Reconnaissance Automatique de Locuteurs à l'aide de Fonctions de Croyance

Simon Petitrenaud, Vincent Jousse, Sylvain Meignier, Yannick Estève

Laboratoire d'Informatique de l'Université du Maine

Reconnaissance des Formes et Intelligence Artificielle (RFIA10), 20-22 janvier 2010

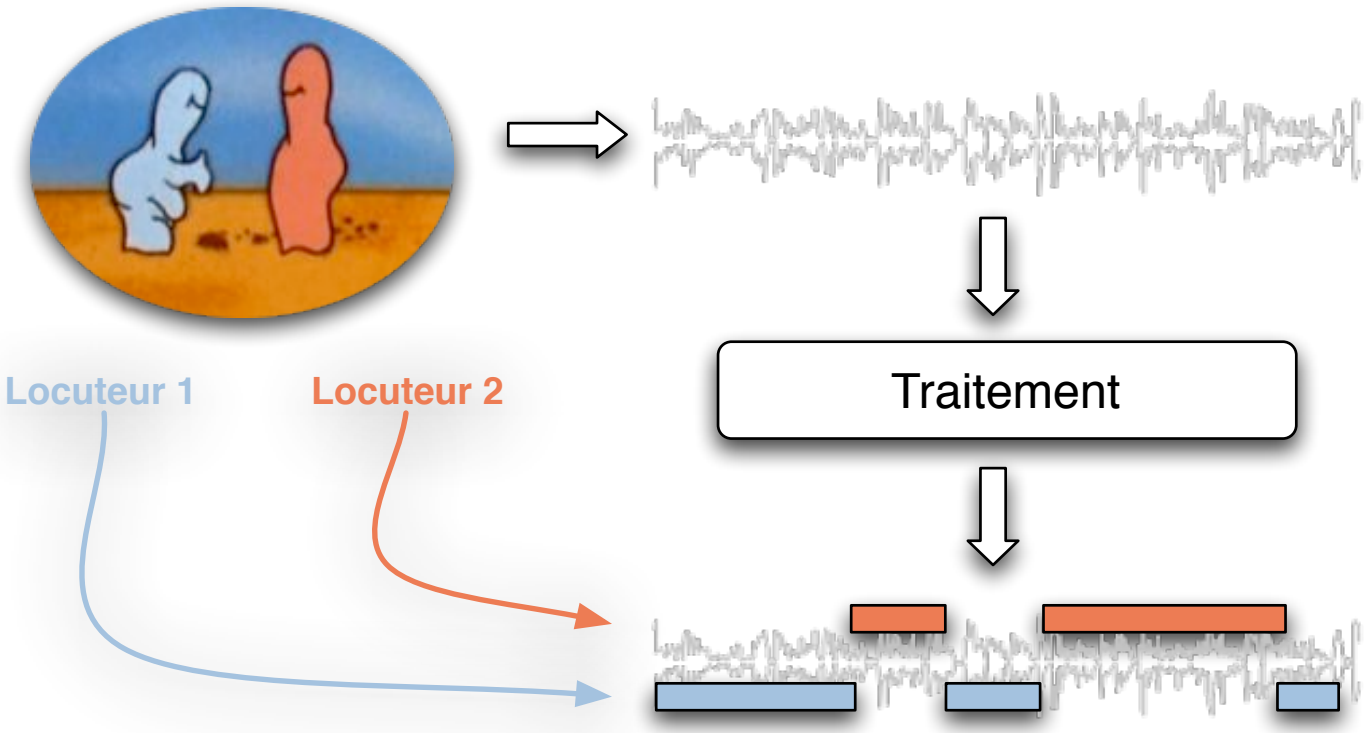
Lignes directrices

1. **Identification nommée de locuteur**
   - Objectif général
   - Identification de locuteur: état de l'art
   - Système de référence
2. **Fonctions de croyance et identification de locuteur**
   - Théorie des fonctions de croyance ("Dempster-Shafer")
   - Utilisation en reconnaissance de locuteur
3. **Evaluation**
   - Campagne ESTER
   - Métriques utilisées
   - Résultats
4. **Conclusion et perspectives**

Lignes directrices

Contexte et buts de l'identification nommée

- Contexte: émissions radiophoniques (ou télévisuelles)
- Objectifs: déterminer qui parle, et à quel moment



Lignes directrices

Etat de l'art

- Méthodes basées sur l'acoustique
  - Reconnaissance automatique du locuteur
  - Enregistrements de chaque locuteur $\Rightarrow$ Difficile à obtenir

Etat de l'art

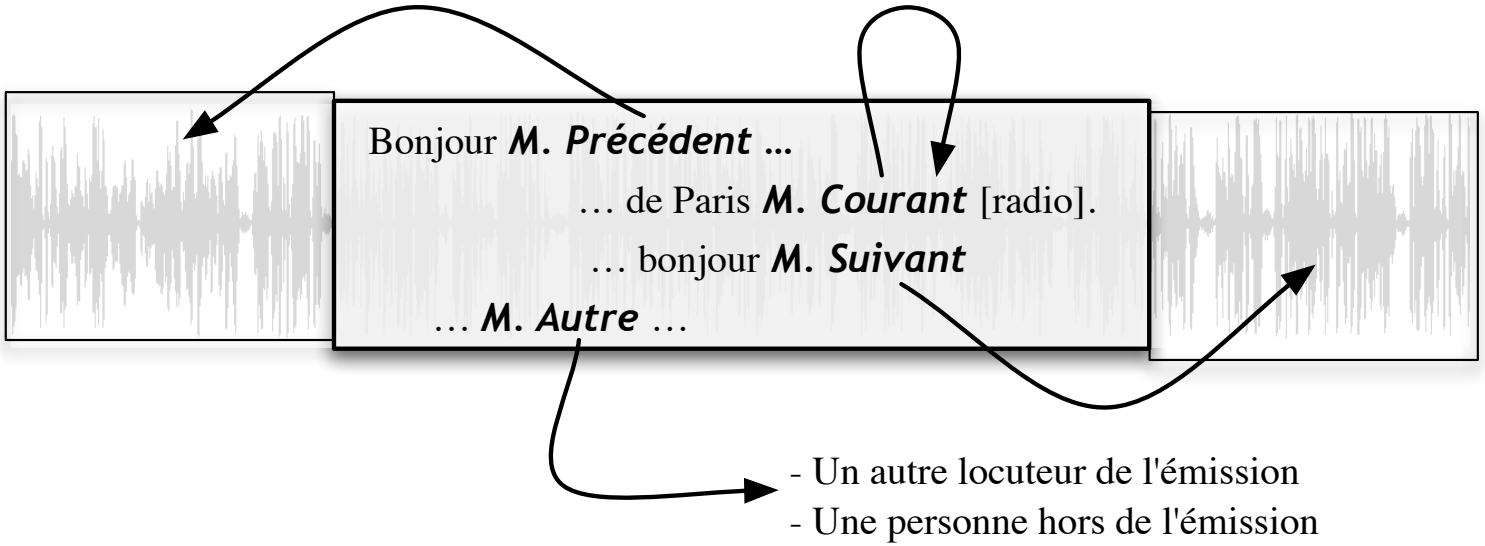
- Méthodes basées sur l'acoustique
  - Reconnaissance automatique du locuteur
  - Enregistrements de chaque locuteur $\Rightarrow$ Difficile à obtenir
- Méthodes utilisant la transcription du signal
  - Hypothèse : les locuteurs s'annoncent
  - Extraction des noms des locuteurs à partir des paroles prononcées
  - Reconnaissance automatique de la parole
  - Détection d'"*entités nommées*" : PERSONNES, Lieux, Radios ...

Etat de l'art

- Méthodes basées sur l'acoustique
  - Reconnaissance automatique du locuteur
  - Enregistrements de chaque locuteur $\Rightarrow$ Difficile à obtenir
- Méthodes utilisant la transcription du signal
  - Hypothèse : les locuteurs s'annoncent
  - Extraction des noms des locuteurs à partir des paroles prononcées
  - Reconnaissance automatique de la parole
  - Détection d'"*entités nommées*" : PERSONNES, Lieux, Radios ...

$\Rightarrow$ **Solution retenue: transcription du signal+ détection de Personnes**

Identification de personnes: principe de base



4 hypothèses sur la personne nommée (Canseco 05)

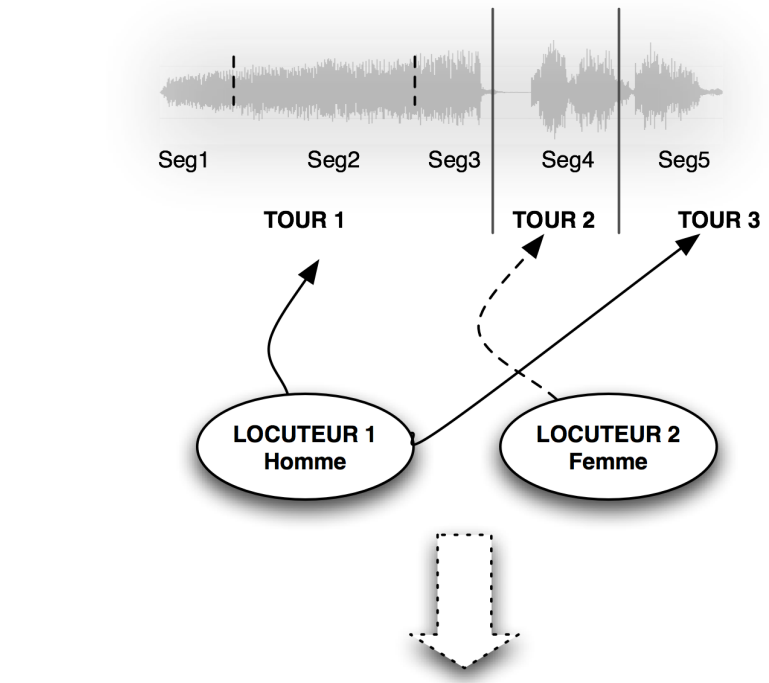
- Précédent: elle vient de parler
- Courant: elle parle
- Suivant: elle va parler
- Autre

Lignes directrices

Système de référence (revue TAL 09, ICASSP'09)

- Traitement acoustique
- Transcription de la parole en mots, détection de noms complets
- Hypothèses sémantiques sur les noms complets
- Méthode de combinaison, propagation des informations
- Processus de décision d'affectation

Transcription enrichie d'un document sonore



Signal sonore

Segments

Tours de parole

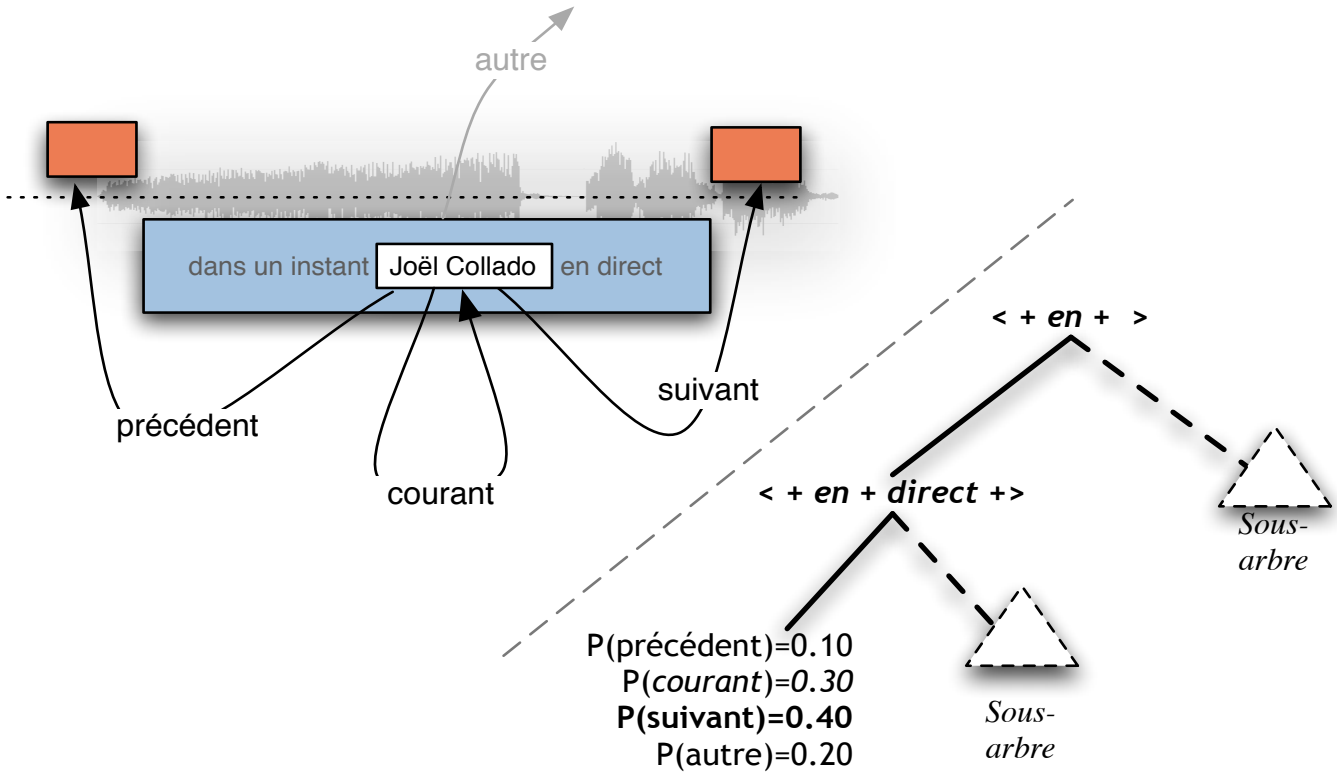
Classes de locuteur anonymes avec leur genre

TOUR 1 (Locuteur 1/Homme) : les grands titres de l'actualité **PERSONNE***(Maude Bayeu)* bonjour
TOUR 2 (Locuteur 2/Femme) : bonjour, la polémique sur la mission de **PERSONNE***(Didier Julia)* pour libérer les otages en **LIEU***(Irak)* [...]
TOUR 3 (Locuteur 1/Homme) : Merci **PERSONNE***(Maude Bayeu)*, maintenant la météo [...]

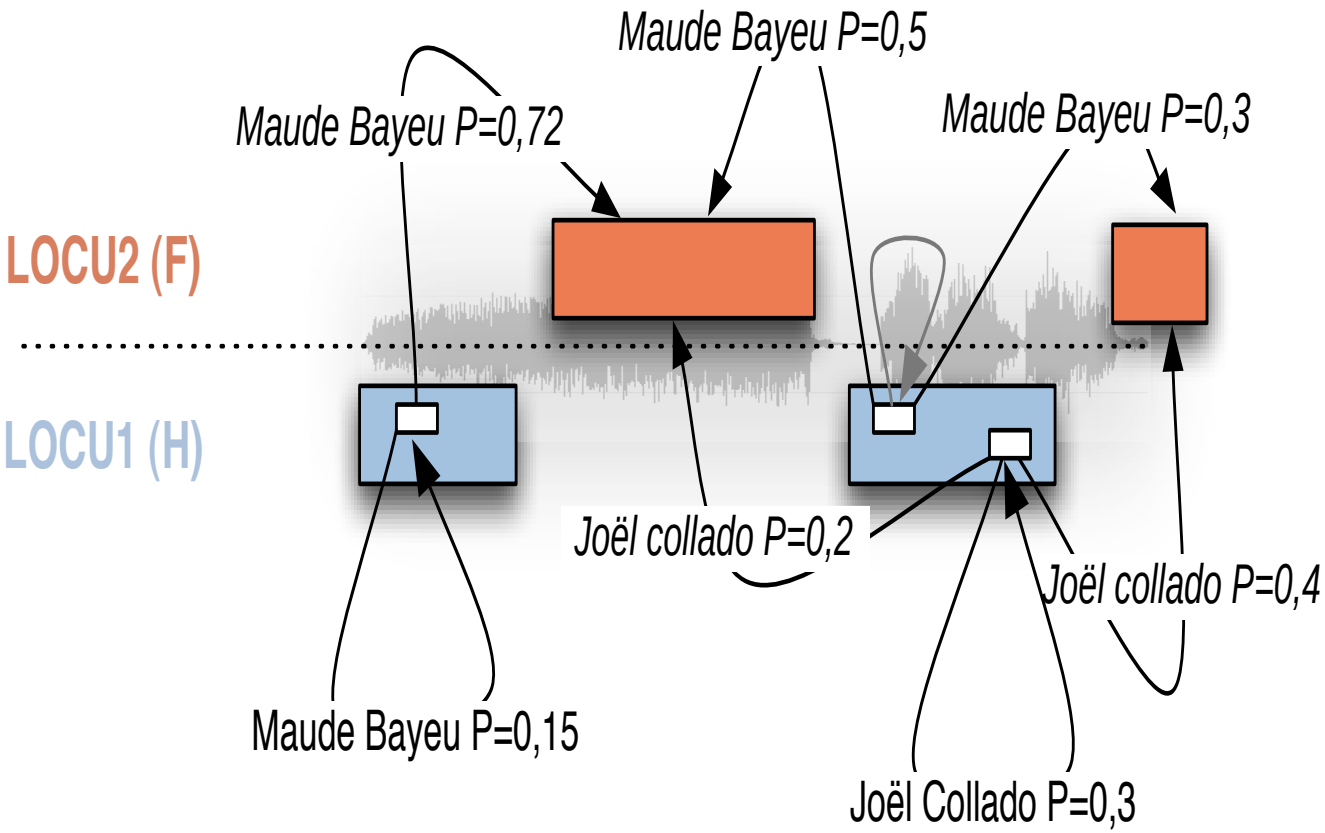
Transcription enrichie annotée en entités nommées

Utilisation du contexte lexical des noms

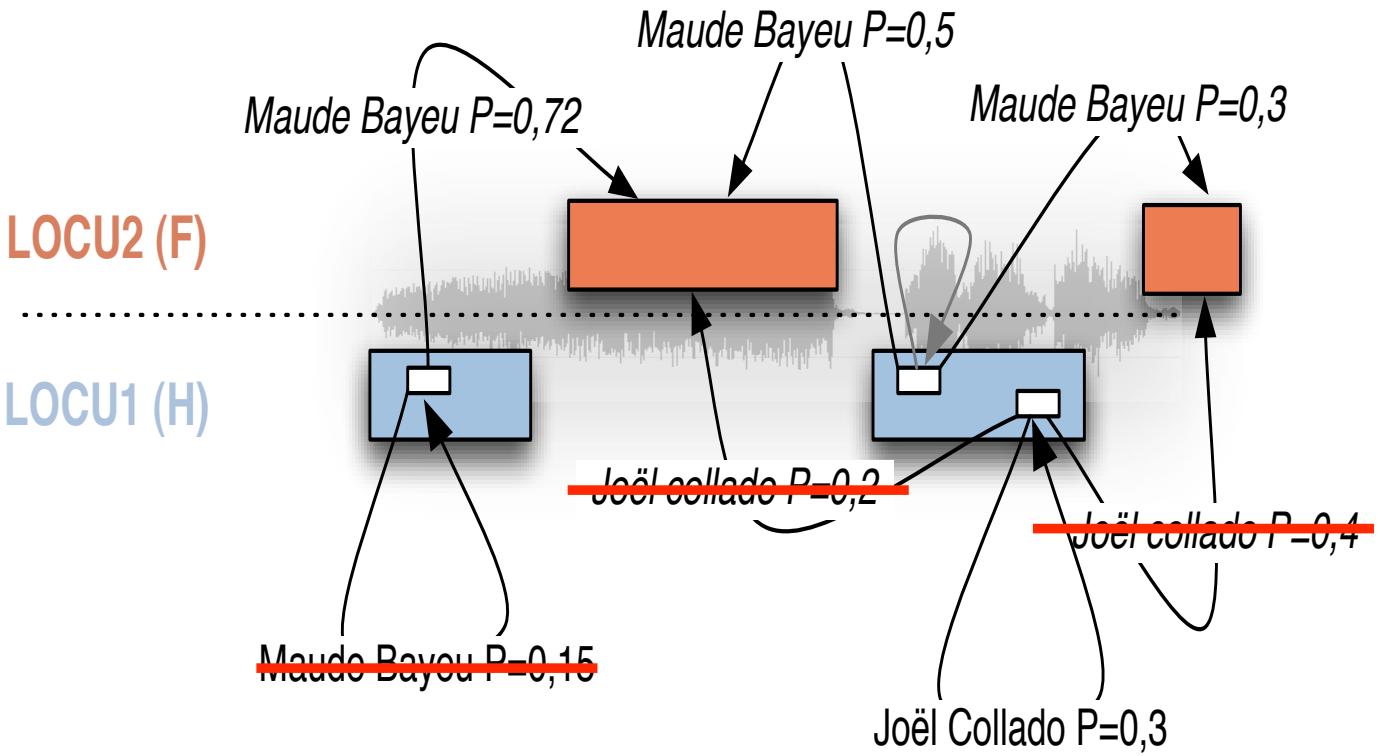
Arbre de Classification Sémantique Probabiliste
Pour chaque occurence: calcul de proba pour les 4 hypothèses



Probabilités issues de l'arbre



Prise en compte du genre



Formalisation

- Occurrence o_j dans un tour t, se réfère au locuteur du tour précédent $(t-1)$, courant ou suivant $(t+1)$
- Probabilités associées $P(o_j, t+r)$, $r = -1, 0, 1$
- Locuteur c_l anonyme (peut regrouper plusieurs tours)
- Noms complets candidats: $\mathcal{E} = \{e_1, \ldots, e_I\}$
- $\alpha_{jl} \in [0, 1]$: degré de "compatibilité" entre les genres de c_l et e_j.

Propagation des informations

- "score" pour chaque nom complet e_i:

$$s_l(e_i) = \sum_{\{(o_j,t)|o_j=e_i,\, t\in c_l\}} \alpha_{jl} P(o_j, t)$$

- Remarque: ce score n'est plus une probabilité

Propagation des informations: exemple

Calcul de scores: 8 occurrences dans un tour t (masculin)

Occurrence o_j	sexe	liste	$P(o_j, t)$	score
Oscar Temaru	M	*non*	*0,29*	–
Hamid Karzaï	M	*non*	*0,29*	–
Jacques Chirac	M	oui	0,29	0,87
Jacques Chirac	M		0,29	
Jacques Chirac	M		0,29	
Jean-Claude Pajak	M	oui	0,29	
Jean-Claude Pajak	M		**0,96**	1,25
Véronique Rebeyrotte	*F*	oui	*0,29*	–

Décision: étape 1

- Pour chaque c_l, choix de l'attribution d'un nom complet e_i^*
- Règle de décision $\mathbf{R_1}$:

$$e_i^* = \arg\max_{e_i \in \mathcal{E}} s_l(e_i)$$

- Etiquetage multiple? $\Rightarrow$ problème de mise en correspondance.

Etiquettes concurrentes (1)

- Exemple: même nom complet attribué initialement à 3 locuteurs différents:

Loc 1: scores

J. Derrida	7,58
N. Demorand	3,99
H. Gardette	1,75
A. Adler	1,12
M. Kravetz	0,54
5 candidats	14,98

Loc 2: scores

J. Derrida	1,67
A. Adler	0,88
2 candidats	2,55

Loc 3: scores

J. Derrida	4,94
O. Duhamel	0,39
2 candidats	5,35

- Quelle(s) solution(s) pour les départager?

Décision finale: 1ère méthode (1)

On prend le plus fort parmi les “gagnants"!

Loc 1: scores

J. Derrida	7,58
N. Demorand	3,99
H. Gardette	1,75
A. Adler	1,12
M. Kravetz	0,54
5 candidats	14,98

Loc 2: scores

J. Derrida	1,67
A. Adler	0,88
2 candidats	2,55

Loc 3: scores

J. Derrida	4,94
O. Duhamel	0,39
2 candidats	5,35

Décision finale : 1ère méthode (2)

Formellement, règle de décision $\mathbf{R_1}$:

1. Pour tout c_l,
$$e_i^* = \arg \max_{e_i \in \mathcal{E}} s_l(e_i)$$
2. Soit $\mathcal{C}_e$: l'ensemble des locuteurs dont l'assignation est e
3. Finalement:
$$c_l^* = \arg \max_{c_l \in \mathcal{C}_{e_i^*}} s(e_i^*, c_l)$$

Etiquettes concurrentes (2)

- Autre solution pour les départager: tenir compte du score relatif

Loc 1: scores

J. Derrida	7,58
N. Demorand	3,99
H. Gardette	1,75
A. Adler	1,12
M. Kravetz	0,54
5 candidats	14,98

Loc 2: scores

J. Derrida	1,67
A. Adler	0,88
2 candidats	2,55

Loc 3: scores

J. Derrida	4,94
O. Duhamel	0,39
2 candidats	5,35

Etiquettes concurrentes (3)

Pondération des scores par τ_{il}: proportion de score allouée à e_i pour l'affectation à c_l

$$\tau_{il} = \frac{s_l(e_i)}{\sum_{q=1}^{I} s_l(e_q)}$$

Loc 1

J. Derrida:	51%
N. Demorand	27%
H. Gardette	12%
A. Adler	7%
M. Kravetz	3%
5 candidats	100%

Loc 2

J. Derrida	65%
A. Adler	35%
2 candidats	100%

Loc 3

J. Derrida	**93**%
O. Duhamel	7%
2 candidats	100%

Décision finale: 2ème méthode

- Pondération des scores par les scores relatifs τ_{il}:
- Règle $\mathbf{R_2}$:

$$SC_l(e_i) = s_l(e_i)\tau_{il}$$

Loc 1	
J. Derrida:	3,84
N. Demorand	1,06
H. Gardette	0,20
A. Adler	0,08
M. Kravetz	0,02
5 candidats	

Loc 2: scores	
J. Derrida	1,09
A. Adler	0,30
2 candidats	

Loc 3: scores	
J. Derrida	4,57
O. Duhamel	0,03
2 candidats	

Récapitulatif

Exemple d'une assignation initiale multiple.

Locuteur	nom complet e_i^*	$s_l(e_i^*)$	τ_{il}	$SC_l(e_i^*)$
Loc 1	Jacques Derrida	$\mathbf{7,58}$	51%	$3,84$
Loc 2	Jacques Derrida	$1,67$	65%	$1,09$
Loc 3	Jacques Derrida	$4,94$	**93**%	$\mathbf{4,57}$

Décision finale

- Algorithme itératif
- Exemple du processus de décision avec deux itérations.

Locuteur	e_i^* 1ère itération	2ème itération
Loc 1	J. Derrida $(3,84)$	**N. Demorand** $(1,06)$
Loc 2	J. Derrida $(1,09)$	**A. Adler** (0,30)
Loc 3	**J. Derrida** $(4,57)$	-
Loc 4	**O. Duhamel** $(1,15)$	-

Critique du modèle précédent

- Clarté: utilisation d'un "score" qui n'est pas une probabilité
- Justification de la pondération des scores? (notion de pureté)
- Cohérence des informations au sein de tours de parole contigus.

Cohérence d'informations dans un tour de parole

Exemple: 5 occurrences effectives dans un tour t (masculin)

Occurrence o_j	sexe	$P(o_j, t)$	score
Jacques Chirac	M	0,29	0,87
Jacques Chirac	M	0,29	
Jacques Chirac	M	0,29	
Jean-Claude Pajak	M	0,29	
Jean-Claude Pajak	M	**0,96**	1,25

- Absence de prise en compte globale des informations du tour: conflit entre 2 concurrents.
- Accumulation potentielle de petites erreurs.

Cohérence d'informations dans un tour de parole

Solutions envisageables:

- Rester dans le cadre probabiliste: formalisation par probabilités conditionnelles : problème, absence d'information *a priori*, complexe.
- Modélisation et combinaison des informations par la Théorie des fonctions de croyance (MCT, "Dempster-Shafer"), adaptée à la gestion du conflit d'informations

Cohérence d'informations dans un tour de parole

Solutions envisageables:

- Rester dans le cadre probabiliste: formalisation par probabilités conditionnelles : problème, absence d'information *a priori*, complexe.
- Modélisation et combinaison des informations par la Théorie des fonctions de croyance (MCT, "Dempster-Shafer"), adaptée à la gestion du conflit d'informations

$\Rightarrow$ Modélisation par la Théorie des fonctions de croyance .

Lignes directrices

Notions sur les Fonctions de croyance

Soit Ω un ensemble fini (ici).

- Fonction de croyance m sur Ω, application m: $2^{\Omega} \rightarrow [0, 1]$ telle que:

$$\sum_{A \subseteq \Omega} m(A) = 1.$$

- Interprétation :
  - traduction d'un état de connaissance sur une variable x dans Ω
  - $m(A) =$ part de croyance allouée à l'hypothèse $x \in A$ et à aucune hypothèse plus restrictive.
- Eléments focaux de m: sous-ensembles A, t.q. $m(A) > 0$.

Combinaison d’informations

- Soient 2 fonctions de croyance m_1 et m_2 issues de 2 sources d’information
- Opérateur de combinaison binaire conjonctif (non normalisé): $\cap \Rightarrow m_{1,2} = m_1 \cap m_2$:

$$\forall A \subseteq \Omega,\ m_{1,2}(A) = \sum_{B \cap C = A} m_1(B) m_2(C).$$

- Opérateur associatif, commutatif.
- Combinaison multiple: $m = m_1 \cap \ldots \cap m_n$
- Degré de conflit: $m(\emptyset)$

Fonctions de croyance et décision

- Décision: conversion d'une fonction de croyance m en une distribution de probabilité, probabilité *pignistique* (smets94):

$$\forall \omega \in \Omega \quad P_m(\{\omega\}) = \sum_{A \subset \Omega} \frac{m(A)}{|A|(1 - m(\emptyset))} \delta_A(\omega),$$

  - $|A|$:cardinal de A,
  - δ_A : fonction indicatrice d'appartenance à A.
- P_m répartit équitablement la masse d'un sous-ensemble de Ω entre ses éléments.

Lignes directrices

Définition des masse de croyance

- Croyance fournie par UNE occurrence de nom complet donnée o_j sur la connaissance du locuteur c_l du tour t
- Cadre de discernement: ensemble des noms complets (liste fermée): $\mathcal{E} = \{e_1, \ldots, e_l\}$
- Information sur le genre de e_i et c_l (M, F, inconnu, incertain)
- Masse de croyance à support simple sur $\mathcal{E}$:

$$\begin{cases} m_t^{jr}(\{e_i\}) = \alpha_{ij} P(o_j, t-r) \text{ si } o_j = e_i \\ m_t^{jr}(\mathcal{E}) = 1 - \alpha_{ij} P(o_j, t-r). \end{cases}$$

$r = -1, 0$ ou 1, selon le tour concerné (suivant, courant, précédent).

Combinaison d'informations

1. Combinaison d'information dans un tour de parole

$$m_t = \bigcap_{r=-1}^{1} \bigcap_{j=1}^{n_{t+r}} m_t^{jr}$$

(n_{t+r} : nombre d'occurrences concernées)

2. Propagation à l'ensemble de l'émission

$$M_l = \bigcap_{t \in c_l} m_t$$

Croyance globale concernant l'étiquetage du locuteur c_l.

Exemple de distribution des masses dans un tour de parole (1)

Exemple: 5 occurrences effectives dans un tour t (masculin)

Occurrence o_j	sexe	$P(o_j, t)$
Jacques Chirac	M	$0,29$
Jacques Chirac	M	$0,29$
Jacques Chirac	M	$0,29$
Jean-Claude Pajak	M	$0,29$
Jean-Claude Pajak	M	**0,96**

Exemple de distribution des masses dans un tour de parole (1)

Exemple: 5 occurrences effectives dans un tour t (masculin)

Occurrence o_j	sexe	$P(o_j, t)$
Jacques Chirac	M	$0,29$
Jacques Chirac	M	$0,29$
Jacques Chirac	M	$0,29$
Jean-Claude Pajak	M	$0,29$
Jean-Claude Pajak	M	**0,96**

$\Rightarrow$ 5 fonctions de croyance à combiner.

Exemple de distribution des masses dans un tour de parole (2)

Eléments focaux	$m_t(\{e_i\})$
Jacques Chirac	0.018
Jean-Claude Pajak	**0.348**
$\emptyset$	**0.624**
$\mathcal{E}$	0.010

$\Rightarrow$ degré de conflit important.

Décision

1. Transformation des masses de croyance M_l en une probabilité pignistique P_{M_l}
2. Règle de décision **R** d'affectation de e_i^* à c_l:

$$e_i^* = \arg\max_{e_i \in \mathcal{E}} \mathrm{P}_{M_l}(e_i)$$

3. Affectation multiple: partage des noms complets comme précédemment.

Décision: exemple

Exemple de décision avec deux itérations

Locuteur	1ère itération	2ème itération
Loc 1	J. Derrida $(0,72)$	**N. Demorand** $(0,17)$
Loc 2	J. Derrida $(0,71)$	**A. Adler** $(0,25)$
Loc 3	**J. Derrida** $(0,99)$	-
Loc 4	**O. Duhamel** $(0,88)$	-

Lignes directrices

Corpora

- Campagne ESTER (2005): émissions radiophoniques francophones
- 6 radios différentes
- Liste fermée de 1008 noms complets possibles
- 3 corpus : apprentissage (76h), développement (30h) et évaluation (10h)

	Nombre d'occurrences	Nombre de tours
Apprentissage	7416	11292
Dév.	2931	4933
Evaluation	1082	1541

Lignes directrices

Evaluation: 5 cas possibles

- Identité proposée correcte (C_1)
- Absence d'identité de la référence(C_2)
- Erreur de substitution (S)
- Erreur de suppression (D)
- Erreur d'insertion (I)

Taux d'erreur

$$Err = \frac{S + I + D}{S + I + D + C_2 + C_1}.$$

Lignes directrices

Expériences

Système	ErrDur	ErrLoc
Référence (règle $\mathbf{R_1}$)	20,6%	20,2%
Référence (règle $\mathbf{R_2}$)	16,6%	19,5%
Proposé (règle $\mathbf{R}$)	$\mathbf{13,7\%}$	$\mathbf{14,9\%}$

- Corpus d'évaluation ESTER
- Résultats utilisant la transcription de référence.
- ErrDur: taux d'erreur en durée
- ErrLoc: taux d'erreur en nombre de locuteurs

Conclusion

Apports des fonctions de croyance dans notre méthode d'identification de locuteur:

- Quantitativement: diminution du taux d'erreur.
- Qualitativement:
  - Clarté et unicité de la décision
  - Cohérence de la prise en compte des informations au sein de tours de parole contigus.

Perspectives

- Parole simultanée (au lieu de tours de parole séquentiels): pb de segmentation
- Adaptation à une liste ouverte des noms possibles
- Fusion d'informations acoustiques (coûteuses) sur les locuteurs?
- Optimisation de la mise en correspondance des locuteurs
- **Vers un système entièrement automatique**:
  - Segmentation/classification en locuteur automatique
  - Transcription automatique

Références sélectives

- L. Canseco-Rodriguez, L. Lamel, J. L. Gauvain. A comparative study using manual and automatic transcriptions for diarization. *Automatic Speech Recognition and Understanding (ASRU)*, 2005.
- V. Jousse, S. Petitrenaud, S. Meignier, Y. Estève, C. Jacquin. Automatic named identification of speakers using diarization and ASR systems. *ICASSP'09, 2009*.
- V. Jousse, S. Meignier, C. Jacquin, S. Petitrenaud, Y. Estève, B. Daille. Analyse conjointe du signal sonore et de sa transcription pour l'identification nommée de locuteur. *Traitement automatique des langues*, 50(1), 2009.