



Identification nommée du locuteur

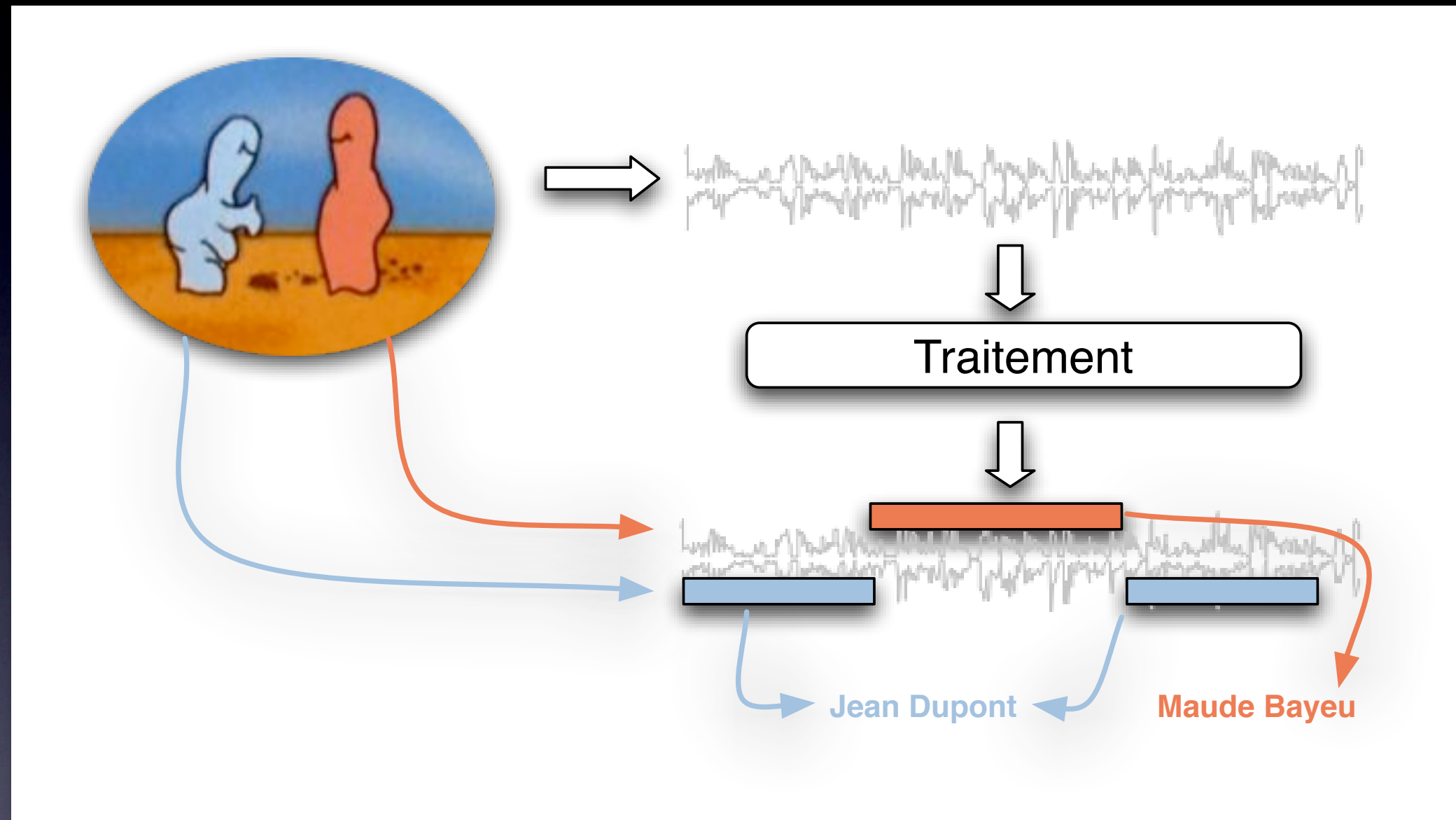
26 juin 2009
Nantes

V. Jousse, S. Meignier, C. Jacquin, B. Daille

Cadre de travail

- Collaboration entre Le Mans (LIUM) et Nantes (LINA)
- Identification nommée du locuteur par analyse conjointe
 - Reconnaissance de la parole (Le Mans)
 - TALN (Nantes)

Nommer les locuteurs



Traiter des émissions radiophoniques ou
télévisuelles

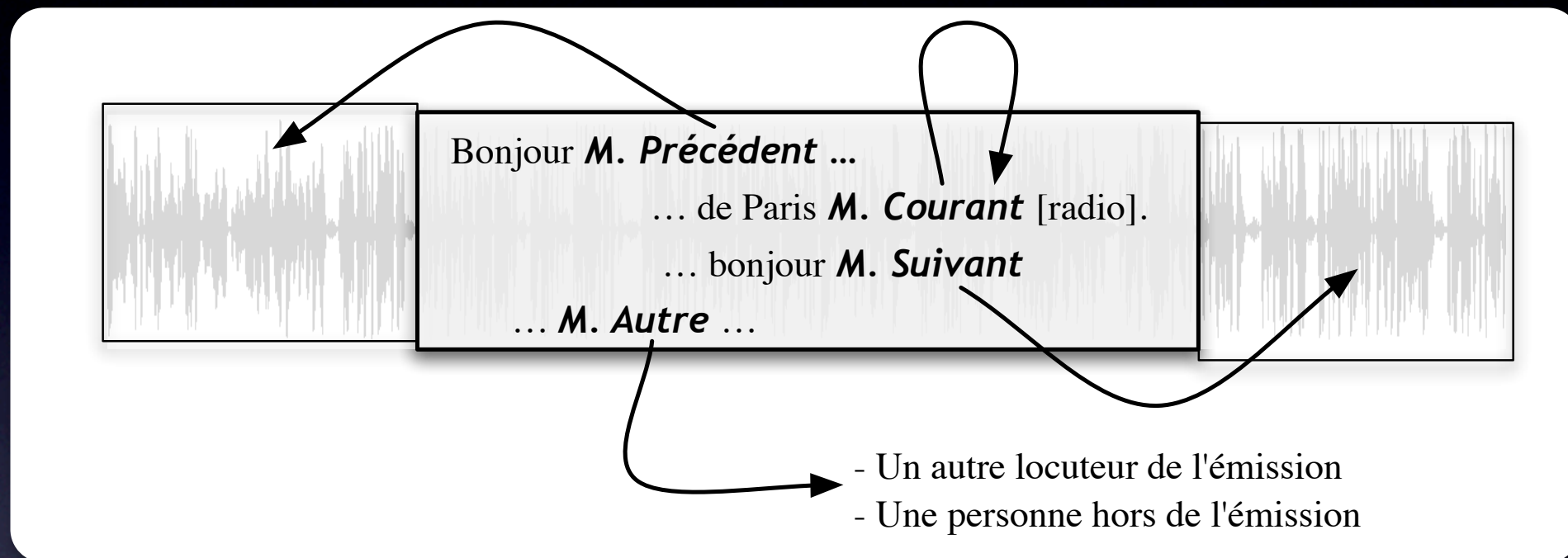
État de l'art (I)

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

État de l'art (2)

- Méthodes utilisant la transcription du signal
 - Hypothèse : les locuteurs s'annoncent
 - Extraire les noms des locuteurs à partir des paroles prononcées
 - Reconnaissance automatique de la parole
 - Détection d'entités nommées

État de l'art (3)



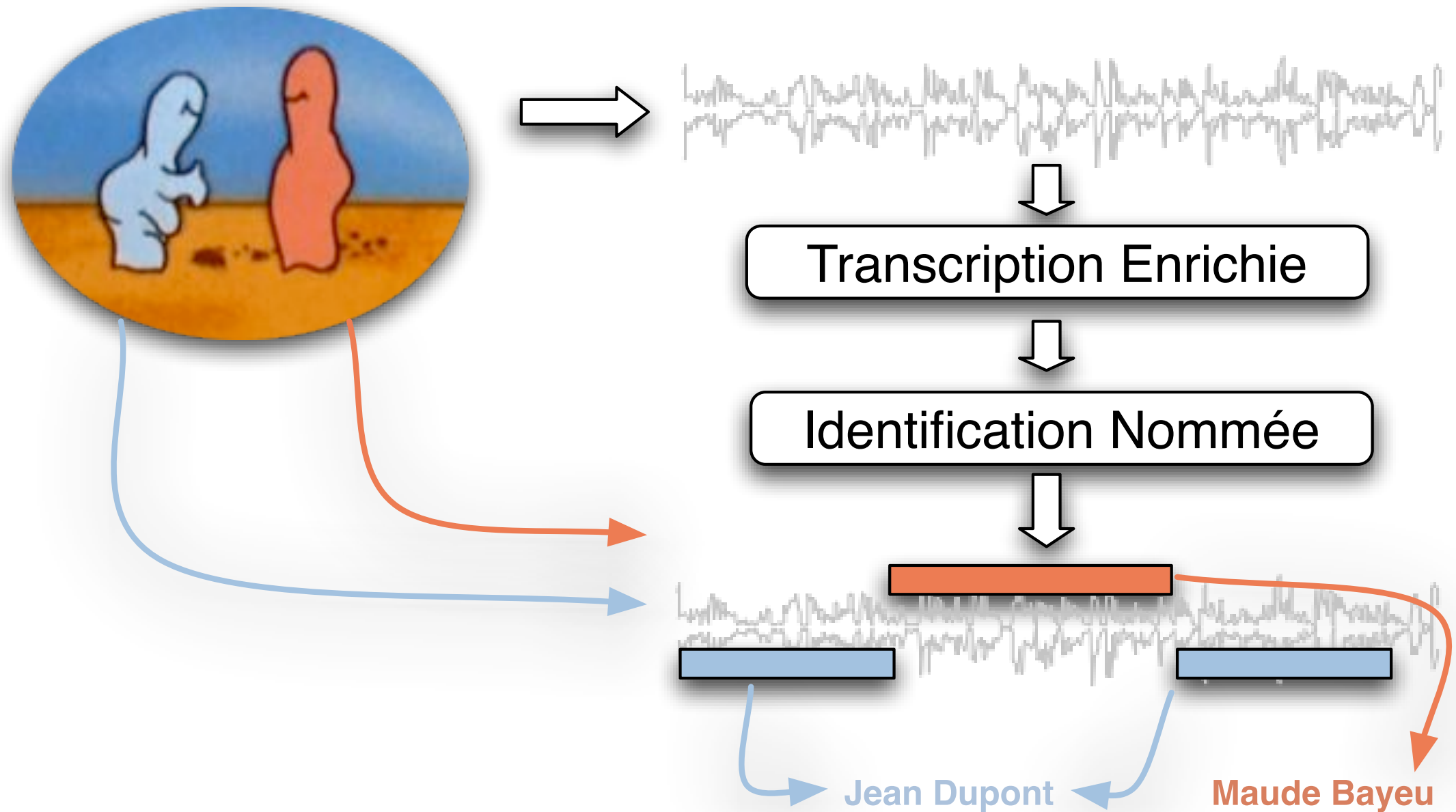
4 étiquettes (*Canseco 05*) :

- *Précédent* : il vient de parler
- *Courant* : il parle
- *Suivant* : il va parler
- *Autre*

Attribution des étiquettes

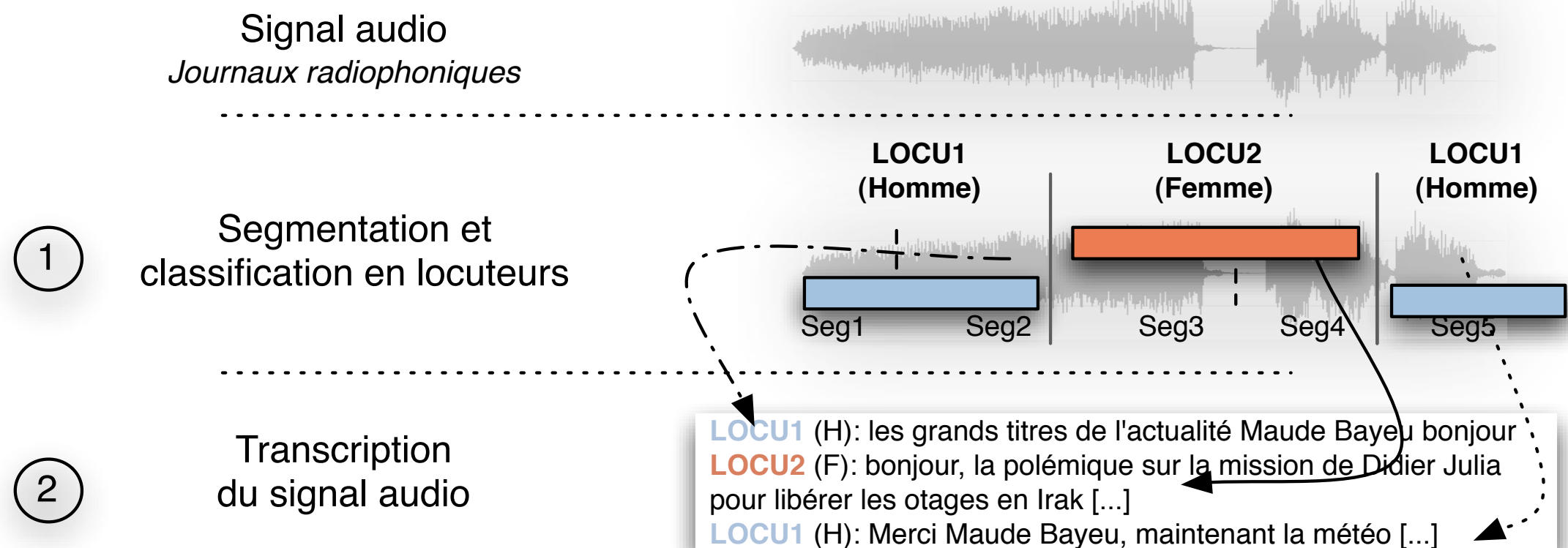
- Utilisation du contexte lexical : * *bonjour*
[PERSONNE] </s>
- État de l'art, 3 méthodes :
 - Règles manuelles LIMSI (*Canseco 05*)
 - Modèle n-grammes Cambridge (*Tranter 06*)
 - Arbre de classification (SCT) LIUM (*Mauclair 06*)
 - Méthode retenue car plus performante que les n-grammes (*Esteve07*)

Identification nommée



Transcription enrichie (I)

Traitement à partir de l'acoustique



Transcription enrichie (2)

Traitement à partir de la transcription

- Détection des entités nommées : Nemesis
- Détection des anthroponymes
- Couple prénom et patronyme

3

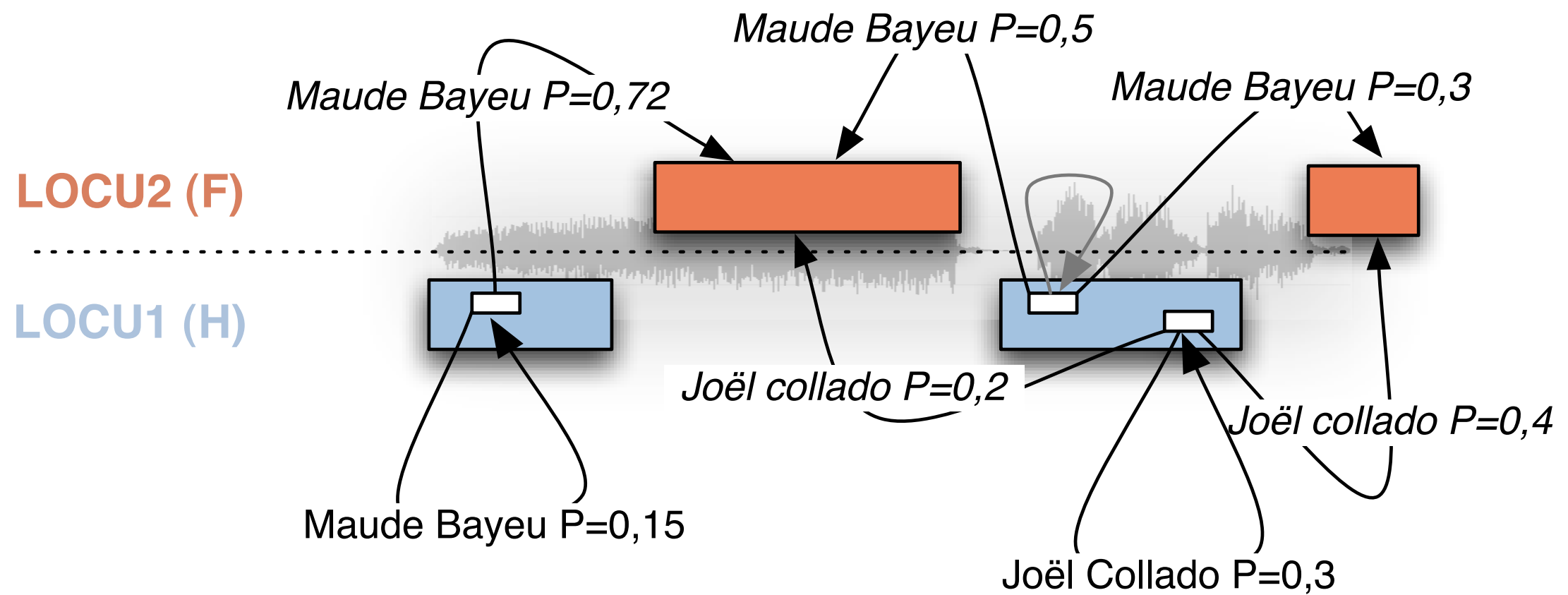
Détection et catégorisation
des entités nommées

LOCU1 (H): les grands titres de l'actualité **PERSONNE**(Maude Bayeu) bonjour

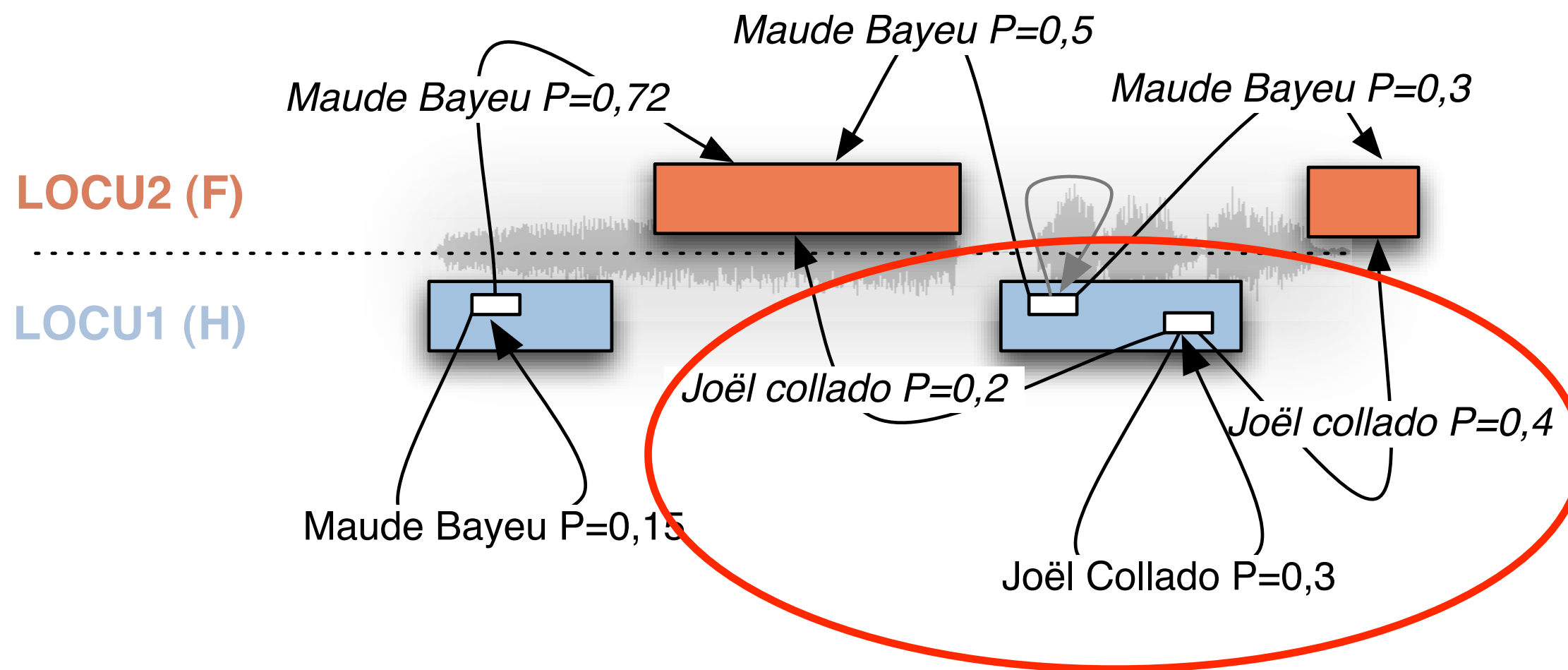
LOCU2 (F): bonjour, la polémique sur la mission de **PERSONNE** (Didier Julia) pour libérer les otages en **LIEU**(Irak) [...]

LOCU1 (H): Merci **PERSONNE**(Maude Bayeu), maintenant la météo [...]

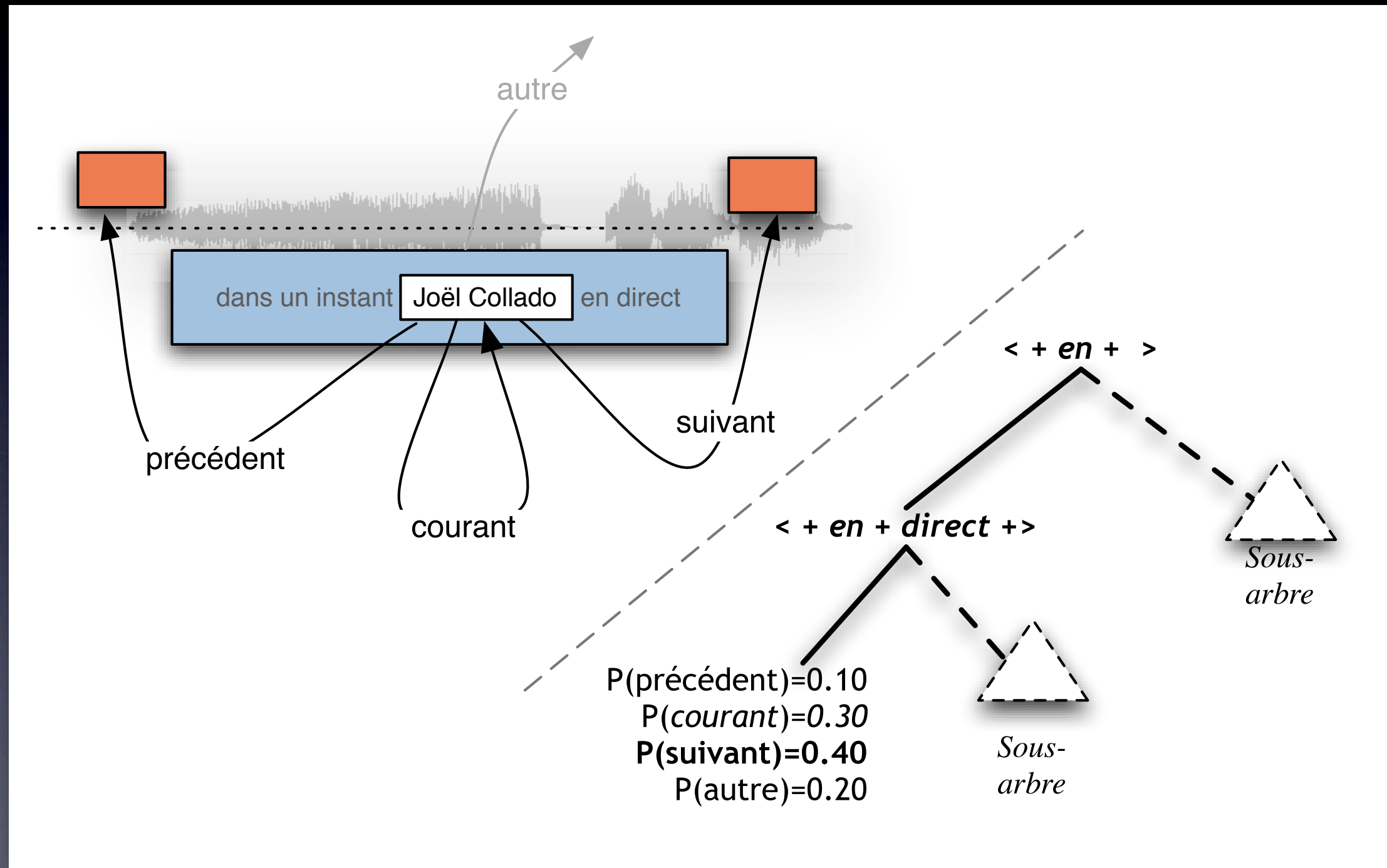
Attribution des étiquettes (I)



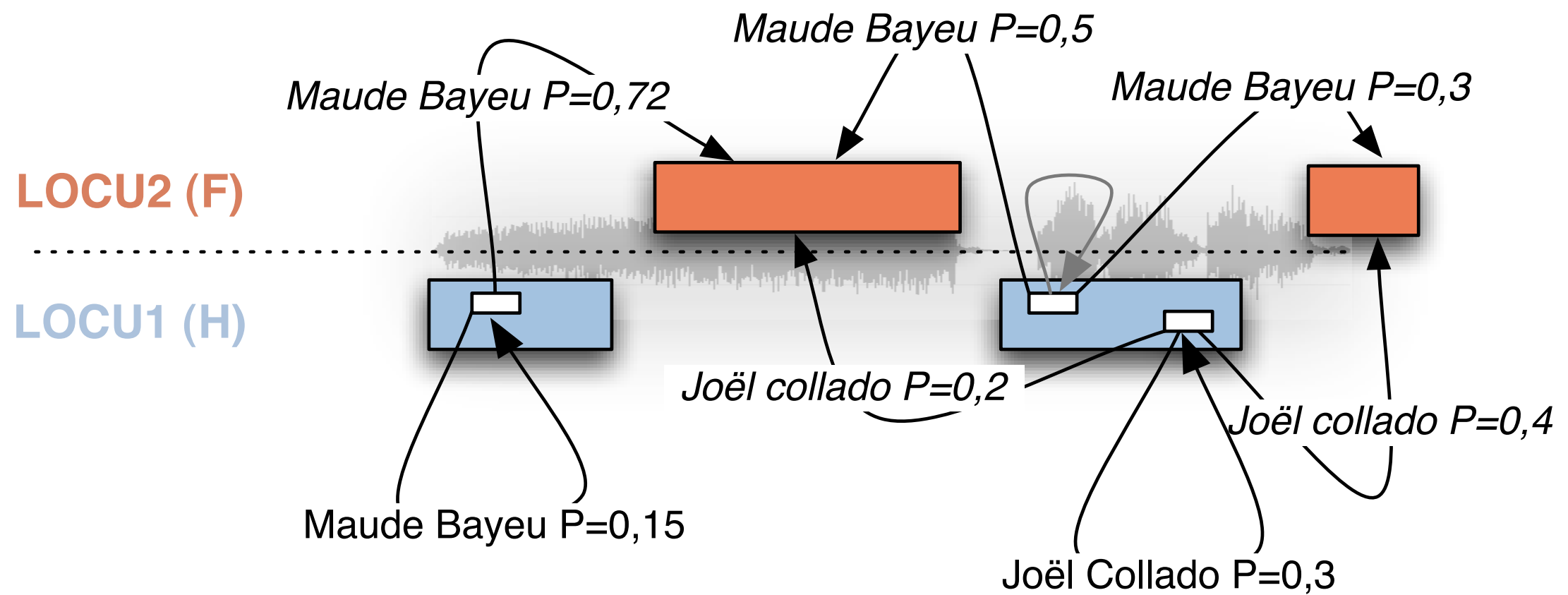
Attribution des étiquettes (I)



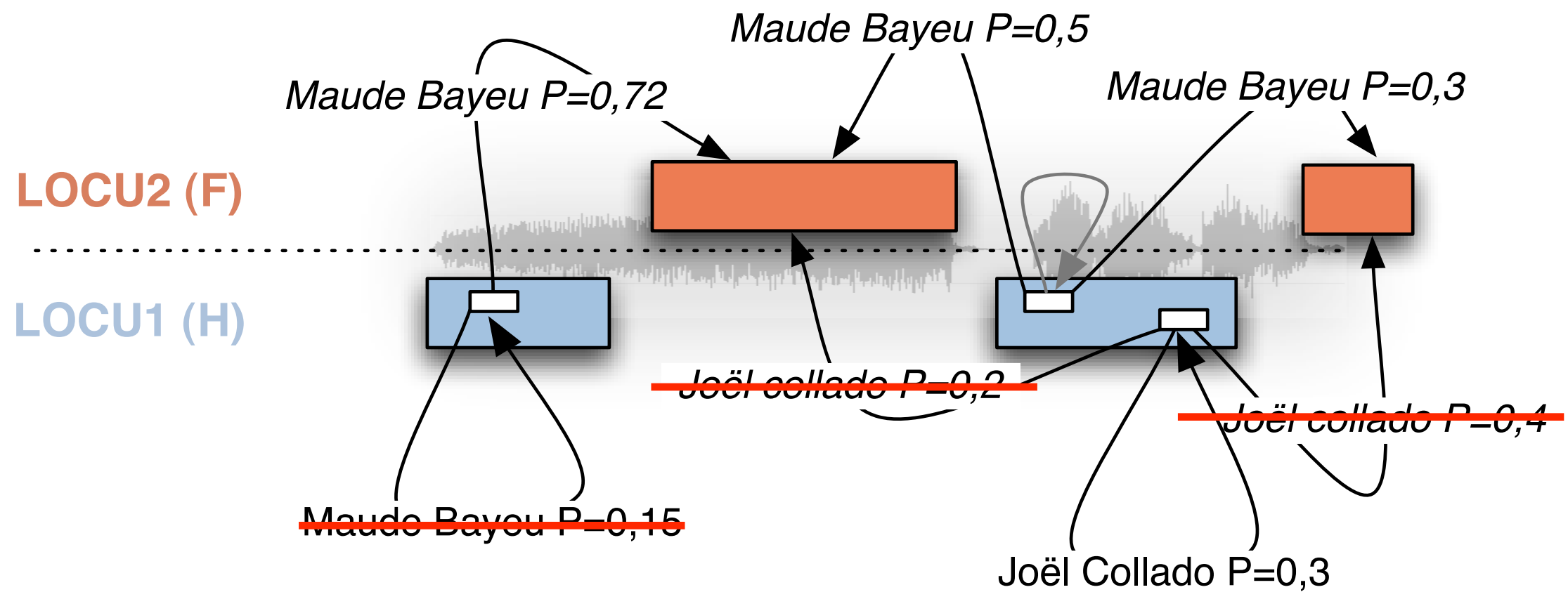
Attribution des étiquettes (2)



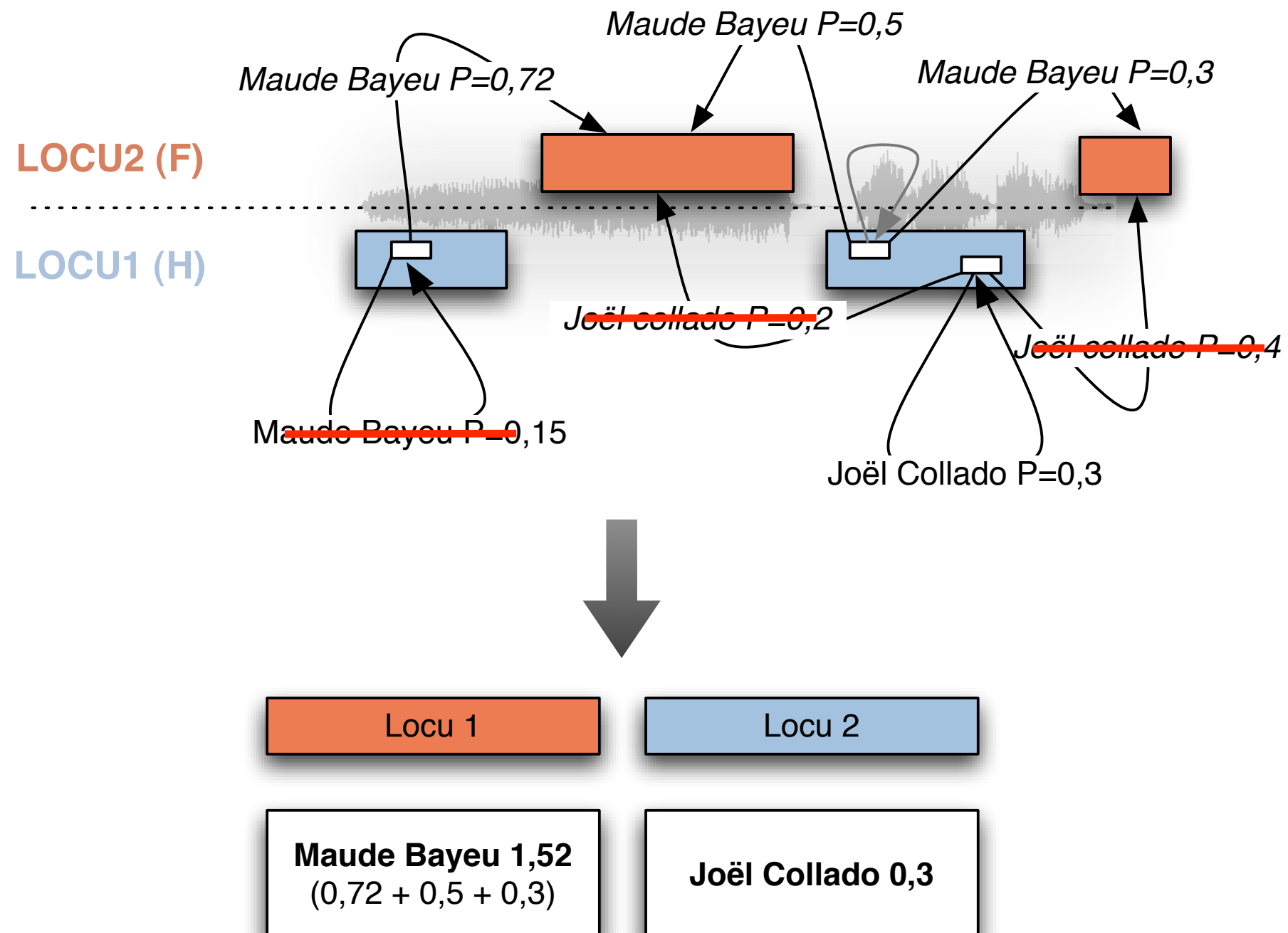
Prise en compte du genre



Prise en compte du genre



Décision globale



Décision globale

Locu 1

Maude Bayeu : 4,2
Anne Françoise : 3,2
Béatrice Leveiller : 3,1
Catherine Monet : 2,1

...

10 candidats

Locu 3

Maude Bayeu : 1,52
Brigitte Dagot : 0,2
Marjorie Roulman : 0,1

3 candidats

Décision globale

- Prendre en compte la contribution du score d'un locuteur
- Multiplier le score par sa contribution à la classe

$$SC(i) = s(i) \times \frac{s(i)}{\sum_{q=1}^I s(q)}$$

Décision globale

Locu 1

Maude Bayeu : 4,2 * 0,24 = 1,01
Anne Françoise : 3,2 * 0,21 = 0,78
Béatrice Leveiller : 3,1 * 0,20 = 0,62
Catherine Monet : 2,1 * 0,12 = 0,25

...

10 candidats / score total cumulé : 17,23

Locu 3

Maude Bayeu : 1,52 * 0,84 = 1,28
Brigitte Dagot : 0,2 * 0,11 = 0,02
Marjorie Roulman : 0,1 * 0,05 = 0,005

3 candidats / score total cumulé : 1,82

Décision globale

Locu 1

~~Maude Bayeu : 4,2 * 0,24 = 1,01~~
Anne Françoise : 3,2 * 0,21 = 0,78
Béatrice Leveiller : 3,1 * 0,20 = 0,62
Catherine Monet : 2,1 * 0,12 = 0,25

...

10 candidats / score total cumulé : 17,23

Locu 3

Maude Bayeu : 1,52 * 0,84 = 1,28
Brigitte Dagot : 0,2 * 0,11 = 0,02
Marjorie Roulman : 0,1 * 0,05 = 0,005

3 candidats / score total cumulé : 1,82

Décision globale

Locu 1

~~Maude Bayeu : 4,2 * 0,24 = 1,01~~
Anne Françoise : 3,2 * 0,21 = 0,78
Béatrice Leveiller : 3,1 * 0,20 = 0,62
Catherine Monet : 2,1 * 0,12 = 0,25

...

10 candidats / score total cumulé : 17,23

Locu 3

Maude Bayeu : 1,52 * 0,84 = 1,28
Brigitte Dagot : 0,2 * 0,11 = 0,02
Marjorie Roulman : 0,1 * 0,05 = 0,005

3 candidats / score total cumulé : 1,82

Décision globale

Locu 1

~~Maude Bayeu : 4,2 * 0,24 = 1,01~~
Anne Françoise : 3,2 * 0,21 = 0,78
Béatrice Leveiller : 3,1 * 0,20 = 0,62
Catherine Monet : 2,1 * 0,12 = 0,25

...

10 candidats / score total cumulé : 17,23

Locu 3

Maude Bayeu : 1,52 * 0,84 = 1,28
Brigitte Dagot : 0,2 * 0,11 = 0,02
Marjorie Roulman : 0,1 * 0,05 = 0,005

3 candidats / score total cumulé : 1,82

Expériences

Corpus

- Campagne ESTER 2005
 - Journaux d'information français
 - 6 radios différentes
- 3 corpus : apprentissage (81h), développement (12,5h) et évaluation (10h)

Cas possibles

- Identité proposée correcte (C1)
- Pas d'identité (C2)
- Erreur de substitution (S)
- Erreur de suppression (D)
- Erreur d'insertion (I)

Métriques

- Précision / Rappel

$$P = \frac{C_1}{C_1 + S + I}$$

$$R = \frac{C_1}{C_1 + S + D}$$

- Taux d'erreur

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

Transcriptions manuelles et liste de locuteurs

Système	En durée			En nb de Locuteur
	Rappel	Précision	ErrDur	ErrLoc
Référence	70,70%	92,59%	26,64%	37,40%
Proposé	83,16%	89,72%	16,66%	19,5%

Tableau 4. *Comparaison système proposé et système de référence sur le corpus d'évaluation ESTER 1 phase II*

*Les résultats sont donnés en utilisant la transcription enrichie de référence. **Rappel** et **Précision** calculés en en durée.*

***ErrDur** : Taux d'erreur en durée.*

***ErrLoc** : Taux d'erreur en nombre de locuteurs.*

Transcriptions manuelles et liste de locuteurs

Système	En durée			En nb de Locuteur
	Rappel	Précision	ErrDur	ErrLoc
Référence	70,70%	92,59%	26,64%	37,40%
Proposé	83,16%	89,72%	16,66%	19,5%

Tableau 4. Comparaison système proposé et système de référence sur le corpus d'évaluation ESTER 1 phase II

*Les résultats sont donnés en utilisant la transcription enrichie de référence. **Rappel** et **Précision** calculés en en durée.*

***ErrDur** : Taux d'erreur en durée.*

***ErrLoc** : Taux d'erreur en nombre de locuteurs.*

Vers un système automatique

		En durée				En nb de Locuteur
Trans.	Seg/Class.	R	P	F	ErrDur	ErrLoc
Système présenté						
M	M	83,2%	89,7%	0,86	16,7%	19,5%
M	A	38,0%	58,2%	0,46	58,3%	-
A	M	31,0%	58,3%	0,40	62,8%	70,0%
A	A	18,4%	42,1%	0,26	75,2%	-

Trans. : Transcription **M**anuelle ou **A**utomatique.

Seg/Class. : segmentation/classification manuelles ou automatiques.

R, P, F : rappel, précision et F-mesure calculés en durée.

ErrDur : Taux d'erreur en durée.

ErrLoc : Taux d'erreur en nombre de locuteurs.

Vers un système automatique

		En durée				En nb de Locuteur
Trans.	Seg/Class.	R	P	F	ErrDur	ErrLoc
Système présenté						
M	M	83,2%	89,7%	0,86	16,7%	19,5%
M	A	38,0%	58,2%	0,46	58,3%	-
A	M	31,0%	58,3%	0,40	62.8%	70,0%
A	A	18,4%	42,1%	0,26	75,2%	-

Trans. : Transcription **M**anuelle ou **A**utomatique.

Seg/Class. : segmentation/classification manuelles ou automatiques.

R, P, F : rappel, précision et F-mesure calculés en durée.

ErrDur : Taux d'erreur en durée.

ErrLoc : Taux d'erreur en nombre de locuteurs.

Conclusion

Conclusion

- Système basé sur une analyse conjointe signal/texte
- Système de décision : deux fois plus performant sur du manuel
- Passage à l'automatique problématique

Perspectives

- Passage à l'automatique : analyse des erreurs
- Segmentation et classification
 - Collaboration plus étroite avec le système
- Transcription
 - Détection des entités nommées
 - Étudier les systèmes de la campagne ESTER 2

Publications

- Analyse conjointe du signal sonore et de sa transcription pour l'identification nommée du locuteur - *Revue TAL, Volume 50 à paraître*
- Automatic named identification of speakers using diarization and asr systems - *ICASSP 09 (IEEE), Taiwan*
- Étude pour l'amélioration d'un système d'identification nommée du locuteur - *JEP 08, Avignon*

FIN

Influence de la transcription

Transcription	En durée			En nb de Locuteur
	Rappel	Précision	ErrDur	ErrLoc
LIUM	30,98%	58,3%	62,8%	69,96%
LIMSI	40,96%	65,07%	53,80%	59,50%

Tableau 7. *Comparaison des résultats avec deux systèmes de transcription différents sur le corpus d'évaluation ESTER 1 phase II*

Rappel et Précision calculés en en durée.

ErrDur : Taux d'erreur en durée.

ErrLoc : Taux d'erreur en nombre de locuteurs.

Influence de la transcription

Transcription	En durée			En nb de Locuteur
	Rappel	Précision	ErrDur	ErrLoc
LIUM	30,98%	58,3%	62,8%	69,96%
LIMSI	40,96%	65,07%	53,80%	59,50%

Tableau 7. *Comparaison des résultats avec deux systèmes de transcription différents sur le corpus d'évaluation ESTER 1 phase II*

Rappel et Précision calculés en en durée.

ErrDur : Taux d'erreur en durée.

ErrLoc : Taux d'erreur en nombre de locuteurs.

Connaissance des noms

Noms complets	En durée			En nb de Locuteur
	Rappel	Précision	ErrDur	ErrLoc
connus	83,16%	89,72%	16,66%	19,5%
inconnus	69,05%	76,48%	31,49%	33,59%

Tableau 5. Résultats avec et sans connaissance a priori sur les noms complets, évaluation faite sur le corpus d'évaluation ESTER 1 phase II

Les résultats sont données en utilisant la transcription enrichie de référence.

***Noms complets connus** : le système connaît les noms complets des locuteurs potentiels.*

***Noms complets inconnus** : le système ne connaît pas les noms complets des locuteurs potentiels.*

***Rappel et Précision** calculés en en durée.*

***ErrDur** : Taux d'erreur en durée.*

***ErrLoc** : Taux d'erreur en nombre de locuteurs.*

Connaissance des noms

Noms complets	En durée			En nb de Locuteur
	Rappel	Précision	ErrDur	ErrLoc
connus	83,16%	89,72%	16,66%	19,5%
inconnus	69,05%	76,48%	31,49%	33,59%

Tableau 5. Résultats avec et sans connaissance a priori sur les noms complets, évaluation faite sur le corpus d'évaluation ESTER 1 phase II

Les résultats sont données en utilisant la transcription enrichie de référence.

***Noms complets connus** : le système connaît les noms complets des locuteurs potentiels.*

***Noms complets inconnus** : le système ne connaît pas les noms complets des locuteurs potentiels.*

***Rappel et Précision** calculés en en durée.*

***ErrDur** : Taux d'erreur en durée.*

***ErrLoc** : Taux d'erreur en nombre de locuteurs.*