

Un Corpus de Flux TV Annotés pour la Prédiction de Genres

Mohamed Bouaziz^{1,2} Mohamed Morchid¹ Richard Dufour¹

Georges Linarès¹ Prosper Correa²

(1) LIA, 339 Chemin des Meinajaries, 84140 Avignon, France

(2) EDD, 28 Boulevard de Port-Royal, 75005 Paris, France

mohamed.bouaziz@alumni.univ-avignon.fr, {mohamed.morchid, richard.dufour, georges.linares}@univ-avignon.fr, pcorrea@edd.fr

RÉSUMÉ

Cet article présente une méthode de prédiction de genres d'émissions télévisées couvrant 2 jours de diffusion de 4 chaînes TV françaises structurés en émissions annotées en genres. Ce travail traite des médias de masse de flux de chaînes télévisées et rejoint l'effort global d'extraction de connaissance à partir de cette grande quantité de données produites continuellement. Le corpus employé est fourni par l'entreprise EDD, anciennement appelée "L'Européenne de Données", une entreprise spécialisée dans la gestion des flux multimédias. Les expériences détaillées dans cet article montrent qu'une approche simple fondée sur un modèle de n-grammes permet de prédire le genre d'une émission selon un historique avec une précision avoisinant les 50 %.

ABSTRACT

A Genre Annotated Corpus of French Multi-channel TV Streams for Genre Prediction.

This paper presents a method for telecast genre prediction covering 4 French TV channels on two consecutive days and structured into genre labelled telecasts. This work falls within the mass media management as regard to television channel streams and joins the global endeavours for extracting meaningful knowledge from the huge quantity of data continuously produced. The used dataset is provided by EDD (previously known as "L'Européenne de Données"), a company specialized in managing multimedia streams. Preliminary experiments conducted in this paper show that an n-gram straightforward approach allows us to predict the TV program genre according to the history with an accuracy of about 50%.

MOTS-CLÉS : Genres Télévisuels, Corpus Multi-chaîne, Structuration de Flux.

KEYWORDS: Telecast Genres, Multi-channel Corpus, Stream Structuring.

1 Introduction

La télévision a pris une place très importante dans la société avec un très grand nombre de chaînes télévisées disponibles qui n'a cessé de croître ces dernières décennies. Il apparaît alors important de pouvoir structurer ces données, les indexer, et les rendre facilement accessibles pour les professionnels ainsi que pour le grand public. Les scientifiques se sont donc fortement intéressés à l'analyse et la structuration du contenu audio dans un contexte multi-chaînes (Bredin *et al.*, 2014; Bouchekif *et al.*, 2013). Malheureusement, peu de corpus de documents audio sont disponibles, d'autant plus quand il s'agit de corpus en langue française. En outre, ces corpus, lorsqu'ils sont disponibles, sont

annotés en ne considérant qu'une seule chaîne (Amaral & Trancoso, 2003) ou en ne traitant qu'un seul aspect du problème de structuration ou une seule tâche spécifique. Il est donc difficile de les utiliser dans un contexte d'étude différent (Wang *et al.*, 2006; Gravier *et al.*, 2012). Les corpus multi-chaînes disponibles font face à deux problèmes majeurs. Premièrement, l'enregistrement est fait à des horodatages différents, ce qui rend difficile, voire impossible, la comparaison entre les différents canaux. Deuxièmement, la taille de l'ensemble des programmes pris en compte sont généralement considérablement réduites. Par exemple, récemment, le corpus REPERE (Giraudel *et al.*, 2012) contient des émissions télévisuelles provenant de différentes chaînes françaises enregistrées à des dates différentes dans la période comprise entre 2012 et 2014. Les émissions ont été extraites à partir de moments précis de la journée et ont été également choisies pour contenir principalement des genres télévisuels spécifiques localisés (journaux, débats, interviews...) excluant les autres programmes diffusés tout au long de la journée (documentaires, films, inter-programmes...).

Cet article présente, dans un premier temps, un corpus audio de flux TV de 2 jours enregistré à partir de la diffusion de 4 chaînes françaises (TF1, M6, France 5 et TV5 Monde). Ensuite, des expérimentations préliminaires d'un système temps-réel d'aide à l'annotation manuelle sont conduites en évaluant différentes représentations d'une séquence de $n - 1$ émissions dans l'objectif de prédire le genre de la $n^{ème}$ émission qui suit. Enfin, nous discutons de l'intérêt de regrouper les flux considérant leur contenu (généraliste/semi-thématique). Certains travaux se sont intéressés à la problématique de structuration de flux TV. (Poli, 2008) est le seul, à notre connaissance, à s'être appuyé sur les anciennes séquences de programmes pour structurer un flux. En apprenant un Modèle de Markov Caché Contextuel sur l'historique de diffusion de chaque chaîne, (Poli, 2008) obtient des grilles précises à 97 %. Les autres travaux utilisent strictement le contenu audio-visuel (Liang *et al.*, 2005; Wang *et al.*, 2008; El-Khoury *et al.*, 2010) ou n'exploitent l'historique d'émissions que pour des raffinements *a posteriori* (Naturel *et al.*, 2007; Manson & Berrani, 2010; Ibrahim & Gros, 2011). Le corpus présenté est généré par l'entreprise EDD, une entreprise française âgée de 35 ans et spécialisée dans la collecte, l'analyse, l'indexation et la redistribution de divers types de journaux, magazines, et données radio et TV. EDD offre également, pour des professionnels et des administrations, plusieurs services comme une plate-forme de panorama de presse, un système d'alerte thématique et un moteur de recherche multimédia. EDD a mis à la disposition du LIA des données audio enregistrées sur 2 jours de diffusion relatives à environ 4 chaînes radio et TV. Avec les données de télévision, nous avons des méta-données (chaque 2 minutes) précisant le titre de l'émission en cours de diffusion. Elles sont enrichies, pour un certain nombre de programmes, par des informations complémentaires (e. g. description, genre, etc.) et une transcription automatique produite par SPEERAL, le système de reconnaissance automatique de la parole du LIA (Linarès *et al.*, 2007).

Ce papier est structuré comme suit. La section 2 présente la taxonomie de genres. La section 3 décrit la procédure de collecte et d'annotation du corpus. La section 4 est dévolue aux expérimentations pour la prédiction du genre télévisé. Enfin, les conclusions ainsi que les travaux futurs sont présentés dans la section 5.

2 Taxonomie des Genres

La classification des émissions en genres est souvent une tâche subjective qui dépend de la perception de l'annotateur. Par conséquent, plusieurs entités compétentes, comme l'Institut National de l'Audio-visuel (INA), ont créé chacune leur propre taxonomie. L'INA dispose des archives de flux de chaînes

radio et TV françaises, dont les émissions sont classées en 52 genres différents (INA, 2002).

La conception d'une taxonomie trop exhaustive peut conduire à un certain nombre de confusions au sein de la nomenclature. En conséquence, la différence entre certains genres peut être mineure. Par exemple, les *films* sont très proches des *courts métrages* et des *téléfilms*. En outre, on ne peut que difficilement distinguer de différences entre un *talk-show* comme "On n'est pas couché" et un *magazine de débat* avec généralement un ou deux présentateurs, accompagnés d'invités (e. g. "Le magazine de la santé").

Vu le grand nombre de genres et la difficulté pour l'annotateur de les distinguer, l'identification d'un genre particulier s'avère être une tâche difficile pour un être humain et d'autant plus pour un système automatique. Nous avons alors établi une taxonomie inspirée de celle proposée par l'INA en procédant à un certain nombre de fusions permettant aux genres d'être à la fois bien définis et distincts entre eux tout en couvrant les genres d'émission les plus fréquemment rencontrés. Le contenu visuel est divisé en 15 genres définis comme suit : *Inter-programmes (IP)*, *Actualité*, *Météo*, *Dessins animés*, *Fiction*, *Documentaire*, *Téléachat*, *Plateau/Débat*, *Magazine de reportage*, *Autres magazines*, *Musique*, *Télé-réalité*, *Programme court*, *Jeu* et *Autres*.

Les *inter-programmes* sont référencés comme un genre particulier dans la mesure où ils apparaissent en général entre deux émissions ou comme pause au sein d'une même émission. Ce genre, dénommé par la suite IP, englobe les *publicités*, *jingles*, *génériques* de début et de fin de programmes, etc. Nous avons choisi de considérer les *génériques* comme des IP vu leurs particularités, principalement acoustiques, pouvant aider à la recherche des IP dans le cadre de la prochaine étape de ce travail, à savoir, la structuration automatique des flux TV. Quant au genre *Fiction*, il inclut les *films*, *courts-métrages*, *séries* ainsi que les *feuilletons*. Pour ce qui est de la catégorie *autres magazines*, celle-ci concerne uniquement les magazines qui ne sont ni sous forme de débat ni constitués d'une suite de reportages (e. g. Dr CAC, Astuces du Chef...). Les programmes moins fréquents et les émissions couvrant un événement particulier comme les *événements sportifs* et les *émissions de variétés* sont rassemblés dans la catégorie *autres*.



FIGURE 1: Un exemple d'une courte séquence de programmes sur TF1.

3 Le Corpus Multi-chaînes

L'utilisation du contenu, dans le cadre des travaux de structuration de flux TV cités dans l'introduction, se focalise essentiellement sur l'information vidéo. Ce corpus se limite à l'information audio vu que nous comptons, dans la suite de ce travail, exploiter l'audio pour la structuration des flux TV. L'utilisation de ce média, nous permettra ensuite d'étendre notre futur système aux flux radio également traités par EDD. Cette section décrit le processus de collecte ainsi que le processus d'annotation de ces données. Le contexte et les statistiques du corpus sont ensuite détaillés. Une discussion sur les catégories de programmes est finalement délivrée.

3.1 Annotation

L'objectif principal de la constitution de ce corpus annoté manuellement en genres, est de concevoir un environnement digital homogène permettant d'expérimenter les représentations séquentielles de programmes télévisés et assurant une mission d'aide à l'annotation manuelle. Le corpus TV est segmenté, non seulement en émissions et inter-programmes, mais également en considérant comme différents segments les parties de programmes séparées par des pauses représentées par le genre *inter-programmes*.

La figure 1 présente un exemple d'une séquence de programmes annotés. La première étape du processus d'annotation manuelle consiste à tirer profit des méta-données relatives à chaque tranche de 2 minutes de retransmission afin d'établir une structure primaire concernant les frontières entre les différentes émissions. En effet, ces méta-données ne sont ni précises ni exhaustives. De plus, elles n'offrent aucune information en ce qui concerne le *timing* des *inter-programmes*. Ces circonstances ont rendu l'annotation manuelle très difficile sachant que nous traitons des données audio.

Concrètement, l'outil libre *WaveSurfer* (Sjölander & Beskow, 2000) de visualisation et de manipulation de son a été utilisé. En procédant à une annotation manuelle effectuée par un seul annotateur, un ensemble d'informations est affecté à chaque segment, *i.e.* l'instant de début et de fin, le genre correspondant provenant de notre taxonomie et, si possible, le nom de l'émission correspondante audit segment. Ces labels sont enregistrés dans un fichier texte pour chaque tranche de 24 heures de flux.

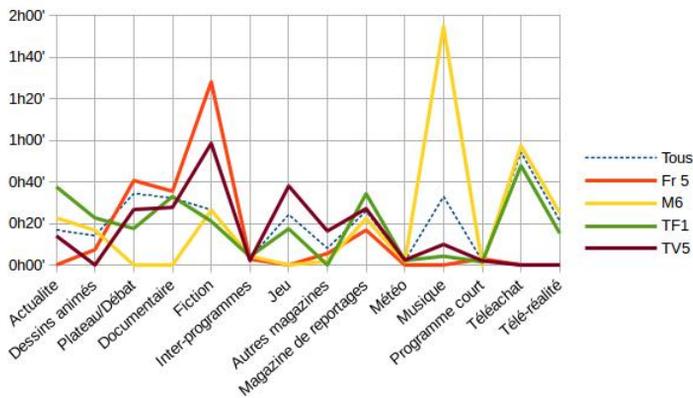


FIGURE 2: La durée moyenne des segments par genre

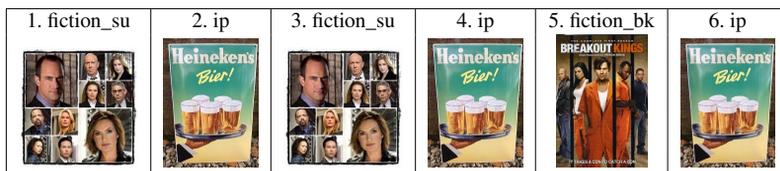


FIGURE 3: Un exemple d'ambiguïté pour deux fictions successives : Special Unit (su) et Breakout Kings (bk) sur TF1. *ip* : *inter-programmes*

Un problème qui se pose est que certaines émissions peuvent combiner deux genres considérés assez éloignés l’un de l’autre. Par exemple, le programme appelé “Petits secrets entre voisins” de la chaîne TF1 associe à la fois le genre *fiction* ainsi que le genre *télé-réalité*¹. Dans de telles situations, la liste des genres concernés est spécifiée en les concaténant et en les séparant par le signe +.

Par ailleurs, les génériques des émissions sont considérés comme des *inter-programmes*. Ceci dit, une ambiguïté se pose toujours pour discerner certains types de génériques (cf. première colonne du tableau 1). La question de les joindre au segment d’*inter-programmes* ou à celui de l’émission voisine se pose. Il a été décidé de créer un label différent pour chaque cas comme montré dans le tableau 1.

Cas	Étiquette
Parole non répétée, sur la musique du générique	générique_parole_non_répétée
Parole répétée, sur la musique du générique	générique_parole_répétée
Parole introductive suivie du générique de début & générique de fin suivi d’une parole de fin	parole_extra_générique

TABLE 1: Annotation des génériques ambigus

3.2 Détails et Statistiques

Le corpus fourni par l’entreprise EDD est composé de retransmissions de 4 chaînes TV françaises (TF1, M6, France 5 et TV5 Monde) pour une période de 2 jours entre le 10 et 12 février 2014.

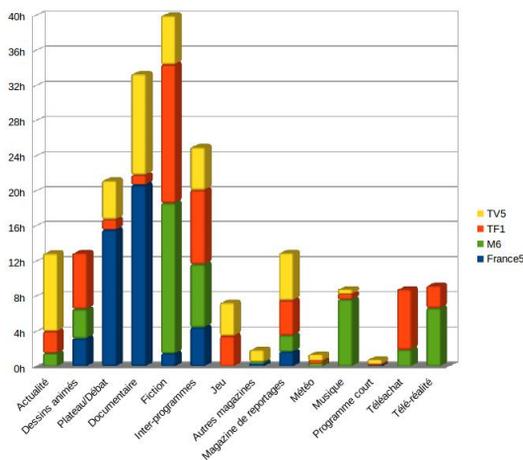


FIGURE 4: Distribution des genres dans le corpus.

Comme illustré dans la figure 4, les 4 chaînes TV ne partagent pas la même distribution de genres, chose prévisible sachant que ces chaînes grands publics ne partagent pas une même politique éditoriale. Pour les deux chaînes généralistes (TF1 et M6), la *fiction* est le genre de loin le plus représenté avec quasiment un tiers de parts de diffusion. En outre, ces deux chaînes sont les seules qui retransmettent des émissions de *télé-réalité* et, étant donné qu’elles sont privées, du *téléachat*. Pour cette raison,

1. Ce type de programme TV est désigné récemment comme “réalité scénarisée” ou *scripted reality*

Genres	Fr5	TV5	TF1	M6	Gen.	Total
Actualité	0	38	4	4	4	46
Cartoon	26	0	17	12	14.5	55
Plateau/Débat	23	10	4	0	2	37
Documentaire	35	25	2	0	1	62
Fiction	1	6	45	39	42	91
Inter-prog.	95	134	124	101	112.5	454
Jeu	0	6	12	0	6	18
Autres mag.	4	5	0	5	2.5	14
Mag. de rep.	6	12	7	5	6	30
Météo	0	18	8	15	11.5	41
Musique	0	3	9	4	6.5	16
Prog. court	3	15	6	0	3	24
Téléachat	0	0	1	2	1.5	3
Télé-réalité	0	0	10	16	13	26
Total	193	272	249	203	226	917

TABLE 2: Nombre de segments pour chaque genre. *FR5* : France 5, *Gen.* : Moyenne des chaînes généralistes, *Inter-prog.* : Inter-programmes, *Mag.* : Magazine, *Rep.* : Reportages, *Prog.* : Programme.

elles disposent des plus grandes parts d'*inter-programmes* parmi les 4 chaînes. En effet, nous pouvons remarquer à partir de la figure 2 que les segments de *fiction* durent environ deux fois plus longtemps dans les chaînes plus spécialisées France 5 et TV5, deux chaînes semi-thématiques, que dans TF1 et M6. De plus, TV5 possède le plus grand nombre d'*inter-programmes*, comme le montre le tableau 2. Cependant, cela ne signifie pas forcément qu'elle possède la plus grande quantité d'*inter-programmes* vu le nombre relativement important d'émissions de courtes durées, comme les *journaux d'actualité*, les *prévisions météo* et *programmes courts*.

Malgré la similarité significative entre les politiques des deux chaînes généralistes, nous pouvons tout de même constater quelques différences. Par exemple, M6 consacre, tard dans la nuit, une portion de retransmission d'environ 4 heures par jour à la *musique*. D'un autre côté, TF1 se distingue par diverses émissions de jeu. Pour ce qui est des deux chaînes semi-thématiques, France 5 est spécialisée dans les *documentaires* et alloue aussi une part représentant quasiment un tiers de la programmation télévisuelle aux émissions de *débat*. Néanmoins, elle ne diffuse ni des *journaux* ni des *prévisions météo*. Ces deux derniers genres constituent par contre le thème principal de TV5 Monde.

4 Expériences

L'objectif de cet article est de construire un système d'aide à l'annotation manuelle. Ce dernier est évalué lors d'expériences de classification automatique en genres des programmes TV, présentées dans cette section, qui sont conduites sur le corpus TV multi-chaînes décrit dans la section 3. Le but de cette tâche est de prédire le genre d'une émission en cours, sachant les émissions qui la précèdent, à l'aide d'un modèle de n-grammes proposé dans la section 4.1, et d'évaluer ensuite la consistance de l'utilisation de plusieurs chaînes durant le processus de prédiction de genres.

4.1 La Prédiction du Genre

Le but de cette expérience est de découvrir le genre d'un programme télévisé en cours sachant les programmes précédents. Cette étude emploie un modèle n-gramme (Brown *et al.*, 1992) évaluant la probabilité *a posteriori* du $i^{\text{ème}}$ programme x_i étant donné les $n - 1$ programmes précédents. Ceci équivaut à évaluer la probabilité d'observer un genre particulier dans le contexte d'un historique restreint aux genres des $n - 1$ émissions précédentes.

$$p(x_i | x_{i-(n-1)}, \dots, x_{i-1}) = \frac{\text{count}(x_{i-(n-1)}, \dots, x_{i-1}, x_i)}{\text{count}(x_{i-(n-1)}, \dots, x_{i-1})} \quad (1)$$

La figure 3 illustre un exemple d'une séquence de deux émissions successives. Il est admis que l'étiquetage automatique des programmes n'est pas une tâche triviale. D'un côté, un modèle n-gramme (avec $n > 3$) va considérer les segments de deux *fictions* successives comme appartenant à une seule émission. D'un autre côté, si nous prenons en compte le label "inter-programmes" dans le calcul des probabilités (*i.e.* comme un genre d'émission de même niveau que les autres) et vu le nombre relativement grand des segments de *fiction*, un modèle n-gramme appris sur un historique assez court ($n < 3$) va produire uniquement une suite de *fictions* et d'*inter-programmes* (cf. extrait de programmation télévisée dans la figure 3).

4.2 Protocole Expérimental

Afin de simplifier cette tâche, des choix ont été effectués. D'une part, nous avons pris en compte uniquement les premiers genres de chaque label. D'autre part, nous avons joint les occurrences des trois sortes de génériques ambigus (cf. tableau 1) au segment d'*inter-programmes* voisin. Nous avons donc spécifié un ensemble de configurations ayant les mêmes données de test, à savoir, la deuxième journée de TF1, et dépendant des paramètres suivants (cf. tableau 3) :

- La taille du modèle n-gramme (de 3-grammes à 6-grammes),
- Les données utilisées lors de l'apprentissage du modèle de n-gramme,
- La prise en compte des *inter-programmes* au sein des données d'apprentissage et de test.

La première phase consiste à apprendre un modèle n-gramme en se restreignant à la première journée de TF1. Ensuite, une première expérience de la modélisation multi-chaînes est réalisée en combinant les données d'apprentissage de plusieurs chaînes. En ajoutant les deux journées de M6 à la première journée de TF1, nous voulons vérifier, sachant qu'il s'agit de deux chaînes concurrentes possédant des lignes éditoriales proches, si un modèle combiné pourra être plus performant. Enfin nous essayons de pousser encore le caractère multi-flux en combinant les deux chaînes restantes, à savoir France 5 et TV5 Monde, malgré leur divergence présumée par rapport aux deux autres chaînes généralistes.

4.3 Résultats

En entraînant le modèle sur la première journée de TF1, la meilleure performance est obtenue en employant les modèles 5-grammes et 6-grammes, et en excluant les *inter-programmes*. Jusqu'alors, le fait de ne pas inclure les *inter-programmes* était bénéfique, ceci étant probablement dû au fait que l'on enlève impact négatif qu'ils peuvent produire sur la taille effective de l'historique (les

inter-programmes sont présents entre chaque deux segments successifs de programmes) et de la masse de probabilité qu’ils peuvent monopoliser (particulièrement en ce qui concerne les probabilités uni-grammes). Ensuite, nous observons la manière dont un modèle multi-chaîne pourrait influencer le comportement du système. Malheureusement, au lieu de renforcer le modèle n-gramme, il semble que l’ajout des autres chaînes, mais aussi l’exclusion des *inter-programmes*, y insère de la “pollution”. Ceci est probablement dû à une divergence importante entre les chaînes utilisées. Dans la suite de ce travail, nous comptons étudier plus en détail ces résultats et conduire davantage d’expériences afin d’acquérir une meilleure compréhension de la tâche de prédiction de genres.

Taille	Inter-prog	Appr.	Précision	Rappel	F-mesure	TR
3g & 4g	inclus	TF1 1/2	0.473	0.484	0.478	0.169
5g			0.481	0.433	0.456	0.381
6g			0.489	0.551	0.518	0.195
3g	non inclus		0.537	0.462	0.497	0.367
4g			0.579	0.508	0.541	0.350
5g & 6g			0.588	0.513	0.548	0.333
3g	inclus	TF1 1/2 + M6	0.406	0.401	0.403	0.305
4g			0.503	0.447	0.473	0.280
5g			0.441	0.453	0.447	0.331
6g			0.476	0.430	0.452	0.280
3g	non inclus		0.402	0.363	0.382	0.383
4g			0.408	0.305	0.349	0.467
5g			0.408	0.305	0.349	0.467
6g			0.410	0.322	0.360	0.450
3g	inclus	TF1 1/2 + M6 + Fr5 + TV5	0.343	0.356	0.349	0.220
4g			0.379	0.384	0.382	0.297
5g			0.321	0.370	0.344	0.305
6g	0.330		0.3910	2.000	0.288	
3g	non inclus		0.317	0.343	0.329	0.450
4g			0.311	0.256	0.281	0.667
5g			0.326	0.288	0.306	0.550
6g			0.335	0.238	0.279	0.650

TABLE 3: Précisions, Rappels, F-mesures et Taux de Réussite (TR) observés pour différentes configurations d’apprentissage. 1/2 : jour 1, *inter-prog.* : inter-programmes, *Appr.* : données d’apprentissage.

5 Conclusion

Nous avons présenté dans cet article un corpus télévisé s’étalant sur une période de 2 jours ainsi que des méthodes fondées sur les n-grammes permettant de prédire l’étiquette d’une émission sachant les programmes précédents. Ce flux multi-chaîne a été, dans un premier temps, annoté manuellement avec le genre des émissions en utilisant une taxonomie que nous avons conçue auparavant. Diverses statistiques et caractéristiques de ce corpus ont été décrites dans ce travail. Par ailleurs, nous avons proposé un système préliminaire temps-réel d’aide à l’annotation manuelle de prédiction de genres d’émissions. Le meilleur modèle n-gramme a obtenu des résultats prometteurs en ce qui concerne la tâche de prédiction de genres malgré quelques déceptions. Les travaux futurs se concentreront sur les détails de l’expérience de prédiction de genre et établiront, dans une prochaine étape, un système temps-réel de structuration automatique des flux TV.

Références

- AMARAL R. & TRANCOSO I. (2003). Topic indexing of tv broadcast news programs. In *Computational Processing of the Portuguese Language*, p. 219–226. Springer.
- BOUCHEKIF A., DAMNATI G. & CHARLET D. (2013). Complementarity of lexical cohesion and speaker role information for story segmentation of french tv broadcast news. In *Statistical Language and Speech Processing*, p. 51–61. Springer.
- BREDIN H., LAURENT A., SARKAR A., LE V.-B., ROSSET S. & BARRAS C. (2014). Person instance graphs for named speaker identification in tv broadcast. In *Proceedings of Odyssey*.
- BROWN P. F., DESOUZA P. V., MERCER R. L., PIETRA V. J. D. & LAI J. C. (1992). Class-based n-gram models of natural language. *Computational linguistics*, **18**(4), 467–479.
- EL-KHOURY E., SÉNAC C. & JOLY P. (2010). Unsupervised segmentation methods of tv contents. *International Journal of Digital Multimedia Broadcasting*, **2010**.
- GIRAUDEL A., CARRÉ M., MAPELLI V., KAHN J., GALIBERT O. & QUINTARD L. (2012). The repere corpus : a multimodal corpus for person recognition. In *LREC*, p. 1102–1107.
- GRAVIER G., ADDA G., PAULSON N., CARRÉ M., GIRAUDEL A. & GALIBERT O. (2012). The etape corpus for the evaluation of speech-based tv content processing in the french language. In *LREC-Eighth international conference on Language Resources and Evaluation*, p.ñ.a.
- IBRAHIM Z. A. A. & GROS P. (2011). Tv stream structuring. *ISRN Signal Processing*, **2011**.
- INA (2002). Connaissance des fonds, histoire des programmes : par genre. **2**.
- LIANG L., LU H., XUE X. & TAN Y.-P. (2005). Program segmentation for tv videos. In *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, p. 1549–1552 : IEEE.
- LINARÈS G., NOCÉRA P., MASSONIE D. & MATROUF D. (2007). The lia speech recognition system : from 10xrt to 1xrt. In *Proceedings of the 10th international conference on Text, speech and dialogue*, p. 302–308 : Springer-Verlag.
- MANSON G. & BERRANI S.-A. (2010). Automatic tv broadcast structuring. *International journal of digital multimedia broadcasting*, **2010**.
- NATUREL X., GRAVIER G. & GROS P. (2007). Fast structuring of large television streams using program guides. In *Adaptive Multimedia Retrieval : User, Context, and Feedback*, p. 222–231. Springer.
- POLI J.-P. (2008). An automatic television stream structuring system for television archives holders. *Multimedia systems*, **14**(5), 255–275.
- SJÖLANDER K. & BESKOW J. (2000). Wavesurfer-an open source speech tool. In *INTERSPEECH*, p. 464–467.
- WANG J., DUAN L., LI Z., LIU J., LU H. & JIN J. S. (2006). A robust method for tv logo tracking in video streams. In *Multimedia and Expo, 2006 IEEE International Conference on*, p. 1041–1044 : IEEE.
- WANG J., DUAN L., LIU Q., LU H. & JIN J. S. (2008). A multimodal scheme for program segmentation and representation in broadcast video streams. *Multimedia, IEEE Transactions on*, **10**(3), 393–408.