



Documentation du corpus AF-NOTICES-V4

contributor, editor Bénédicte PINCEMIN
contributor Serge HEIDEN
contributor Matthieu DECORDE

Copyright © avril 2022 ENS de Lyon

<https://textometrie.org>



Ce document est publié sous licence Creative Commons Attribution-NonCommercial-ShareAlike 4.0

<https://creativecommons.org/licenses/by-nc-sa/4.0>

Contenu et version du corpus

Ce corpus représente la collection complète des notices documentaires INA des journaux métropolitains hebdomadaires des *Actualités françaises*, de janvier 1945 à février 1969, soit 1261 journaux (notices sommaires) composés de 10 776 sujets (notices sujets).

L'import de ce corpus a été réalisé par l'équipe TXM en avril 2022. Ses sources sont le tableau des notices INA du 27 avril 2022 bénéficiant des synchronisations manuelles de Matthieu Frey et Bénédicte Pincemin. Le corpus TXM intègre des liens aux vidéos des émissions en ligne sur Okapi (il faut avoir installé l'extension Media Player de TXM pour pouvoir les utiliser, et une authentification INA est requise pour l'accès aux vidéos).

Le corpus compte 2,2 millions de mots.

Schéma simplifié

- **<text>** = un journal hebdomadaire
 - id** = identifiant INA
 - date-de-diffusion-tri** = AAAA-MM-JJ
 - date-de-diffusion-annee** = AAAA
 - ...
 - **<div>** = un sujet
 - id** = identifiant INA
 - date-de-diffusion-emission-tri** = AAAA-MM-JJ
 - date-de-diffusion-tri** = AAAA-MM-JJ
 - date-de-diffusion-annee** = AAAA
 - n** = position du sujet dans le journal (1, 2, etc.)
 - ...
 - **<titre-propre>** = le titre du sujet
 - contenu textuel : paragraphe **<p>** avec des **[mot]**
 - **<resume>** = résumé INA
 - contenu textuel : paragraphes **<p>** avec des **[mot]**
 - **<sequences>** = description plan à plan
 - contenu textuel : paragraphes **<p>** avec des **[mot]**
 - **<descripteurs-aff-lig>** = une **<list>** de mots-clés INA
 - **<item>** = un descripteur INA
 - type** = code de nature du descripteur (DET, DEL, DEI, etc.)
 - contenu textuel formé de **[mot]**
 - **<generique-aff-lig>** = une **<list>** de personnes impliquées à divers titres
 - **<item>** = un nom de personne tiré du référentiel INA
 - type** = code de nature du descripteur (PAR, OPV, etc.)
 - contenu textuel formé de **[mot]**

Relation aux autres corpus Antract

Le corpus AF-NOTICES-V4 a le même contenu textuel que le corpus AF-NOTICES-V3, ce dernier en était une version provisoire : les deux corpus ont la même structure et contiennent les mêmes notices, mais les données d'AF-NOTICES-V4 sont complètes au niveau des time-codes, ce qui permet que tous sujets soient ordonnés chronologiquement au sein des journaux. Par ailleurs les sujets s'avérant une unité de composition de sous-corpus très importante, ils sont dotés de deux nouvelles propriétés (*date-de-diffusion-emission-tri* et *n*, cf. § Métadonnées des sujets). Enfin, le corpus AF-NOTICES-V4 est doté de nouveaux liens de lecture des vidéos.

Le corpus AF-NOTICES-V4 se rapproche, dans sa structure et son contenu, du corpus AF-VOIX-OFF. Comme ce dernier, il se présente comme un ensemble d'émissions (et non plus comme un ensemble de notices, ce qui était le cas d'AF-NOTICES-V2). Une conséquence de ce choix est qu'AF-NOTICES-V4 ne comprend plus toutes les notices des Actualités françaises, puisqu'il exclut les notices dites « isolées » correspondant aux éditions régionales ou aux sujets non diffusés. Ainsi les mots de ce corpus correspondent à ce qu'on avait dans AF-NOTICES-V2 lorsque l'on faisait un sous-corpus des notices sujets (pour se rapprocher d'AF-VOIX-OFF) : dans AF-NOTICES-V4 les mots relèvent tous des notices sujets, les notices sommaires n'apportent que des métadonnées et contribuent à structurer le corpus sans apporter de contenu lexical propre.

Par ailleurs, le corpus AF-NOTICES-V4 corrige un problème de lacunes dans les données (certaines périodes manquent) qui a été découvert tardivement dans AF-NOTICES-V2. Le sous-corpus *sujets-pas-dans-v2* rassemble ces 244 « nouveaux » sujets,¹ il permet de faire des recherches en vue de compléter des sélections faites dans AF-NOTICES-V2.

Voici un tableau récapitulatif des dernières évolutions du corpus AF-NOTICES :

	AF-NOTICES-V2 (octobre 2018)	AF-NOTICES-V3 (octobre 2021)	AF-NOTICES-V4 (avril 2022)
Périmètre	Toutes les notices documentaires INA correspondant aux AF : notices sommaires + notices sujets + notices isolées.	Émissions diffusées en métropole. (Cela exclut les notices isolées.)	Émissions diffusées en métropole. (Cela exclut les notices isolées.)
Complétude	Les tableaux INA source sont lacunaires pour les dernières années (il manque le 2 ^e semestre 1968 et le dernier trimestre de 1966). Ce doit être dû à un incident technique resté longtemps inaperçu.	Les données sont complètes (descriptions documentaires, sans les transcriptions), mais plus de 200 sujets ne sont pas synchronisés. ²	Les données sont complètes. (descriptions documentaires, sans les transcriptions). On bénéficie de plus d'un accès aux vidéos ³ (depuis tout mot du contenu de la notice on accède au début du sujet dans la vidéo de l'émission).
Structuration	Les notices sont groupées par année (<text>). À l'intérieur de l'année, elles sont les unes après les autres par ordre chronologique de leur date de diffusion (<notice>).	Le corpus est composé d'émissions (<text>), et chaque émission contient ses sujets (<div>). Les émissions sont dans l'ordre chronologique, les sujets synchronisés sont ordonnés chronologiquement à l'intérieur de l'émission, précédés ou/et suivis des sujets non-synchronisés (triés par time-code de début et identifiant).	Le corpus est composé d'émissions (<text>), et chaque émission contient ses sujets (<div>). Les émissions sont dans l'ordre chronologique, les sujets sont ordonnés chronologiquement à l'intérieur de l'émission.

1 Requête utilisée : `<text>[_ .text_date-de-diffusion-tri="((1966-(09|10|11|12))|(1968-(06|07|08|09|10|11|12)))-.." & _ .text_date-de-diffusion-tri!="1966-09-07|1968-06-(05|12)"] expand to text`

2 Voir par exemple la concordance de : `<div>[_ .div_antract-duree="00:00:00:00|"]`

3 Il faut avoir installé l'extension Media Player pour pouvoir mettre en œuvre cette fonctionnalité.

Structuration et ordre

Les textes (structure <text>) du corpus sont les 1261 **journaux** hebdomadaires.

L'ordre des textes est **chronologique** (ce qui ordonne les pages des textes affichées par TXM -commande Édition- et donne sens à la commande Progression sur l'ensemble du corpus).

Le texte de chaque journal est un ensemble de **fiches documentaire de sujet** (<div>), selon un **ordre qui suit le déroulement de l'émission**.

Chaque sujet est décrit par plusieurs **champs textuels réguliers** : le titre du sujet (<titre-propre>), son résumé dans la base INA (<resume>), sa description plan à plan (<sequences>), ses mots clés (<descripteurs-aff-lig>), et les personnes impliquées à divers titres dans le sujet (<generique-aff-lig>).

Les trois premiers champs sont composés de paragraphes (<p>). Le titre est lui-même un seul paragraphe, mais les résumés et séquences peuvent être formés d'une succession de paragraphes.

Concernant les résumés et séquences, il est important de noter que les séquences ne sont pas toujours renseignées ; le contenu du résumé prend parfois la forme d'une description de type séquences. En l'état actuel des données il vaut donc mieux généralement **interroger ensemble résumé et séquences**.

Les deux derniers champs (descripteurs et générique) se présentent sous forme d'une liste (<list>) composée d'éléments (<item>). Chaque élément est un descripteur issu des référentiels documentaires INA. La propriété **type** de la structure <item> renseigne sur la catégorie du descripteur, un code formé de trois lettres majuscules. Pour les mots-clés du champ <descripteurs-aff-lig>, les catégories sont : DET (descripteur de thème), DEL (descripteur de lieu), DEI (ce qui apparaît à l'image), DSO (ce qu'on entend dans la bande son). Dans le générique, ces catégories sont des rôles, tels que PAR (participant), OPV (opérateur de prise de vue), etc. Ces codes rôles sont nombreux, on en trouve une présentation dans le fichier `Code_roles.xls` sur le sharedocs du projet Antract (dans le répertoire SP2 Corpus & Recherche technologique).

Notons que lorsque l'on visualise les textes du corpus dans TXM (avec la fonction Édition, ou par retour au texte depuis la Concordance), seuls les contenus des champs textuels font partie des mots que l'on peut chercher et qui sont comptabilisés. Sont affichés pour la lecture et l'interprétation, mais ne font pas partie des mots sur lesquels porte l'analyse textométrique :

- la liste des métadonnées, qui précède les champs textuels ;
- les intitulés des champs textuels (« Titre propre », « Résumé », etc.) ;
- les codes de catégorie affichés devant chaque descripteur (mots-clés et noms de personnes au Générique).

Les mots du corpus sont tous soit dans un paragraphe (<p>) (pour le titre, le résumé et les séquences), soit dans un élément de liste (<item>) (pour les descripteurs et le générique).

Métadonnées des journaux hebdomadaires

Chaque journal des Actualités françaises correspond à une structure `text`, dotée de propriétés dont :

- `id` : l'identifiant de la notice INA (ex. AFE86004415)
- `date-de-diffusion-tri` : notée AAAA-MM-JJ (ex. : 1945-01-25 pour le 25 janvier 1945).

On notera aussi la disponibilité de variantes d'expression de la date de diffusion, potentiellement utiles pour certaines interrogations (construction de partitions, affichage, etc.) : `date-de-diffusion-annee`, `date-de-diffusion-mois`, `date-de-diffusion-jour-semaine`, `date-de-diffusion-jour`.

Si besoin, un inventaire complet des métadonnées des journaux est consultable via la commande Propriétés, onglet Détails, section Propriétés des structures, point `text`. Ces propriétés sont :

- soit directement issues des bases de données INA, en reprenant le nom du champ (en minuscules, sans accents, sans parenthèses, avec des tirets entre les mots) ;
- soit produites par le projet Antract (nom commençant par "antract-"), notamment les times-codes positionnant le début et la fin du sujet dans l'émission (`antract-debut`, `antract-fin`) ;

- soit ajoutées par l'import TXM, comme les variantes d'écriture de date ci-dessus, mais aussi d'autres propriétés pour des considérations techniques (ex. `text-order`).

Métadonnées des sujets

Chaque sujet (structure `<div>`) est d'abord identifié par la propriété `id`, qui reprend l'identifiant de la notice INA.

Les `date-de-diffusion-tri` des sujets ne sont pas toujours égales à celle du journal (elles peuvent être quelques jours antérieures), c'est pourquoi on dispose aussi de la métadonnée `date-de-diffusion-emission-tri` qui reprend, au niveau de chaque sujet, la date de l'émission dans laquelle il est inclus. On a au niveau du sujet les mêmes propriétés dérivées de la date de diffusion qu'au niveau des émissions (`date-de-diffusion`, `date-de-diffusion-annee`, etc.)

L'ordre dans lequel sont présentés les sujets à l'intérieur de chaque journal est noté dans la propriété `n`.

Les propriétés des sujets apportent des informations de catalogage INA : les notes du titre (`notes-du-titre`), des indications sur la production (`nature-de-production`, `producteurs-aff`) et sur la diffusion effective (`type-de-date`), un classement en genre (`genre`) et des catégories thématiques (`thematique`), l'indication de la langue lorsque ce n'est pas le français (`langue-vo-ve`).

La durée du sujet est présente deux fois, telle qu'elle a été notée par les documentalistes (`duree`) et telle qu'elle est mesurée sur l'empan de vidéo associé au sujet (`antract-duree`) (cette dernière information manque dans certains sujets, car elle n'a pas été recalculée après certaines corrections manuelles).

Les propriétés préfixées par `antract-` sont des informations techniques de synchronisation entre les sujets et les vidéos produites par le projet ANTRACT. On a également l'information technique des fichiers vidéo dans le fonds INA (`nom-fichier-segmente-info`).

Si besoin, un inventaire complet des propriétés disponibles pour les sujets est consultable via la commande Propriétés, onglet Détails, section Propriétés des structures, point `div`.

Propriétés des contenus textuels

Nous avons vu que les mots du corpus sont tous soit dans un paragraphe (`<p>`) (pour le titre, le résumé et les séquences), soit dans un élément de liste (`<item>`) (pour les descripteurs et le générique).

Pour les mots, les principales propriétés sont :

- `word` : forme graphique, telle qu'elle se présente dans la transcription ASR fournie ;
- `fr lemma` : lemme, c'est-à-dire l'entrée de dictionnaire correspondante (singulier, masculin, infinitif, etc.)
- `fr pos` : catégorie grammaticale (liste disponible dans le manuel TXM) (`pos= part-of-speech = partie-du-discours`)

Les lemmes et les catégories grammaticales ont été affectés automatiquement par le logiciel TreeTagger avec le principal modèle de langue pour le français moderne écrit. Cela enrichit les interrogations possibles mais il peut y avoir des erreurs.

Pour les besoins de l'analyse sur la grammaire cinématographique (et en particulier de la production du corpus AF-PLANS), certains mots du corpus AF-NOTICES ont été annotés automatiquement selon des catégories de valeurs de plan enregistrées dans une propriété `plan`. Nous donnons en annexe les éléments de définition des catégories utilisées (étiquette, signification et occurrences visées, équation de sélection dans le corpus).

Si besoin, les inventaires complets des propriétés des mots sont consultables via la commande Propriétés, onglet Détails, section Propriétés des unités lexicales. (En revanche, la commande Propriétés ne liste pas toutes les valeurs pour chacune des propriétés disponibles : pour cela on peut appliquer au corpus la commande Lexique.)

Retour à la vidéo

Le corpus AF-NOTICES-V4 fournit un accès au visionnage de la vidéo des émissions : à partir d'un mot d'une notice, on accède au début du **sujet** correspondant à la notice. Pour retrouver un moment précis à l'intérieur d'un sujet, il reste à déplacer soi-même le curseur de lecture au niveau voulu. Pour se repérer, on trouve le temps de

début de chaque sujet dans les références de concordances (colonne de gauche), ainsi que dans les métadonnées listées au début de la page d'édition de chaque sujet.

Cette fonctionnalité suppose d'avoir installé l'extension MediaPlayer de TXM.

L'accès à la vidéo se fait depuis tout mot du texte des notices, dans l'Édition (affichage du texte complet) ou la Concordance : faire un clic-droit sur un mot, et lancer la commande « Jouer le média ».

Selon les environnements (système d'exploitation, version de TXM, etc.), il peut être nécessaire de faire attention aux points suivants :

- dans l'édition il faut avoir sélectionné le mot avant de cliquer-droit dessus (sinon la commande ne se lance pas) ;
- il faut prévoir un affichage assez large de la vidéo pour avoir le curseur qui permet de se positionner où l'on veut dans la vidéo.

Références de concordance

Les références de concordance sont composées par défaut de :

- la date de l'émission (au format tel que le tri soit chronologique), puis
- l'identifiant de la notice du sujet (qui est différent de celui de l'émission, qui sert de nom au fichier vidéo), et
- le temps du début du sujet dans l'émission.

Il est toujours possible de choisir un autre affichage en faisant un clic-droit dans la colonne des références et en réglant les « Options d'affichage des références ».

Comment récupérer et mettre dans AF-NOTICES-V4 un sous-corpus (sélection de sujets) déjà constitué

- Si le sous-corpus à récupérer est dans Okapi : utilitaire `ImporterCorpusOkapi`.
- Si le sous-corpus à récupérer est dans un autre corpus TXM : utilitaires `ListerIdentifiantsSujets` puis `ImporterCorpusOkapi`.

L'installation et la mise en œuvre des utilitaires sont décrites dans cette documentation en ligne : https://groupes.renater.fr/wiki/txm-info/public/chantier_antract#tutoriel_d_utilisation_des_outils_d_echanges_entre_txm_et_okapi

Noter également que le sous-corpus `sujets-pas-dans-v2` peut être utilisé pour chercher d'éventuels sujets pouvant compléter un sous-corpus qui aurait été initialement construit par des requêtes dans AF-NOTICES-V2 : on lance les mêmes requêtes sur ce sous-ensemble complémentaire pour ne pas avoir à retraiter des sujets déjà vus.

Annexe : Documentation des étiquettes de valeur de plan

Remarque sur l'algorithme d'étiquetage

L'étiquetage a été réalisé à l'aide de l'utilitaire TXM `CQLList2WordProperties`⁴ appliqué au tableau de définition ci-après.⁵

Les étiquetages sont effectués dans l'ordre des lignes du tableau. Les étiquettes « 00DP », affectées très largement à tous les mots « plan(s) » ou « vue(s) » au début du processus d'étiquetage, sont ensuite en bonne partie affinées en fonction du contexte, soit pour être écartées en hors-sujet (01HS), soit pour être déclinées selon divers types de

⁴ Il s'agit de la version du 21 septembre 2021, intégrée à TXM à partir de TXM 0.8.2 beta.

⁵ Le fichier du tableau au format demandé par l'utilitaire, `cql_valeurs_de_plan_211007a.tsv`, est disponible sur le sharedocs du projet Antract, dans le répertoire dédié au corpus AF-NOTICES-V4 : ANR > ANTRACT > SP2 Corpus & Recherche technologique > TXM > corpus > notices documentaires > version4 > src

plans plus précis. Donc au terme du traitement, **les occurrences étiquetées par un code sont un sous-ensemble des occurrences couvertes par les équations CQL** associées au code.

Pour en avoir plus sur cet étiquetage, voir l'article : Bénédicte Pincemin, Serge Heiden, Franck Mazuet - « The Textometric Concept of Active Corpus. Illustration by an Analysis Scenario based on Annotation then Projection », soumis à la *16th International Conference on Statistical Analysis of Textual Data (JADT 2022)*, Naples, juillet 2022.

Code	Type	Exemples de réalisations	Équation CQL de sélection initiale⁶
00DP	divers plans	DP, DV	[word=" - *(DP DV) "%c]
00DP	divers plans	plan/vue(s)	[(resume sequences) & word=" - *(plan vue)s? "%c]
01HS	hors sujet	en arrière plan, au premier plan, sur le deuxième plan, en avant plan, en A V plan, (Au) 2ème plan, Au/en 1er plan, sur le 1er plan, au/sur le dernier plan...	(([word=" - *(premier 1er avant second deuxième 2ème arrière dernier) "%cd] ([word="A"][word="V"])) @[word="plan" "%c])
01HS	hors sujet	commissaire (général(e)) au plan, Secrétaire d'Etat ... au plan, et autres expressions	([word="au" "%c] @[word="plan" "%c])
01HS	hors sujet	sur le plan international, sur le plan	([word!="surimpressionnée?s" "%c] [word="sur" "%c] [word="le"] @[word="plan" "%c])
01HS	hors sujet	devant un/les plan(s)	([word="devant" "%c] [word="un les des"] @[word="plans?" "%c])
01HS	hors sujet	plan Marshall,...	([(resume sequences) & word="plan" "%c] [word="."]? [word="Monnet Marshall? Schuman Courant Pinay-Rueff Challe" "%c])
01HS	hors sujet	plan britannique, plan cadastral, plan mural, plan(s) incliné(s)...	([(resume sequences) & word="plans?" "%c] [word="britannique algérien cadastra(l ux) architectura(l ux) mura(l ux) inclinés? anciens?" "%c])
01HS	hors sujet	vues microscopiques, plans cinématographiques	([(resume sequences) & word=" - *(plan vue)s? "%c] [word="(microscopique photographique cinématographique)s?" "%c])
01HS	hors sujet	miroir plan	([word="miroirs?" "%c] [word=" - "]? @[word="plans?" "%c])
01HS	hors sujet	plan (carte), plan en coupe/relief	([word="plan" "%c] [word="en \("] [word="carte coupe relief" "%c])
01HS	hors sujet	plan d'eau, plan de vol	([word="plan" "%c] [word="d."] [word="eau vol" "%c])
01HS	hors sujet	plan(s) d'architecte	([word="plans" "%c] [word="d."] [word="architectes?" "%c])
01HS	hors sujet	longue vue	([(resume sequences) & word="longues?" "%c] @[word="vues?" "%c])
01HS	hors sujet	à perte de vue, échanges de vues, point de vue,	([word="perte échanges? point gardé?e?" "%c] [word="de à"] @[word="vues?" "%c])

6 Voir l'explication donnée sur l'algorithme d'étiquetage, juste avant le tableau.

		gardé(e) à vue	
01HS	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	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	hors sujet	plan indiquant	([word="plans? "%c] []? [frlemma="indiquer"])
01HS	hors sujet	avoir une vue, avoir vue (patricipe passé)	([frlemma="avoir"] []? @[word="vues? "])
10PG	plan général	VG, PG, GVG, CG	[word=" - *(VG PG GVG CG) "%c]
10PG	plan général	plan/vue(s) général(e)(s)	([word=" - *(plan vues?) "%c] [word="générale? s? "%c])
10PG	plan général	PANORAMA	[word=" - *PANORAMA "%c]
10PG	plan général	VE	[word=" - *(PE VE) "%c]
10PG	plan général	plan/vue d'ensemble	([word=" - *(plan vue)s? "%c] [] [word="ensemble "%c])
11PL	plan large	PL, PSG, VSG	[word=" - *(PL PSG VSG) "%c]
11PL	plan large	plan/vue large	([word=" - *(plan vue) "%c] [word="large "%c])
11PL	plan large	plan/vue(s) semi(-)général(e)(s)	([word=" - *(plan vue)s? "%c] [word="semi.* "%c])
12PM	plan moyen	PM, VM	[word=" - *(PM VM) "%c]
12PM	plan moyen	plan moyen	([word=" - *plans? "%c] [word="moyens? "%c])
13PR	plan rapproché	PA, PR, VR	[word=" - *(PA PR VR) "%c]
13PR	plan rapproché	plan/vue rapproché(e)	([word=" - *(plan vue)s? "%c] [word="rapprochée? s? "%c])
13PR	plan rapproché	plan proche	([word=" - *plans? "%c] [word="proches? "%c])
13PR	plan rapproché	plan américain	([word=" - *plan "%c] [word="américain "%c])
14GP	gros plan	GP, GPP	[word=" - *(GP GPP) "%c]
14GP	gros plan	gros plan(s)	([word=" - *gros "%c] @[word="plans? "%c])
15TGP	très gros plan	TGP	[word=" - *TGP "%c]
15TGP	très gros plan	très gros plan(s)	([word=" - *très "%c] [word="gros "%c] @[word="plans? "%c])
20PLON	plongée	VP	[word=" - *VP "%c]
20PLON	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	plongée	plongée sur	([word!=" - *contre"%c]@[word="plongée"%cd] [word="sur"])
20PLON	plongée	vue plongeante	([word=" - *(vues? v.)"%c] [word!="nageu.*"]? @[word="plongeant.*"%c])
21CPL	contre-plongée	CPL	[word=" - *CPL"%c]
21CPL	contre-plongée	contre-plongée, contreplongée	[word=" - *contre-?plongée"%cd]
21CPL	contre-plongée	contre plongée	([word=" - *contre"%c]@[word="plongée"%cd])
30TRAV	travelling	trav(elling)	([word="trav(el(ling)?)?"%c] [word!="agen.* service"%c])
31PANO	panoramique	Panoramique, PANO	[(resume sequences) & word=" - *pano(ramiqué?e)?"%c]
32PP	plan porté	PP	[word!="au"%c]@[word=" - *PP"%c]
40VA	vue aérienne	VA	([(resume sequences) & word=" - *VA"] [word="PANO" word!="[A-Z]+"])
40VA	vue aérienne	plan/vue(s) aérien(ne)(s)	([(resume sequences) & word=" - *(plan vue)s?"%c] [word="aérien(ne)?s?"%cd])
50VE	vue extérieure	plan/vue(s) extérieur(e)(s)	([word=" - *(plan vue)s?"%c] [word="extérieure?s?"%cd])
51VI	vue intérieure	plan/vue(s) intérieur(e)(s)	([word=" - *(plan vue)s?"%c] [word="intérieure?s?"%cd])
60ZOOM	zoom	zoom	[word=" - *ZOOM"%c]
60ZOOM	zoom	ZAV, ZAR	[word=" - *(ZAV ZAR)"%c]
70TI	titre	titre(s) : xxx, titre(s) "xxx"	([(resume sequences) & word=" - *titres?"%c] [] {0,3} [word="[:\]"])
70TI	titre	titre(s) en début de séquences	(<sequences> [] {0,10} @[word=" - *titres?"%c])
71BT	banc-titre	BT	[word=" - *BT"%c]
71BT	banc-titre	banc(s) titre(s)	([(resume sequences) & word="bancs?"%c] [word="titres?"%c])
71BT	banc-titre	banc(s)-titre(s)	[(resume sequences) & word="bancs?-titres?"%c]
72GR	graphique	graphique(s)	[(resume sequences) & word="graphiques?"%c]
73FLOU	flou	flou(e)(s)	[frlemma = "flou"]