

24^{ème} édition du Festival des Rendez-vous de l'Histoire
« Le travail »
Blois, 6-10 octobre 2021



Travail et cinéma : la fabrique des Actualités Françaises (1945-1969)



Jean CARRIVE (INA, Bry-sur-Marne)
Pascale GOETSCHER (CHS, Univ. Paris 1)
Pasquale LIENA (EURECOM, Sophia-Antipolis)
Franck MAZUET (CHS, Univ. Paris 1)
Sylvain MEIGNIER (LIUM, Univ. Le Mans)
Bénédicte PINCEMIN (IHRIM, U. Lyon, CNRS)



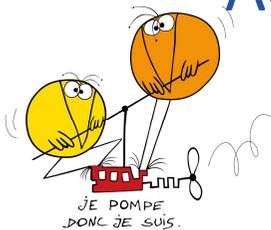


Analyse Transdisciplinaire des Actualités filmées (1945-1969)



Contexte et objectifs

- Un corpus exceptionnel d'actualités cinématographiques
- 1262 journaux, 20 232 sujets, tapuscrits des commentaires, notices documentaires
- Une équipe transdisciplinaire d'historiens du contemporain et de chercheurs en informatique
- proposer une analyse transdisciplinaire (historique et technologique), d'un corpus volumineux des images, des textes et des sons produits par la société Les Actualités Françaises entre 1945 et 1969



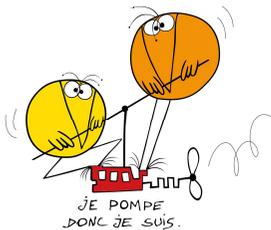
ANTRACT

Problématiques historiques

- Foules et émotions dans la 2nde moitié du XX^e siècle
- Prisonniers de la seconde guerre mondiale
- Stars du cinéma français
- Expositions d'art : attitudes, habitudes du public

Aujourd'hui

- Le monde du travail aux Actualités Françaises
- Le travail des opérateurs (La standardisation des sujets des Actualités Françaises à partir d'une grammaire cinématographique récurrente)
- Le travail de médiatisation des grandes figures de l'actualité



ANTRACT

Le Corpus

La Société « Les Actualités Françaises »

- 4 janvier 1945 – 25 février 1969
- 1262 journaux filmés (un journal par semaine)
- Environ 10 minutes
- Une dizaine de sujets par journal

Documentation très riche

- « Notices documentaires »
- Tapuscrits des voix
- Description de l'image « plan par plan »



ANTRACT

Un extrait (transcrit automatiquement)



Actualités Françaises, « La femme et le travail : celles qui font des métiers d'homme », 30 mars 1966



ANTRACT

Tapuscrits et sommaires

LES ACTUALITÉS FRANÇAISES
« REGARDS SUR LE MONDE »

JOURNAL N° 17

LE SPORT
1. Le Relais à travers à Paris 31m 30

LA SEMAINE
1. A Cannes, l'ouverture du 6ème Festival du film 16m 50
2. A Paris, Golette grand Officier de la Légion d'Honneur 13m 10

JEU DE MAINS
1. A Paris, une perspicacité qui paie *Concours d'ouvrages du "FIGARO"* 14m 20

DEMAIN
1. Les "castors de Thiais" 27m 60
2. Choisy le Roi aura sa cité moderne 21m 10
3. A Montlouis, les fours solaires 21m

REGARDS SUR LE MONDE
1. A Paris, l'élégance des mains 29m 90
2. Au Château de Windsor, l'anniversaire de la Reine 15m 50
3. Les fêtes de la création du soleil sur la colline de Sion (exclusif) 22m 50
29m 70
9m 20

GENÉRIQUE ET FIN
DURÉE: 9' 8" TOTAL: 251m 60

Sortie le 23 Avril 1953

--:--:--

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 14m 60
2. Radar 16m 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

ÉTRANGER
1. Départ du paquebot "Flandre" 12m
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 vigneron à l'Elysée 17m

Journal 17/53 2

REGARDS SUR LE MONDE
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 Seng-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 Château de Windsor, berceau des souverains britanniques, la Reine Elizabeth, refaisait 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 vœux dans les trois "hours" 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. Percant les nuées, annoncé au son des trompes de David, l'astre, tel qu'il sortit des mains divines apparaît au peuple élu qui, manifeste sa joie en embrassant les rouleaux de la loi. De grandes réjouissances commencent alors, pour célébrer cet événement 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 de 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

OCR des tapuscrits

LA FEMME ET LE TRAVAIL 66-070.

Sa première visite à l'étranger a conduit Madame Indira Gandhi à l'Elysée où le Président de la République a eu une longue conversation avec le Premier Ministre de l'Inde. Entretien facilité par la parfaite connaissance de notre langue de celle qui, a courageusement accepté de prendre en mains la destinée de 400 millions d'Indiens... Oui, il est passé beaucoup d'eau sous les ponts de Delhi et de Londres depuis le temps où les suffragettes, derrière Mrs Pankhurst, réclamaient l'égalité des droits civiques pour les femmes. Depuis cette époque héroïque, le féminisme a fait tache d'huile et la petite histoire a retenu l'exemple de ce village français où Madame le Maire et les Conseillères Municipales gèrent, à la satisfaction générale, les affaires de la commune. Dans les pays de l'Est certains emplois réputés masculins sont aujourd'hui couramment pratiqués par des femmes. Mais que dire de cette promotion exceptionnelle : chef de gare principale ... Il y a de l'espoir pour les poinçonneuses!

LA FEMME ET LE TRAVAIL 66-070.

Sa première visite à l'étranger a conduit Madame Indira Gandhi à l'Elysée où le Président de la République a eu une longue conversation avec le Premier Ministre de l'Inde. Entretien facilité par la parfaite connaissance de notre langue de celle qui, a courageusement accepté de prendre en mains la destinée de 400 millions d'Indiens... Oui, il est passé beaucoup d'eau sous les ponts de Delhi et de Londres depuis le temps où les suffragettes, derrière Mrs Pankhurst, réclamaient l'égalité des droits civiques pour les femmes. Depuis cette époque héroïque, le féminisme a fait tache d'huile et la petite histoire a retenu l'exemple de ce village français où Madame le Maire et les Conseillères Municipales gèrent, à la satisfaction générale, les affaires de la commune, Dans les pays de



ANTRACT



ASR →
OCR →



ANTRACT

REGARDS SUR LE MONDE

Journal I7/53 2

1. A Pén Mun Jom; la signature des accords 53.166

A Pén 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 vœux 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. Percant les nuées, annoncé au son des trompes de David, l'astre, tel qu'il sortit des mains divines apparaît au peuple élu qui, manifeste sa joie en embrassant les rouleaux de la loi. De grandes réjouissances commencent alors, pour célébrer cet événement 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 de 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 9 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

« Plan à plan »

Titre " La femme et le travail " surimpressionné sur PM de Madame Indira GANDHI, nouveau Premier ministre d'Inde posant sur les marches de l'Elysée

- Drapeau français flottant sur le toit de l'Elysée avec lanterne en 1er plan
- PM de Madame GANDHI s'entretenant avec le président De GAULLE dans un salon de l'Elysée
- Gardes républicains rendant les honneurs dans la cour de l'Elysée
- PM de Madame GANDHI serrant la main du président De GAULLE et de Madame De GAULLE en présence de Monsieur et

Madame POMPIDOU, sur le perron de l'Elysée

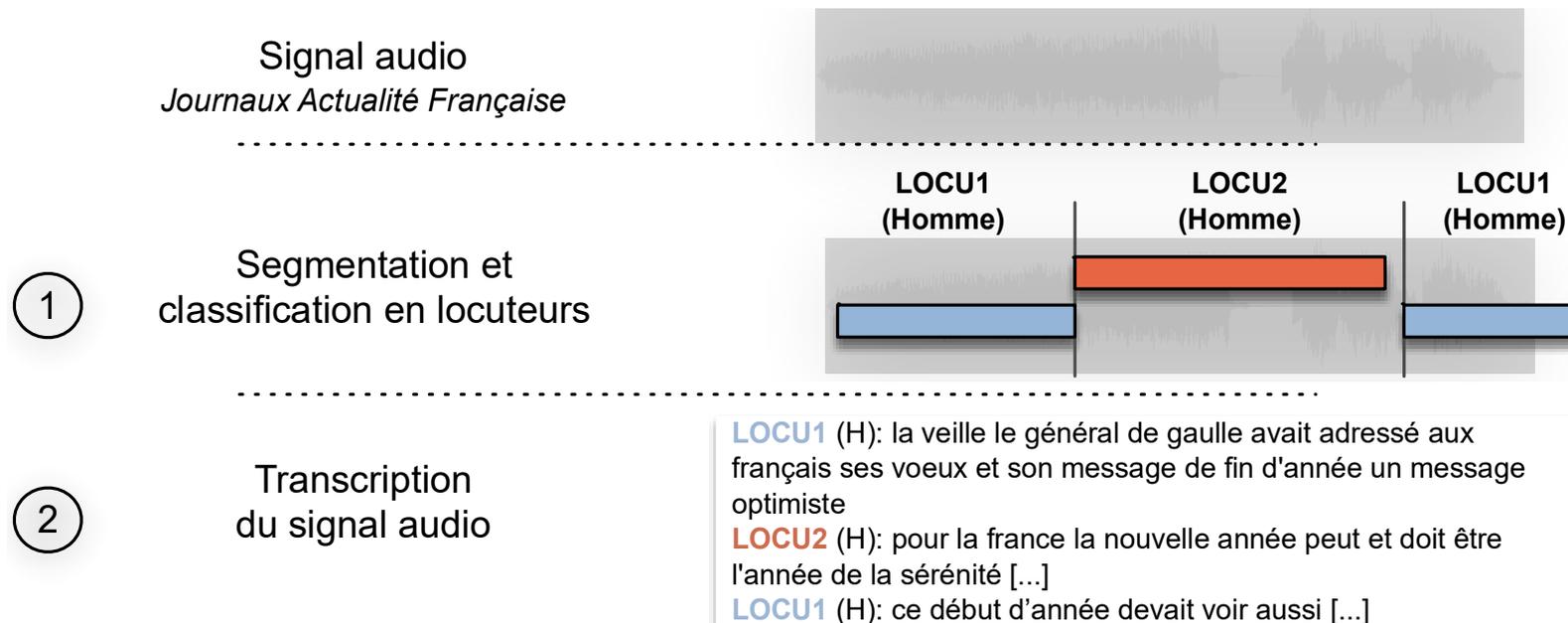
- GP du visage de Madame GANDHI
- VG en plongée d'un grand défilé de manifestants avec pancartes
- VG d'un défilé de femmes brandissant un drapeau (suffragettes de 1915 environ)
- 2 PM de suffragettes : brandissant une banderole, parlant dans un porte-voix, jouant du tambour (en Angleterre ?)
- VG d'une église de village

- VG d'une salle de classe dans laquelle une femme -maire et ses conseillères tiennent une réunion

PM de la femme -maire, ceinte de son écharpe, prononçant un discours avec panoramique sur ses conseillères. écoutant



Transcription enrichie de la parole



- Évaluation sur 12 émissions
 - 6% de la durée de l'émission est mal segmenté
 - 10% des mots sont erronés



Transcription enrichie de la parole

Signal audio
Journaux Actualité Française



2

Transcription
du signal audio

LOCU1 (H): la veille le général de gaulle avait adressé aux français ses voeux et son message de fin d'année un message optimiste

LOCU2 (H): pour la france la nouvelle année peut et doit être l'année de la sérénité [...]

LOCU1 (H): ce début d'année devait voir aussi [...]

3

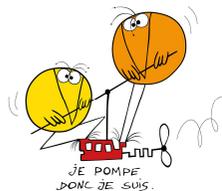
Ponctuations et majuscules

LOCU1 (H): La veille, le Général De Gaulle avait adressé aux français ses voeux et son message de fin d'année, un message optimiste.

LOCU2 (H): Pour la France, la nouvelle année peut et doit être l'année de la sérénité. [...]

LOCU1 (H): Ce début d'année devait voir aussi [...]

INTRAC



Transcription enrichie de la parole

3

Ponctuations et majuscules

LOCU1 (H): La veille, le Général De Gaulle avait adressé aux français ses voeux et son message de fin d'année, un message optimiste.

LOCU2 (H): Pour la France, la nouvelle année peut et doit être l'année de la sérénité. [...]

LOCU1 (H): Ce début d'année devait voir aussi [...]

4

Entités nommées

LOCU1 (H): La veille, le **FONCTION(Général) PERSONNE(De Gaulle)** avait adressé aux français ses voeux et son message de **DATE(fin d'année)**, un message optimiste.

LOCU2 (H): Pour la **ORGANISATION(France)**, **DATE(la nouvelle année)** peut et doit être l'année de la sérénité. [...]

LOCU1 (H): Ce **DATE(début d'année)** devait voir aussi [...]

- Entité nommée

- Wikipédia : une expression linguistique référentielle, souvent associée aux noms propres et aux descriptions définies
- Relativement floue et mouvante dans ses contours
- Lien fort avec la compréhension



Transcription enrichie de la parole

Signal audio
Journaux Actualité Française



4

Entités nommées

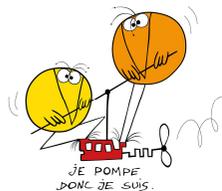
LOCU1 (H): La veille, le **FONCTION(Général) PERSONNE(De Gaulle)** avait adressé aux français ses voeux et son message de **DATE(fin d'année)**, un message optimiste.
LOCU2 (H): Pour la **ORGANISATION(France)**, **DATE(la nouvelle année)** peut et doit être l'année de la sérénité. [...]
LOCU1 (H): Ce **DATE(début d'année)** devait voir aussi [...]

5

*Identification des personnes
(modèle de voix a priori)*

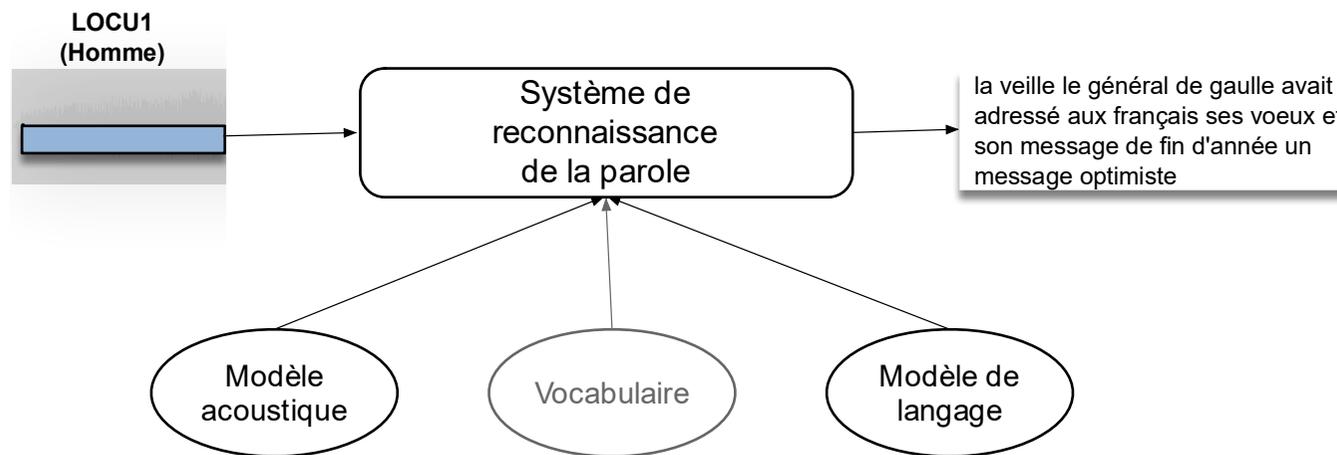
Présentateur (H): La veille, le **FONCTION(Général) PERSONNE(De Gaulle)** avait adressé aux français ses voeux et son message de **DATE(fin d'année)**, un message optimiste.
Charles De Gaulle (H): Pour la **ORGANISATION(France)**, **DATE(la nouvelle année)** peut et doit être l'année de la sérénité. [...]
Présentateur (H): Ce **DATE(début d'année)** devait voir aussi [...]

ANTRAC



Architecture : reconnaissance de la parole

*Système automatique et statistique,
il prédit la suite de mots la plus probable*

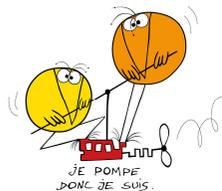


Probabilité des phonèmes

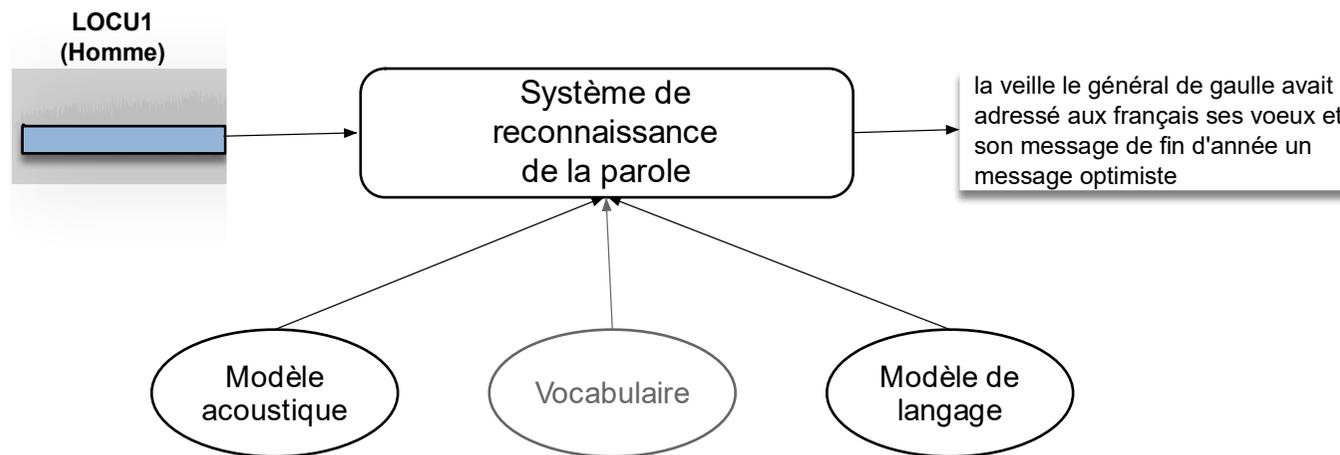
Suite de phonèmes pour chaque mot

Probabilité d'une suite de mots

ANTRAC



Architecture : reconnaissance de la parole



- Données d'apprentissage

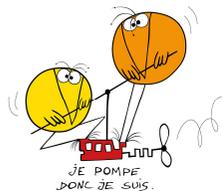
- Modèle acoustique (300 heures)

- Émissions actuelles transcrites manuellement (France inter, BFM, LCP, ...)

- Vocabulaire (160 000 mots) et modèle de langage (1,6 milliard de mots)

- Sélection d'une partie des tapuscrits et des notices documentaires
 - Transcriptions manuelles d'émissions actuelles, journaux « papier » (Le Monde, ...)

ANTRAC



Architecture : parole enrichie

3

Ponctuations et majuscules

LOCU1 (H): La veille, le Général De Gaulle avait adressé aux français ses voeux et son message de fin d'année, un message optimiste.

LOCU2 (H): Pour la France, la nouvelle année peut et doit être l'année de la sérénité. [...]

LOCU1 (H): Ce début d'année devait voir aussi [...]

4

Entités nommées

LOCU1 (H): La veille, le **FONCTION(Général) PERSONNE(De Gaulle)** avait adressé aux français ses voeux et son message de **DATE(fin d'année)**, un message optimiste.

LOCU2 (H): Pour la **ORGANISATION(France)**, **DATE(la nouvelle année)** peut et doit être l'année de la sérénité. [...]

LOCU1 (H): Ce **DATE(début d'année)** devait voir aussi [...]

• Données d'apprentissage

- Ponctuation et majuscules : 500 millions de mots
 - Transcriptions manuelle d'émissions actuelles, journaux « papier » (Le Monde, ...)
- Entités nommées : 1,6 millions de mots
 - Transcriptions manuelle d'émissions actuelles (200 heures)



Entités nommées : limites

4

Entités nommées

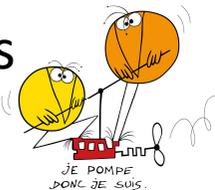
LOCU1 (H): La veille, le **FONCTION(Général) PERSONNE(De Gaule)** avait adressé aux français ses vœux et son message de **DATE(fin d'année)**, un message optimiste.

LOCU2 (H): Pour la **ORGANISATION(France)**, **DATE(la nouvelle année)** peut et doit être l'année de la sérénité. [...]

LOCU1 (H): Ce **DATE(début d'année)** devait voir aussi [...]

- 7 catégories
 - personne, fonction, lieu, organisation, production, date, quantité
- Les entités nommées « Antract » sont peu fréquentes dans le corpus d'apprentissage
- Elles sont mal modélisées (modèles de langage) ou même absentes
- Souvent, source d'erreur de transcription. Donc, elles sont mal détectées

ANTRAC



Erreurs

10% des mots sont en erreur, 90% sont correctement transcrit

- Détection du début et des fins de tours de parole
 - Des mots ne seront pas transcrits
 - Omission de majuscules et de ponctuation
- Mots inconnus : Mots absents du vocabulaire / du modèle de langage
 - Le système n'invente pas des mots, il met les mots les plus proches phonétiquement
- Mots proches phonétiquement
 - Le système substitue un mot par un autre (foule → poule)
 - Le système supprime un mot (acoustique ou mot peu fréquent dans le contexte)



Erreurs

10% des mots sont erronés, 90% sont correctement transcrit

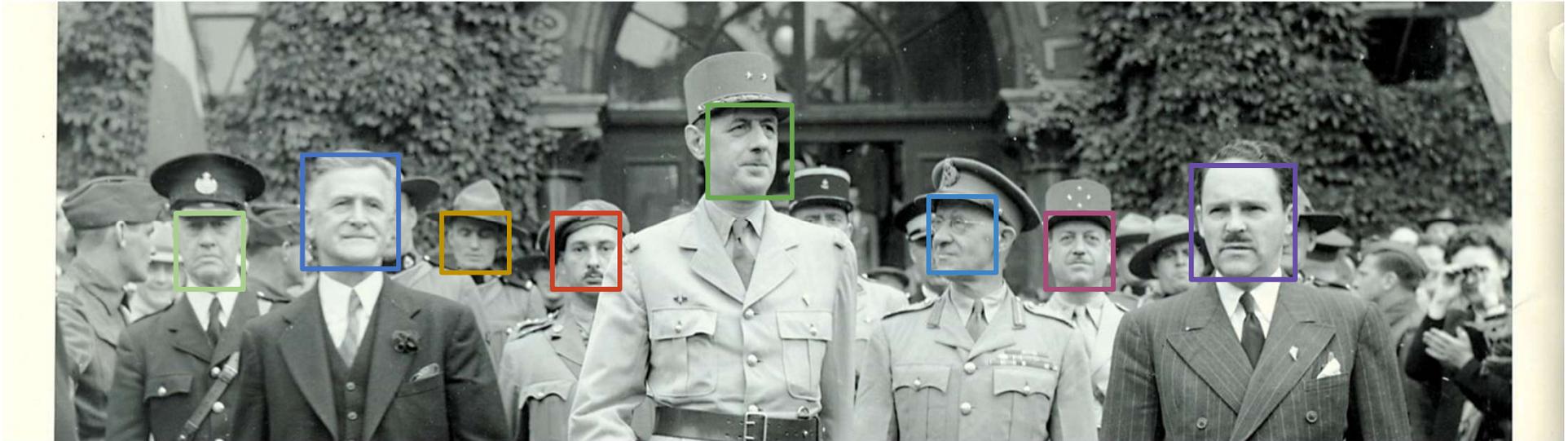
- Ces erreurs sont-elles graves ?
 - Elles portent souvent sur des noms de personne et de lieux (entités nommées)
- Cas de la recherche d'information par mots clés
 - Retrouve qu'une partie des extraits pertinents
 - Si le mot est mal transcrit, l'extrait n'est pas retrouvé
 - (Propose les extraits non pertinents)
- Les erreurs de la machine semblent évidentes pour un humain
- Accepter de ne pas être exhaustifs



“Qui figure dans la vidéo ?”

Les archives nécessitent de annotations

- Annotations manuelles ?
 - Coûteux et non scalable
- Reconnaissance automatique ?
 - Comment obtenir les images de référence ?



Just Google it!

- Le web contient une quantité énorme de photos de personnes connues (célébrités).
- Pour d'autres tâches de recherche, le web est déjà utilisé comme source d'information
 - par exemple, en linguistique informatique, dans les systèmes de recommandation, ...

The screenshot shows a Google search for "eisenhower". The search bar contains the text "eisenhower". Below the search bar, there are navigation options: "Tutti", "Immagini" (highlighted), "Notizie", "Video", "Maps", "Altro", "Impostazioni", "Strumenti", and "Ra". Below these options, there are several image thumbnails with captions:

- Thumbnail 1: "mamie" (caption: mamie)
- Thumbnail 2: "34th president" (caption: 34th president)
- Thumbnail 3: "dwight d" (caption: dwight d)
- Thumbnail 4: "matrice" (caption: matrice)
- Thumbnail 5: "mamie eisenhower" (caption: mamie eisenhower)

Below these thumbnails, there are several image results with captions:

- Image 1: Dwight D. Eisenhower ... fr.wikipedia.org
- Image 2: Présidence de Dwight ... fr.wikipedia.org
- Image 3: Dwight David Eisenho... wikimanche.fr
- Image 4: Dwight D. Eisenhower ... fr.wikipedia.org
- Image 5: https://static.lexpress.fr/medias... lexpress.fr
- Image 6: Dwight Eisenhower Photos et image... gettyimages.fr
- Image 7: Fichier:General Dwight... fr.m.wikipedia.org
- Image 8: Eisenhower : un Général po... perfony.com
- Image 9: Comment prioriser vos tâches avec la Matric... esprit-dentreprenre.com

At the bottom, there are more image thumbnails, including one with the number "2" in the bottom right corner.



images du web

ajustement par des experts humains



annotation des visages

Eisenhower

FRAMES 18-29

BOX [11,23, 200,
176]

CONFIDENCE 0.76

Presidents de la republique



Nom	N. Apparitions	Plan (en moyenne)	Durée (en moyenne)
Coty , René	14	14 068 (plan taille)	6.7
Auriol , Vincent	8	2 861 (plan d'ensemble)	3.2

Presidents du conseil



Nom	N. Apparitions	Plan (en moyenne)	Durée (en moyenne)
Ramadier , Paul	5	32 861 gros plan	2.8
Schuman , Robert	13	20 447 plan poitrine	7.3
Queuille , Henri	3	11 420 plan taille	1.6
Bidault , Georges	6	17 383 plan taille	6.6
Pleven , René	6	12 006 plan taille	7.0
Pinay , Antoine	5	13 969 plan taille	5.6
Mayer , René	3	11 917 plan taille	2.6
Laniel , Joseph	3	23 766 plan poitrine	2.6
Mendès France , Pierre	18	26 697 plan poitrine	4.8
Mollet , Guy	8	9 975 plan d'ensemble	11.6
Gaillard , Félix	3	3 921 plan d'ensemble	2.3
Pflimlin , Pierre	2	5 696 plan d'ensemble	2.0
Gaulle , Charles de	7	17 814 plan taille	2.7

Presidents du conseil



Nom	1946	1947	1948	1949	1950
Schuman , Robert		3	5	1	2
Mendès France , Pierre					
Mollet , Guy					
Gaulle , Charles de	1	3			

Nom	1951	1952	1953	1954	1955	1956
Schuman , Robert	1	1				
Mendès France , Pierre	1			14	1	2
Mollet , Guy						8
Gaulle , Charles de	1	1			1	

On cherche des syndicalistes



1947-11-27 "Léon Blum tente de dénouer la crise" (Louis Jouhaux)



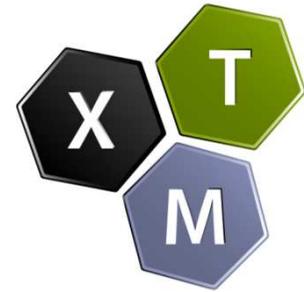
1 apparition en 1117 videos (plus de 180 heures)



a group of people walking down a street



Exploration des AF avec TXM



- Le logiciel open-source TXM permet l'analyse de données textuelles selon l'approche **textométrique** (statistique mais aussi qualitative)
 - <http://textometrie.org>
- Cette expérimentation a reposé sur de nouveaux développements au sein de la plateforme TXM, conçus et implémentés par **Serge Heiden** et **Matthieu Decorde**, membres du projet Antract.
- L'exposé sera en deux temps :
 - Comment avons-nous présenté nos données dans TXM (le texte, les catégories de plan, l'annotation par requêtes)
 - Quels types d'analyses pouvons-nous effectuer (3 exemples d'observations)

Vue du sujet « Rugby » dans TXM

The screenshot displays the TXM software interface with three main panes:

- Left Pane (Structure):** A tree view showing the document's structure with categories like 'word' and 'plan'. The 'plan' column lists various categories such as 10PG, 20PLON, and UNDEF.
- Center Pane (Document View):** Shows the document content with annotations. The text is organized into sections: 'TITRE PROPRE' (La tournée des Springboks en France...), 'RÉSUMÉ' (Résumé du second test-match...), and 'SÉQUENCES'. Words are annotated with categories like 'VG' (highlighted in red) and 'GP'. A search bar at the bottom of this pane shows the query: `frpos=NOM, frlemma=plan, n=1014, ref=1968-11-20, AFE86001312, plan=00DP`.
- Right Pane (SÉQUENCES):** A list of sequence identifiers such as 10PG 20PLON, 10PG 00DP, 14GP 31PANO, etc.

At the bottom, a console window displays log messages, including: 'Ouverture du navigateur de AF-NOTICES-V3-2021-09-30...' and 'Concordance de <[_div_identifiant-de-la-notice="AFE86001312" & plan!="_UNDEF_[01HS"]> dans le corpus AF-NOTICES-V3-2021-09-30... 55 occurrences.'

Lecture verticale du texte, avec les annotations pour chaque mot

Le texte de la notice documentaire, les mots annotés par une catégorie cinématographique sont surlignés

On en dérive un texte réduit aux seules annotations (projection)

Les catégories d'analyse

The screenshot shows the TXM software interface with several panes. The main pane displays search results for the query '[plan="10PG"]'. The results are organized into columns: 'ref', 'Pivot', and 'Contexte droit'. The console window at the bottom shows the search process and results for various queries.

plan	Fréquence
10PG	13
14GP	13
70TI	10
00DP	9
12PM	6
31PANO	3
20PLOW	1

plan	Fréquence
10PG	23846
14GP	16497
00DP	15255
32PP	10530
12PM	9722
31PANO	7302
13PR	2969
70TI	1983
01HS	1668
40VA	1602
20PLOW	1092
30TRAV	506
60ZOOM	467
11PL	230
50VE	212
21CPL	168
51VI	64
71GR	41
15TGP	38
72FLOU	25

word	Fréquence
VG	21712
Vue	861
PG	636
vue	198
panorama	98
VE	76
Vues	49
Plan	39
vues	37
Panorama	29
PE	27
-VG	23
GVG	21
plans	7
vg	7
plan	6
PANORAMA	5
VG	4
CG	3
-PG	2
-Vue	2
Pa	1

Console output:

```

Surtit standard
Index de <[_div identifiant-de-la-notice="AFE86001312" & plan!="UNDEF_"]>, propriété @plan, dans le corpus AF-NOTICES-V3-2021-09-30...
7 items pour 55 occurrences.
Index de <[plan!="UNDEF_"]>, propriété @plan, dans le corpus AF-NOTICES-V3-2021-09-30...
10 items pour 94 217 occurrences.
Index de <[plan="10PG"]>, propriété @word, dans le corpus AF-NOTICES-V3-2021-09-30...
25 items pour 23 846 occurrences.
Concordance de <[plan="10PG"]> dans le corpus AF-NOTICES-V3-2021-09-30...
23 846 occurrences.
Concordance de <[plan="01HS"]> dans le corpus AF-NOTICES-V3-2021-09-30...
1 666 occurrences.
    
```

Dans la notice Rugby

Dans tout le corpus

Les mots sous la catégorie Plan général...

... avec leur contexte (on voit toute l'expression qui justifie l'annotation)

Cas de la catégorie « Hors sujet » (exclue des analyses)

Exemples de requêtes CQL

pour l'annotation automatique des catégories

- Les requêtes pour la catégorie **Plan général (10PG)** :

```
[word="-* (VG|PG|GVG|CG) "%c]
```

```
[word="-* (plan|vues?) "%c] [word="générale?s?" "%c]
```

```
[word="-*PANORAMA "%c]
```

```
[word="-* (PE|VE) "%c]
```

```
[word="-* (plan|vue) s?" "%c] [] [word="ensemble" "%c]
```

- Un exemple pris dans la catégorie **Hors-sujet (01HS)**, illustrant davantage de possibilités du langage :

```
[frlemma="consulter|déplier|discuter|étudier|examiner|montrer|regarder|tracer" "%c] ( [] {0,3}
```

```
[frpos=".*DET.*" "%c] ) ?@ [word="plans?" "%c]
```

- « Le général MAST au milieu d'architectes discutant plans en mains »
- « GP du général MAST examinant un plan sur un tréteau »
- « Salle avec techniciens du repérage, écouteurs aux oreilles, assis autour d'une table, consultant des cartes et des plans »
- « GP du vieux monsieur qui trace un plan »
- « PM du sculpteur traçant des traits sur le plan d'une maison »

Analyse des plans ouvrant un sujet

The screenshot displays the TXM software interface with three main panels and a console window.

Panel 1 (Left): Word frequency list for the query `<{resume}>|<sequences>[]`. The table shows the following data:

word	Fréquence
10PG	4449
70TI	1493
00DP	1454
32PP	808
14GP	691
12PM	484
31PANO	417
40VA	334
13PR	224
20PLON	79
30TRAV	44
50VE	42
11PL	16
60ZOOM	11
51VI	6
15TGP	2
71GR	2
21CPL	1
72FLOU	1

Panel 2 (Middle): Word frequency list for the query `<[]>@w...`. The table shows the following data:

word	Fréquence
10PG	23846
14GP	16497
00DP	15255
32PP	10530
12PM	9722
31PANO	7302
13PR	2969
70TI	1983
40VA	1602
20PLON	1092
30TRAV	506
60ZOOM	467
11PL	230
50VE	212
21CPL	168
51VI	64
71GR	41
15TGP	38
72FLOU	25

Panel 3 (Right): Statistical table of specificities for the sub-corpus `AF-PLANS-V2-2021-09-30/1erplansujet t=10558`. The table shows the following data:

Unités	Fréquence T 92549	AF-PLANS-V2-2021-09-30/1erplansujet t=10558	indice
10PG	23846	4449	1 000,0
70TI	1983	1493	1 000,0
40VA	1602	334	27,3
50VE	212	42	3,6
51VI	64	6	-0,4
72FLOU	25	1	-0,7
15TGP	38	2	-0,8
71GR	41	2	-0,9
30TRAV	506	44	-1,5
11PL	230	16	-1,8
20PLON	1092	79	-5,6
21CPL	168	1	-7,5
13PR	2969	224	-12,1
60ZOOM	467	11	-12,6
00DP	15255	1454	-15,7
32PP	10530	808	-40,6
31PANC	7302	417	-66,8
12PM	9722	484	-118,3
14GP	16497	691	-277,3

Console: The console window shows the execution of the queries and the resulting statistics.

```
Sortie standard
Index de <{resume}>|<sequences>[]>, propriété @word, dans le corpus AF-PLANS-V2-2021-09-30...
19 items pour 10 558 occurrences.
Index de <[]>, propriété @word, dans le corpus AF-PLANS-V2-2021-09-30...
19 items pour 92 549 occurrences.
Sous-corpus 1erplansujet de AF-PLANS-V2-2021-09-30 <{resume}>|<sequences>[]>...
Terminé.
Spécificités du sous-corpus AF-PLANS-V2-2021-09-30/1erplansujet...
Terminé.
```

Les catégories en début de sujet (1^{ère} évoquée)

L'ensemble des catégories (n'importe où : au début ou ailleurs)

Le calcul statistique des SPÉCIFICITÉS met en évidence les sur- et sous-emplois qui peuvent vraiment attirer notre attention

Séquence de plans en début de sujet

A-t-on des constructions-types, récurrentes ?

The screenshot shows the TXM software interface with several panes. On the left, a search results table lists words and their frequencies. The central pane shows a list of search results with columns for 'ref', 'Pivot', and 'Contexte'. The right pane displays the content of a document, including a 'RÉSUMÉ' section and a 'SÉQUENCES' section. A video player is also visible, showing a football match.

word	Fréq...
10PG 10PG 10PG	342
10PG 32PP 32PP	198
70TI 10PG 10PG	175
10PG 10PG 14GP	159
70TI 10PG 32PP	159
10PG 10PG 00DP	157
10PG 00DP 10PG	150
70TI 10PG 00DP	150
10PG 31PANO 10PG	139
10PG 14GP 14GP	135
10PG 32PP 10PG	134
10PG 00DP 00DP	132
10PG 14GP 10PG	129
00DP 00DP 00DP	128
10PG 10PG 12PM	117
10PG 10PG 32PP	115
10PG 12PM 14GP	114
10PG 12PM 12PM	110
32PP 32PP 32PP	110
10PG 12PM 10PG	100
10PG 00DP 14GP	91
10PG 10PG 31PANO	91
14GP 14GP 14GP	73

ref	Pivot	Contexte
1945-02-02, AFE86002966	10PG 12PM 14GP	14GP 00D
1945-02-23, AFE86002993	10PG 12PM 14GP	00DP 10P
1945-03-09, AFE86003009	10PG 12PM 14GP	40VA 10P
1945-04-20, AFE86003061	10PG 12PM 14GP	10PG 00D
1945-04-20, AFE86003061	10PG 12PM 14GP	10PG 00D
1945-05-25, AFE86003107	10PG 12PM 14GP	12PM 10P
1945-08-10, AFE86003213	10PG 12PM 14GP	14GP 10P
1945-08-17, AFE86003224	10PG 12PM 14GP	10PG 00D
1945-08-31, AFE86003230	10PG 12PM 14GP	14GP 14G
1945-09-28, AFE86003262	10PG 12PM 14GP	10PG 10P
1945-12-21, AFE86003380	10PG 12PM 14GP	10PG 10P
1946-06-27, AFE85001535	10PG 12PM 14GP	10PG 50V
1955-11-01, AFE85006441	10PG 12PM 14GP	14GP 12P
1955-11-11, AFE85006460	10PG 12PM 14GP	10PG 31P
1955-12-13, AFE85006510	10PG 12PM 14GP	14GP 31P
1956-03-28, AFE85006685	10PG 12PM 14GP	10PG 12P
1956-04-04, AFE85006688	10PG 12PM 14GP	12PM 12P
1956-04-25, AFE85006723	10PG 12PM 14GP	14GP 00D
1956-05-02, AFE85006732	10PG 12PM 14GP	12PM 14G
1956-05-02, AFE85006736	10PG 12PM 14GP	14GP 12P
1956-05-09, AFE85006744	10PG 12PM 14GP	10PG 14G
1956-05-23, AFE85006765	10PG 12PM 14GP	10PG 14G
1956-06-06, AFE85006795	10PG 12PM 14GP	10PG 12P

RÉSUMÉ

Aux Etats-Unis, la finale inter-armes de football américain oppose l'armée à la marine. Dans les tribunes deux supporters importants : le Général MARSHALL pour l'armée ; l'Amiral LEAHY pour la Navy.

SÉQUENCES

- Vue générale d'une grande parade militaire sur un stade
- Nombreux militaires montant en courant les gradins du stade
- Plan moyen de soldats maintenant les animaux mascottes
- Foule de militaires de la Marine américaine, acclamant leurs collègues
- Gros plan du général MARSHALL parmi les spectateurs du côté de l'armée

MATCH DE FOOTBALL AMÉRICAIN : ARMÉE US

02/02/1945

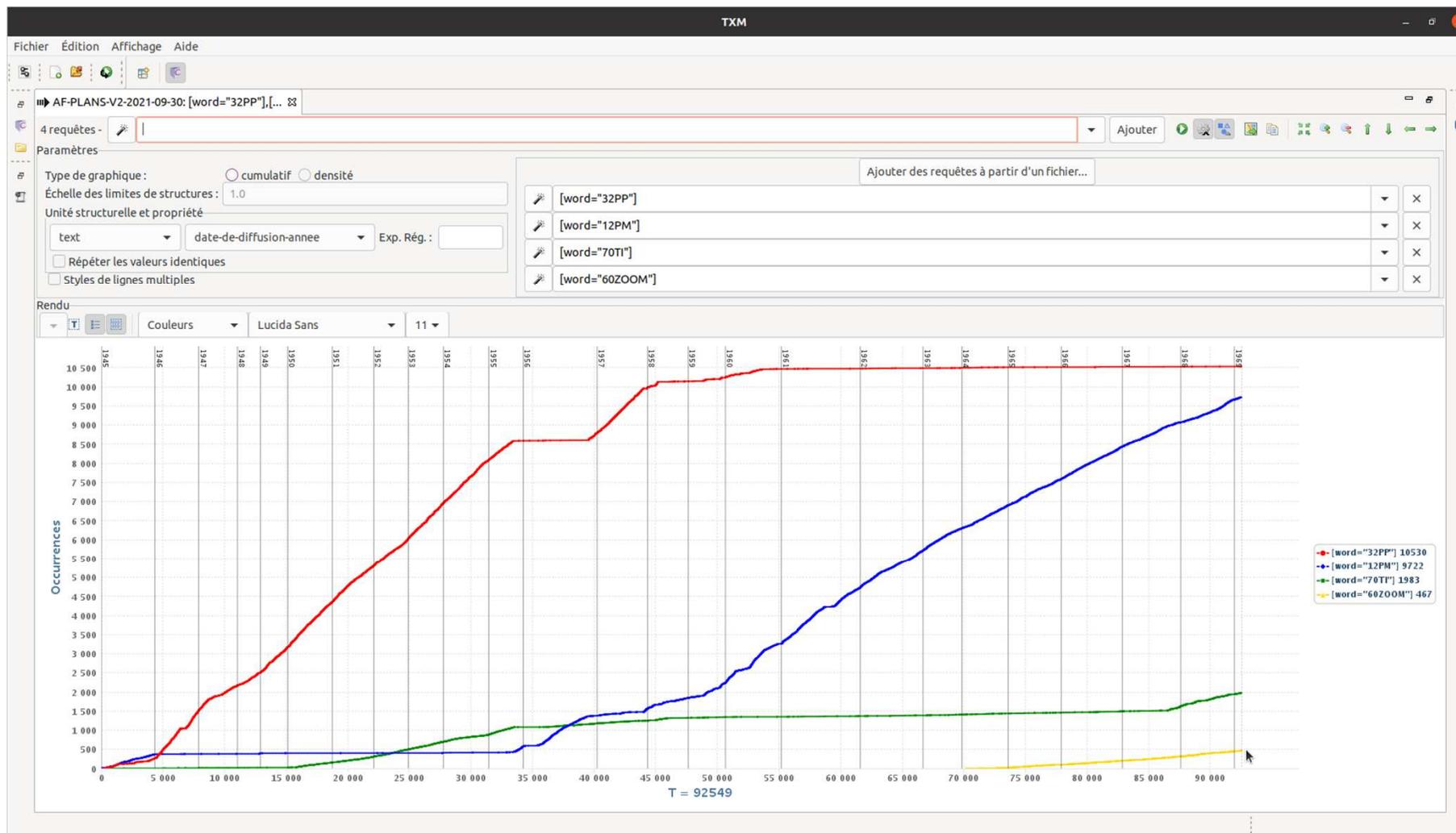
Aux états-unis pour la finale Inter armes de football

Les séquences les plus fréquentes en début de sujet

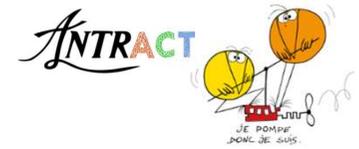
Nous pouvons lister tous les sujets débutant par l'« entonnoir » Général-Moyen-Gros

Pour chaque sujet (ici le 1^{er}), nous pouvons visualiser comment cet « entonnoir » Général-Moyen-Gros se traduit : description documentaire, commentaire audio, vidéo

Visualisation au fil du temps



- **32PP** (Plan proche, rouge) et **12PM** (Plan moyen, bleu) montrent des profils complémentaires : Remplacement ? Équivalence ? Consigne de catalogage ?...
- **70TI** (Titres, vert) : des périodes « magazine » où les titres semblent plus à la mode ?
- Apparition de **60ZOOM** (Zoom, jaune) : nouveauté technique



TXM

Fichier Édition Affichage Aide

Requête Propriétés : word

word	Fréquence
travail	498
travaillant	445
Travail	58
travailleurs	49
travaillent	29
travailler	23
travaille	19
travailleur	14
TRAVAIL	7
Travailleurs	5
travaillé	4
travailliste	4
travaillait	2
Travaillant	2
TRAVAILLEURS	2
travaillant	1
travaillaient	1
travaillez	1
Travailliste	1
TRAVAILLISTE	1
TRAVAILLISTES	1
Travillons	1

Requête Propriétés : frlemma

frlemma	Fréquence
travail	563
travailler	508
travailleur	70
travaillant	19
travailliste	7
travaillant	1

Requête "travail.*%c

Cooccurrent	Fréquence	CoFréquence	Indice	Distance moyenne
ouvriers	569	162	178	1,3
au	11660	353	68	1,3
atelier	304	65	63	4,0
Ouvriers	164	52	60	1,0
construction	730	65	38	4,4
sculpteur	80	26	31	2,2
ouvrier	335	41	30	1,7
laboratoire	216	30	23	4,3
usine	681	47	23	4,3
ouvrières	61	18	20	1,2
travail	498	35	17	5,7
chantier	313	28	17	4,0
réfection	17	10	15	2,8
plans	10717	201	14	4,6
bureau	687	36	14	3,1
soudeur	10	8	14	2,4
un	32488	477	13	4,4
dans	17580	287	13	2,9
Plusieurs	3178	83	13	5,0
d'	44355	615	13	4,8
fête	460	28	12	1,1
BACON	13	8	12	2,6
montage	109	15	12	3,9
ouvrière	48	11	11	2,7

1 t1168, v 22, fmin 1, fmax 498

1 t1168, v 6, fmin 1, fmax 563

t pivot 1168, v cooc 441, t cooc 9225, T corpus 2168048

Console

Sortie standard

```
Index de <"travail.*%c">, propriété @word, dans le corpus AF-NOTICES-V3-2021-09-30...
22 items pour 1 168 occurrences.
Index de <"travail.*%c">, propriété @frlemma, dans le corpus AF-NOTICES-V3-2021-09-30...
6 items pour 1 168 occurrences.
Cooccurrents de <"travail.*%c">, propriété @word 9 9, >=2 >=2.0, dans le corpus AF-NOTICES-V3-2021-09-30...
441 cooccurrents pour 1 168 occurrences du pivot.
```





TXM

Fichier Édition Affichage Aide

AF-NOTICES-V3-2021-09-30/<{frpos="NOM}|... 412

Requête `al("travaillant")` Propriétés : word Éditer

word	Fréquence
ouvriers travaillant	51
Ouvriers travaillant	30
ouvriers au travail	23
ouvrier travaillant	12
photographes au travail	7
ouvrières au travail	6
Ouvriers au travail	6
sculpteur travaillant	6
PM photographes au travail	5
Marie CURIE travaillant	4
atelier au travail	3
cameraman au travail	3
cinéastes au travail	3
enfants travaillant	3
femmes travaillant	3
femme travaillant	3
filles travaillant	3
gens travaillant	3
homme travaillant	3
mains travaillant	3
Photographes au travail	3
sculpteurs travaillant	3
soudeur au travail	3
techniciens travaillant	3
aidés travaillant	2

t604, v 412, fmin 1, fmax 51

AF-NOTICES-V3-2021-09-30/<{word="travail... 408

Requête `ER:ppre"?[frpos="NOM"]` Propriétés : word Éditer

word	Fréquence
travaillant à la construction	16
travaillant à la reconstruction	10
travaillant à la réfection	7
travaillant dans les champs	7
travaillant dans son atelier	7
travaillant dans un champ	7
travaillant à la fabrication	6
travaillant à la mise	6
travail dans un atelier	5
travaillant aux fouilles	5
travaillant la pierre	5
travaillant sur une machine	5
travail dans une usine	4
travaillant au sommet	4
travaillant dans un laboratoire	4
travaillant le taureau	4
travaillant sur un chantier	4
travaillant sur un échafaudage	4
travail de déblaiement	3
travaillant à l'achèvement	3
travaillant à l'extraction	3
travaillant à la chaîne	3
travaillant au montage	3
travaillant au sol	3
travaillant dans son bureau	2

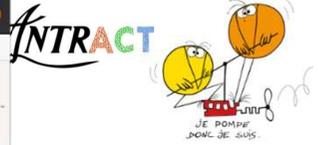
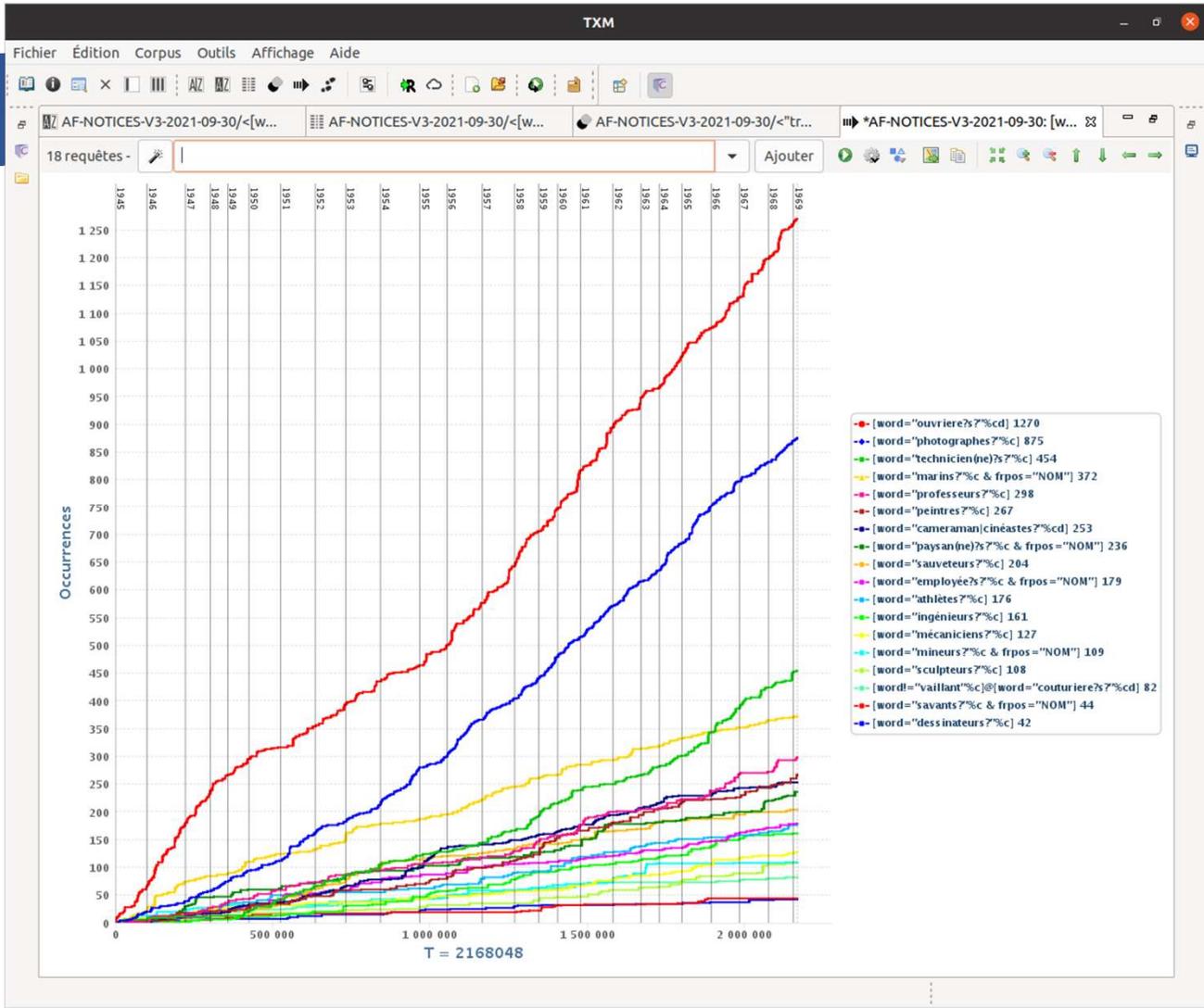
t566, v 408, fmin 1, fmax 16

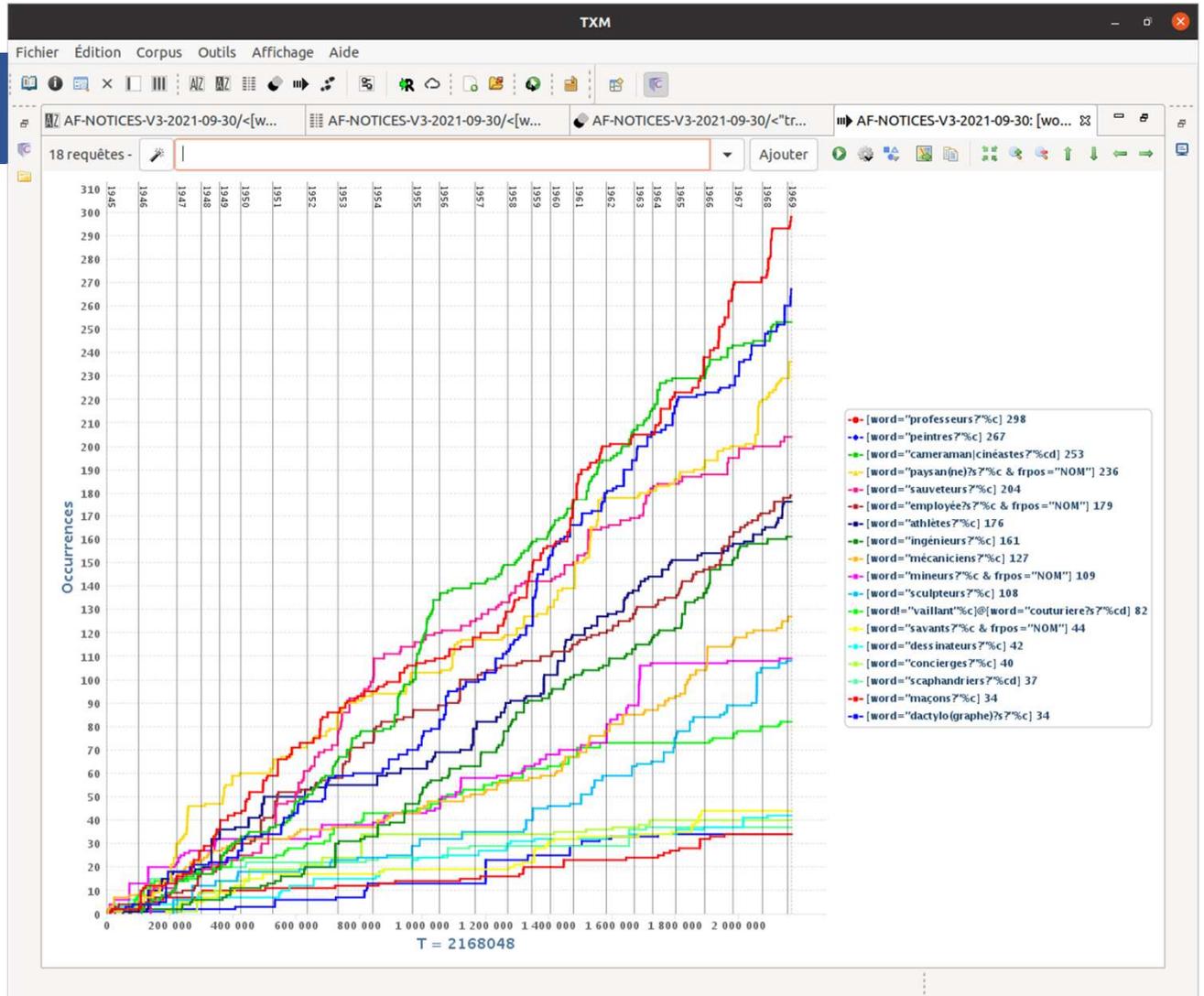
Console

Sortie standard

```
Index de <{frpos="NOM|NAM" & word!="plans"?|vues?|V[GAPEMFR]|P[PMGRALE]|D[PV]|PV|SG|ZA[VR]|GVG|GPP?|CG|TGP"}[frpos="PRP.*|DET.*|ADJ|VER:pper"]{1,3}>
412 items pour 604 occurrences.
Index de <{word="travail.*%c}[frpos="PRP.*"?[frpos="DET.*"?[frpos="ADJ|VER:pper"?[frpos="NOM"]}>, propriété @word, dans le corpus AF-NOTICES-V3-20
408 items pour 566 occurrences.
```







TXM

Fichier Édition Affichage Aide

AF-VOIX-OFF-V4-2021-05-19/<[frpos="NOM"]... [Requête: ')]][frlemma="travailler"] Propriétés: word Éditer

word	Fréquence
ouvriers travaillent	4
France travaille	2
ouvriers travaillant	2
ouvriers travaille	2
acier , travaillé	1
activité au travail	1
ancienne qui travaillait	1
années au travail	1
années fait travailler	1
artisans travaille	1
billancourt travaille	1
bosse en travaillant	1
chatou Travaille	1
chorégraphe Travaille	1
colomb béchar expérimente travaille	1
cuisines ils travaillent	1
direction qui travaille	1
distance puis travaille	1
domicile au travail	1
écoliers au travail	1
ennemi travaillent	1
envie de travailler	1
équipes étaient au travail	1
état au travail	1
Europe qui travaillent	1

t111, v 105, fmin 1, fmax 4

AF-VOIX-OFF-V4-2021-05-19/<[word="trava... [Requête: ER:ppre"]?[frpos="NOM"] Propriétés: word Éditer

word	Fréquence
travail d' hercule	5
travail à la chaîne	2
travail au fond	2
travail d' une équipe	2
travail de patience	2
travail de sape	2
travail des champs	2
travail grâce	2
travailler sans relâche	2
travail à l' oeuvre	1
travail à la portée	1
travail à la préfecture	1
travail à notre courage	1
travail au domicile	1
travail au milieu	1
travail à une équipe	1
travail au sol	1
travail aux fellahs	1
travail aux usines	1
travail avec haydn	1
travail avec occupation	1
travail cinquante ans	1
travail d' émulation	1
travail d' équipe	1
travail d' esprit	1

t194, v 182, fmin 1, fmax 5

Console

Sortie standard

Index de <[frpos="NOM|NAM"]{1,2}]?(((word="au")[word="travail"])[frlemma="travailler"]>, propriété @word, dans le corpus AF-VOIX-OFF-V4-2021-05-19
105 items pour 111 occurrences.

Index de <[word="travail.*%c][frpos="PRP.*"]?[frpos="DET.*"]?[frpos="ADJ|VER:ppre"]?[frpos="NOM"]>, propriété @word, dans le corpus AF-VOIX-OFF-V4-2
182 items pour 194 occurrences.

