Combinaison d'approches pour la reconnaissance du rôle des locuteurs

Richard Dufour[1] Antoine Laurent [2, 3] Yannick Estève [2]
(1) France Telecom R&D - Orange Labs
(2) LIUM - Université du Maine
(3) Spécinov - Trélazé
richard.dufour@orange.com, antoine.laurent@lium.univ-lemans.fr,
yannick.esteve@lium.univ-lemans.fr

RÉSUMÉ

Dans cet article, nous nous intéressons à la reconnaissance des rôles de locuteurs dans des émissions radiophoniques d'information. Des travaux antérieurs mettent en avant l'existence d'une relation entre la spontanéité de la parole et les rôles des locuteurs. Un système de détection automatique de la parole spontanée a déjà été appliqué pour reconnaître des rôles, sans adaptation ni modification de la méthode (Dufour et al., 2011). Nous proposons d'améliorer cette méthode en ajoutant des marqueurs propres à la détection du rôle. Ainsi, des caractéristiques extraites à partir de l'Analyse des Réseaux Sociaux (SNA) ont été utilisées dans le processus de décision. Cette nouvelle source d'informations a permis un gain aux deux niveaux de décision de la méthode, en améliorant les performances de 3,2 et 1,9 points en absolu, respectivement pour la décision *locale* puis *globale*. Au final, la méthode proposée permet d'associer le rôle correct à 76,3 % des locuteurs, pour un nombre de rôle supérieur à celui régulièrement retenu dans ce type de travaux.

ABSTRACT

Combination of approaches for speaker role recognition

In this article, we are particularly interested in recognizing speaker role inside broadcast news shows. Previous studies highlighted a link between speech spontaneity and speaker roles. An automatic spontaneous speech detection system has already been applied to recognize speaker roles, without any change in the method process (Dufour et al., 2011). We propose to improve this method by adding specific speaker role features. Thus, features from Social Network Analysis (SNA) are used in the method. This new information allowed to improve the two decision steps oh the method, with a gain of 3.2 and 1.9 points in absolute, respectively for the *local* then the *global* decision. Finally for a larger number of focused roles than usually retained, the proposed method allowed to associate the correct role to 76.3% of the speakers.

MOTS-CLÉS : indexation automatique, analyse des réseaux sociaux, parole spontanée, rôle du locuteur.

KEYWORDS: document indexing, social network analysis, spontaneous speech, speaker role recognition.

Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 1: JEP, pages 827–834,
Grenoble, 4 au 8 juin 2012.

1 Introduction

La recherche d'information à partir de masses de données audio nécessite l'extraction de son contenu linguistique et peut être affinée par une structuration automatique du document. Face au nombre croissant de documents multimédia hétérogènes disponibles sur Internet, ce domaine devient de plus en plus important mais augmente également en complexité. Pouvoir fournir des informations supplémentaires à ces données brutes pourrait, par exemple, être utile dans le cadre d'indexation automatique de documents (Amaral et Trancoso, 2003). Dans cet article, nous nous intéressons particulièrement à la reconnaissance des rôles de locuteurs dans des émissions radiophoniques d'information. Des solutions ont déjà été proposées pour cette tâche spécifique, comme celles initiées par (Barzilay *et al.*, 2000).

Dans ce travail, nous proposons d'identifier 10 rôles de locuteurs, puisqu'au sein d'émissions radiophoniques d'information se rencontrent des rôles divers variant en fonction du type d'émissions (débats, interviews, chroniques...). Nous continuons les travaux initiés dans (Dufour *et al.*, 2011) qui ont mis en avant l'existence d'une relation entre la parole spontanée et les rôles des locuteurs. Le système de détection automatique de la parole spontanée développé dans (Dufour *et al.*, 2009) a permis de montrer qu'une reconnaissance des rôles était possible au moyen de cet outil, sans modification majeure de la méthode. Bien que cette méthode ait permis d'atteindre des performances acceptables, il est probable que la prise en compte d'éléments provenant de travaux déjà initiés dans le domaine de la reconnaissance des rôles permettrait encore d'améliorer les résultats. En effet, la méthode de détection issue de la parole spontanée utilise des bases de connaissances totalement différentes de celles généralement manipulées pour les rôles. Nous proposons alors d'ajouter des caractéristiques issues de l'Analyse des Réseaux Sociaux (Vinciarelli, 2007), approche déjà appliquée avec succès dans la reconnaissance des rôles des locuteurs.

La partie suivante présente les travaux antérieurs réalisés dans ce domaine. Nous verrons dans la partie 3 la méthode que nous proposons pour détecter les 10 rôles de locuteurs étudiés, puis dans la partie 4 les résultats obtenus.

2 Travaux antérieurs

Deux niveaux d'information sont généralement utilisés pour identifier les rôles de locuteurs, à savoir l'utilisation de caractéristiques acoustiques / prosodiques (Salamin *et al.*, 2009; Bigot *et al.*, 2010) ou de caractéristiques lexicales (Barzilay *et al.*, 2000; Damnati et Charlet, 2011).

Ces travaux proposent généralement un processus de classification automatique dans l'optique d'assigner un rôle à chaque locuteur. Dans (Barzilay *et al.*, 2000), les auteurs proposent d'associer à chaque locuteur un des trois rôles définis (*Présentateur*, *Journaliste* et *Invité*) au moyen d'un algorithme de *boosting* et un modèle d'entropie maximum manipulant des caractéristiques lexicales (n-grammes de mots) et la durée des segments. Ces caractéristiques ont été extraites à partir de transcriptions automatiques fournies par un système de reconnaissance de la parole. À cela s'ajoute également l'utilisation d'informations contenues au voisinage du segment à catégoriser (n-grammes de mots, durées...). En utilisant des transcriptions automatiques et des tours de parole étiquetés manuellement, la méthode proposée par (Barzilay *et al.*, 2000) a permis d'atteindre un taux de bonne classification de 80% des rôles au niveau des segments de parole. Une approche similaire a été récemment développée par (Damnati et Charlet, 2011), où les auteurs ont également cherché à catégoriser trois rôles dans des émissions d'informations télévisées. Les expériences ont, cette fois-ci, été réalisées à partir d'un système totalement

automatique, que ce soit au niveau de la transcription, de la segmentation et du regroupement en locuteurs. Un taux de bonne classification de 86% des tours de parole a été atteint. Dans (Liu, 2006), les auteurs utilisent des modèles de Markov cachés et un modèle d'entropie maximum sur des transcriptions, des tours de parole ainsi que des rôles manuellement annotés. La combinaison des deux modèles a permis de correctement catégoriser 80% des tours de parole. Les auteurs dans (Salamin *et al.*, 2009) proposent d'utiliser des caractéristiques temporelles ainsi que des caractéristiques extraites d'un réseau d'affiliation sociale construit à partir d'un regroupement en locuteurs obtenu automatiquement. Les travaux réalisés par (Bigot *et al.*, 2010) choisissent de catégoriser jusqu'à cinq rôles en utilisant des caractéristiques temporelles et en ajoutant des caractéristiques acoustiques et prosodiques. Ces caractéristiques associées à un algorithme de classification supervisée ont permis d'attribuer le bon rôle à 92% des locuteurs.

La structure globale d'une émission est également prise en compte pour la détection des rôles des locuteurs. Dans (Vinciarelli, 2007), les auteurs introduisent le concept de l'Analyse des Réseaux Sociaux – *Social Network Analysis* (SNA) – pour la reconnaissance du rôle des locuteurs. Dans (Vinciarelli, 2007; Salamin *et al.*, 2009), SNA est combiné avec différentes approches utilisant des durées d'interaction associées à chaque locuteur. Grâce à cette combinaison, 85% du temps d'intervention des locuteurs a été assigné à un rôle correct (six rôles et onze locuteurs). De plus, l'approche SNA a également été appliquée avec succès dans (Garg *et al.*, 2008), où les auteurs l'ont combinée avec un algorithme de classification (*AdaBoost*) utilisant des informations lexicales. L'approche a permis de correctement catégoriser 70% des rôles en termes de durée. Notons que la plupart de ces études cherchent à reconnaître de trois (Barzilay *et al.*, 2000; Damnati et Charlet, 2011) à six (Salamin *et al.*, 2009) rôles au maximum.

3 Méthode proposée

3.1 Détection automatique de la parole spontanée

Une méthode permettant de détecter automatiquement la parole spontanée dans des documents audio a été proposée par (Dufour *et al.*, 2009). L'objectif de cet outil est d'associer à chaque segment de parole une des trois classes de spontanéité : parole *préparée*, *faiblement spontanée* et *fortement spontanée*. Des caractéristiques acoustiques (durées des voyelles, durées phonémiques, pitch...) et linguistiques (nombre de répétitions et de noms propres, taille du découpage syntaxique...) sont extraites pour chaque segment à partir d'un système de transcription automatique. La détection est réalisée au moyen d'une méthode statistique à deux niveaux :

– *Décision locale* : les caractéristiques acoustiques et linguistiques extraites pour chaque segment de parole sont utilisées dans un processus de classification, associant une classe de spontanéité à chaque segment. L'outil de classification automatique utilisé est *IcsiBoost* [1], un outil *open-source* proche du programme de classification *Boostexter* (Schapire et Singer, 2000), classifieur à large marge reposant sur la méthode de *boosting*.
– *Décision globale* : utilisation d'un modèle contextuel probabiliste prenant en compte les segments voisins. Des machines à états-finis sont utilisées pour réestimer les probabilités de classification obtenues pour chaque segment, en prenant en compte les résultats de classification des segments précédents et suivants.

L'outil de détection permet d'atteindre une précision de 69,3 % et un rappel de 74,6 % sur les segments de parole *fortement spontanée*.

1. `http://code.google.com/p/icsiboost/`

Dans (Dufour *et al.*, 2011), une analyse a permis de mettre en lumière le lien existant entre le type de parole et le rôle d'un locuteur. Ainsi, par exemple, un présentateur a tendance à préparer son discours, au contraire d'un invité, dont la parole est beaucoup moins fluide, que l'on qualifie plutôt de fortement spontanée. Ces travaux proposent d'appliquer directement l'outil de détection automatique de la parole spontanée pour détecter les rôles des locuteurs dans des émissions radiophoniques, sans modification ni adaptation de la méthode (caractéristiques et niveaux de décision identiques). La seule différence se situe au niveau de la taille des données à catégoriser : la détection ne se fait plus au niveau du segment (durée moyenne de 20 secondes) mais au niveau de l'ensemble des interventions d'un même locuteur.

3.2 Analyse des Réseaux Sociaux (SNA)

L'analyse des réseaux sociaux (SNA) consiste à déterminer la position de chaque locuteur dans le dialogue. L'idée défendue par les SNA est qu'un rôle spécifique (un *acteur*) interagit avec d'autres *acteurs* pendant des *événements*. Ces interactions peuvent aider à l'identification du rôle associé à chaque locuteur impliqué dans le réseau. L'objectif de cette méthode est d'être capable de déterminer la *centralité* de chaque locuteur (Vinciarelli, 2007) par rapport aux autres locuteurs dans une émission, en considérant que le locuteur i dialogue avec le locuteur j si j intervient juste après i dans la transcription. En nous inspirant de (Vinciarelli, 2007), nous proposons de calculer la centralité selon l'équation :

$$C_i = \frac{\sum_{j=1}^{nb} \chi D_{i,j}}{\sum_{j=1}^{nb} D_{i,j}} \quad (1)$$

Avec $\chi = 1$ si $D_{i,j} = 1$, et $\chi = 0$ sinon, C_i la centralité de i, nb le nombre de locuteurs et $D_{i,j}$ la distance entre i et j. Cette distance est exprimée en nombre de liens (orientés) à parcourir pour atteindre chaque noeud. Dans l'exemple de réseau social affiché dans la figure 1, $D_{1,2} = 1$ et $D_{1,3} = 2$. Certains noeuds ont une distance "infini" avec les autres : c'est le cas du noeud correspondant au locuteur 5 dans l'exemple. Dans ce cas, la distance entre ce noeud et les autres est égale à $nb + 1$.

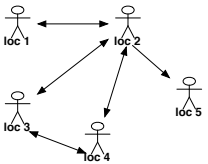

FIGURE 1 – Exemple d'un réseau social

3.3 Intégration du SNA

Nous avons choisi d'intégrer les caractéristiques extraites de l'analyse des réseaux sociaux au niveau de la décision *locale* de la méthode de détection de la parole spontanée. En effet, les caractéristiques du SNA peuvent être facilement intégrées dans le processus de classification, au même niveau que les caractéristiques de la parole spontanée : au lieu de combiner *a posteriori* les deux approches, l'algorithme de *boosting* définira et pondèrera les règles utiles à partir de l'ensemble des caractéristiques proposées pour catégoriser les rôles. La centralité sera ajoutée, ainsi que la *couverture* et le temps de parole de chaque locuteur. La couverture du locuteur correspond au temps écoulé entre sa première et sa dernière intervention dans l'émission.

4 Expériences

4.1 Corpus

Le projet EPAC [2] (Estève *et al.*, 2010) concerne le traitement de données audio non structurées. L'objectif principal de ce projet a été de proposer des méthodes d'extraction d'information et de structuration de documents spécifiques aux données audio. Les données audio traitées durant le projet EPAC proviennent d'émissions radiophoniques enregistrées entre 2003 et 2004 au sein de trois radios françaises : France Info, France Culture et RFI. Au cours de ce projet, 100 heures de données audio ont été manuellement annotées, avec principalement des émissions contenant une forte proportion de parole spontanée (interviews, débats, talk shows...). Le corpus EPAC inclut des informations supplémentaires au niveau des locuteurs et des émissions transcrites. Plus précisément, le rôle, la fonction et la profession de chaque locuteur ont été manuellement annotés selon la disponibilité des informations. Ainsi, un même locuteur a un seul rôle général (par exemple *Invité*, *Interviewé*, *Commentateur*...) mais qui peut être affiné avec un maximum de deux autres étiquettes (par exemple, pour un *Invité* : *politicien* / *premier ministre*). L'intégralité du corpus a été manuellement annotée en rôles par un expert linguiste.

TABLE 1 – Détails du corpus EPAC manuellement annoté en rôles de locuteurs

Rôle du locuteur	**# Locuteurs**	**# Tours de parole**	**Durée**
Auditeur	238	424	3h09
Chroniqueur	135	182	4h19
Envoyé spécial	85	113	1h38
Expert	151	1,527	16h23
Invité	134	2,813	26h46
Interviewé	116	438	4h02
Intervieweur	31	227	0h30
Journaliste	11	18	0h10
Présentateur	191	5,223	19h46
Autre	45	220	1h47
Total	1,137	11,125	78h30

Le tableau 1 présente la proportion (en nombre de locuteurs, nombre de tours de parole et durée) des 10 rôles étudiés et annotés manuellement dans le projet EPAC. Nous pouvons voir que les rôles *Expert*, *Invité*, et *Présentateur* sont les plus représentés en termes de nombre de locuteurs, de tours de parole et de durée. Ce constat semble normal puisque ce corpus contient principalement des données issues d'émissions radiophoniques d'information. La grande variété de genres des émissions de ce corpus (*journal*, *reportage*, *chronique*, *débat*...) a conduit l'annotateur à définir des rôles très précis. Les définitions suivantes sont données pour chaque rôle :

– **Auditeur** : intervenant occasionnellement pendant une émission au téléphone, dans un environnement bruyant, et prépare peu ses interventions, avec une parole plutôt spontanée.
– **Chroniqueur** : rapporte et analyse des événements qui font l'actualité pendant une émission. Les interventions d'un chroniqueur sont très largement préparées.
– **Envoyé spécial** : journaliste particulier contribuant sur des sujets à distance (par téléphone par exemple). Ce rôle n'apparaît que dans le cadre de journaux d'information.
– **Expert** : possède des connaissances poussées dans un domaine particulier. Il apporte de nouvelles informations, souvent techniques, sur un sujet précis.

2. `http://projet-epac.univ-lemans.fr`

– **Interviewé** : répond aux questions soulevées par un intervieweur ou un présentateur.
– **Intervieweur** : dirige une interview. Dans le corpus EPAC, peu de locuteurs ont été identifiés en tant qu'intervieweur : le rôle de *Présentateur* est souvent préféré.
– **Invité** : intervient pour discuter autour de sujets de société. Il peut apparaître dans une émission en parallèle d'un expert, mais n'a pas une connaissance poussée du sujet.
– **Journaliste** : réalise des reportages sur des sujets ou faits divers particuliers. Au cours d'une émission, un journaliste ne rapporte des évènements que sur un sujet précis.
– **Présentateur** : possède de multiples fonctions au sein d'une émission et prépare ses interventions : il peut animer un débat, interagir avec les auditeurs, présenter les informations...
– **Autre** : tous les rôles restants et qui ne sont pas étudiés (manque de données).

4.2 Résultats

Afin d'évaluer la performance du système de reconnaissance de rôles de locuteurs, nous avons suivi la méthode du *Leave One Out* sur les 121 fichiers du corpus EPAC : 120 fichiers ont été utilisés pour l'apprentissage, 1 fichier pour l'évaluation, et le processus a été répété jusqu'à évaluation de tous les fichiers. La segmentation manuelle et le découpage en locuteurs de référence ont été utilisés : nous savons exactement qui parle et quand. Les résultats présentés évalueront seulement le processus de reconnaissance des rôles, les problèmes induits par la segmentation automatique et le regroupement en locuteurs ne sont pas étudiés dans ces travaux. L'évaluation se fera donc en fonction de l'attribution du rôle pour chaque locuteur. Les transcriptions automatiques ainsi que l'extraction des paramètres acoustiques et linguistiques nécessaires à la reconnaissance des rôles ont été fournis par le système de transcription automatique du LIUM (Deléglise *et al.*, 2009).

Les expériences menées cherchent à évaluer la performance de la détection de rôles au moyen de caractéristiques issues de l'analyse des réseaux sociaux associées à des caractéristiques de la parole spontanée. Le tableau 2 présente les performances de la décision *locale* du système de reconnaissance des rôles (rappel et précision) en utilisant les caractéristiques issues du réseau social seules (*SNA*), les caractéristiques de la parole spontanée seules (*Sponta*) et enfin l'association des deux (*Sponta+SNA*).

TABLE 2 – Performances de la décision locale en utilisant les caractéristiques issues du réseau social seules (*SNA*), les caractéristiques de la parole spontanée seules (*Sponta*) et enfin les deux combinées (*Sponta+SNA*)

Décision	**SNA**		**Sponta**		**Sponta+SNA**	
Locale	Rappel	Précision	Rappel	Précision	Rappel	Précision
Auditeur	88,2	74,7	92,4	92,8	**94,1**	89,6
Chroniqueur	37,1	34,3	60,7	57,8	**66,6**	**66,6**
Envoyé spécial	15,3	22,0	56,5	53,3	**62,4**	**61,6**
Expert	78,2	69,4	73,5	71,2	76,8	**80,0**
Interviewé	32,8	31,4	51,7	45,5	**56,0**	**50,8**
Intervieweur	9,7	23,1	61,3	57,6	**61,4**	**61,3**
Invité	20,9	26,7	61,2	66,1	**62,7**	62,7
Journaliste	18,2	66,7	18,2	50,0	**36,4**	**66,7**
Présentateur	71,7	65,9	93,2	90,8	**95,3**	**91,0**
Autre	18,2	19,4	17,8	34,8	**20,0**	**40,9**

Nous remarquons tout d'abord que l'approche *SNA* permet de correctement classifier les rôles les plus représentés (*Auditeur*, *Expert* et *Présentateur*). Bien entendu, ces caractéristiques seules ne sont pas suffisantes pour reconnaître correctement tous les rôles étudiés, ce que l'approche

Sponta permet avec des caractéristiques plus étendues (acoustiques+linguistiques). Au final, nous constatons que, pour la majorité des rôles, la combinaison de ces approches *Sponta+SNA* permet d'améliorer les performances. Les améliorations sont particulièrement visibles pour les rôles *Journaliste*, *Expert*, *Envoyé spécial* ou *Chroniqueur*. Notons une légère baisse de la précision pour les rôles *Auditeur* et *Invité* qui est cependant compensée par une amélioration du rappel. Au niveau de cette décision locale, un gain de 3,2 points en absolu est constaté grâce au SNA, passant de 71,2 % de locuteurs assignés avec son rôle correct (*Sponta*) à 74,4 % (*Sponta+SNA*). Ce gain est principalement dû au fait que le SNA apporte des informations inexploitées au niveau de la décision locale : il permet l'analyse des interactions des locuteurs dans tout le document, alors que les caractéristiques issues de la parole spontanée ne s'intéressent qu'au locuteur courant.

Nous nous sommes ensuite intéressés à la décision globale de la méthode, prenant en compte les décisions prises au niveau des locuteurs voisins. Le tableau 3 présente les performances de la décision globale du système de reconnaissance des rôles utilisant les caractéristiques de la parole spontanée seules (*Sponta*) et intégrant l'analyse des réseaux sociaux (*Sponta+SNA*).

TABLE 3 – Performances de la décision globale des locuteurs en utilisant d'une part les caractéristiques de la parole spontanée seules (*Sponta*) et d'autre part en intégrant l'analyse du réseau social (*Sponta+SNA*)

Décision Globale	**Sponta**		**Sponta+SNA**	
	Rappel	Précision	Rappel	Précision
Auditeur	91,6	95,6	**92,4**	94,0
Chroniqueur	63,7	59,3	**65,2**	**67,7**
Envoyé spécial	52,9	56,3	**56,5**	**60,8**
Expert	82,1	72,1	**83,4**	**80,3**
Interviewé	56,0	52,9	**62,1**	**56,7**
Intervieweur	64,5	95,2	**71,0**	75,9
Invité	69,4	65,0	**70,2**	61,4
Journaliste	9,1	33,0	**18,2**	**50,0**
Présentateur	96,3	92,0	**97,4**	91,6
Autre	22,2	45,5	20,0	42,9

Nous constatons toujours une amélioration des performances, mais dans une proportion moindre. En effet, de nombreux rôles voient toujours leurs performances s'améliorer (*Chroniqueur*, *Envoyé spécial*, *Expert*...) que ce soit en termes de précision ou de rappel. L'impact des caractéristiques du SNA est cependant faible, voire nul, pour les rôles *Auditeur* ou *Présentateur*, et même négatif pour le rôle *Intervieweur*, avec une chute de la précision. Ces constats peuvent s'expliquer par le fait que les performances de certains rôles étaient déjà très élevées mais aussi parce que la décision globale utilise déjà des informations au niveau de la "structure" du document : l'enchaînement des tours de parole des locuteurs est prise en compte. Globalement, le système passe de 74,4 % de locuteurs assignés avec le rôle correct (*Sponta*) à 76,3 % *(Sponta+SNA*). Notons que la méthode *Sponta+SNA*, avec une décision locale seule, permet d'atteindre la même performance globale que la méthode *Sponta* mais où les deux niveaux de décision (locale + globale) sont nécessaires.

5 Conclusion et perspectives

Dans cet article, nous avons proposé une méthode de reconnaissance automatique de rôles de locuteur dans des émissions radiophoniques d'information (journaux, débats, interviews...). Nous avons ainsi continué les travaux initiés dans (Dufour *et al.*, 2011), proposant d'appliquer

directement une méthode de détection de la parole spontanée pour la détection de rôles, en enrichissant les caractéristiques au moyen d'une analyse des réseaux sociaux (SNA). L'utilisation de cette nouvelle source d'informations a permis un gain aux deux niveaux de décision de la méthode, en améliorant de 3,2 points en absolu au niveau de la décision *locale* et 1,9 points en absolu au niveau de la décision *globale*. Les caractéristiques de l'analyse des réseaux sociaux permettent de fournir des informations au niveau de l'interaction des rôles de locuteur dans tout le document, ce qui n'était auparavant pas exploité dans la décision locale de l'approche initiale. Lors du passage à la phase de décision globale de la méthode de détection des rôles, les gains sont moins importants puisque des informations au voisinage de chaque tour de parole étaient déjà exploitées. Dans des travaux futurs, nous nous intéresserons à la mise en place d'une méthode de détection des rôles complètement automatique, avec une segmentation et un regroupement en locuteurs automatiques. Nous pouvons également penser à intégrer d'autres caractéristiques encore inexploitées, telles que les interactions relatives pouvant être extraites du SNA.

Références

Amaral, R. et Trancoso, I. (2003). Segmentation and indexation of broadcast news. *In ISCA Workshop on Multilingual Spoken Document Retrieval (MSDR)*, pages 31–36, Hong Kong, Chine.

Barzilay, R., Collins, M., Hirschberg, J. et Whittaker, S. (2000). The rules behind roles : Identifying speaker role in radio broadcasts. *In AAAI*, pages 679–684.

Bigot, B., Ferrané, I., Pinquier, J. et André-Obrecht, R. (2010). Speaker role recognition to help spontaneous conversational speech detection. *In Searching Spontaneous Conversational Speech*, Italie.

Damnati, G. et Charlet, D. (2011). Robust speaker turn role labeling of tv broadcast news shows. *In ICASSP*, Prague, République Tchèque.

Deléglise, P., Estève, Y., Meignier, S. et Merlin, T. (2009). Improvements to the LIUM French ASR system based on CMU Sphinx : what helps to significantly reduce the word error rate ? *In Interspeech*, pages 2123–2126, Brighton, Angleterre.

Dufour, R., Estève, Y. et Deléglise, P. (2011). Investigation of spontaneous speech characterization applied to speaker role recognition. *In Interspeech*, Florence, Italie.

Dufour, R., Estève, Y., Deléglise, P. et Béchet, F. (2009). Local and global models for spontaneous speech segment detection and characterization. *In ASRU*, Merano, Italie.

Estève, Y., Bazillon, T., Antoine, J.-Y., Béchet, F. et Farinas, J. (2010). The EPAC corpus : manual and automatic annotations of conversational speech in French broadcast news. *In LREC*, Valletta, Malte.

Garg, P. N., Favre, S., Salamin, H., Hakkani-Tür, D. et Vinciarelli, A. (2008). Role recognition for meeting participants : an approach based on lexical information and social network analysis. *In ACM Multimedia Conference (MM'08)*, pages 693–696, Vancouver, Canada.

Liu, Y. (2006). Initial study on automatic identification of speaker role in broadcast news speech. *In Human Language Technology Conference of the NAACL*, pages 81–84, New York, USA.

Salamin, H., Favre, S. et Vinciarelli, A. (2009). Automatic role recognition in multiparty recordings : Using social affiliation networks for feature extraction. *In IEEE Transactions on Multimedia*, volume 11.

Schapire, R. E. et Singer, Y. (2000). BoosTexter : A boosting-based system for text categorization. *Machine Learning*, 39:135–168.

Vinciarelli, A. (2007). Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. *IEEE Transaction on Multimedia*, 9(6):1215–1226.