The Textometric Concept of Active Corpus. Illustration by an Analysis Scenario based on Annotation then Projection.

Bénédicte Pincemin[1], Serge Heiden[2], Franck Mazuet[3]

[1]University of Lyon, CNRS, IHRIM UMR5317 – benedicte.pincemin *at* ens-lyon.fr

[2] University of Lyon, ENS Lyon, IHRIM UMR5317 – slh *at* ens-lyon.fr

[3]University of Paris 1, CHS UMR8058 – fmazuet *at* gmail.com

APPENDIX
Supplementary materials

Publication deposit URL in the HAL-SHS archive:
https://halshs.archives-ouvertes.fr/halshs-03667319

Contents

1. INA descriptive forms for the *Actualités françaises* films

1.1. Source file: an xlsx spreadsheet

Every newsreel and every news report are described in a form.

A form content is delivered as a row, and fields are coded by columns: thus, a cell records the value of the column field for the newsreel or news report row.

99 fields are available. However, some are redundant (several layouts for the same piece of information), some are constant throughout the entire *Actualités françaises* collection, and some do not seem relevant for our research. The AF-NOTICES corpus exploits the following 23 fields:

- Date de diffusion
- **Descripteurs (Aff. Lig.)**
- Durée
- **Générique (Aff. Lig.)**
- Genre
- Identifiant de la notice
- Langue VO / VE
- Lien notice principale
- Nature de production
- Nom fichier segmenté (info)
- Notes du titre
- Producteurs (Aff.)
- **Résumé**
- **Séquences**
- Thématique
- **Titre propre**
- Type de date
- Type de notice
- antract_debut
- antract_fin
- antract_duree
- antract_tc_date
- antract_tc_type

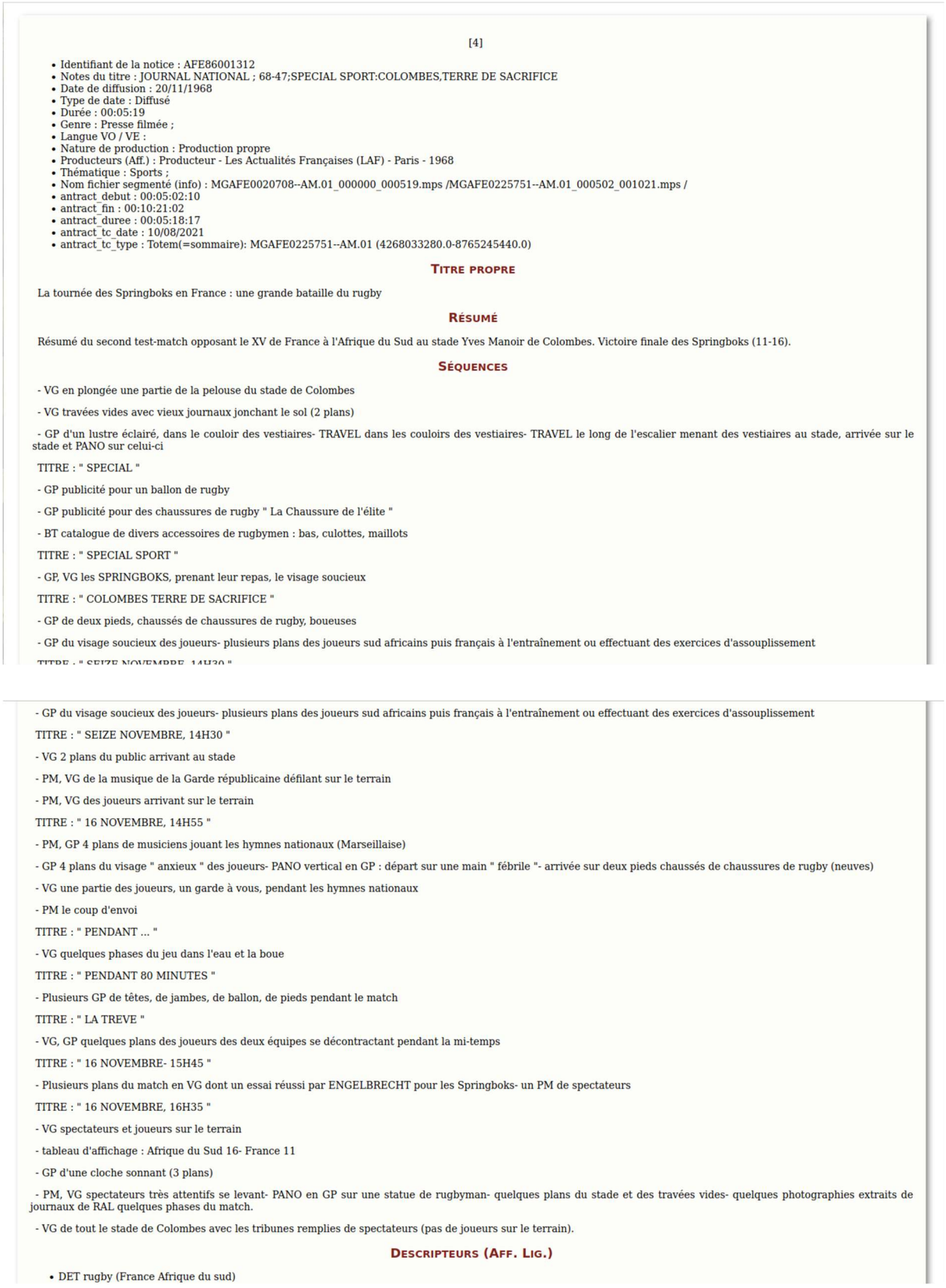
Bolded values are used as text content for textometric analysis, and other values are used as metadata.

The workflow in order to convert the xlsx source file into a TXM corpus is documented in the txm-users wiki:

https://groupes.renater.fr/wiki/txm-users/public/antract/antract_corpus_notices

1.2. Example of a documentary form

Here is the content of the form for the AFE86001312 news report, as shown in TXM:



[4]

- Identifiant de la notice : AFE86001312
- Notes du titre : JOURNAL NATIONAL ; 68-47;SPECIAL SPORT:COLOMBES,TERRE DE SACRIFICE
- Date de diffusion : 20/11/1968
- Type de date : Diffusé
- Durée : 00:05:19
- Genre : Presse filmée ;
- Langue VO / VE :
- Nature de production : Production propre
- Producteurs (Aff.) : Producteur - Les Actualités Françaises (LAF) - Paris - 1968
- Thématique : Sports ;
- Nom fichier segmenté (info) : MGAFE0020708--AM.01_000000_000519.mps /MGAFE0225751--AM.01_000502_001021.mps /
- antract_debut : 00:05:02:10
- antract_fin : 00:10:21:02
- antract_duree : 00:05:18:17
- antract_tc_date : 10/08/2021
- antract_tc_type : Totem(=sommaire): MGAFE0225751--AM.01 (4268033280.0-8765245440.0)

TITRE PROPRE

La tournée des Springboks en France : une grande bataille du rugby

RÉSUMÉ

Résumé du second test-match opposant le XV de France à l'Afrique du Sud au stade Yves Manoir de Colombes. Victoire finale des Springboks (11-16).

SÉQUENCES

- VG en plongée une partie de la pelouse du stade de Colombes

- VG travées vides avec vieux journaux jonchant le sol (2 plans)

- GP d'un lustre éclairé, dans le couloir des vestiaires- TRAVEL dans les couloirs des vestiaires- TRAVEL le long de l'escalier menant des vestiaires au stade, arrivée sur le stade et PANO sur celui-ci

TITRE : " SPECIAL "

- GP publicité pour un ballon de rugby

- GP publicité pour des chaussures de rugby " La Chaussure de l'élite "

- BT catalogue de divers accessoires de rugbymen : bas, culottes, maillots

TITRE : " SPECIAL SPORT "

- GP, VG les SPRINGBOKS, prenant leur repas, le visage soucieux

TITRE : " COLOMBES TERRE DE SACRIFICE "

- GP de deux pieds, chaussés de chaussures de rugby, boueuses

- GP du visage soucieux des joueurs- plusieurs plans des joueurs sud africains puis français à l'entraînement ou effectuant des exercices d'assouplissement

TITRE : " SEIZE NOVEMBRE, 14H30 "

- VG 2 plans du public arrivant au stade

- PM, VG de la musique de la Garde républicaine défilant sur le terrain

- PM, VG des joueurs arrivant sur le terrain

TITRE : " 16 NOVEMBRE, 14H55 "

- PM, GP 4 plans de musiciens jouant les hymnes nationaux (Marseillaise)

- GP 4 plans du visage " anxieux " des joueurs- PANO vertical en GP : départ sur une main " fébrile "- arrivée sur deux pieds chaussés de chaussures de rugby (neuves)

- VG une partie des joueurs, un garde à vous, pendant les hymnes nationaux

- PM le coup d'envoi

TITRE : " PENDANT ... "

- VG quelques phases du jeu dans l'eau et la boue

TITRE : " PENDANT 80 MINUTES "

- Plusieurs GP de têtes, de jambes, de ballon, de pieds pendant le match

TITRE : " LA TREVE "

- VG, GP quelques plans des joueurs des deux équipes se décontractant pendant la mi-temps

TITRE : " 16 NOVEMBRE- 15H45 "

- Plusieurs plans du match en VG dont un essai réussi par ENGELBRECHT pour les Springboks- un PM de spectateurs

TITRE : " 16 NOVEMBRE, 16H35 "

- VG spectateurs et joueurs sur le terrain

- tableau d'affichage : Afrique du Sud 16- France 11

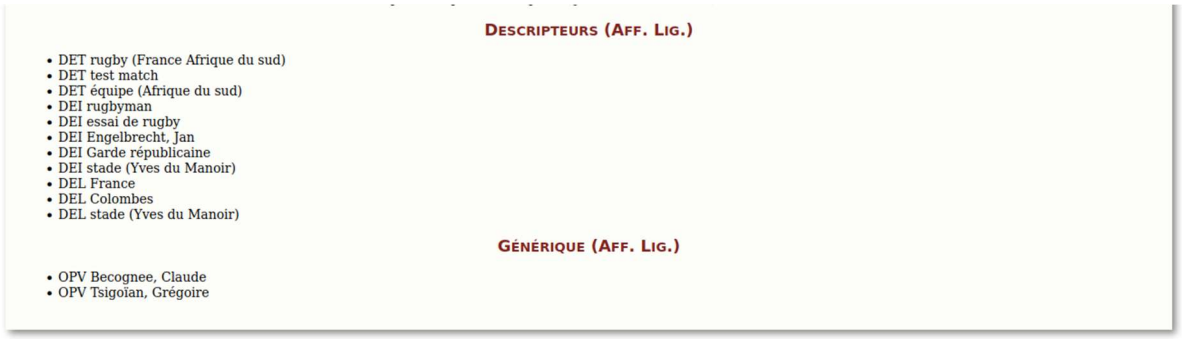
- GP d'une cloche sonnant (3 plans)

- PM, VG spectateurs très attentifs se levant- PANO en GP sur une statue de rugbyman- quelques plans du stade et des travées vides- quelques photographies extraits de journaux de RAL quelques phases du match.

- VG de tout le stade de Colombes avec les tribunes remplies de spectateurs (pas de joueurs sur le terrain).

DESCRIPTEURS (AFF. LIG.)

- DET rugby (France Afrique du sud)



DESCRIPTEURS (AFF. LIG.)

- DET rugby (France Afrique du sud)
- DET test match
- DET équipe (Afrique du sud)
- DEI rugbyman
- DEI essai de rugby
- DEI Engelbrecht, Jan
- DEI Garde républicaine
- DEI stade (Yves du Manoir)
- DEL France
- DEL Colombes
- DEL stade (Yves du Manoir)

GÉNÉRIQUE (AFF. LIG.)

- OPV Becognee, Claude
- OPV Tsigoïan, Grégoire

2. Corpus quantitative description

2.1. Counts for AF-NOTICES-V3-2021-10-11 corpus (documentary descriptive forms)

Number of texts (newsreels):	1,261	
Number of news reports (topics that compose the newsreels):	10,783	
Number of tokens (words and punctuations):	2,234,187	
Number of words (tokens without punctuations):	1,888,759	
Number of unique words (types):	78,825	
Number and percentage of hapax (words with frequency=1):	29,691	(38 %)
Maximum frequency for a word:	132,928	(*de*)
Number of unique lemmas (lemma types):	46,848	
Number and percentage of hapax (lemmas with frequency=1):	19,082	(41 %)
Maximum frequency for a lemma:	183,418	(*de*)

Comment

The low proportion of hapax could be explained by the fact that some fields that we consider as textual data are lists of controlled descriptors, taken from the INA Thesaurus. Controlled vocabulary actually aims at avoiding singular uses and favoring repetition. If we focus on *Titre propre*, *Résumé* and *Séquences* sections, and exclude *Descripteurs* and *Générique* sections, hapax rate is 28,950 / 66,697 = 43 % for lexical forms, 18364 / 43167 = 43 % for lemmas.

2.2. Counts for AF-PLANS-V2-2021-10-11 corpus (projection of camera shot tags)

Number of texts (newsreels):	1,259
Number of news reports (topics that compose the newsreels):	10,303

Note: Counts for newsreel and news reports are lower in AF-PLANS than in AF-NOTICES, because a news report (or a newsreel) without any identified mention for camera shot or angle type cannot be present in AF-PLANS corpus as it contains no word.

Number of tokens (words and punctuations):	91,898	
Number of words (tokens without punctuations):	91,898	

Note: No punctuation in the AF-PLANS corpus (punctuation was not projected).

Number of unique words (types):	20	
Number and percentage of hapax (words with frequency=1):	0	(0 %)
Maximum frequency for a word:	23,721	(*10PG*)

Note: No lemmatization in the AF-PLANS corpus (words are not lexical words but codes for categories).

3. Focus on an annotated excerpt

We still consider the first two paragraphs for the AFE86001312 news report.

3.1. Non-technical view

Tags are appended to words and enclosed in curly brackets.

- VG_{10PG} en plongée\{20PLON} une partie de la pelouse du stade de Colombes
- VG_{10PG} travées vides avec vieux journaux jonchant le sol (2 plans_{00DP})

3.2. XML view from TXM pivot format

```
<p>
<w id="w_AFE86004168_854"><txm:form>-</txm:form><txm:ana resp="#txm"
type="#frpos">PUN</txm:ana><txm:ana resp="#txm" type="#frlemma">-
</txm:ana><txm:ana resp="#src" type="#n">854</txm:ana><txm:ana
resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>
<w id="w_AFE86004168_855"><txm:form>VG</txm:form><txm:ana
resp="#txm" type="#frpos">NOM</txm:ana><txm:ana resp="#txm"
type="#frlemma">VG</txm:ana><txm:ana resp="#src"
type="#n">855</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20,
AFE86001312</txm:ana><ana type="#plan"
resp="#bpincemi">10PG</ana></w>
<w id="w_AFE86004168_856"><txm:form>en</txm:form><txm:ana
resp="#txm" type="#frpos">PRP</txm:ana><txm:ana resp="#txm"
type="#frlemma">en</txm:ana><txm:ana resp="#src"
type="#n">856</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20,
AFE86001312</txm:ana></w>
<w id="w_AFE86004168_857"><txm:form>plongée</txm:form><txm:ana
resp="#txm" type="#frpos">NOM</txm:ana><txm:ana resp="#txm"
```

type="#frlemma">plongée</txm:ana><txm:ana resp="#src" type="#n">857</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana><ana type="#plan" resp="#bpincemi">**20PLON**</ana></w>

<w id="w_AFE86004168_858"><txm:form>**une**</txm:form><txm:ana resp="#txm" type="#frpos">DET:ART</txm:ana><txm:ana resp="#txm" type="#frlemma">un</txm:ana><txm:ana resp="#src" type="#n">858</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_859"><txm:form>**partie**</txm:form><txm:ana resp="#txm" type="#frpos">NOM</txm:ana><txm:ana resp="#txm" type="#frlemma">partie</txm:ana><txm:ana resp="#src" type="#n">859</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_860"><txm:form>**de**</txm:form><txm:ana resp="#txm" type="#frpos">PRP</txm:ana><txm:ana resp="#txm" type="#frlemma">de</txm:ana><txm:ana resp="#src" type="#n">860</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_861"><txm:form>**la**</txm:form><txm:ana resp="#txm" type="#frpos">DET:ART</txm:ana><txm:ana resp="#txm" type="#frlemma">le</txm:ana><txm:ana resp="#src" type="#n">861</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_862"><txm:form>**pelouse**</txm:form><txm:ana resp="#txm" type="#frpos">NOM</txm:ana><txm:ana resp="#txm" type="#frlemma">pelouse</txm:ana><txm:ana resp="#src" type="#n">862</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_863"><txm:form>**du**</txm:form><txm:ana resp="#txm" type="#frpos">PRP:det</txm:ana><txm:ana resp="#txm" type="#frlemma">du</txm:ana><txm:ana resp="#src" type="#n">863</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_864"><txm:form>**stade**</txm:form><txm:ana resp="#txm" type="#frpos">NOM</txm:ana><txm:ana resp="#txm" type="#frlemma">stade</txm:ana><txm:ana resp="#src" type="#n">864</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_865"><txm:form>**de**</txm:form><txm:ana resp="#txm" type="#frpos">PRP</txm:ana><txm:ana resp="#txm" type="#frlemma">de</txm:ana><txm:ana resp="#src" type="#n">865</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_866"><txm:form>**Colombes**</txm:form><txm:ana resp="#txm" type="#frpos">NOM</txm:ana><txm:ana resp="#txm" type="#frlemma">colombe</txm:ana><txm:ana resp="#src" type="#n">866</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_867"><txm:form>**-**</txm:form><txm:ana resp="#txm" type="#frpos">PUN</txm:ana><txm:ana resp="#txm" type="#frlemma">-</txm:ana><txm:ana resp="#src" type="#n">867</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_868"><txm:form>**VG**</txm:form><txm:ana resp="#txm" type="#frpos">NAM</txm:ana><txm:ana resp="#txm" type="#frlemma">VG</txm:ana><txm:ana resp="#src" type="#n">868</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana><ana type="#plan" resp="#bpincemi">**10PG**</ana></w>

<w id="w_AFE86004168_869"><txm:form>**travées**</txm:form><txm:ana resp="#txm" type="#frpos">NOM</txm:ana><txm:ana resp="#txm" type="#frlemma">travée</txm:ana><txm:ana resp="#src" type="#n">869</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_870"><txm:form>**vides**</txm:form><txm:ana resp="#txm" type="#frpos">ADJ</txm:ana><txm:ana resp="#txm" type="#frlemma">vide</txm:ana><txm:ana resp="#src" type="#n">870</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_871"><txm:form>**avec**</txm:form><txm:ana resp="#txm" type="#frpos">PRP</txm:ana><txm:ana resp="#txm" type="#frlemma">avec</txm:ana><txm:ana resp="#src" type="#n">871</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_872"><txm:form>**vieux**</txm:form><txm:ana resp="#txm" type="#frpos">ADJ</txm:ana><txm:ana resp="#txm" type="#frlemma">vieux</txm:ana><txm:ana resp="#src" type="#n">872</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_873"><txm:form>**journaux**</txm:form><txm:ana resp="#txm" type="#frpos">NOM</txm:ana><txm:ana resp="#txm" type="#frlemma">journal|journau</txm:ana><txm:ana resp="#src" type="#n">873</txm:ana><txm:ana resp="#src" type="#ref">1968-11-20, AFE86001312</txm:ana></w>

<w id="w_AFE86004168_874"><txm:form>**jonchant**</txm:form><txm:ana resp="#txm" type="#frpos">VER:ppre</txm:ana><txm:ana resp="#txm"

```
type="#frlemma">joncher</txm:ana><txm:ana                resp="#src"
type="#n">874</txm:ana><txm:ana  resp="#src"  type="#ref">1968-11-20,
AFE86001312</txm:ana></w>

<w            id="w_AFE86004168_875"><txm:form>le</txm:form><txm:ana
resp="#txm"    type="#frpos">DET:ART</txm:ana><txm:ana    resp="#txm"
type="#frlemma">le</txm:ana><txm:ana                      resp="#src"
type="#n">875</txm:ana><txm:ana  resp="#src"  type="#ref">1968-11-20,
AFE86001312</txm:ana></w>

<w           id="w_AFE86004168_876"><txm:form>sol</txm:form><txm:ana
resp="#txm"      type="#frpos">NOM</txm:ana><txm:ana      resp="#txm"
type="#frlemma">sol</txm:ana><txm:ana                     resp="#src"
type="#n">876</txm:ana><txm:ana  resp="#src"  type="#ref">1968-11-20,
AFE86001312</txm:ana></w>

<w id="w_AFE86004168_877"><txm:form>(</txm:form><txm:ana resp="#txm"
type="#frpos">PUN</txm:ana><txm:ana                       resp="#txm"
type="#frlemma">(</txm:ana><txm:ana                       resp="#src"
type="#n">877</txm:ana><txm:ana  resp="#src"  type="#ref">1968-11-20,
AFE86001312</txm:ana></w>

<w id="w_AFE86004168_878"><txm:form>2</txm:form><txm:ana resp="#txm"
type="#frpos">NUM</txm:ana><txm:ana                       resp="#txm"
type="#frlemma">@card@</txm:ana><txm:ana                  resp="#src"
type="#n">878</txm:ana><txm:ana  resp="#src"  type="#ref">1968-11-20,
AFE86001312</txm:ana></w>

<w          id="w_AFE86004168_879"><txm:form>plans</txm:form><txm:ana
resp="#txm"      type="#frpos">NOM</txm:ana><txm:ana      resp="#txm"
type="#frlemma">plan</txm:ana><txm:ana                    resp="#src"
type="#n">879</txm:ana><txm:ana  resp="#src"  type="#ref">1968-11-20,
AFE86001312</txm:ana><ana                                type="#plan"
resp="#bpincemi">00DP</ana></w>

<w id="w_AFE86004168_880"><txm:form>)</txm:form><txm:ana resp="#txm"
type="#frpos">PUN</txm:ana><txm:ana                       resp="#txm"
type="#frlemma">)</txm:ana><txm:ana                       resp="#src"
type="#n">880</txm:ana><txm:ana  resp="#src"  type="#ref">1968-11-20,
AFE86001312</txm:ana></w>

</p>
```

4. High definition pictures

4.1. High definition version of the figures that are published in the paper

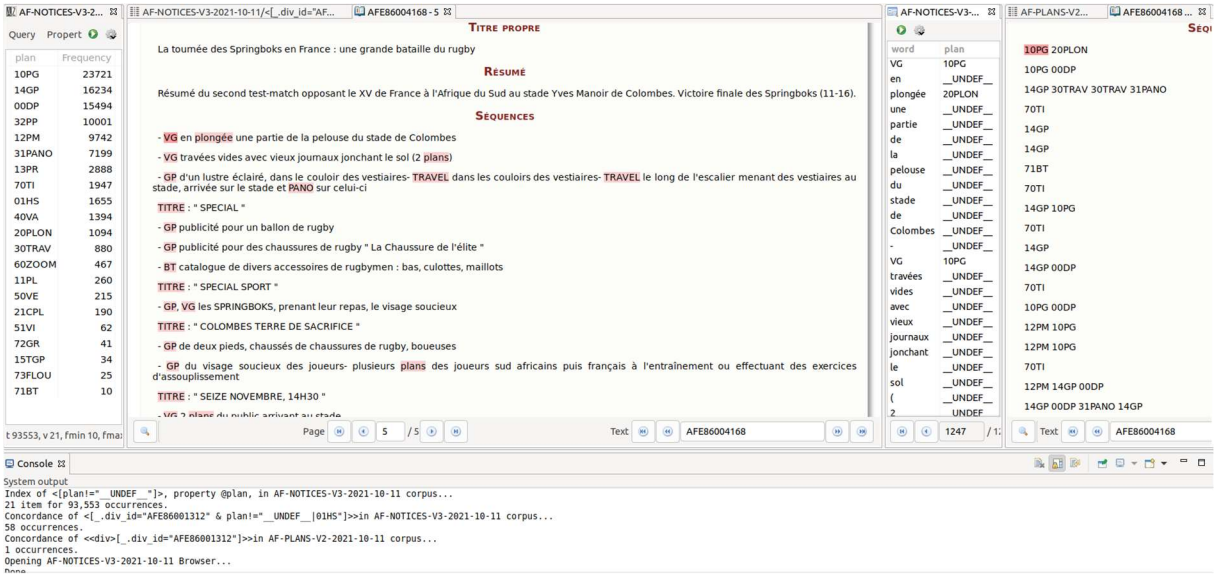


Figure 1. *The* Les Actualités françaises *corpus in TXM, with shot type annotations. From left to right: (i) annotated shot tags and their frequencies in the corpus; (ii) view of a documentary description with all annotated words highlighted; (iii) view of the shot tag assigned to each word of this text; (iv) same text in its projected version. Below: log of executed commands.*

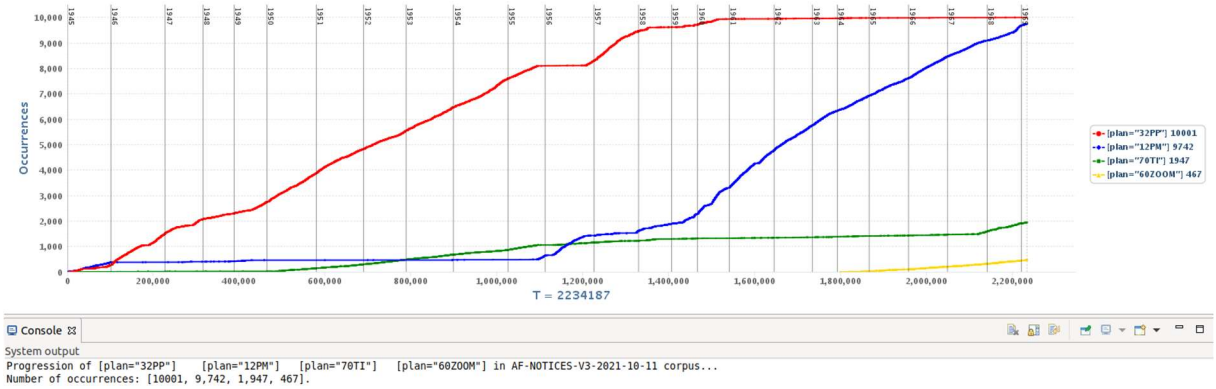


Figure 2. *Progression chart showing the diachronic evolution of 4 shot types. By decreasing frequency: Handheld Shot (red), Medium Shot (blue), Title & Credits (green), Zoom In & Out (yellow)*

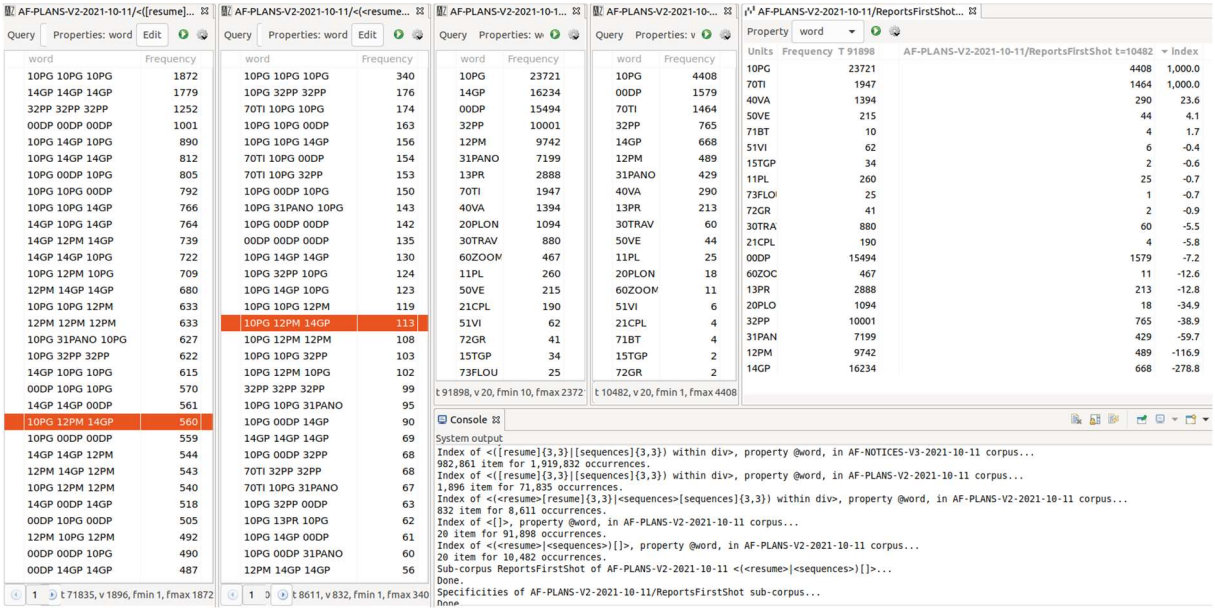


Figure 3. *Study of the film grammar in* Les Actualités françaises.
From left to right: (i) frequency list of 3-shot patterns and (ii) same thing for initial 3-shot patterns only; (iii) frequency list of shots and (iv) same thing for first shot in a report only; (iv) specificity score of shot types for the first position in reports. Bottom right panel: log of executed commands.

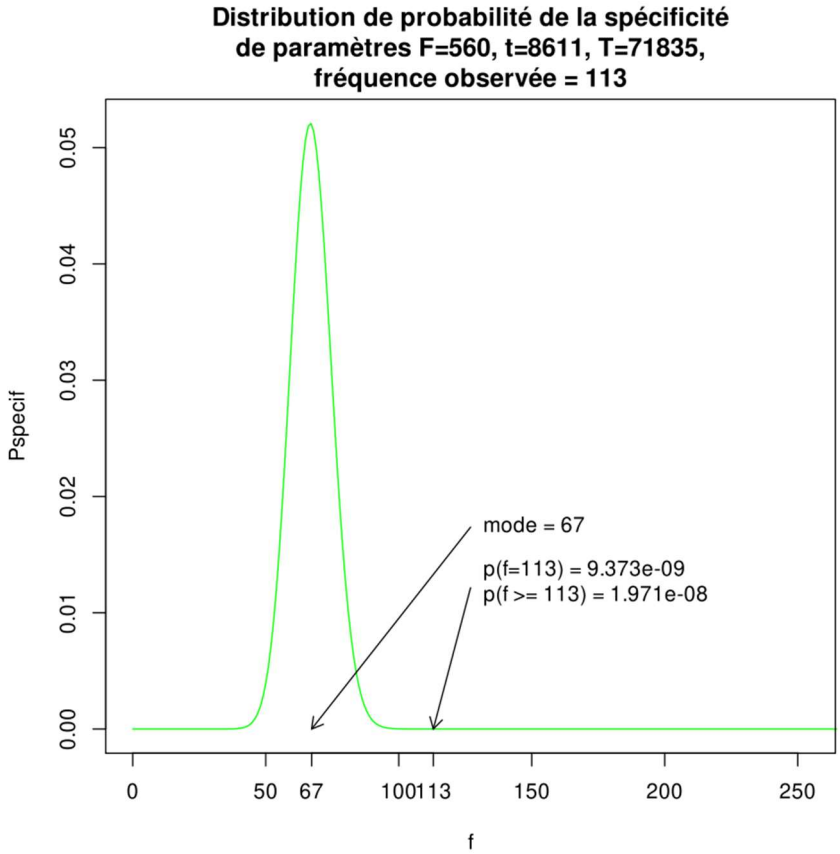


Figure 4. *Specificity result +7.7 for the funnel pattern10GP 12PM 14GP (Extreme Wide Shot, Medium Shot, close-Up) as first shot sequence in a report (TXM PlotSpecif utility output).*

4.2. Supplementary figures

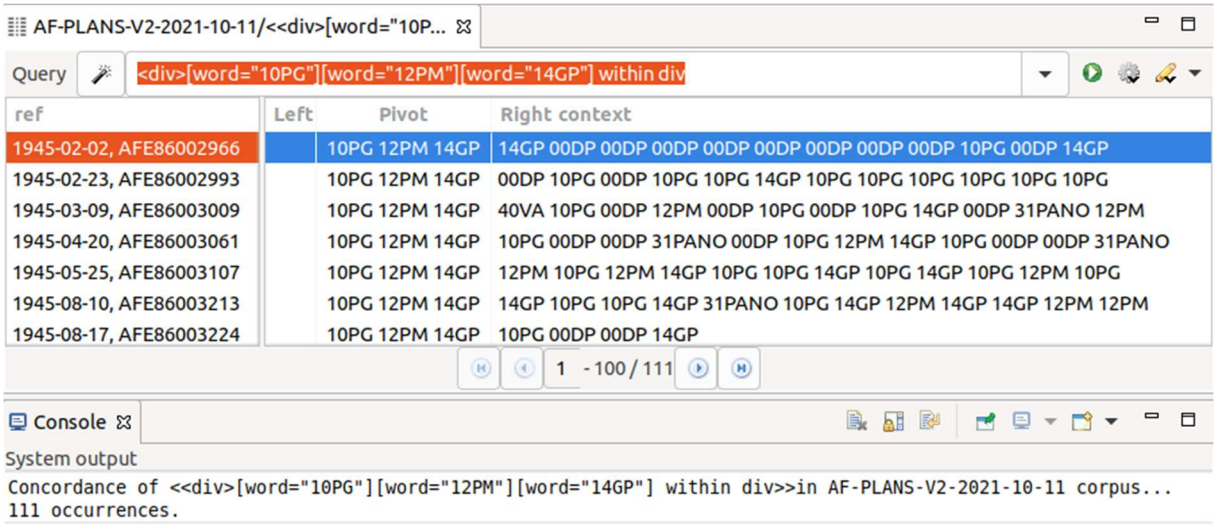


Figure 5a. *Example of analytical path (1 / 3). Concordance in the projected corpus AF-PLANS. Searching for the funnel pattern 10GP 12PM 14GP (Extreme Wide Shot, Medium Shot, close-Up) as first shot sequence in a report.*

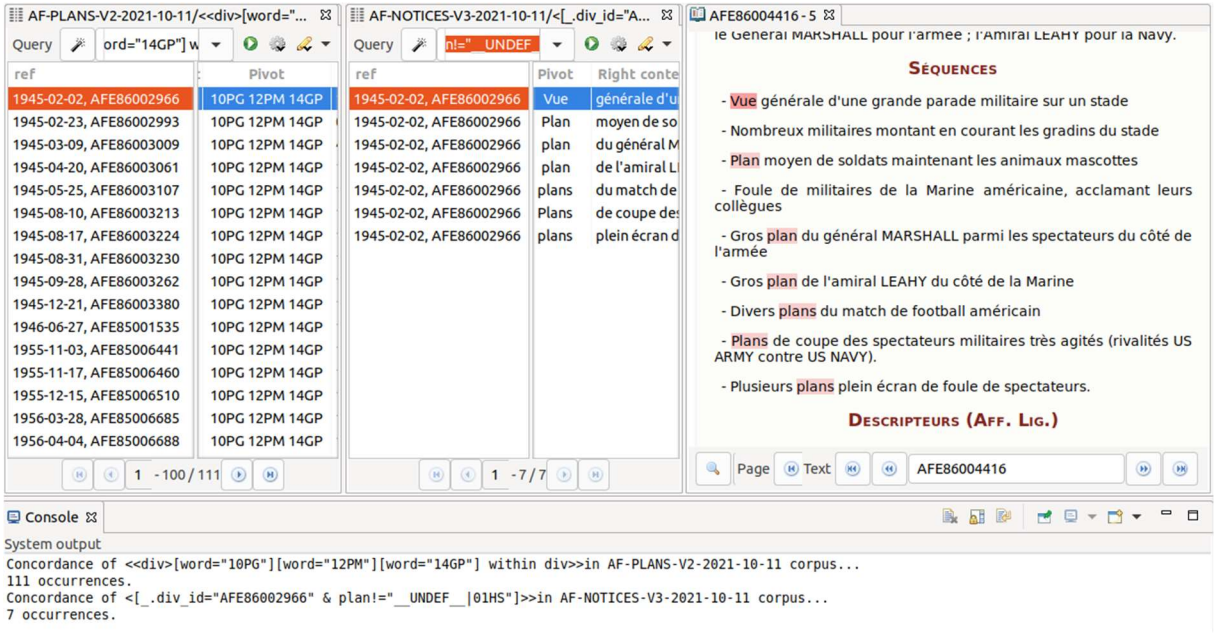


Figure 5b. *Example of analytical path (2 / 3). From the first 10GP 12PM 14GP pattern found, looking back to the context in the corresponding documentary form.*

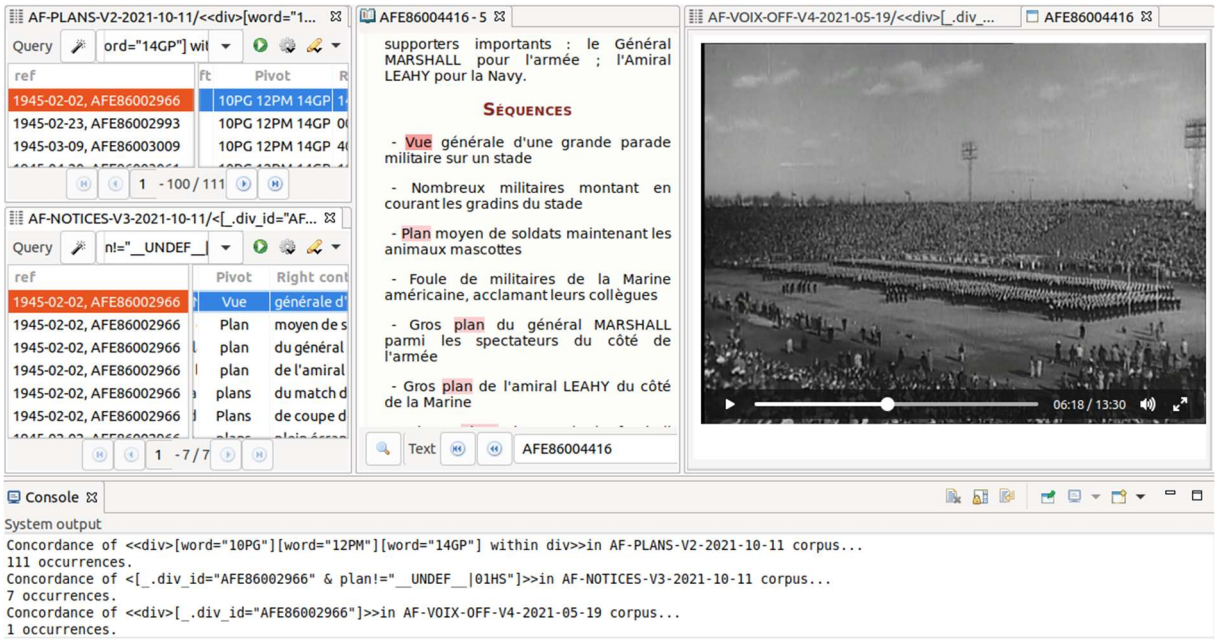


Figure 5c. *Example of analytical path (3 / 3).*
Use of the AF-VOIX-OFF corpus to access the video for the same news report.

Note: Back-to-media feature and AF-VOIX-OFF corpus are described in the following papers:

Carrive J., Beloued A., Goetschel P., Heiden S., Laurent A., Lisena P., Mazuet F., Meignier S., Pincemin B., Poels G., Troncy R. (2021). Transdiciplinary Analysis of a Corpus of French Newsreels: The ANTRACT Project. *Digital Humanities Quarterly*, 15 (1). http://digitalhumanities.org/dhq/vol/15/1/000523/000523.html

Pincemin B., Heiden S., Decorde M. (2020). Textometry on Audiovisual Corpora: Experiments with TXM software. *In* Ratinaud P. and Marchand P., editors, *Proc. of JADT 2020*. Univ. Toulouse 3. http://lexicometrica.univ-paris3.fr/jadt/JADT2020/jadt2020_pdf/PINCEMIN_HEIDEN_DECORDE_JADT2020.pdf

5. Annotation strategy and material

To complete our inventory of all expressions relating to shot types, we listed then checked in context (KWIC view, with context sorting) every word that looked like a shot notation[1], and every phrase based on “plan” (shot) or “vue” (view)[2], as this is a common pattern to shot name.

The annotation was processed sequentially. First, any mention of “plan” or “vue” was tagged as miscellaneous. Then, context-sensitive queries made it possible to recode occurrences into identified types, either one of the 19 shot types or an off-topic type. Additional queries dealt with abbreviated shot names (like “VG” or “PM”) and with other specialized terms (like “ZOOM” or “contre-plongée”). This ordered annotation strategy implied, in phrasal mentions, choosing the word “plan” or “vue” to bear the shot-type annotation[3], whereas the semantic value of the mention was rather provided by the adjective or noun qualifying this noun.

The automatic application of the queries was operated with the CQLList2WordProperties utility. We used an updated version of this utility, with respect to the one released within TXM 0.8.1. Current version is available in TXM sources repository: http://forge.cbp.ens-lyon.fr/svn/txm/TXM/trunk/org.txm.groovy.core/src/groovy/org/txm/macro/annotation/

If necessary, after the automatic annotation step, a small amount of manual edition could have been documented and applied through concordance views. In our case, it was possible to achieve such a formal modelization of our typology based on queries (obviously this may not be possible for any kind of semantic description), and we did not need to proceed to manual post-editing, as our analyses could cope with an effective very low number of wrong annotations.

The AF-NOTICES-V3-2021-10-11 corpus counts 2,234,187 words, in which *Résumé* and *Séquences* sections (sections where shot mentions can occur) totalize 1,953,088 words. Our process annotated 91,898 shot mentions (93,553 words if we include off-topic mentions).

[1] CQL queries : `"-" "[A-Z]{2,}"` and `"-[A-Z]{2,}"`

[2] CQL queries : `[word="-*(plans?)"%c]` and `[word="-*(vues?)"%c]`

[3] In a multiple word mention, the use of the target operator (@) in the query allows pointing the token that will bear the annotation. By default, without target mark, the annotation is assigned to the first token.

***Table 1**. Queries used to annotate shot types in the INA documentary descriptions of* Les Actualités françaises *reports.*

label	***Shot type*** ***EN (FR)***	***Examples of matching terms***	***Query (CQL equation)***
00DP	**miscellaneous** (divers plans)	DP, DV	`[word="-*(DP|DV)"%c]`
00DP	miscellaneous (divers plans)	plan/vue(s)	`[(resume|sequences) & word="-*(plan|vue)s?"%c]`
01HS	**off-topic** (hors sujet)	en arrière plan, au premier plan, sur le deuxième plan, en A V plan, (Au) 2ème plan, au/sur le dernier plan…	`(([word="-*(premier|1er|avant|second|deuxième|2ème|arrière|dernier)"%cd] | ([word="A"][word="V"])) @[word="plan"%c])`
01HS	off-topic (hors sujet)	commissaire (général(e)) au plan, Secrétaire d'Etat … au plan	`([word="au"%c] @[word="plan"%c])`
01HS	off-topic (hors sujet)	sur le plan international, sur le plan	`([word!="surimpressionnée?s?"%c][word="sur"%c][word="le"]@[word="plan"%c])`
01HS	off-topic (hors sujet)	devant un/les plan(s)	`([word="devant"%c][word="un|les|des"]@[word="plans?"%c])`
01HS	off-topic (hors sujet)	plan Marshall,…	`([(resume|sequences) & word="plan"%c][word="."]?[word="Monnet|Marshall?|Schuman|Courant|Pinay-Rueff|Challe"%c])`
01HS	off-topic (hors sujet)	plan algérien, plan cadastral, plan mural, plan(s) incliné(s)…	`([(resume|sequences) & word="plans?"%c][word="britanique|algérien|cadastra(l|ux)|architectura(l|ux)|mura(l|ux)|inclinés?|anciens?"%c])`
01HS	off-topic (hors sujet)	vues microscopiques, plans cinématographiques	`([(resume|sequences) & word="-*(plan|vue)s?"%c] [word="(microscopique|photographique|cinématographique)s?"%c])`
01HS	off-topic (hors sujet)	miroir plan	`([word="miroirs?"%c][word="-"]?@[word="plans?"%c])`
01HS	off-topic (hors sujet)	plan (carte), plan en coupe/relief	`([word="plan"%c] [word="en|\("][word="carte|coupe|relief"%c])`
01HS	off-topic (hors sujet)	plan d'eau, plan de vol	`([word="plan"%c][word="d."][word="eau|vol"%c])`
01HS	off-topic (hors sujet)	plan(s) d'architecte	`([word="plans"%c][word="d."][word="architectes?"%c])`

label	*Shot type* *EN (FR)*	*Examples of matching terms*	*Query (CQL equation)*
01HS	off-topic (hors sujet)	longue vue	`([(resume|sequences) & word="longues?"%c] @[word="vues?"%c])`
01HS	off-topic (hors sujet)	à perte de vue, échanges de vues, point de vue, gardé(e) à vue	`([word="perte|échanges?|point|gardé?e?"%c][word="de|à"]@[word="vues?"%c])`
01HS	off-topic (hors sujet)	discutant plans en mains, examinant un plan, consultant des cartes et des plans, trace un plan, etc.	`([frlemma="consulter|déplier|discuter|étudier|examiner|montrer|regarder|tracer"%c][]{0,3}[frpos=".*DET.*"%c])?@[word="plans?"%c])`
01HS	off-topic (hors sujet)	prenant une vue, qui prennent des vues, (se) masquant la vue	`([frlemma="prendre|masquer"%c & word!="prises?"%c][frpos=".*DET.*"%c]@[word="vues?"%c])`
01HS	off-topic (hors sujet)	plan indiquant	`([word="plans?"%c][]?[frlemma="indiquer"])`
01HS	off-topic (hors sujet)	avoir une vue, avoir vue (participe passé)	`([frlemma="avoir"][]?@[word="vues?"])`
10PG	**Extreme Wide Shot** (plan général)	VG, PG, GVG, CG	`[word="-*(VG|PG|GVG|CG)"%c]`
10PG	Extreme Wide Shot (plan général)	plan/vue(s) général(e)(s)	`([word="-*(plan|vues?)"%c][word="générale?s?"%c])`
10PG	Extreme Wide Shot (plan général)	PANORAMA	`[word="-*PANORAMA"%c]`
10PG	Extreme Wide Shot (plan général)	VE	`[word="-*(PE|VE)"%c]`
10PG	Extreme Wide Shot (plan général)	plan/vue d'ensemble	`([word="-*(plan|vue)s?"%c][][word="ensemble"%c])`
11PL	**Wide Shot** (plan large)	PL, PSG, VSG	`[word="-*(PL|PSG|VSG)"%c]`
11PL	Wide Shot (plan large)	plan/vue large	`([word="-*(plan|vue)"%c][word="large"%c])`
11PL	Wide Shot (plan large)	plan/vue(s) semi(-)général(e)(s)	`([word="-*(plan|vue)s?"%c][word="semi.*"%c])`

label	*Shot type* *EN (FR)*	*Examples of matching terms*	*Query (CQL equation)*
12PM	**Medium Shot** (plan moyen)	PM, VM	`[word="-*(PM|VM)"%c]`
12PM	Medium Shot (plan moyen)	plan moyen	`([word="-*plans?"%c][word="moyens?"%c])`
13PR	**Medium Close Up** (plan rapproché)	PA, PR, VR	`[word="-*(PA|PR|VR)"%c]`
13PR	Medium Close Up (plan rapproché)	plan/vue rapproché(e)	`([word="-*(plan|vue)s?"%c][word="rapprochée?s?"%c])`
13PR	Medium Close Up (plan rapproché)	plan proche	`([word="-*plans?"%c][word="proches?"%c])`
13PR	Medium Close Up (plan rapproché)	plan américain	`([word="-*plan"%c][word="américain"%c])`
14GP	**Close Up** (gros plan)	GP, GPP	`[word="-*(GP|GPP)"%c]`
14GP	Close Up (gros plan)	gros plan(s)	`([word="-*gros"%c]@[word="plans?"%c])`
15TGP	**Extreme Close Up** (très gros plan)	TGP	`[word="-*TGP"%c]`
15TGP	Extreme Close Up (très gros plan)	très gros plan(s)	`([word="-*très"%c][word="gros"%c]@[word="plans?"%c])`
20PLON	**High Angle Shot** (plongée)	VP	`[word="-*VP"%c]`
20PLON	High Angle Shot (plongée)	plan/vue(s) etc. en plongée	`([word="-*(vg|p[pmg]|gpp?|plans?|vues?|pano(ramique)?s?|panorama|trav(el(l?ing)?)?|zoom)"%c | word="V[APEMFR]|P[RALE]|D[PV]|[PV]SG|ZA[VR]|GVG|CG|TGP"][]{0,2}[word="en"%c]@[word="plongée"%cd])`
20PLON	High Angle Shot (plongée)	plongée sur	`([word!="-*contre"%c]@[word="plongée"%cd][word="sur"])`
20PLON	High Angle Shot (plongée)	vue plongeante	`([word="-*(vues?|v.)"%c][word!="nageu.*"]?@[word="plongeant.*"%c])`
21CPL	**Low Angle Shot** (contre-plongée)	CPL	`[word="-*CPL"%c]`
21CPL	Low Angle Shot (contre-plongée)	contre-plongée, contreplongée	`[word="-*contre-?plongée"%cd]`

label	*Shot type* *EN (FR)*	*Examples of matching terms*	*Query (CQL equation)*
21CPL	Low Angle Shot (contre-plongée)	contre plongée	`([word="-*contre"%c]@[word="plongée"%cd])`
30TRAV	**Tracking Shot** (travelling)	trav(elling)	`([word="trav(el(ling)?)?"%c][word!="agen.*|service"%c])`
31PANO	**Pan Shot** (panoramique)	Panoramique, PANO	`[(resume|sequences) & word="-*pano(ramiqué?e)?"%c]`
32PP	**Handheld Shot** (plan porté)	PP	`[word!="au"%c]@[word="-*PP"%c]`
40VA	**Aerial Shot** (vue aérienne)	VA	`([(resume|sequences) & word="-*VA"][word="PANO" | word!="[A-Z]+"])`
40VA	Aerial Shot (vue aérienne)	plan/vue(s) aérien(ne)(s)	`([(resume|sequences) & word="-*(plan|vue)s?"%c][word="aérien(ne)?s?"%cd])`
50VE	**Exterior Shot** (vue extérieure)	plan/vue(s) extérieur(e)(s)	`([word="-*(plan|vue)s?"%c][word="extérieure?s?"%cd])`
51VI	**Interior Shot** (vue intérieure)	plan/vue(s) intérieur(e)(s)	`([word="-*(plan|vue)s?"%c][word="intérieure?s?"%cd])`
60ZOOM	**Zoom**: Zoom In - Zoom Out (zoom)	zoom	`[word="-*ZOOM"%c]`
60ZOOM	Zoom: Zoom In - Zoom Out (zoom)	ZAV, ZAR	`[word="-*(ZAV|ZAR)"%c]`
70TI	**Title - Credits** (titre)	titre(s) : xxx, titre(s) "xxx"	`([(resume|sequences) & word="-*titres?"%c][]{0,3}[word="[:\"]"])`
70TI	Title - Credits (titre)	titre(s) en début de séquences	`(<sequences>[]{0,10}@[word="-*titres?"%c])`
71BT	**Rostrum camera** (banc-titre)	BT	`[word="-*BT"%c]`
71BT	Rostrum camera (banc-titre)	banc(s) titre(s)	`([(resume|sequences) & word="bancs?"%c][word="titres?"%c])`
71BT	Rostrum camera (banc-titre)	banc(s)-titre(s)	`[(resume|sequences) & word="bancs?-titres?"%c]`
72GR	**Graphics** (graphique)	graphique(s)	`[(resume|sequences) & word="graphiques?"%c]`
73FLOU	**Out of focus** (flou)	flou(e)(s)	`[frlemma = "flou"]`