moteurs de recherche en utilisant leur nom complet comme terme de recherche. Dans ANTRACT, l'idée a été d'exploiter ces images pour identifier les visages des célébrités dans les archives vidéo.

Figure 2. De Gaulle et Adenauer ensemble dans une vidéo de 1959.



© INA

- 23 De nombreux progrès ont été réalisés au cours de la dernière décennie concernant la reconnaissance automatique des personnes. Elle comprend généralement deux étapes : il faut d'abord détecter les visages (c'est-à-dire savoir quelle région de l'image est susceptible de contenir un visage de personne), puis les reconnaître (c'est-à-dire savoir à quelle personne chaque visage appartient).
- 24 L'algorithme de Viola-Jones (Viola, 2004) pour la détection des visages et les caractéristiques des motifs binaires locaux [LBP] (Ahonen, 2006) pour le regroupement et la reconnaissance des visages étaient les techniques les plus utilisées jusqu'à l'avènement de l'apprentissage profond et des réseaux de neurones dits convolutionnels [CNN]. Aujourd'hui, deux approches principales sont utilisées pour détecter les visages dans les vidéos et toutes deux utilisent des CNN. La bibliothèque Dlib (King, 2009) offre de bonnes performances pour les vues de face, mais elle nécessite une étape supplémentaire d'alignement (qui peut également être effectuée à l'aide de la bibliothèque Dlib) avant de pouvoir procéder à la reconnaissance des visages. L'approche plus récente *Multi-task Cascaded Convolutional Networks* [MTCNN] fournit des performances encore meilleures en utilisant une approche image-pyramide et intègre la détection des points de repère des visages afin de réaligner les visages détectés sur la vue de face (Zhang, 2016).
- 25 Après avoir repéré la position et l'orientation des visages dans les images vidéo, le processus de reconnaissance peut être effectué dans de bonnes conditions. Plusieurs stratégies ont été détaillées dans la littérature pour réaliser la reconnaissance. Actuellement, l'approche la plus pratique consiste à effectuer une comparaison de visages à l'aide d'un espace de transformation dans lequel les visages similaires sont proches les uns des autres, et à utiliser cette représentation pour identifier la bonne personne. De tels espaces de « plongement » (*embeddings*), calculés sur de grandes collections de visages, sont disponibles pour la communauté de recherche (Schroff, 2015).

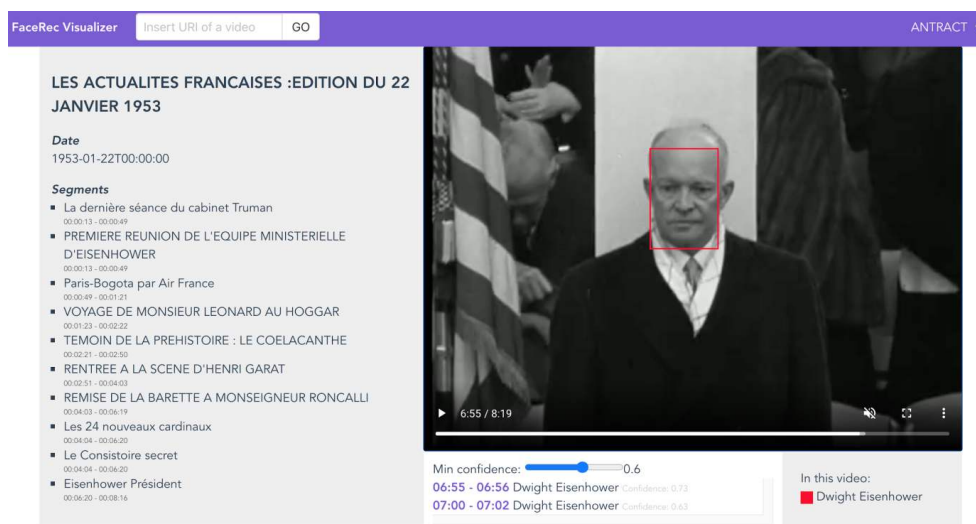
26 Au sein d'ANTRACT, nous avons développé un système *open source* de reconnaissance faciale des célébrités. Cette application est composée des modules suivants :

- un robot d'exploration du Web qui, étant donné le nom d'une personne, télécharge automatiquement *via* Google un ensemble de k photos qui seront utilisées pour l'entraînement d'un modèle de visage particulier. Dans nos expériences, nous utilisons généralement $k = 50$. Parmi les résultats, les images ne contenant aucun visage ou contenant plus d'un visage sont écartées. En outre, les utilisateurs finaux (par exemple, les experts du domaine) peuvent exclure manuellement les résultats non pertinents, qui, par exemple, ne correspondent pas à la personne recherchée ;
- un module d'entraînement où les photographies récupérées peuvent être converties en noir et blanc, recadrées et redimensionnées afin d'obtenir des images contenant uniquement un visage, en utilisant l'algorithme MTCNN (Zhang, 2016). Un modèle Facenet (Schroff, 2015) préentraîné avec une architecture Inception ResNet v1 entraînée sur VGGFace2dataset (Cao, 2018), est appliqué afin d'extraire les caractéristiques visuelles des visages. Les plongements obtenus sont utilisés pour entraîner un classifieur SVM ;
- un module de reconnaissance qui prend en entrée une vidéo d'actualité et en extrait une image toutes les d images (dans nos expériences, nous avons généralement fixé $d = 25$, soit une image par seconde). Pour chaque image, les visages sont détectés (en utilisant l'algorithme MTCNN) et les *embeddings* sont calculés (Facenet). Le classifieur SVM décide si le visage correspond à l'un des visages des images d'entraînement. *Simple Online and Realtime Tracking* [SORT] est un algorithme de suivi d'objets, qui peut suivre plusieurs objets en temps réel (Bewley, 2016). Son implémentation est inspirée du code de suggestion de Linzaer⁵. L'algorithme utilise la détection de la boîte de délimitation MTCNN et la suit à travers les images. Nous avons introduit ce module pour augmenter la robustesse du traitement. En utilisant ce module, tout en faisant l'hypothèse que les visages ne changent quasiment pas de coordonnées entre deux images consécutives, nous visons à obtenir une prédiction plus cohérente ;
- enfin, le dernier module regroupe les résultats provenant du classifieur et des modules de suivi. Nous observons que même si le visage à reconnaître reste le même sur plusieurs images consécutives, la prédiction du visage change parfois. Pour cette raison, nous sélectionnons pour chaque suivi la prédiction la plus fréquente, en prenant également en compte le score de confiance donné par le classifieur. De cette façon, le système fournit une prédiction commune pour toutes les images impliquées dans un suivi, ainsi qu'un score de confiance agrégé. Un seuil t peut être appliqué à ce score afin d'écarter les prédictions peu fiables. D'après nos expériences, $t = 0,6$ donne un bon compromis entre la précision et le rappel.

27 Afin de rendre le logiciel disponible en tant que service, nous l'avons intégré dans une API web RESTful⁶. Le service prend en entrée l'URI d'une ressource vidéo, telle qu'elle apparaît dans Okapi, à partir duquel il récupère l'objet média encodé en MPEG-4. Deux formats de sortie sont disponibles : un format JSON personnalisé et un format de sérialisation en RDF utilisant la syntaxe Turtle et la syntaxe Media Fragment URI (Troncy, 2012), avec la durée de lecture normale exprimée en secondes pour situer les fragments temporels et les coordonnées *xywh* pour définir le rectangle encadrant le visage dans l'image. Un troisième format, toujours selon la syntaxe Turtle, sera bientôt implémenté afin que les résultats puissent être directement intégrés dans le graphe de connaissances Okapi. Un système de cache léger est également mis en place afin de pouvoir fournir des résultats précalculés, sauf si le paramètre *no cache* est activé.

- 28 Nous avons mené des expériences en utilisant le modèle de visage de Dwight D. Eisenhower sur une sélection de segments vidéo extraits d'Okapi, parmi ceux qui ont été annotés avec la présence du président américain selon les propriétés *ina:imageContient* et *ina:aPourParticipant* dans le graphe de connaissances. En l'absence d'une vérité terrain, nous avons effectué une analyse qualitative de notre système sur trois vidéos. Pour chaque personne détectée, nous avons évalué manuellement si la bonne personne était trouvée ou non. Sur les 90 segments sélectionnés, le système a correctement identifié Eisenhower dans 33 d'entre eux. Cependant, nous ne sommes pas sûrs que Eisenhower soit effectivement présent visuellement dans les 90 segments (il peut avoir été indexé pour une apparition plus tôt ou plus tard dans le même sujet par exemple). Nous avons ensuite produit une vérité terrain qui nous a permis d'évaluer la précision et le rappel du système (Lisena, 2022).
- 29 En outre, nous avons fait les observations suivantes :
- notre logiciel ne parvient généralement pas à détecter les personnes lorsqu'elles sont en arrière-plan ou lorsque le visage est masqué ;
 - lorsque les visages sont parfaitement de face, ils sont plus faciles à détecter. Des améliorations de l'algorithme d'alignement sont prévues dans les travaux futurs.
- 30 En fixant un seuil de confiance élevé, nous ne rencontrons pas de cas de confusion entre deux célébrités. La plupart des erreurs consistent plutôt à confondre un visage inconnu avec celui d'une célébrité.
- 31 Afin de visualiser facilement les résultats et de faciliter le retour d'information des historiens, nous avons développé une application web qui affiche les résultats directement sur la vidéo, en tirant parti des fonctionnalités HTML5⁷. L'application fournit également un résumé des différentes prédictions, permettant à l'utilisateur de passer directement à la partie relative de la vidéo où la célébrité apparaît. Un curseur permet de modifier la valeur du seuil de confiance, afin de mieux étudier les résultats jugés peu fiables.

Figure 3. Le visualiseur du système de reconnaissance des visages de célébrités.



Analyse textométrique interactive

- 32 Dans ANTRACT, l'exploration et l'analyse des données textuelles sont proposées aux historiens selon une approche textométrique (Lebart, 1998). La textométrie combine à la fois des outils quantitatifs – statistiques –, et des outils qualitatifs – de recherche, de lecture et d'annotation de textes. Les fonctionnalités statistiques comprennent des calculs de spécificités lexicales, de cooccurrences (mots associés), de classification hiérarchique et d'analyse des correspondances. Cela représente un gain significatif en matière de possibilités d'analyse par rapport aux fonctions habituelles d'annotation, de recherche et de décompte des logiciels de transcription audiovisuelle tels que CLAN (MacWhinney, 2000) ou ELAN (ELAN, 2018). Quant à l'analyse qualitative, elle est réalisée par des concordances avancées, avec un accès hypertexte aux documents sources, et avec des possibilités d'annotation dynamique du corpus en cours d'analyse. Un tel aspect qualitatif est marginal, voire absent, dans les applications classiques de fouille de textes (Hotho, 2005 ; Feinerer, 2008 ; Weiss, 2015) : la plupart d'entre elles traitent du texte brut, en commençant si besoin par éliminer les marques de structuration du texte, et elles cherchent à produire une visualisation synthétique qui remplace la lecture attentive du texte (au lieu de garder une relation constante au texte original étudié).
- 33 La textométrie est ici mise en œuvre avec la plateforme logicielle TXM (Heiden, 2010). TXM est développé de façon *open-source* et intègre plusieurs composants spécialisés : R (R Core Team, 2014) pour les calculs statistiques, CQP comme moteur de recherche en texte intégral (Christ, 1994), TreeTagger (Schmid, 1994) pour le traitement du langage naturel (étiquetage morphosyntaxique et lemmatisation). TXM s'inscrit dans les pratiques de la science ouverte au niveau de la standardisation et du partage des données et du code informatique, et a notamment été conçu pour gérer des corpus richement structurés et annotés, tels que des données XML et des textes encodés suivant les recommandations de la TEI⁸. Pour les données textuelles ANTRACT, TXM importe des données tabulées (export Excel de tableaux depuis les bases documentaires de l'INA) et des fichiers au format XML Transcriber fournis par un logiciel de transcription de la parole (voir Section « Analyse audio automatique »). TXM est un outil pour l'analyse textuelle, mais il permet également de gérer les représentations multimédias associées aux textes, qu'il s'agisse d'images scannées des documents sources, d'enregistrements audio ou vidéo : en effet, ces représentations participent à l'interprétation des résultats des traitements en les remplaçant dans leur contexte sémiotique complet.
- 34 En 2018, nous avons commencé par la construction du corpus TXM AF-NOTICES en important les notices documentaires de l'Ina : chaque sujet est représenté par plusieurs champs textuels (titre, résumé, séquences plan par plan) et plusieurs champs lexicaux (listes de descripteurs documentaires de différents types tels que sujets, personnes ou lieux, et générique des noms des personnes montrées ou des cameramen). Chaque sujet est également caractérisé par une dizaine de métadonnées (identifiant Ina, date de diffusion, producteur du film, genre du film, etc.) utiles pour le contextualiser ou le catégoriser.
- 35 À partir de 2019, nous avons aussi réalisé le corpus TXM AF-VOIX-OFF fondé sur les transcriptions du commentaire audio (voir Section « Analyse audio automatique ») et

synchronisé au mot près aux vidéos. Les champs documentaires des notices Ina sont intégrés au corpus en tant que métadonnées décrivant les transcriptions.

- 36 Ces corpus pourraient encore être étendus par l'ajout de nouvelles données textuelles : les textes issus des tapuscrits des commentaires qui ont été scannés et océrisés (reconnaissance optique des caractères), les annotations sur les vidéos (annotations manuelles ajoutées par les historiens via la plateforme Okapi) [voir Section « Analyse sémantique interactive »], ainsi que les annotations automatiques générées par les logiciels de reconnaissance d'images (voir Section « Analyse visuelle automatique »), etc.
- 37 L'une des innovations techniques réalisées dans le cadre du projet a été la consolidation du module de retour à la vidéo de TXM (Pincemin, 2020), de sorte que tout mot ou passage de texte trouvé dans le résultat d'un calcul textométrique puisse être consulté dans la vidéo originale ; nous avons également mis en œuvre un accès contrôlé par mot de passe aux vidéos en ligne sur le serveur média d'Okapi, ce qui s'est avéré être un développement clé pour la mise à disposition des vidéos, étant donné la taille de stockage importante requise pour ces enregistrements et les contraintes de sécurité sur ces archives cinématographiques.
- 38 Les captures d'écran qui suivent illustrent des étapes types d'une analyse textométrique telle que menée dans le cadre du projet ANTRACT.
- 39 Dans les figures 4 et 5, nous étudions le contexte d'utilisation du mot « foule » à l'aide d'une concordance. Un double-clic sur une ligne de concordance ouvre une nouvelle fenêtre (à droite) qui affiche la transcription complète dans laquelle apparaît le mot. Ensuite, un clic sur le symbole des notes de musique au début du paragraphe permet de lire la vidéo correspondante. Une boîte de dialogue demande des identifiants pour accéder à la vidéo sur le serveur en ligne d'Okapi. Cette possibilité de confronter l'analyse textuelle à la source audiovisuelle est d'autant plus importante ici que les données textuelles ont été générées par un outil automatique de reconnaissance de la parole, dont la sortie n'a pas pu être entièrement vérifiée. De plus, la vidéo peut apporter des éléments de contexte significatifs qui complètent le contenu textuel.

Figure 4. Confrontation de l'analyse textuelle à la source audiovisuelle, étape 1.



CONCORDANCE du mot « foule » dans le corpus de voix off (fenêtre de gauche), ÉDITION de la page de transcription correspondant à la ligne de concordance sélectionnée (fenêtre de droite), et boîte de dialogue d'authentification permettant d'accéder au serveur vidéo Okapi pour lire la vidéo au temps 0:00:06 (fenêtre supérieure gauche).

Figure 5. Confrontation de l'analyse textuelle à la source audiovisuelle, étape 2.

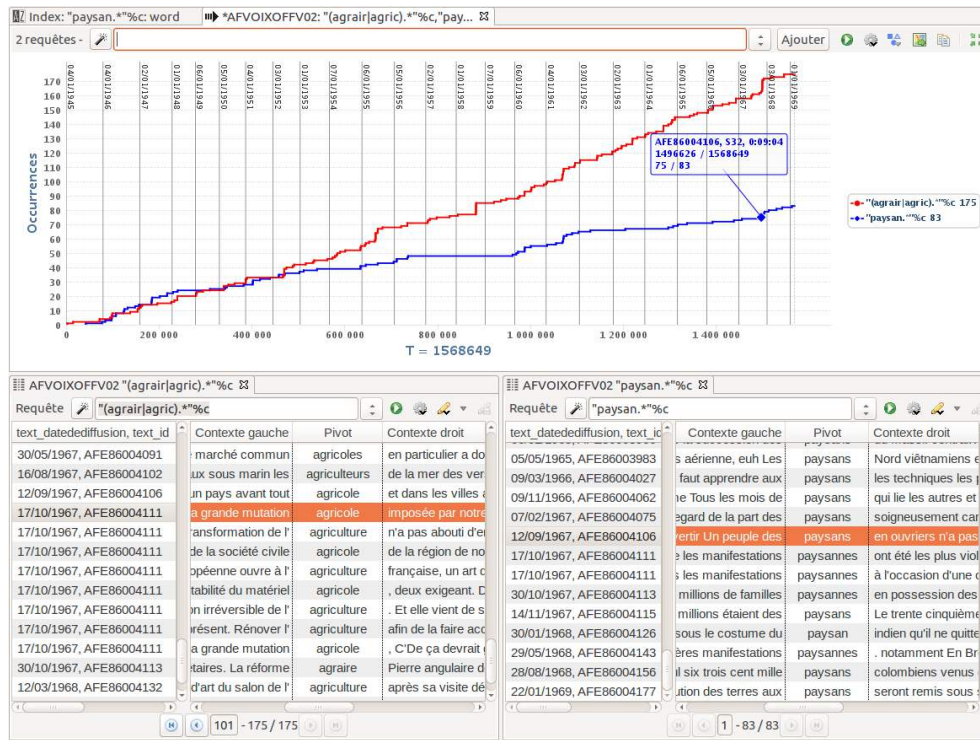
The screenshot displays a software interface for text analysis and video synchronization. It consists of three main panels:

- Left Panel (Concordance Table):** A table showing the occurrences of the word 'foule' in a text corpus. The columns are 'datedediffusion', 'Left context', 'Pivot', and 'Right context'. The table lists various dates and corresponding text snippets, with 'foule' highlighted in red in the pivot column.
- Middle Panel (Transcription):** A window titled 'AFE86003459 - 2' showing the transcription of the video. It includes a search bar with the query 'frlemma = "foule"', a list of results, and a section for 'generique_aff_lig' with a search bar and a list of results. The word 'foule' is highlighted in red in the transcription.
- Right Panel (Video Player):** A window titled 'AFE86003459' showing a video player. The video shows a large crowd of people, likely a religious event. The player includes a search bar, a list of results, and a section for 'RESUM' with a search bar and a list of results.

Fenêtres liées entre elles et présentant des résultats pour le mot « foule » : CONCORDANCE (fenêtre de gauche), ÉDITION de la transcription (fenêtre du milieu) et lecture vidéo synchronisée (fenêtre de droite).

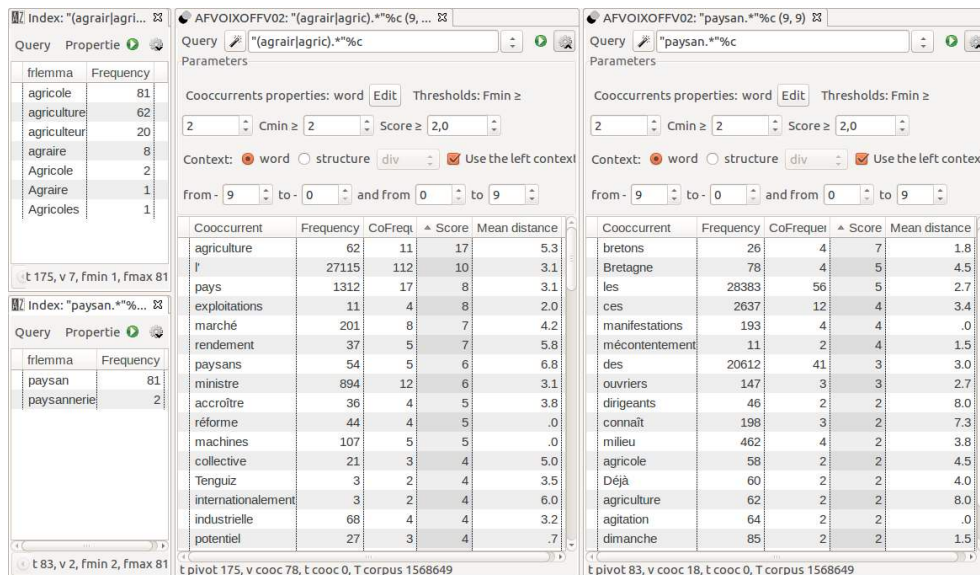
- 40 Notre deuxième exemple concerne la place de l'agriculture et des agriculteurs dans les *Actualités Françaises* et la manière dont ce sujet est abordé. Ce cas illustre comment on peut observer si un mot donné a le même sens dans les notices documentaires et dans les commentaires audio, ou si des mots différents sont utilisés pour traiter du même sujet. Nous obtenons d'abord (Figure 6) un aperçu comparatif de l'évolution quantitative des occurrences de deux familles de mots, les dérivés des radicaux de « paysan » et « agricole »/« agriculture » (voir la liste détaillée des mots dans la Figure 7, fenêtre de gauche). Nous complétons l'analyse par un examen des contextes d'emploi à travers une vue en concordance (voir Figure 6, fenêtre inférieure) et un calcul de cooccurrence (voir Figure 7). Nous remarquons que « paysan » devient moins utilisé à partir de 1952 et qu'il est préféré à « agriculteur » pour parler des individus présents dans les extraits d'actualités ; inversement, « agricole »/« agriculture » sont utilisés de manière plus abstraite, pour traiter des nouveaux équipements agricoles et de la transformation socio-économique de ce secteur d'activité.

Figure 6. Examen des contextes d'emploi de deux familles de mots.



Graphique de PROGRESSION (fenêtre supérieure) et CONCORDANCES associées (fenêtre inférieure), pour comparer deux familles de mots liées à l'agriculture.

Figure 7. Calcul de cooccurrence de deux familles de mots.

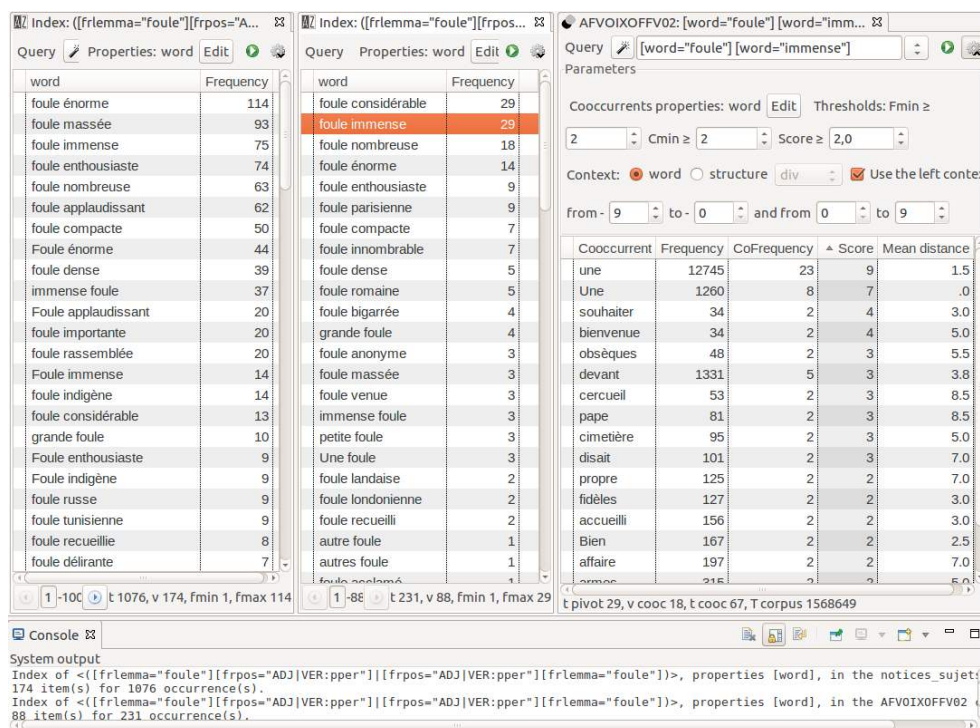


Résultats d'INDEX détaillant le contenu de deux familles de mots (marge gauche) et analyse statistique de COOCCURRENCES pour caractériser leurs contextes.

- 41 La combinaison des listes de mots (INDEX) et des informations morphosyntaxiques est très efficace pour résumer les formulations dans lesquelles les mots sont employés. Par exemple, dans la figure 8, nous pouvons comparer quels adjectifs qualifient « foule » dans les notices documentaires et quels adjectifs qualifient « foule » dans les discours

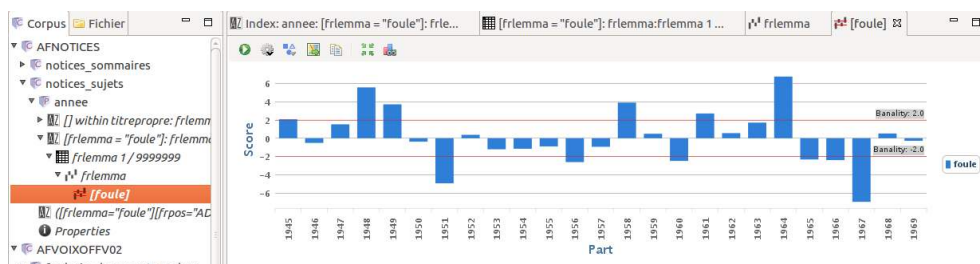
en voix off. Pour une expression donnée (« foule immense ») prononcée dans le commentaire audio, nous calculons ses cooccurrences afin d'identifier dans quels types de circonstances cette expression est généralement utilisée (ici des funérailles et des rassemblements religieux). Dans TXM, la recherche en texte intégral bénéficie du moteur de recherche CQP (Christ, 1994), qui permet des requêtes très précises, y compris avec des conditions de contexte.

Figure 8. Examen de l'emploi d'un mot précédé ou suivi d'un adjectif.



INDEX de « foule » précédé ou suivi d'un adjectif, dans les notices documentaires (fenêtre de gauche) ou dans les transcriptions de la voix off (fenêtre du milieu). COOCCURRENCES de « foule immense » dans les transcriptions de la voix off (fenêtre de droite).

Figure 9. Graphique de SPECIFICITE pour « foule » au fil des années.



- 42 Pour les recherches chronologiques, le logiciel permet de diviser le corpus en périodes de manière très flexible, par exemple année par année ou en définissant des groupes d'années. Toute information disponible codée dans le corpus peut être utilisée pour définir les subdivisions du corpus. Ensuite, la commande SPÉCIFICITÉ mesure statistiquement pour chaque mot l'équilibre de sa répartition à travers les parties et met en évidence ses éventuels sur- ou sous-emplois dans certaines parties. La fonction peut également être utilisée pour lister l'ensemble des termes spécifiques à une période

donnée (ou à une partie quelconque qu'on définit dans le corpus). Par exemple, la figure 9 s'intéresse au mot « foule » au fil des années. Les années aux scores les plus importants révèlent des événements politiques décisifs (par exemple, la Libération de la France après la Seconde Guerre mondiale, l'avènement de la Cinquième République), qui correspondent à la forte exposition du général de Gaulle. Cependant, on note aussi que les moments de plus forte présence du mot ne correspondent pas nécessairement à des bouleversements politiques.

Figure 10. Exemple d'analyse de résonance (Salem, 2004).

The figure consists of two screenshots of a software interface for word analysis. Both windows have a title bar 'word' and a 'Property' dropdown set to 'word'. The top window displays results for the query 'foule_in_documentary_desc t=396023'. It shows a table with columns: Units, Frequency T 1568649, foule_in_documentary_desc t=396023, and index. The bottom window displays results for the query 'foule_in_documentary_desc n voice_without_gaulle_president t=264308'. It shows a table with columns: Units, Frequency T 1568649, foule_in_documentary_desc n voice_without_gaulle_president t=264308, and index.

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

Units	Frequency T 1568649	foule_in_documentary_desc n 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

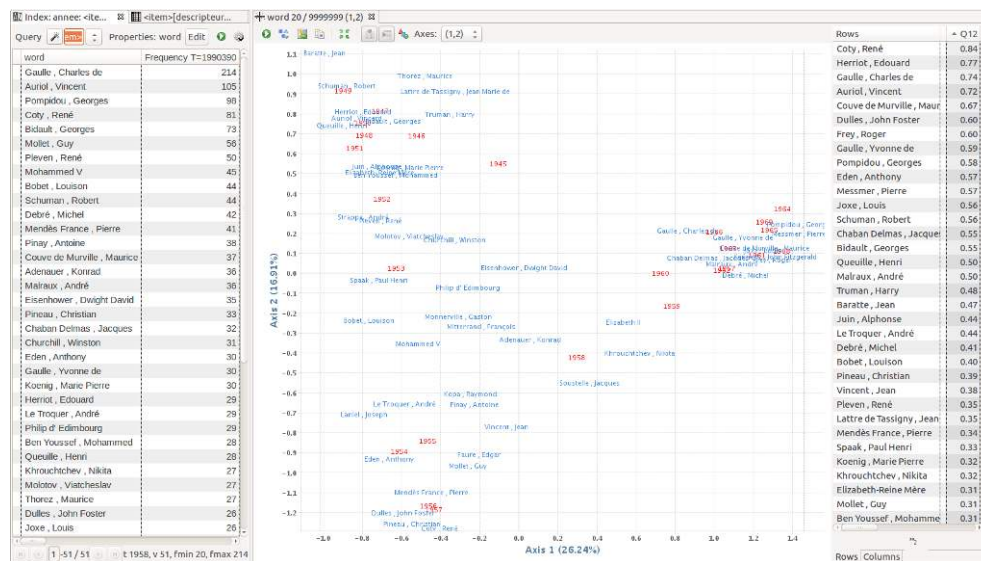
Termes SPÉCIFIQUES dans les commentaires de la voix off pour les sujets montrant une foule (selon les notices documentaires) [fenêtre supérieure] ; puis, termes SPÉCIFIQUES dans les transcriptions de la voix off pour les sujets montrant une foule et n'ayant aucune mention de « De Gaulle » ou de « président » (fenêtre inférieure).

- 43 Avec la figure 10, nous appliquons une analyse de résonance (Salem, 2004). Lorsqu'une foule est montrée (d'après la notice documentaire), quels sont typiquement les mots prononcés dans le commentaire en voix off ? « Président » et « [le général de] Gaulle » représentent ainsi le principal contexte d'emploi (Figure 10, fenêtre supérieure). Dans un second temps, nous supprimons tous les sujets contenant l'un de ces deux mots et nous nous focalisons sur les sujets restants pour faire émerger de nouveaux types de contextes associés à la mise à l'image d'une foule (Figure 10, fenêtre inférieure), tels que le sport, les commémorations, les manifestations, les fêtes, etc. La mention récurrente de la « foule » dans les commentaires en voix off favorise le sentiment d'appartenance à une communauté de destin. D'un point de vue méthodologique, ce type d'interrogation croisée, combinée à une comparaison statistique entre les notices documentaires et les transcriptions des commentaires audio, permet d'étudier les corrélations ou les divergences entre ce qui est montré à l'image et ce qui est dit dans

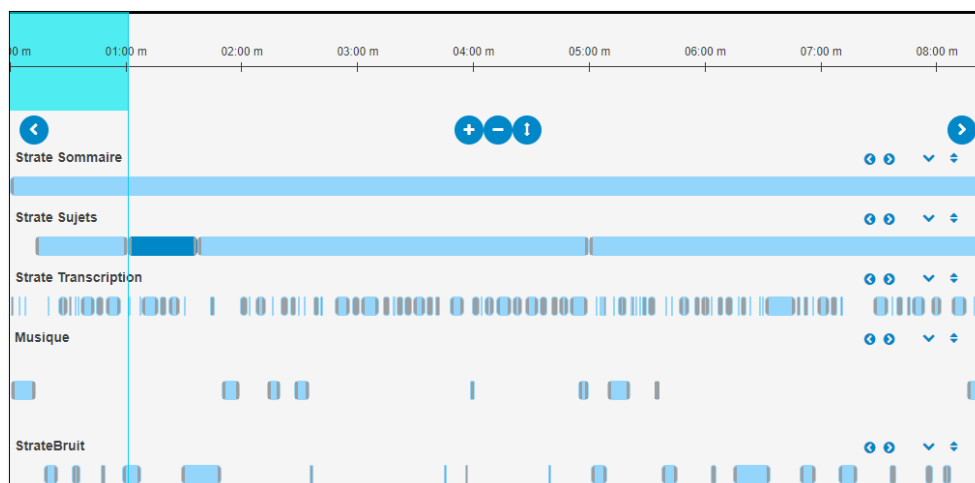
les commentaires. Une telle analyse croisée de différents médias est rarement fournie dans les logiciels d'analyse.

- 44 La figure 11 donne un premier aperçu des résultats d'une analyse des correspondances : nous avons calculé une carte bidimensionnelle des noms des personnes présentes dans plus de 20 sujets, en relation avec les années où elles sont mentionnées. Nous obtenons ainsi une vue synthétique de la relation entre les personnes et le temps dans les sujets des *Actualités Françaises*. En termes de modélisation statistique, comme la textométrie traite souvent de tableaux de fréquences croisant des mots et des parties de corpus (ici nous avons croisé des noms de personnes et des années), elle opte pour l'analyse factorielle des correspondances, car ce type d'analyse multidimensionnelle est particulièrement bien adapté aux tableaux de contingence (Lebart, 1998).

Figure 11. Carte bidimensionnelle des noms des personnes et des années de leur mention.



- 46 Okapi fournit un ensemble d'outils pour la segmentation et la description sémantique de contenus multimédias (vidéo, image, son, texte et 3D) en s'appuyant sur des ontologies de domaine. Le logiciel fournit également des services pour la constitution de corpus hiérarchiques et thématiques à partir d'extraits annotés pour l'ensemble de ces modalités.
- 47 Okapi gère les entités nommées comme des graphes de connaissances et fournit des services destinés à les rechercher, les partager et les présenter comme des données ouvertes. Ces entités peuvent être mises en relation avec d'autres entités dans des bases de connaissances comme DBpedia et Wikidata, ce qui rend Okapi interopérable avec l'écosystème des données ouvertes liées [LOD : *Linked Open Data*] (Bizer, 2009). Les entités nommées peuvent être de différents types qui varient en fonction du domaine étudié, par exemple personnes physiques et morales, lieux géographiques, événements, concepts.
- 48 Le système de gestion de contenu d'Okapi prend en compte les caractéristiques du domaine étudié et les préférences utilisateurs pour générer des interfaces web telles que des portails adaptés au domaine, sans requérir aucune compétence technique de leurs auteurs. Un auteur peut également réaliser une publication thématique sous forme d'un ensemble d'éléments multimédias interconnectés (vidéo, image, son) auxquels il peut ajouter du contenu éditorial. L'outil de publication applique ensuite un ensemble de règles de publication sur ces éléments et génère un mini-portail.
- 49 Dans le cadre du projet ANTRACT, la plateforme Okapi est utilisée par les historiens pour constituer des corpus thématiques, visualiser et, dans certains cas, corriger les métadonnées originelles issues des notices Ina (l'ancrage temporel des sujets Ina non documentés) ainsi que les résultats issus des algorithmes automatiques (détection et reconnaissance de visages, transcription de la parole, OCR sur les tapuscrits et sur les vidéos, etc.) Les sections suivantes montrent quelques exemples d'utilisation de la plateforme Okapi sur la collection des *Actualités Françaises*.
- 50 La description d'un média dans Okapi peut être réalisée manuellement par des annotateurs ou automatiquement par des algorithmes d'indexation selon plusieurs axes (thématique, sonore, visuel, etc.) comme le montre la figure 12. La *timeline* Amalia (Hervé, 2015) est utilisée pour visualiser ces axes de description sous forme de strates et représenter la progression temporelle de chacun d'entre eux. Plusieurs types de strates sont proposés pour la représentation temporelle des données ANTRACT, chacune étant dédiée à un type d'annotation : les strates « sommaire » et « sujets » sont consacrées aux métadonnées originelles issues, respectivement, des notices émissions et notices sujets de l'INA, la strate « visage » est dédiée aux métadonnées extraites par l'algorithme de détection et d'identification des visages (Lisena, 2022) [Section « Analyse visuelle automatique »] et la strate « transcription » aux métadonnées issues de l'algorithme de la transcription de la parole (Section « Analyse audio automatique »). Ces métadonnées sont portées par des objets de type « segment » qui délimitent leurs portées temporelles au sein d'une strate (Figure 12) et peuvent être structurées, suivant leur type, en plusieurs annotations. Par exemple, la description thématique (strate intitulée « sujets ») de l'émission *Journal Les Actualités Françaises : émission du 10 juillet 1968* (Figures 12 et 13) consiste à identifier les thèmes abordés dans cette émission, leur portée temporelle et une description détaillée du sujet abordé, des lieux où l'action se déroule et des personnes impliquées.

Figure 12. *Timeline* de description.

- 51 L'utilisateur peut créer une strate pour ajouter une nouvelle dimension de description, identifier les portées temporelles des annotations en créant des segments et renseigner les informations associées. Okapi fournit une boîte à outils pour ajuster finement l'ancrage temporel de chaque segment. Cette boîte à outils a été utilisée notamment dans le projet ANTRACT pour corriger l'ancrage temporel des sujets mal timecodés dans les notices Ina.
- 52 La description d'un média consiste à affecter un graphe de connaissances à chaque segment de la *timeline*. Cette opération, un peu complexe, a été simplifiée dans Okapi et ramenée à l'édition d'un simple formulaire. Le logiciel Okapi suggère pour chaque champ du formulaire un ensemble pertinent de valeurs en interprétant l'ensemble de l'axiomatique présente dans l'ontologie du domaine. Les formules de l'axiomatique OWL2 sont interprétées comme des contraintes contextuelles qui permettent de réduire cet ensemble de valeurs et donc de réduire la charge cognitive de l'utilisateur en le guidant dans son travail d'annotation. Ces contraintes étant récursives, elles permettent également de définir contextuellement des classes anonymes, cela permet de réduire le nombre de classes nommées et de propriétés devant être déclarées dans l'ontologie du domaine. Prenons comme exemple le deuxième segment de la strate « Sujets » (sélectionné sur la *timeline* en figure 12) où l'on parle des « sports nautiques (concept) » en « Angleterre (Lieu) », en particulier les aventures du navigateur solitaire « Alec Rose (Personne) ». Ces annotations sont structurées autour des thèmes dont on parle, des lieux où cela se passe et des personnes qui y sont impliquées. Ces métadonnées sont représentées par un mini-graphe de connaissances et présentées à l'utilisateur sous forme d'un formulaire éditable (Figure 13). Ces concepts, lieux et personnes sont un sous-ensemble de la base des connaissances qui sont suggérées par Okapi pour compléter la description du segment en interprétant les contraintes posées sur les propriétés « thème », « à l'image » et « lieu ».

Figure 13. Formulaire de métadonnées de segment.

SOMMAIRE GÉNÉRALITÉS DESCRIPTEURS

résumé

- VG du voilier du navigateur solitaire, Alec ROSE, naviguant dans le port de Portsmouth, escorté par d'autres bateaux
- VG PANO en plongée sur une foule dense de spectateurs massés sur les quais du port, certains agitant des drapeaux anglais
- VG avec ZAV sur le voilier "LIVELY LADY" ; Alec ROSE sur le pont, prêt à lancer les amarres
- VG de la foule venue l'accueillir
- PM du navigateur solitaire Alec ROSE, en costume et casquette de marin, embrassant sa femme sur le quai de Portsmouth et faisant un geste de salut
- VG de nombreux spectateurs agitant la main
- VG du maire de Portsmouth, près de Alec ROSE et de sa femme, donnant le signal des "Hurrah !"
- foule acclamant.

thème

sport nautique(Schéma Noms communs)

à l'image

Rose, Alec

lieu

Grande Bretagne

Royaume Uni

Angleterre

Français

Français

Français

Français

Français

Français

- 53 Les autres strates de description (transcription, détection de la musique, détection de la parole homme/femme, etc.) sont calculées automatiquement par des algorithmes. Les métadonnées produites viennent enrichir les métadonnées originelles des notices documentaires de l'Ina ou celles créées manuellement par les utilisateurs d'Okapi. L'ensemble de ces métadonnées sont utilisées par la plateforme Okapi pour générer un portail riche qui apporte de la valeur au contenu multimédia et offre plusieurs possibilités d'accès et de navigation dans ce contenu, comme le montre la figure 14.

Figure 14. Page du portail Okapi de l'émission « *Journal Les Actualités Françaises : émission du 10 juillet 1968* ».

Journal Les Actualités Françaises : émission du 10 juillet 1968

Rechercher

SOMMAIRE GÉNÉRALITÉS DESCRIPTEURS

ANALYSE

00:01:04.06
00:08:26.04

STRATE SOMMAIRE STRATE SUJETS MUSIQUE VOIX HOMME

Au Stade Charlety, la confrontation des athlètes Américains et Français
durée: 00:00:48:0

Alec Rose, après 354 jours sur un bateau : "la terre est ronde"
durée: 00:00:36:0

résumé

- VG du voilier du navigateur solitaire, Alec ROSE, naviguant dans le port de Portsmouth, escorté par d'autres bateaux - VG PANO en plongée sur une foule dense de spectateurs massés sur les quais du port, certains agitant des drapeaux anglais - VG avec ZAV sur le voilier "LIVELY LADY" ; Alec ROSE sur le pont, prêt à lancer les amarres - VG de la foule venue l'accueillir - PM du navigateur solitaire Alec ROSE, en costume et casquette de marin, embrassant sa femme sur le quai de Portsmouth et faisant un geste de salut - VG de nombreux spectateurs agitant la main - VG du maire de Portsmouth, près de Alec ROSE et de sa femme, donnant le signal des "Hurrah !"

thème

sport nautique

à l'image

Rose, Alec

lieu

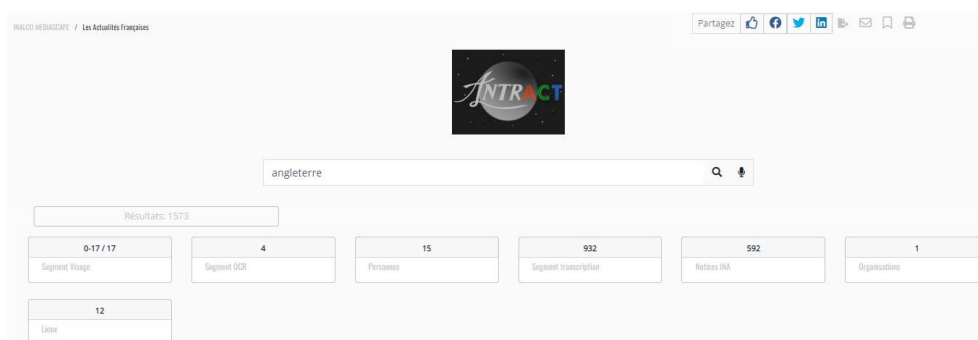
Angleterre

Grande Bretagne

Royaume Uni

- 54 Ces métadonnées viennent également alimenter les index Okapi pour la recherche en texte intégral sur les objets de la base. Ainsi, les métadonnées de transcription de la parole permettent de rechercher des passages dans le flux audio où l'on prononce certains mots-clés, celles de l'OCR sur les vidéos de retrouver les extraits vidéo où certains mots-clés sont affichés à l'écran, les informations extraites sur les visages de retrouver les extraits vidéo où l'on voit à l'écran le visage d'une personnalité donnée. Les résultats d'une recherche en texte intégral sont classés en fonction de leurs natures dans des catégories différentes (voir Figure 15) pour faciliter leur lecture et compréhension.

Figure 15. Recherche sur le texte intégral.



- 55 Ces métadonnées peuvent être aussi utilisées comme des critères avancés pour une recherche fine et sémantique de contenus. La figure 16 montre un exemple de recherche avancée d'extraits vidéo dans lequel on parle de « Sports nautiques » en « Angleterre ». Une recherche avancée est réifiée dans Okapi par un objet de type « requête sémantique » qui est stocké dans la base comme un graphe de connaissances et peut être retrouvé à l'aide du moteur de recherche, édité à l'aide d'un formulaire avant d'être transformé en requête SPARQL et exécuté par le serveur. Les résultats de cette requête, illustrés par la figure 17, peuvent être utilisés pour créer et/ou enrichir un corpus.

Figure 16. Exemple d'une requête Okapi.

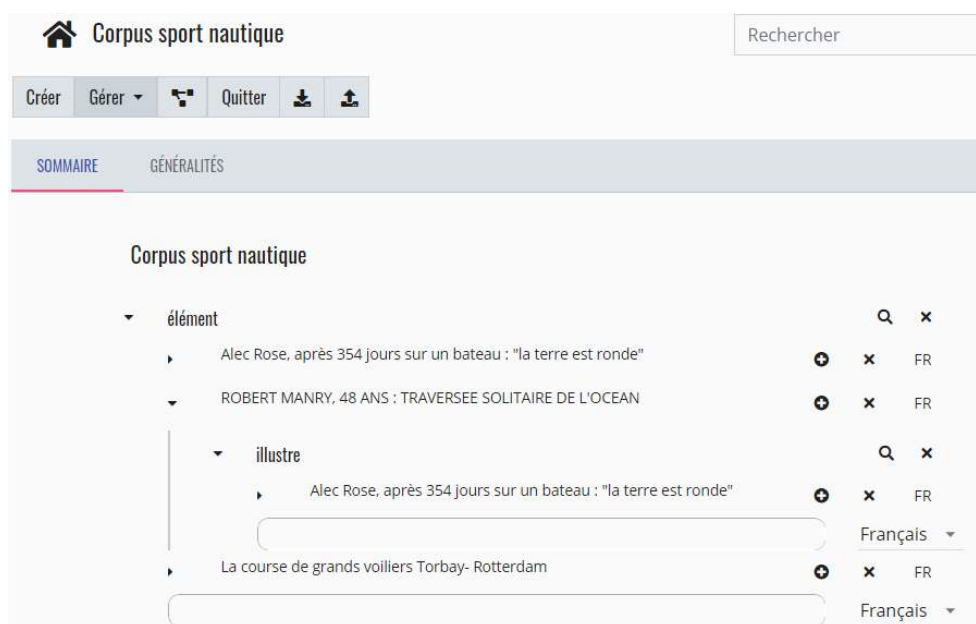
The screenshot shows the Okapi advanced search interface. At the top, there's a 'Notice sujet' header and a 'Rechercher' button. Below the header, there's a navigation bar with 'Créer', 'Gérer', 'Rechercher', 'Valider', and 'Quitter'. The main section is titled 'DESCRIPTEURS' and contains two search criteria: 'thème' and 'lieu'. The 'thème' criterion is set to 'sport nautique (Schéma Noms communs)(Schéma Noms communs)' with a language dropdown set to 'Français'. The 'lieu' criterion is set to 'Angleterre' with a language dropdown set to 'Français'.

Figure 17. Exemple de résultats d'une requête.



- 56 Les outils de suggestions de données et de recherche sémantique présentés dans les paragraphes précédents permettent d'alimenter des corpus utilisateurs. Un corpus dans Okapi est un regroupement hiérarchique de contenus multimédias (extraits vidéo/audio, parties d'image, extraits PDF, points de vue caméra d'une scène 3D) qui partagent une certaine thématique. En fonction de cette thématique et de la taille du corpus, celui-ci peut être constitué de plusieurs sous-corpus, chacun abordant par exemple une sous-thématique. Une fonctionnalité de glisser-déplacer (*drag'n drop*) a été mise en place pour faciliter la réorganisation d'un corpus en déplaçant certains éléments entre ses sous-corpus.
- 57 Un corpus est aussi un objet de la base qui peut être annoté. L'utilisateur peut ajouter de nouvelles métadonnées sur le corpus lui-même ou sur ses éléments. Il peut également tirer des relations rhétoriques (exemplification, illustration, etc.) et discursives entre ces éléments. La figure 18 montre un corpus composé de trois extraits, récupérés à partir de la requête présentée dans le paragraphe précédent. Elle montre également une relation rhétorique entre les deux segments : « Robert Manry, 48 ans : Traversée solitaire de l'océan » qui illustre l'autre segment « Alec Rose, après 354 jours sur un bateau : "la terre est ronde" ». Toutes ces métadonnées peuvent être utilisées pour créer un portail thématique centré sur le contenu du corpus ou intégré à un récit par l'inclusion de contenu éditorial et de parcours de lecture.
- 58 La plateforme Okapi expose un *endpoint* SPARQL sécurisé et une API qui permet aux autres outils ANTRACT, en particulier à la plateforme TXM (Section « Analyse textométrique interactive »), d'interroger et d'enrichir certains objets de la base de connaissances. Par exemple, un utilisateur TXM peut récupérer un corpus par le biais du point d'accès Okapi, tirer parti des capacités de textométrie de l'outil TXM pour enrichir ce corpus et le renvoyer vers Okapi *via* son API. L'utilisateur peut ensuite reprendre ce corpus sur Okapi pour le compléter, le réorganiser et le publier sur le portail ou sous forme d'une publication auteur.

Figure 18. Corpus thématique « Sports nautiques ».



Conclusion

- 59 Présenté tout au long de ce chapitre, le défi du projet ANTRACT est de familiariser les chercheurs en sciences humaines et sociales avec l'analyse automatique et les nouvelles possibilités de recherche sur les grands corpus audiovisuels en contexte numérique. En rassemblant des instruments spécialisés dans l'analyse d'images, d'audio et de textes dans un environnement multimodal conçu pour corréliser leurs résultats, le projet développe un modèle de recherche transdisciplinaire destiné à ouvrir de nouvelles perspectives dans l'étude de sources mono ou multiformat.
- 60 Une grande partie des travaux du projet a été consacrée au développement et au réglage des outils d'analyse automatique de contenu ainsi qu'à l'application de leurs résultats à l'organisation et à l'amélioration des données du corpus en lien avec les recherches fournies par les historiens d'ANTRACT (Goetschel, 2019 ; Carrive, Goetschel et Mazuet, à paraître). Des études de cas ont été menées en utilisant la plateforme de textométrie TXM et la plateforme d'annotation et de publication Okapi qui permettent à leurs utilisateurs d'exploiter les données produites par les instruments développés pour le projet.
- 61 D'un point de vue technologique, la transcription vérifiée réalisée en fin de projet est une nouvelle ressource qui ouvre d'importantes perspectives pour entraîner, adapter et évaluer les outils d'analyse automatique de contenu à la spécificité d'un tel corpus d'archives, comme son contexte historique, son vocabulaire, son format et sa qualité d'image.
- 62 À la fin du projet, un corpus complet, *Les Actualités Françaises*, complété par ses métadonnées ainsi que les résultats de la recherche obtenus par des outils d'analyse de contenu automatique et des annotations manuelles, ont été mis à la disposition de la communauté scientifique via la plateforme en ligne Okapi (pour la consultation et l'analyse en ligne) et via l'entrepôt de données Dataset de l'Ina (pour l'accès aux

données et aux corpus TXM)⁹. À cette fin, des didacticiels Okapi ont été réalisés et TXM continue d'être disponible en tant que logiciel libre pour faciliter l'analyse des corpus utilisés dans les nouvelles études de cas. Le code source d'Okapi doit être prochainement distribué en *open source* afin que d'autres développeurs puissent contribuer à son amélioration.

- 63 En ce qui concerne les sciences humaines et sociales, les outils et la méthodologie d'ANTRACT offrent aux historiens, mais aussi à des spécialistes d'autres disciplines – sociologie, anthropologie, sciences politiques –, la possibilité de disposer d'un corpus enrichi de très grande qualité, composé non seulement des sujets des *Actualités Françaises* entre 1945 et 1969, mais aussi des données obtenues grâce à l'usage des différents outils d'analyse automatique de contenu, considérablement améliorés au fil du projet. Plus globalement, les outils et la méthodologie d'ANTRACT ouvrent de nouvelles perspectives pour l'analyse multidisciplinaire de ce type de corpus.

BIBLIOGRAPHIE

Timo AHONEN, Abdenour HADID et Matti PIETIKAINEN, "Face description with local binary patterns: application to face recognition", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 28.12, 2006, p. 2037-2041.

Scott ALTHAUS, Kaye USRY, Stanley RICHARDS, Bridgette VAN THUYLE, Isabelle ARON, Lu HUANG, Kalev LEETARU, Monica MUEHLFELD, Karissa SNOUFFER, Seth WEBER, Yuji ZHANG et Patricia PHALEN, "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, p. 147-167.

Nathan S. ATKINSON, "Newsreels as Domestic Propaganda: Visual Rhetoric at the Dawn of the Cold War", *Rhetoric & Public Affairs*, 14.1, 2011, p. 69-100.

Ulrike BARTELS, *Die Wochenschau im Dritten Reich. Entwicklung und Funktion eines Massenmediums unter besonderer Berücksichtigung völkisch-nationaler Inhalte*, Francfort-sur-le-Main, Peter Lang, 2004.

Abdelkrim BELOUED, Peter STOCKINGER et Steffen LALANDE, « Studio Campus AAR : Une plateforme sémantique pour l'analyse et la publication de corpus audiovisuels », dans *Intelligence collective et archives numériques*, Hoboken, NJ, John Wiley & Sons Inc, 2017, p. 85-133.

Alex BEWLEY, Zongyuan GE, Lionel OTT, Fabio RAMOS et Ben UPCROFT, "Simple online and realtime tracking", *Conférence internationale de l'IEEE sur le traitement des images [ICIP]*, 2016, p. 3464-3468.

Chritian BIZER, Tom HEATH & Tim BERNERS-LEE, "Linked data – the story so far", *International Journal on Semantic Web and Information Systems*, 5, 2009, p. 1-22.

Gary BRADSKI, "The OpenCV Library", *Dr. Dobb's Journal of Software Tools*, 2000.

Pierre-Alexandre BROUX, Florent DESNOUS, Anthony LARCHER, Simon PETITRENAUD, Jean CARRIVE et Sylvain MEIGNIER, "S4D : Speaker Diarization Toolkit in Python", *Interspeech*, Hyderabad, Inde, 2018.

- Qiong CAO, Li SHEN, Weidi XIE, Omkar M. PARKHI et Andrew ZISSERMAN, "VGGFace2. A dataset for recognising faces across pose and age", *13^e conférence internationale de l'IEEE sur la reconnaissance automatique des visages et des gestes (FG)*, 2018, p. 67-74.
- Jean CARRIVE, Pascale GOETSCHER et Franck MAZUET (dir.), *Pour une histoire outillée d'un corpus d'actualités filmées. Les Actualités Françaises (1945-1969)*. INA, L'Harmattan, coll. « Les Médias en Actes », à paraître en 2024.
- Ciara CHAMBERS, Mats JÖNSSON et Roel VANDE WINKEL (dir.), *Researching Newsreels. Études de cas locales, nationales et transnationales*, Global Cinema, Palgrave Macmillan, Londres, 2018.
- Jean-Hugues CHENOT et Gilles DAIGNEAULT, "A large-scale audio and video fingerprints-generated database of TV repeated contents", *12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, Klagenfurt, Autriche, 2014.
- Oliver CHRIST, "A modular and flexible architecture for an integrated corpus query system", in Ferenc KIEFER et al. (dir.), *3rd International Conference on Computational Lexicography*, Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, 1994, p. 23-32.
- Marilyn DEEGAN et Willard MCCARTY, *Collaborative Research in the Digital Humanities*. Ashgate, Farnham, Burlington, 2012.
- ELAN (Version 5.2) [Logiciel informatique]. Institut Max Planck de psycholinguistique, Nimègue, 2018. Récupéré sur <https://archive.mpi.nl/tla/elan>
- Seth FEIN, "New Empire into Old: Making Mexican Newsreels the Cold War Way", *Histoire diplomatique*, 28.5, 2004, p. 703-748.
- Seth FEIN, "Producing the Cold War in Mexico. The Public Limits of Covert Communications", dans Gilbert M. JOSEPH et Daniela SPENSER (dir.), *In from the Cold : Latin America's New Encounter with the Cold War*, Duke University Press, Durham, 2008, p. 171-213.
- Ingo FEINERER, Kurt HORNIK et David MEYER, "Text Mining Infrastructure in R", *Journal of Statistical Software*, 25.5, 2008, p. 1-54.
- Pascale GOETSCHER et Christophe GRANGER (dir.), « Faire l'événement, un enjeu des sociétés contemporaines », *Sociétés & Représentations*, 32, 2011, p. 7-23.
- Pascale GOETSCHER, « Les Actualités Françaises (1945-1969) : le mouvement d'une époque », *ANTRACT Analyse transdisciplinaire des actualités filmées*, 1, 2019, <https://antract.hypotheses.org/127>
- Davis E. KING, "Dlib-ml: A machine learning toolkit", *Journal of Machine Learning Research*, 10, 2009, p. 1755-1758.
- Serge HEIDEN, "The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme", dans Ryo OTOGURO, Kiyoshi ISHIKAWA, Hiroshi UMEMOTO, Kei YOSHIMOTO, Yasunari HARADA (dir.), *24th Pacific Asia Conference on Language, Information and Computation*, Institute for Digital Enhancement of Cognitive Development, Waseda University, 2010.
- Nicolas HERVÉ, Pierre LETESSIER, Mathieu DERVAL et Hakim NABI, "Amalia.js : an Open-Source Metadata Driven HTML5 Multimedia Player", *Open-Source Software Competition*, ACM Multimedia Conference 2015 (MM), October 2015, Brisbane, Australia.
- Andreas HOTH, Andreas NÜRNBERGER et Gerhard PAASS, "A brief survey of text mining", *LDV Forum*, 20.1, 2005, p. 19-62.

Kornelia IMESCH, Sigrid SCHADE et Samuel SIEBER (dir.), *Constructions of cultural identities in newsreel cinema and television after 1945*, transcript-Verlag, MediaAnalysis, 17, 2016.

Ludovic LEBART, André SALEM et Lisette BERRY, *Exploring textual data. Text, speech, and language technology*, 4, Kluwer Academic, Dordrecht, Boston, 1998.

Sylvie LINDEPERG, *Clio de 5 à 7 : les actualités filmées à la Libération, archive du futur*, Paris, CNRS, 2000.

Sylvie LINDEPERG, « Spectacles du pouvoir gaullien : le rendez-vous manqué des actualités filmées », dans Jean-Pierre BERTIN-MAGHIT (dir.), *Une histoire mondiale des cinémas de propagande*, Paris, Nouveau Monde Éditions, 2008, p. 497-511.

Pasquale LISENA, Jorma LAAKSONEN et Raphaël TRONCY, “Understanding Videos with Face Recognition: A Complete Pipeline and Applications”, *Multimedia Systems*, Special Issue on Data-driven Personalisation of Television Content, 28, 2022, p. 2147-2159.

Brian MCWHINNEY, *The Childes Project. Tools for Analyzing Talk*, L. Erlbaum Associates, Mahwah, N.J., 2000.

Sarah MAITLAND, “Culture in translation. The case of British Pathé News”, *Culture and news translation, Perspectives. Études sur la théorie et la pratique de la traduction*, 23.4, 2015, p. 570-585.

Boris MOTIK, Peter PATEL-SCHNEIDER et Bijan PARSIA, *OWL 2 Ontology Language: Structural Specification and Functional-Style Syntax (seconde édition)*, Recommendation du W3C, 11 décembre 2012.

Bénédicte PINCEMIN, Serge HEIDEN et Matthieu DECORDE, “Textometry on Audiovisual Corpora. Experiments with TXM software”, *15th International Conference on Statistical Analysis of Textual Data (JADT)*, Toulouse, 2020.

Daniel POVEY, Arnab GHOSH, Gilles BOULIANNE, Lukáš BURGET, Ondrej GLEMBEK, Nagendra GOEL, Mirko HANNEMANN, Petr MOTLICEK, Yanmin QIAN, Petr SCHWARZ, Jan SILOVSKY, Georg STEMMER et Karel VESELY, “The kaldi speech recognition toolkit”, *IEEE 2011 workshop on automatic speech recognition and understanding*, IEEE Signal Processing Society, Hilton Waikoloa Village, Big Island, Hawaii, US, 2011.

Daniel POVEY, Vijayaditya PEDDINTI, Daniel GALVEZ, Pegah GHAHREMANI, Vimal MANOHAR, Xingyu NA, Yiming WANG et Sanjeev KHUNDANPUR, “Purely sequence-trained neural networks for ASR based on lattice-free MMI”, *Interspeech*, San Francisco, 2016, p. 2751-2755.

Vladimir POZNER, « Les actualités soviétiques Durant la Seconde Guerre Mondiale : nouvelles sources, nouvelles approches », dans Jean-Pierre BERTIN-MAGHIT (dir.), *Une histoire mondiale des cinémas de propagande*, Paris, Nouveau Monde Éditions, 2008, p. 421-444.

R Core Team, “R : A Language and Environment for Statistical Computing”, R Foundation for Statistical Computing, Vienne, Autriche, 2014.

André SALEM, « Introduction à la résonance textuelle », dans Gérald PURNELLE *et al.* (dir.), *7^{es} Journées internationales d'Analyse statistique des Données Textuelles*, Louvain, Presses universitaires de Louvain, 2004, p. 986-992.

Helmut SCHMIDT, “Probabilistic Part-of-Speech Tagging Using Decision Trees”, *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK, 1994.

Florian SCHROFF, Dmitry KALENICHENKO et James PHILBIN, “Facenet: A unified embedding for face recognition and clustering”, *Actes de la conférence de l'IEEE sur la vision par ordinateur et la reconnaissance des formes*, 2015, p. 815-823.

- Christian SZEGEDY, Vincent VANHOUCKE, Sergey IOFFE, Jonathon SHLENS et Zbigniew WOJNA, “Rethinking the Inception Architecture for Computer Vision”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 2016.
- Raphaël TRONCY, Erik MANNENS, Silvia PFEIFFER et Davy VAN DEURSEN, *Media Fragments URI 1.0 (basic)*, Recommandation du W3C, 2012.
- Laurent VERAY, *Les Films d’actualités français de la Grande Guerre*, Paris, SIRPA/AFRHC, 1995.
- Paul VIOLA et Michael JONES, “Robust real-time face detection”, *International Journal of Computer Vision*, 57.2, 2004, p. 137-154.
- Sholom M. WEISS, Nitin INDURKHYA et Tong ZHANG, *Fundamentals of Predictive Text Mining*, Springer-Verlag, Londres, 2015.
- Kaipeng ZHANG, Zhanpeng ZHANG, Zhifeng LI et Yu QIAO, “Joint face detection and alignment using multitask cascaded convolutional networks”, *IEEE Signal Processing Letters*, 23.10, 2016, p. 1499-1503.

NOTES

1. ANTRACT : ANALyse TRansdisciplinaire des ACTualités filmées (1945-1969).
 2. Corpus ESTER 1 & 2, EPAC, ETAPE, et REPERE disponibles dans les catalogues ELRA (<http://www.elra.info/>).
 3. ETAPE et QUAERO, corpus disponibles dans les catalogues ELRA (<http://www.elra.info/>).
 4. Challenge REPERE, données de test.
 5. <https://github.com/Linzaer/Face-Track-Detect-Extract>.
 6. Disponible à l’adresse <http://facerec.eurecom.fr/>.
 7. L’application est accessible au public à l’adresse <http://facerec.eurecom.fr/visualizer/?project=antract>.
 8. Text Encoding Initiative, <https://tei-c.org>.
 9. Voir <http://okapi.ina.fr>.
-

AUTEURS

JEAN CARRIVE

Institut national de l’audiovisuel [Ina]

ABDELKRIM BELOUED

Institut national de l’audiovisuel [Ina]

PASCALE GOETSCHER

Centre d’histoire sociale des mondes contemporains, UMR 8058 (université Paris 1/CNRS)

SERGE HEIDEN

Institut d'histoire des représentations et des idées dans les modernités [IHRIM], UMR 5317
(université de Lyon)

STEFFEN LALANDE

Institut national de l'audiovisuel [Ina]

PASQUALE LISENA

EURECOM

FRANCK MAZUET

Centre d'histoire sociale des mondes contemporains, UMR 8058 (université Paris 1/CNRS)

SYLVAIN MEIGNIER

Laboratoire d'informatique de l'université du Mans [LIUM]

BÉNÉDICTE PINCEMIN

Institut d'histoire des représentations et des idées dans les modernités [IHRIM], UMR 5317
(université de Lyon)

RAPHAËL TRONCY

EURECOM