



# Identification nommée du locuteur

*exploitation conjointe du signal sonore et de sa transcription*

4 mai 2011  
LIUM - Le Mans

Vincent Jousse



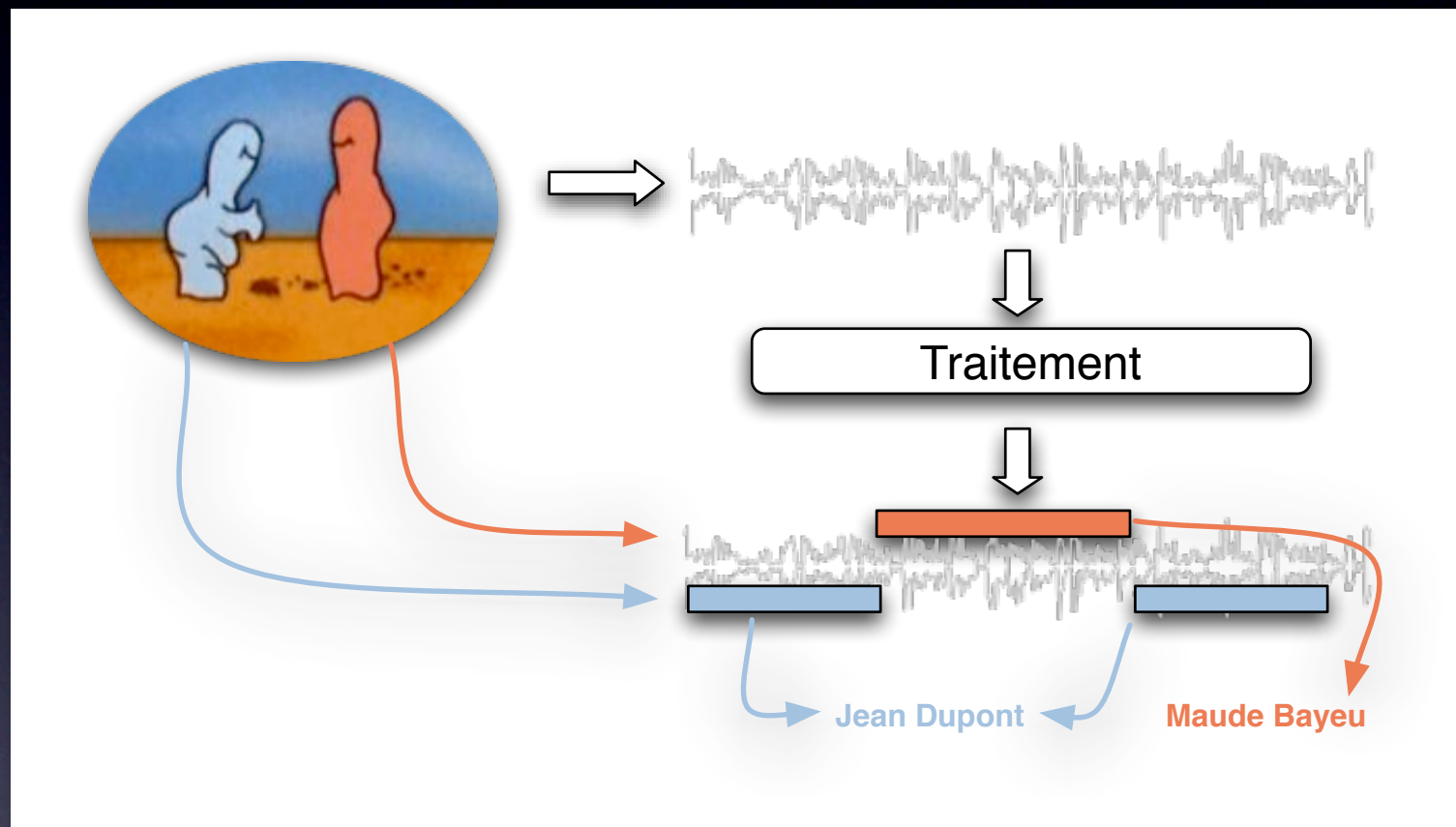
# Cadre de travail

- Collaboration entre Le Mans (LIUM) et Nantes (LINA)
- Deux domaines de recherche
  - Reconnaissance de la parole (Le Mans)
  - TALN (Nantes)



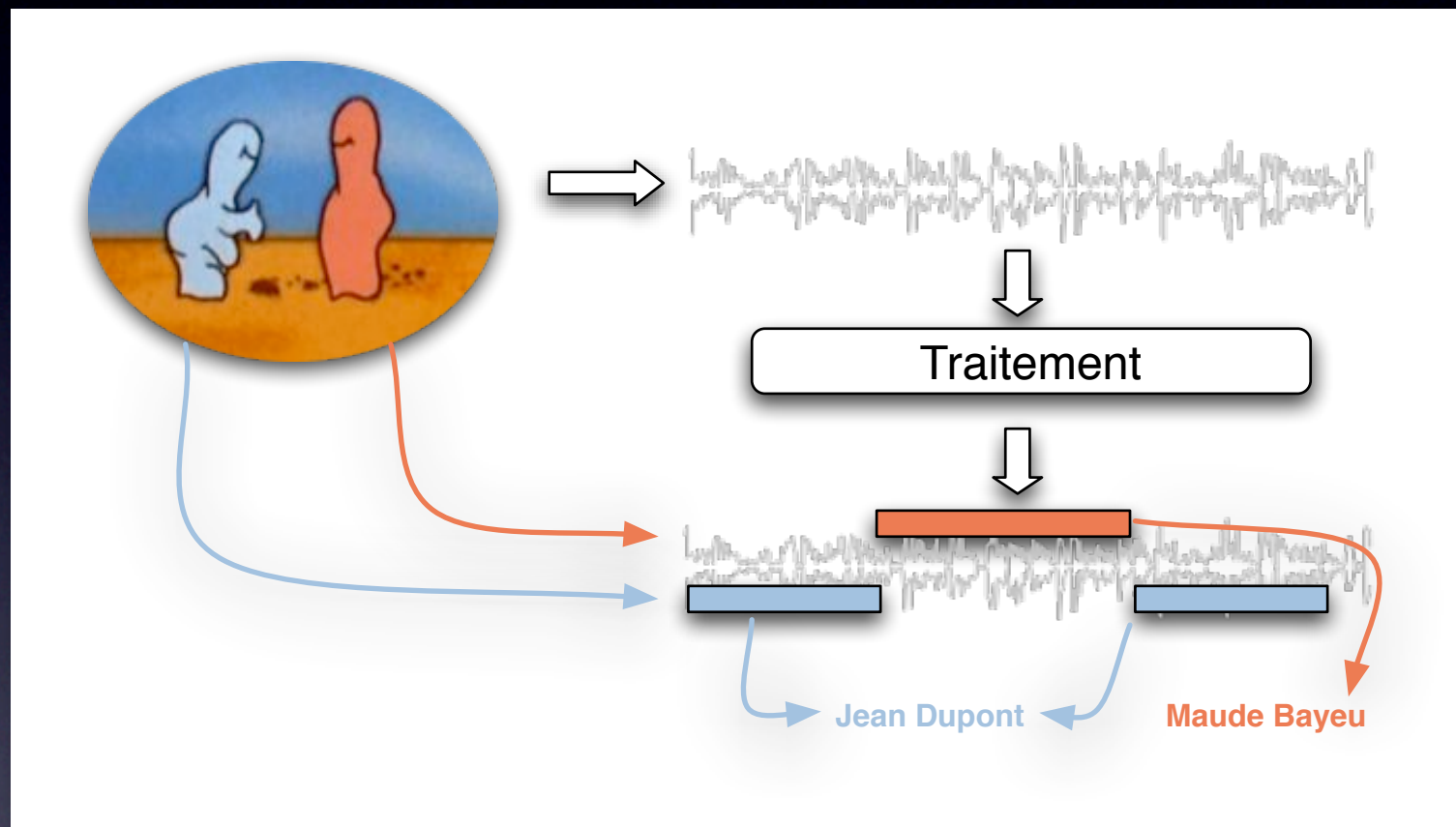
# Nommer les locuteurs

## Qui parle, et quand ?



- Émissions radiophoniques ou télévisées
- Minimiser les connaissances a priori

# Contexte



- Domaine assez peu étudié
- Étape importante pour la recherche documentaire



# Propositions (I)

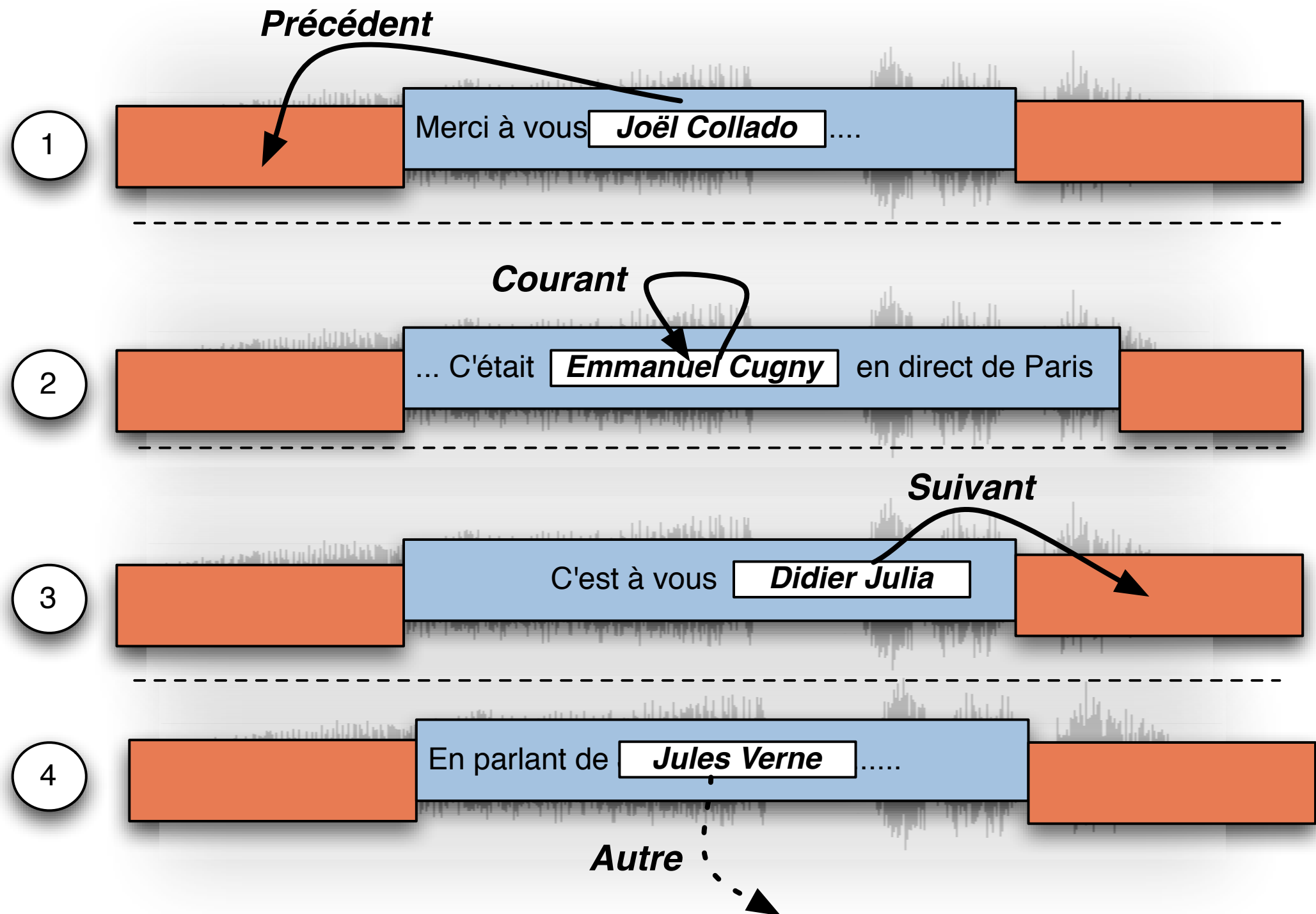
- Méthodes utilisant des modèles acoustiques
  - Reconnaissance automatique du locuteur
  - Enregistrements de chaque locuteur
    - ➔ Difficile à obtenir



# Propositions (2)

- Méthodes utilisant la transcription du signal
- Extraction des noms de locuteurs
  - Reconnaissance automatique de la parole
  - Détection d'entités nommées

# Attributions locales



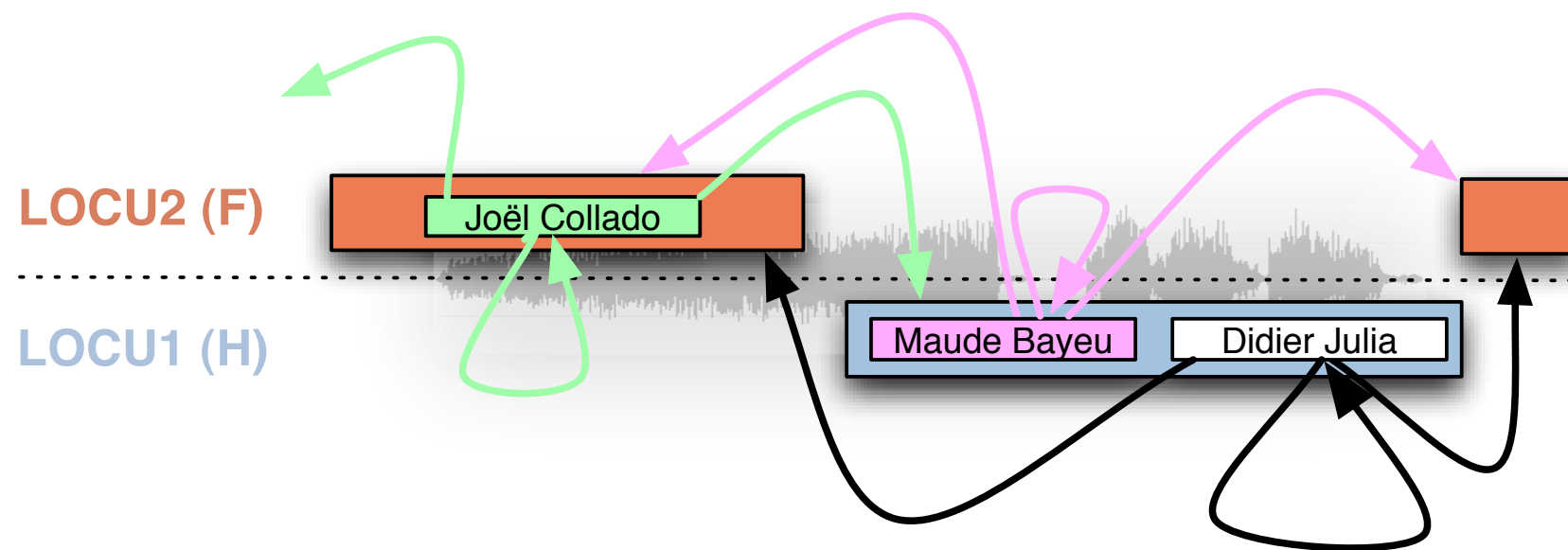


# 4 possibilités

- Pour chaque prénom / patronyme détecté
  - Locuteur *suivant*
  - Locuteur *courant*
  - Locuteur *précédent*
  - Autre *locuteur*

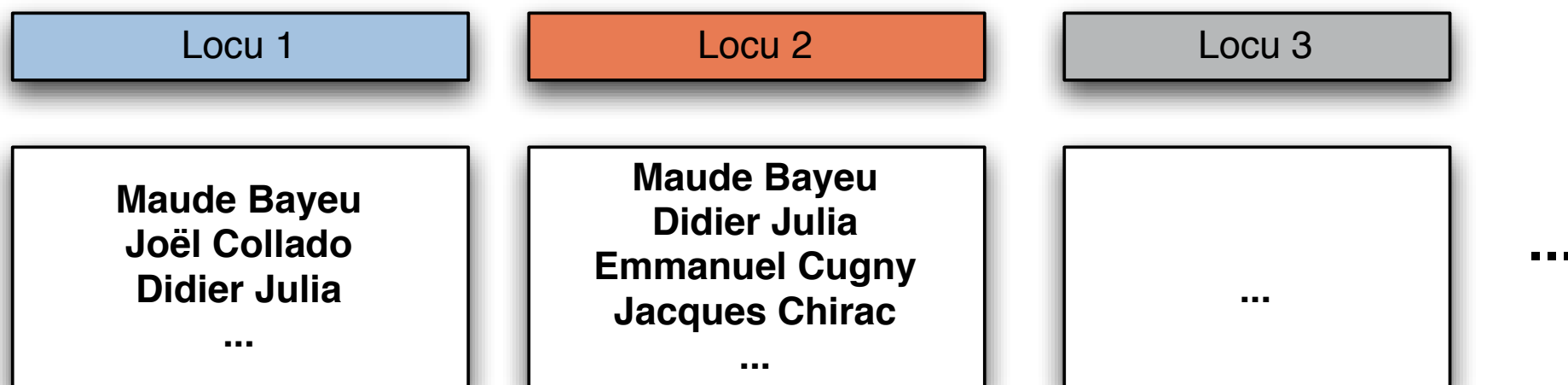
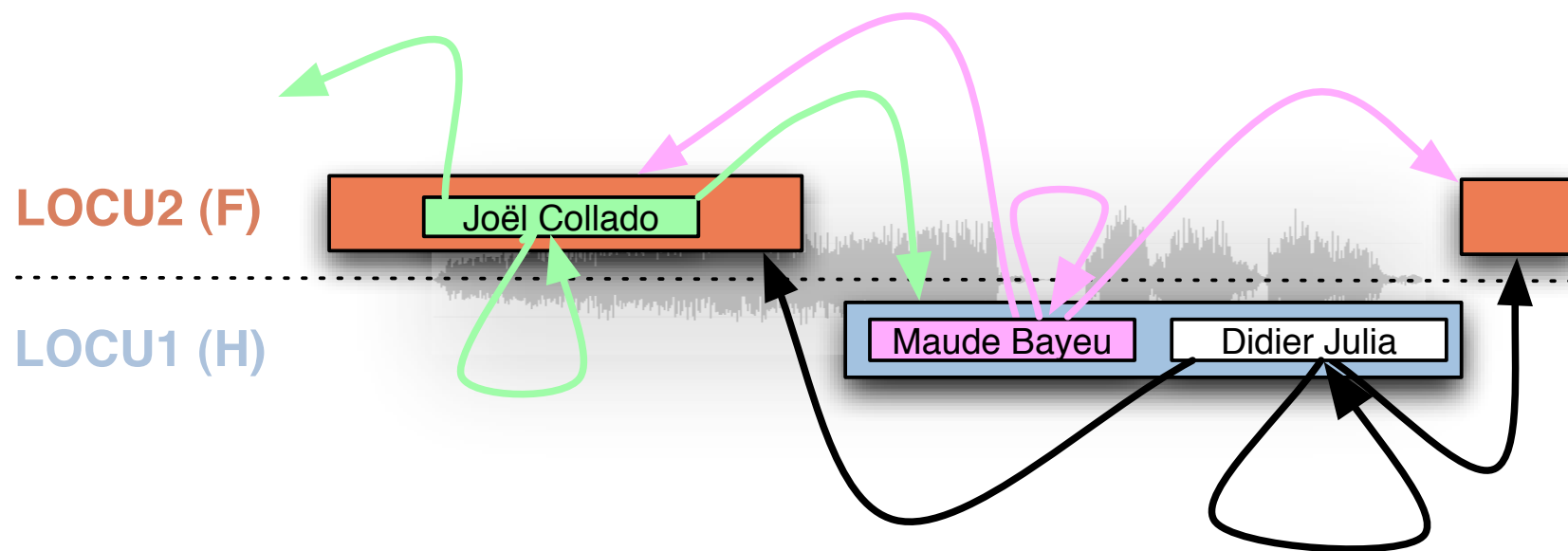


# Aperçu global





# Aperçu global





# Difficultés

- Attribuer un couple prénom/nom à un des tours de paroles (étiquettes)
- Prendre en compte les conflits
- Affecter un seul couple prénom/nom à un locuteur



# Hypothèses

- Locuteurs annoncés par leur prénom/nom
- Prénoms/noms correctement transcrits
- Contexte lexical exploitable



# Identification nommée

- État de l'art, 3 méthodes :
  - Règles manuelles LIMSI (*Canseco 05*)
  - Modèle n-grammes Cambridge (*Tranter 06*)
  - Arbre de classification (SCT) LIUM (*Mauclair 06*)



# Règles manuelles

## Canseco 2005

- Travaux précurseurs en langue anglaise
- Classes sémantiques : *[location]*, *[title]*
- Caractère joker :
  - \* *[person]* reporting from *[location]* \*
  - => It was **John Smith** reporting from **Bagdad**
- Pas de décision globale



# Modèles n-grammes

## Tranter 2006

- N-grammes autour des noms de locuteurs
- Utilisation de classes sémantiques étiquetées manuellement
- Combinaison simple des règles
- Modèle entropie (Chenguyan 07)



# Arbres de classification sémantique (SCT)

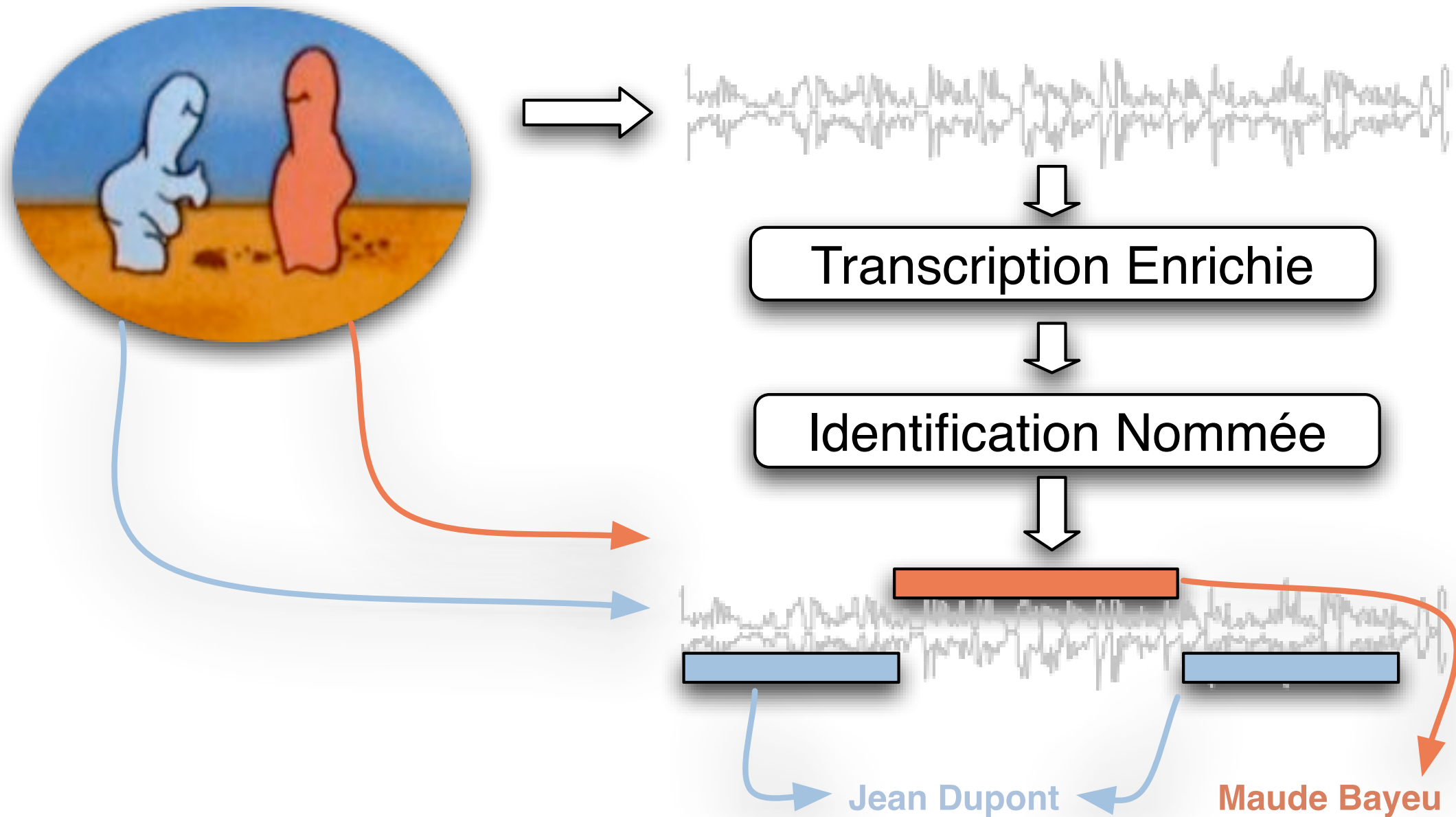
Mauclair 2006

- Expressions régulières simples autour des noms de locuteurs
- Détection d'entités nommées pour les classes sémantiques
- Plus robuste que les n-grammes (Estève 07)



# Systeme du LIUM

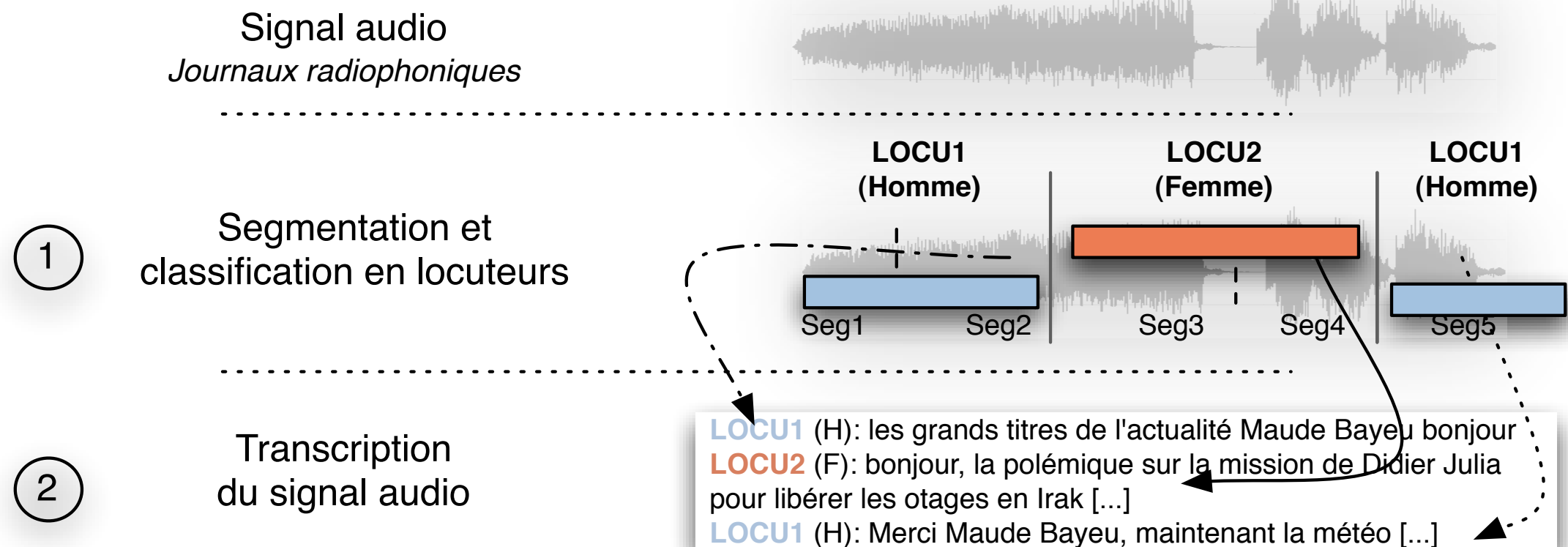
# Identification nommée





# Transcription enrichie (I)

*Traitement à partir de l'acoustique*



# Transcription enrichie (2)

*Traitement à partir de la transcription*

- Détection des entités nommées
  - Classes sémantiques
  - Couple prénom et patronyme

3

Détection et catégorisation  
des entités nommées

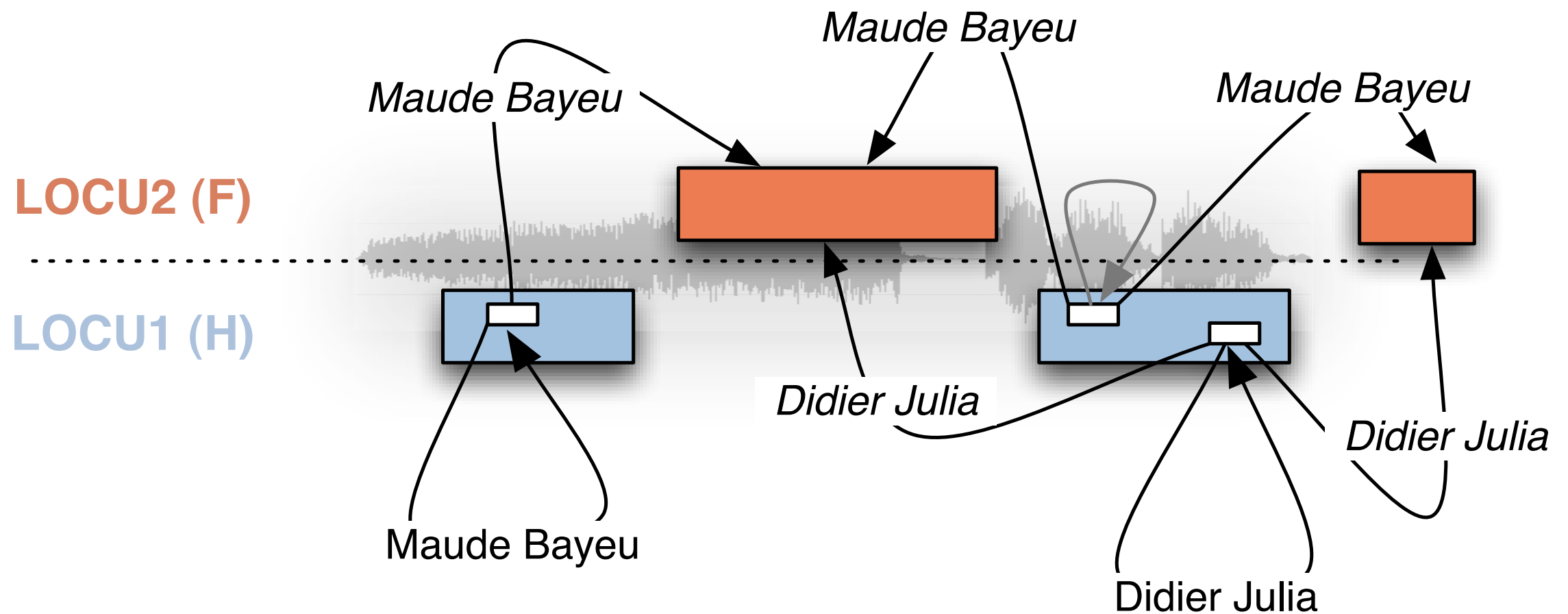
**LOCU1** (H): les grands titres de l'actualité **PERSONNE**(Maude Bayeu) bonjour

**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 [...]

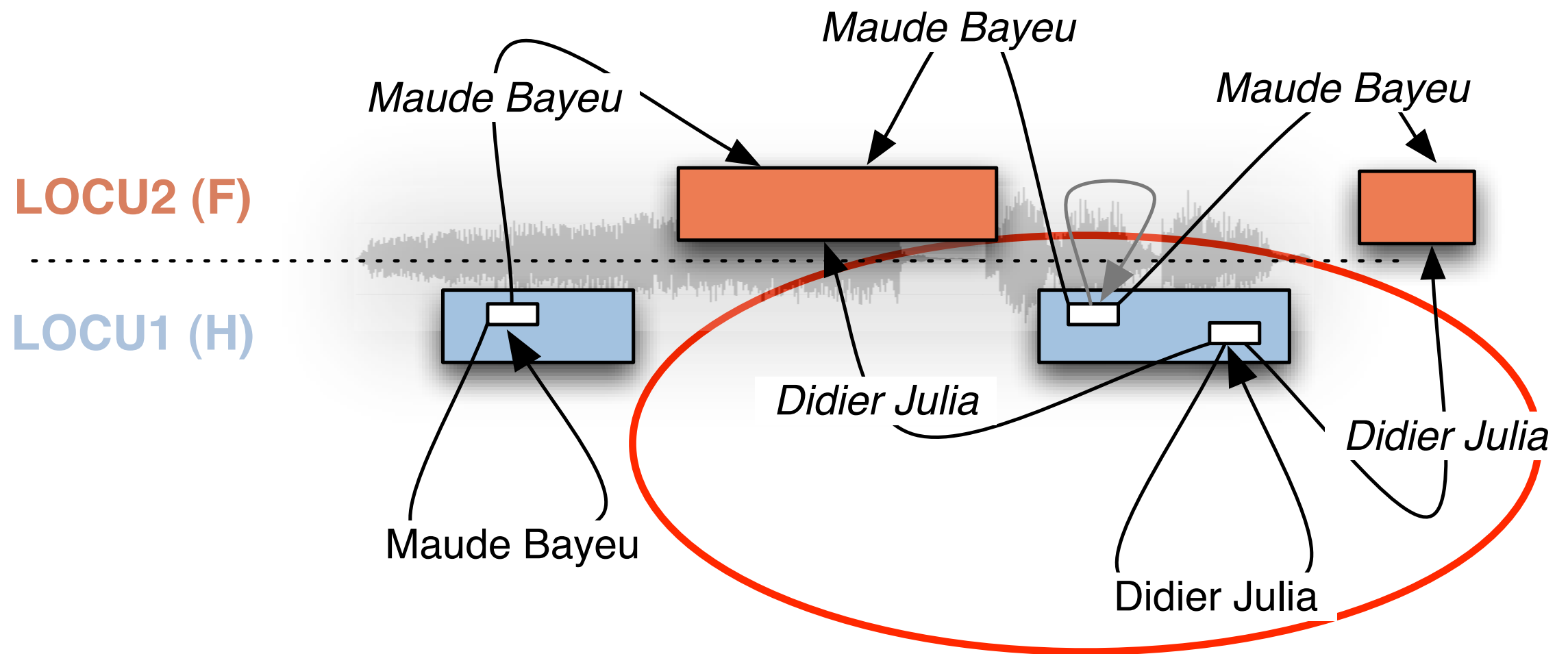


# Attribution des étiquettes (I)



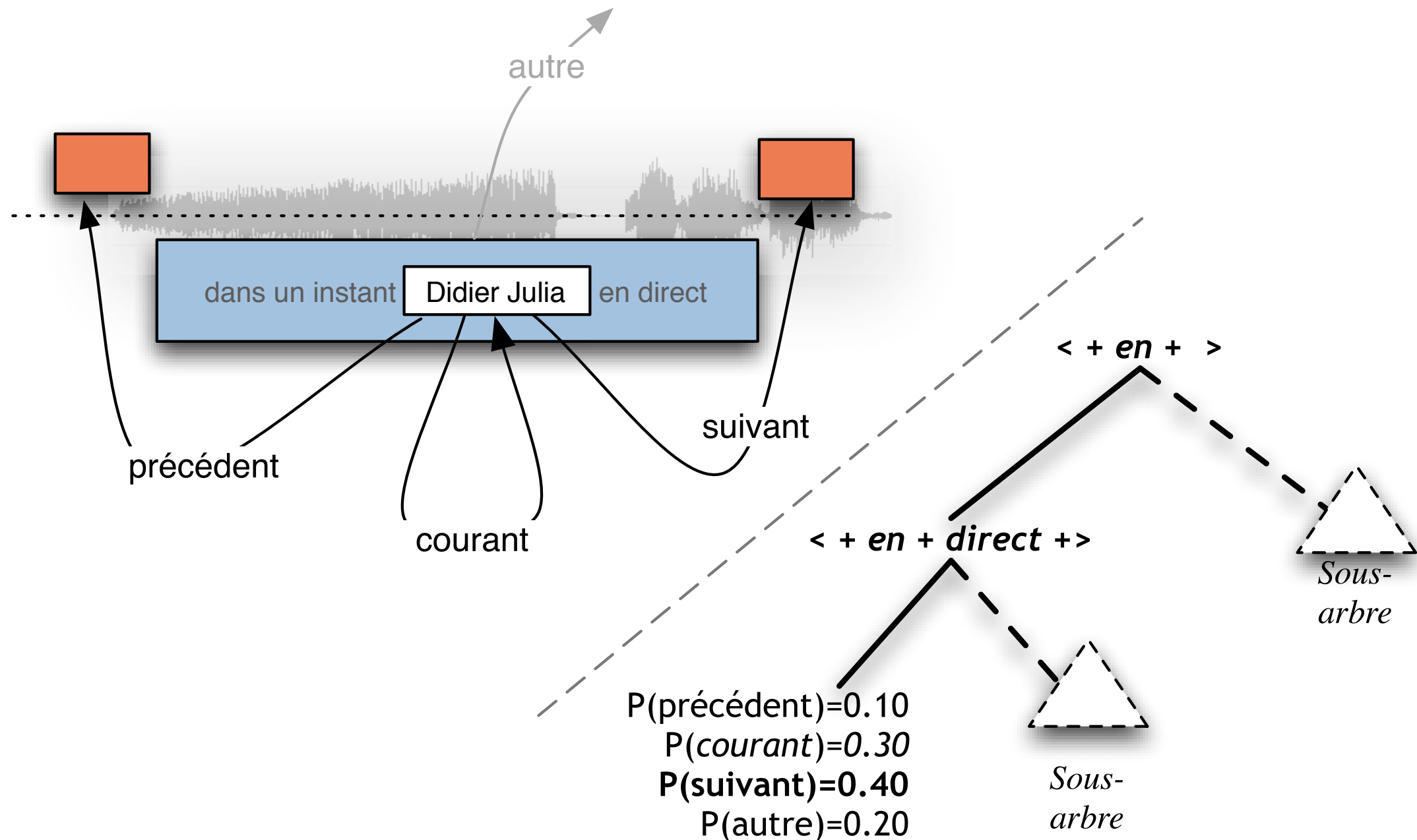


# Attribution des étiquettes (I)



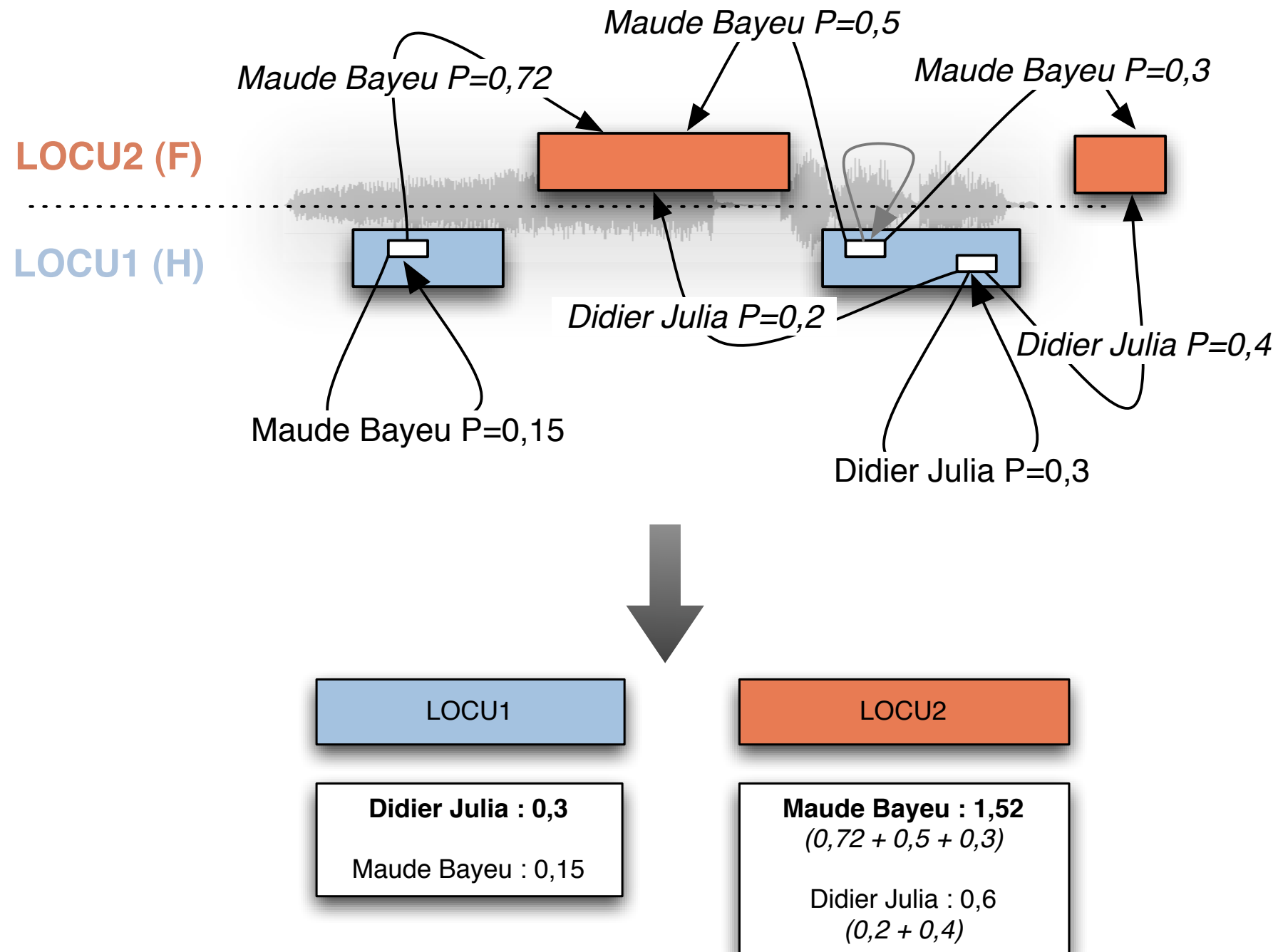


# Attribution des étiquettes (2)

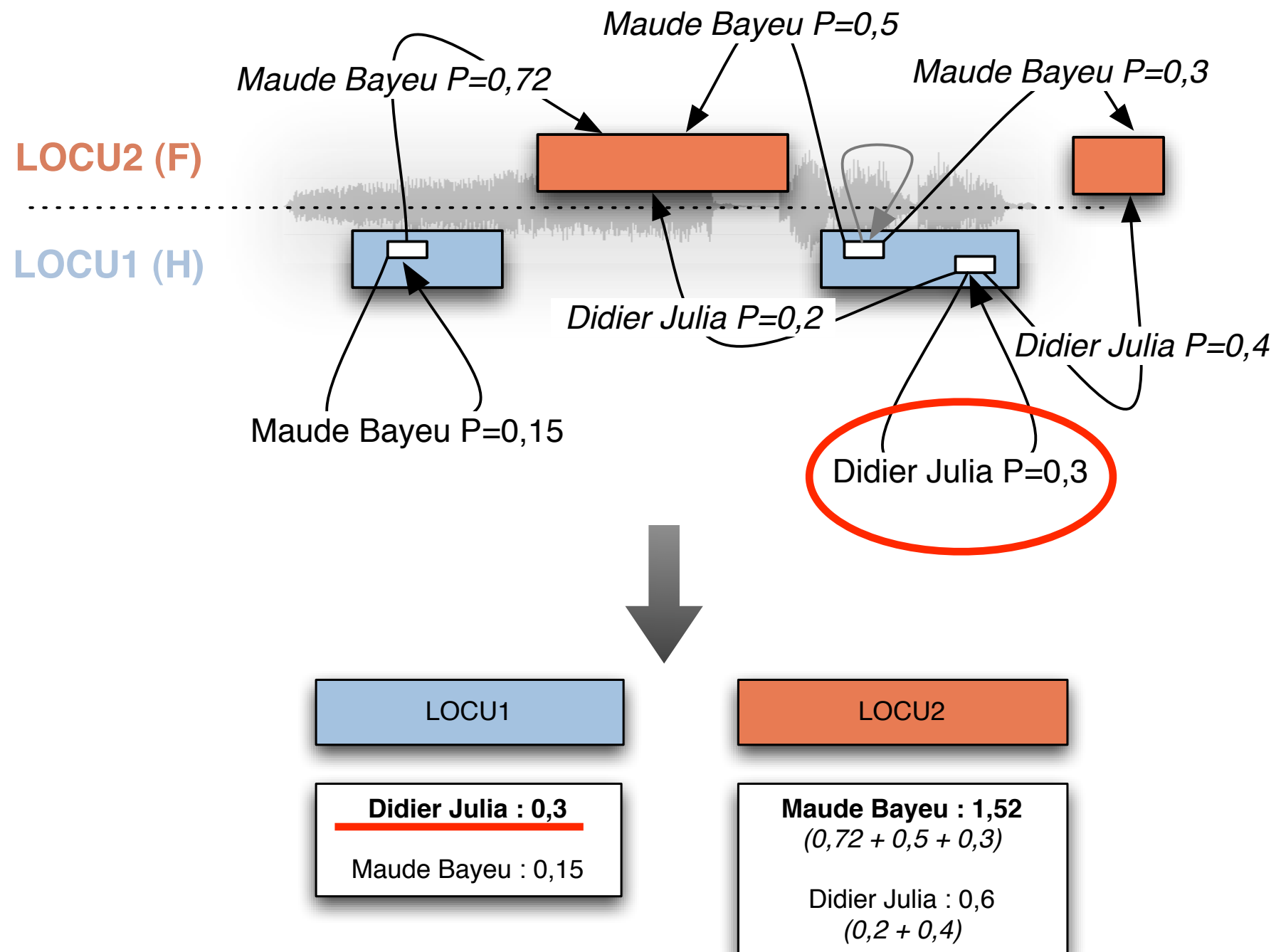




# Décision globale

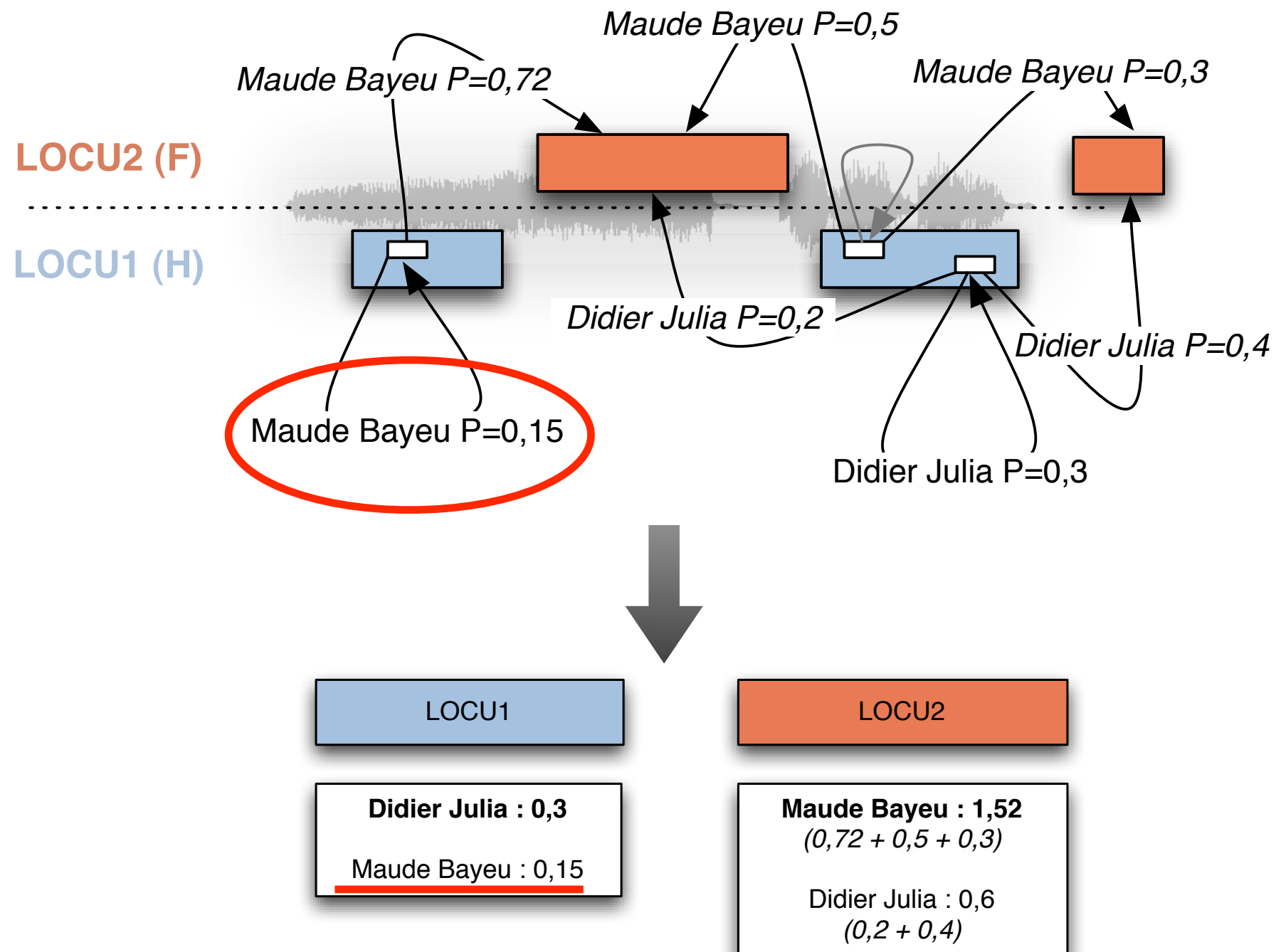


# Décision globale

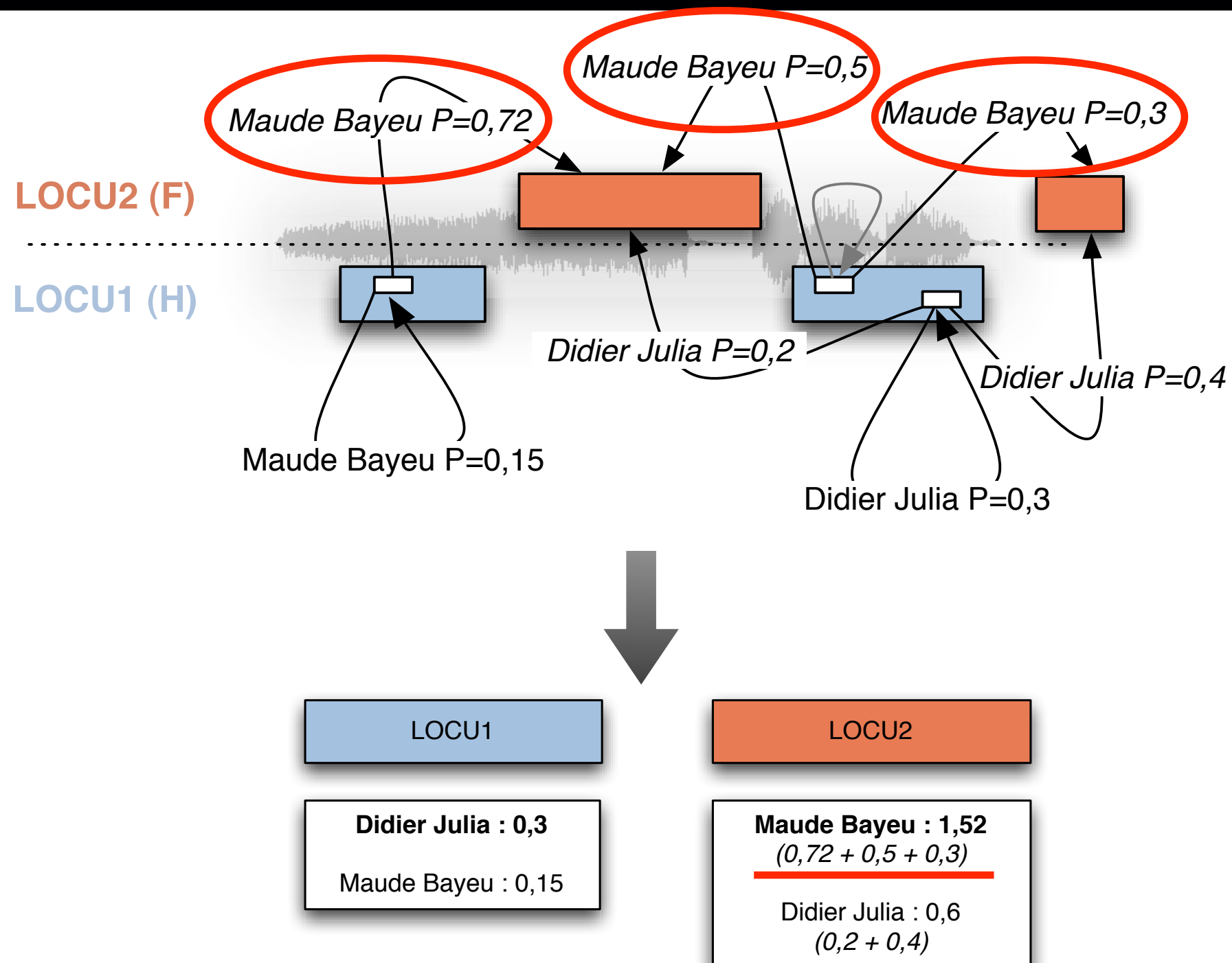




# Décision globale

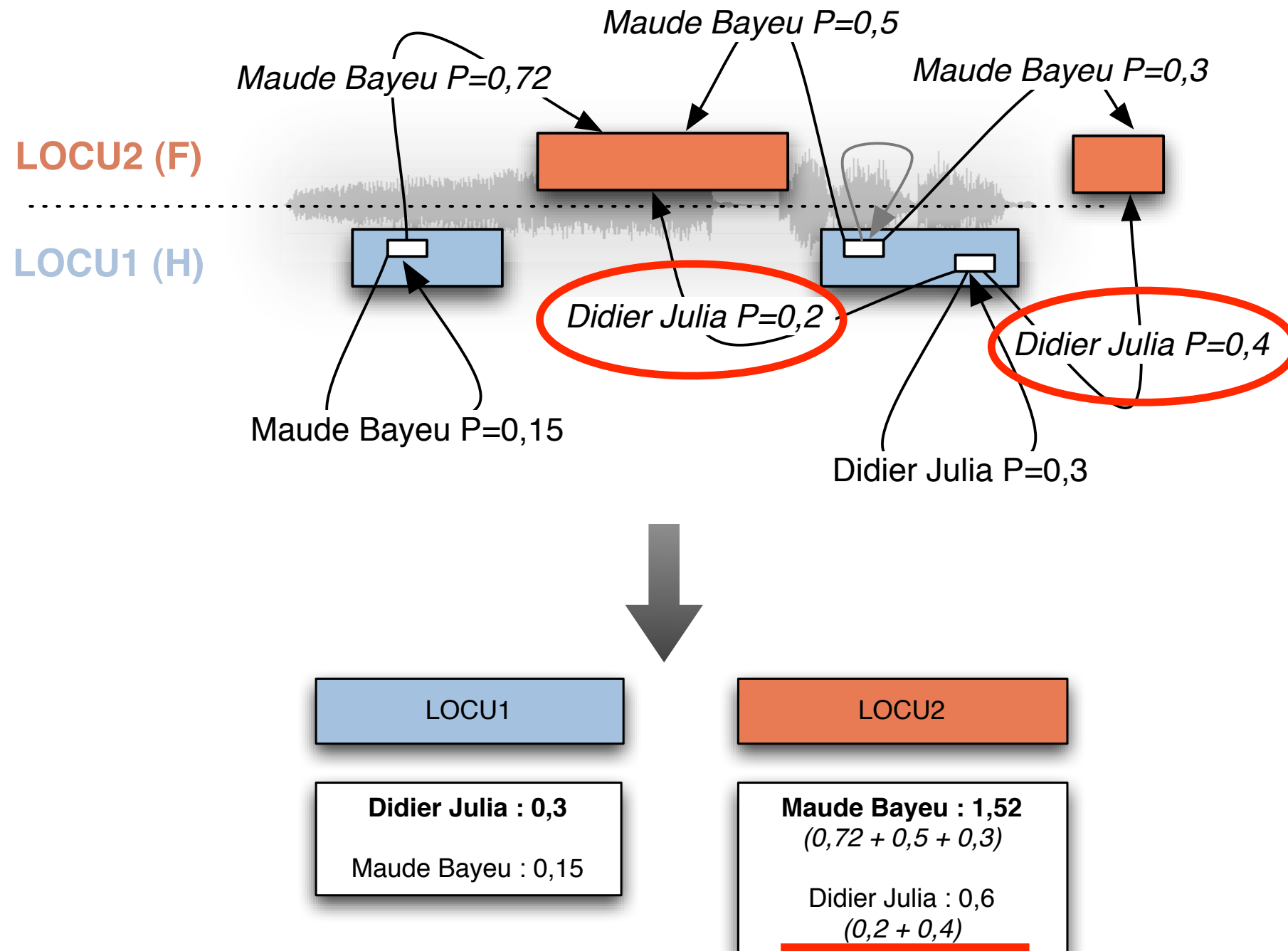


# Décision globale





# Décision globale



# Existant : système du LIUM à base de SCT

- Utilisation de Nemesis pour détecter les entités nommées
- Analyse des erreurs



# Analyse des erreurs

transcriptions manuelles, corpus développement

Problèmes	Nombre	% total
Entités nommées	16	18,6%
Arbre / Décision	62	72,1%
Hypothèse non vérifiée	3	2,6%
Autres problèmes (transcription, noms partiels)	5	6,7%

Nombre total d'EN détectées : 1445



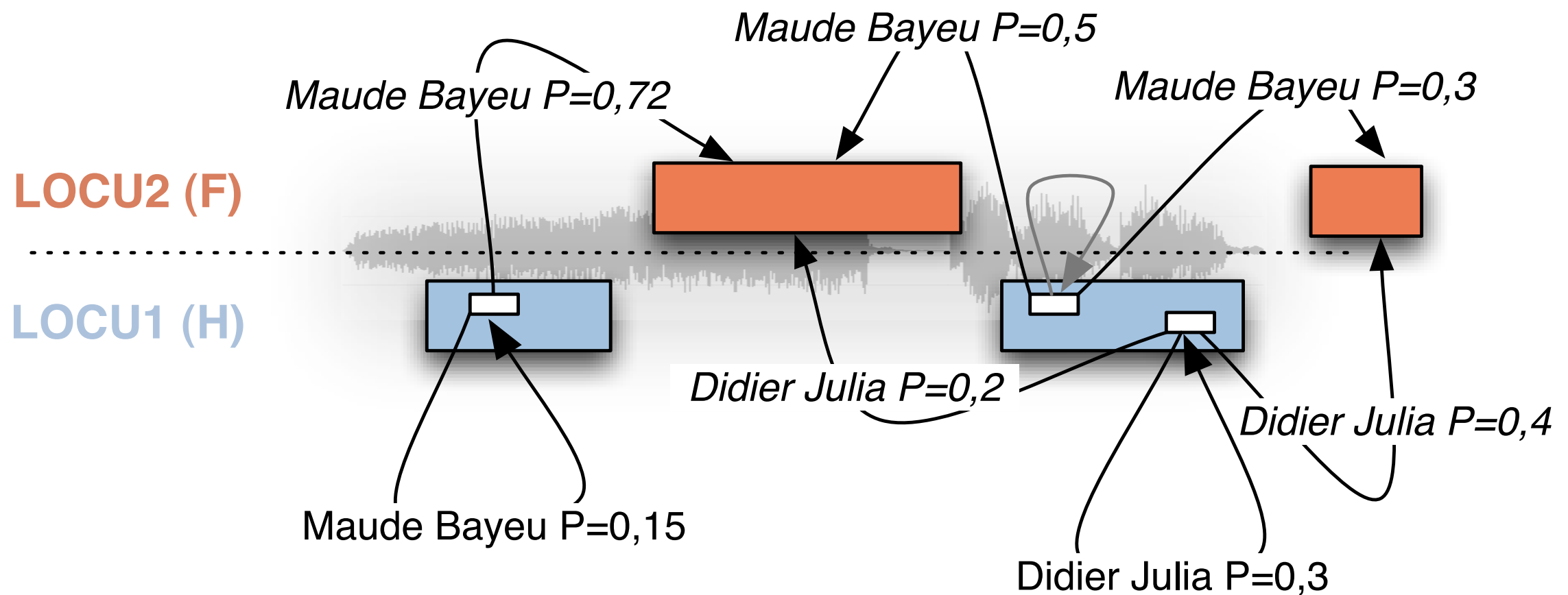
# Nouveau système *Milesin*



# Apports

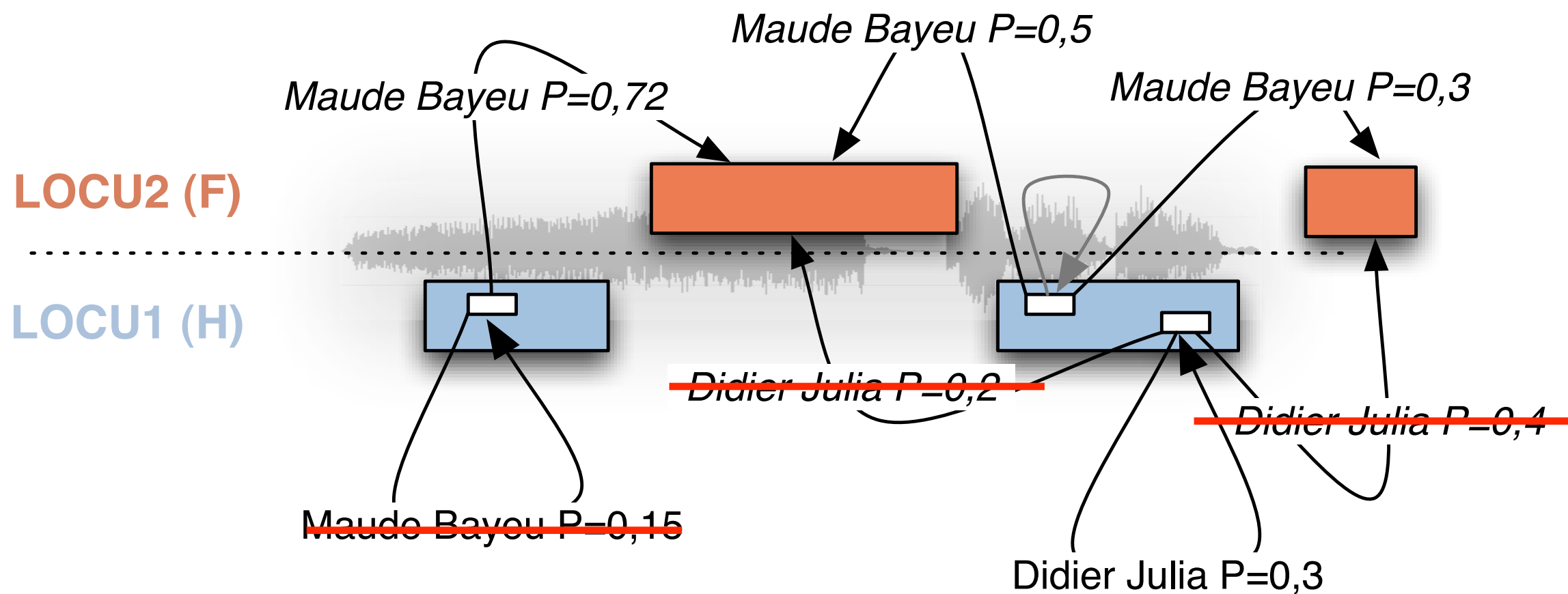
- Utilisation de LIA\_NE : meilleur système de la campagne ESTER 2
- Prise en compte du genre des locuteurs
- Nouveau processus de décision

# Prise en compte du genre

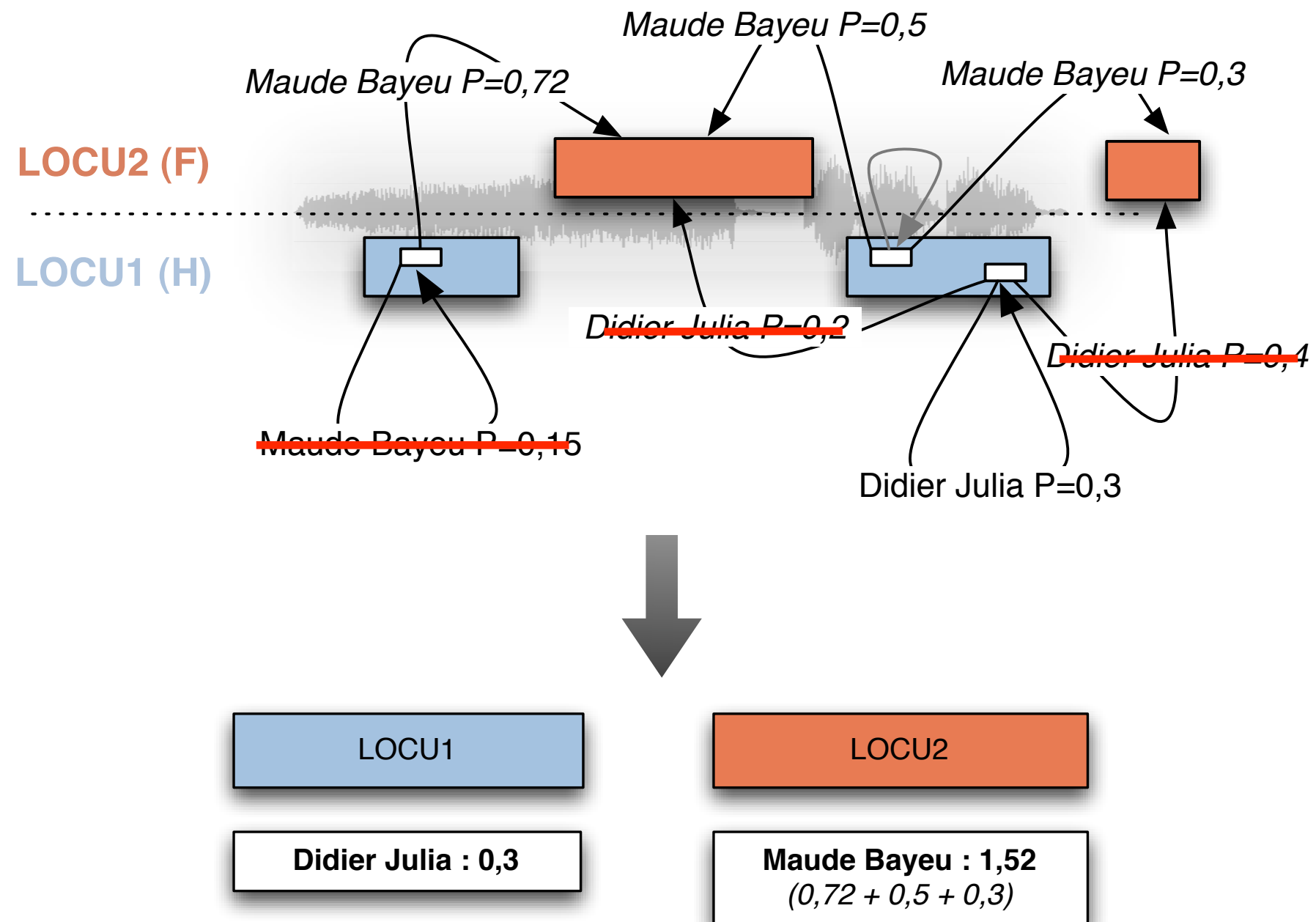




# Prise en compte du genre



# Décision globale





# Difficultés de la décision

- Informations incomplètes qui peuvent se renforcer
- Conflits lors de l'affectation d'un nom



# Fonctions de croyance (I)

- Théorie des fonctions de croyances
- Introduite par Dempster & Shafer (76)
  - Modélise la croyance, l'ignorance et le conflit
  - Permet de raisonner avec des connaissances imparfaites



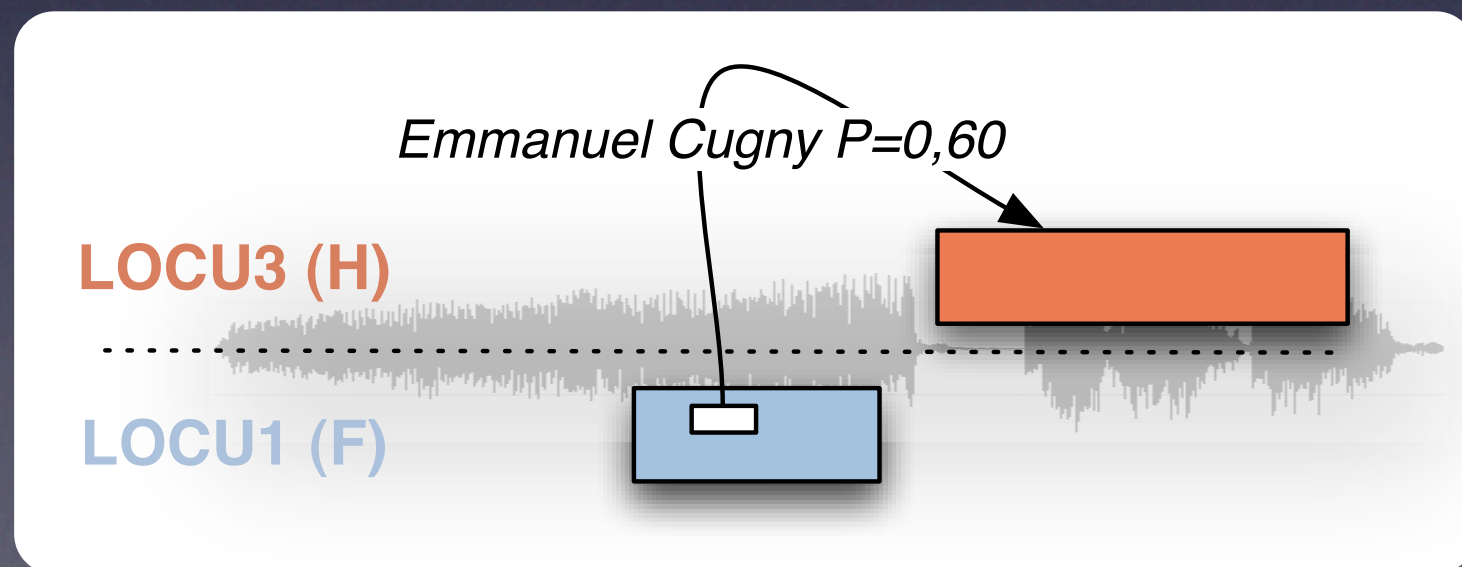
# Fonctions de croyance (2)

- Fonction de croyance  $m$  sur  $\Omega$  (ensemble fini)
  - Application  $m: 2^\Omega \rightarrow [0, 1]$  t.q.  $\sum_{A \subseteq \Omega} m(A) = 1.$
- État de connaissance sur une variable dans  $\Omega$ 
  - $m(X)$  = part de croyance allouée à l'hypothèse  $X$
  - $m(\Omega)$  = degré d'ignorance
  - $m(\emptyset)$  = degré de conflit



# Exemple (I)

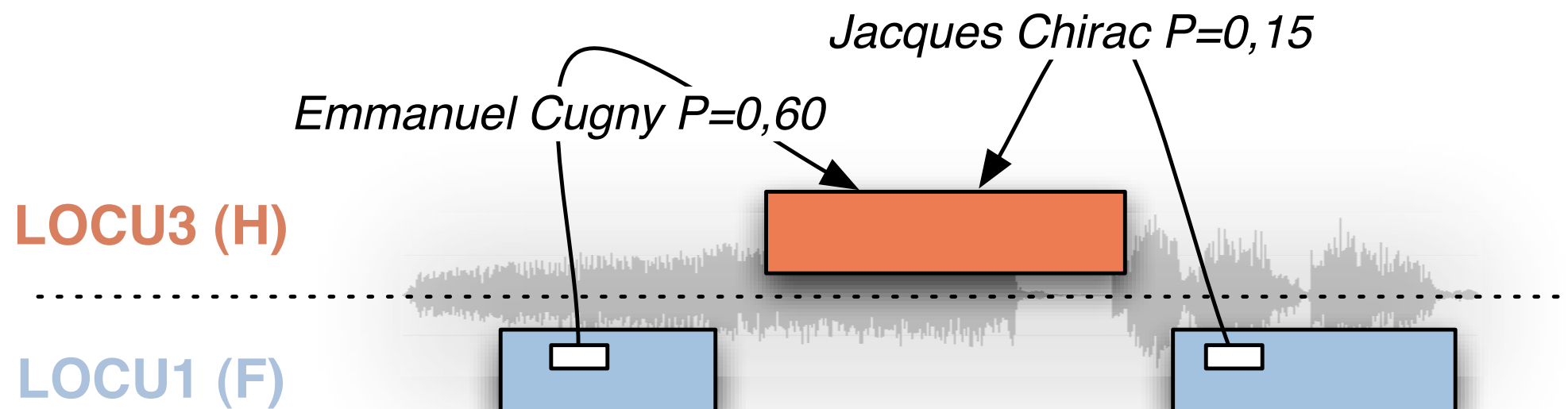
- $\Omega = \{ \text{Emmanuel Cugny, Jacques Chirac, ...} \}$
- Pour le locuteur LOCU3
  - Score SCT : E. Cugny  $\Rightarrow$  suivant = 0,60
  - $m_I(\text{Emmanuel Cugny}) = 0,60$
  - $m_I(\Omega) = 0,40$
  - $m_I(\emptyset) = 0$





# Exemple (2)

- J. Chirac => précédent = 0,15
- $m_2(\text{Jacques Chirac}) = 0,15$
- $m_2(\Omega) = 0,85$
- $m_2(\emptyset) = 0$



# Combinaison (I)

- Soient 2 fonctions de croyances  $m_1$  et  $m_2$
- Opérateur de combinaison  $m_{1,2} = m_1 \cap m_2$

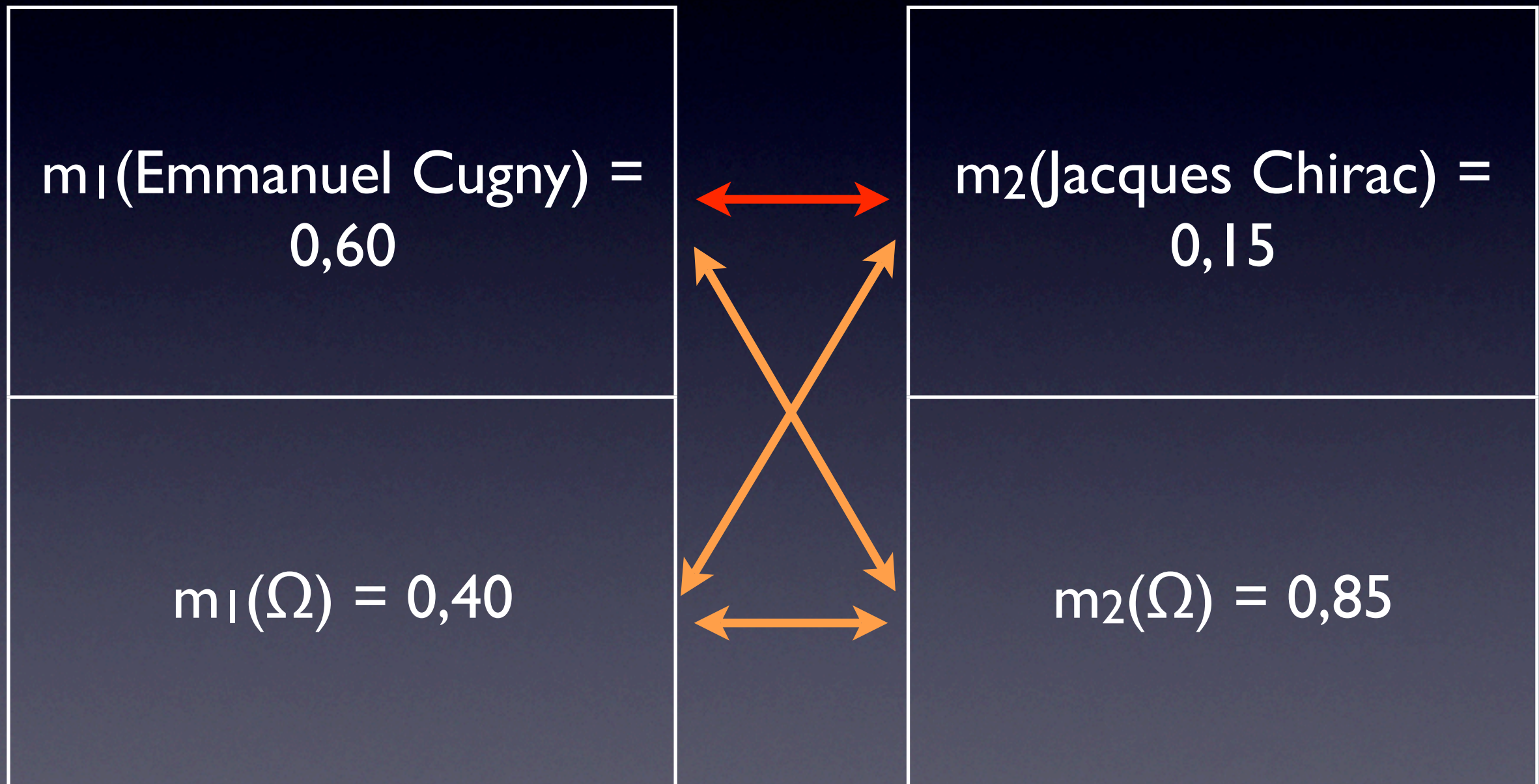
$$\forall A \subseteq \Omega, m_{1,2}(A) = \sum_{B \cap C = A} m_1(B) m_2(C).$$



# Combinaison (2)

$m_1$

$m_2$





# Combinaison (3)

$m_{1,2}$

$m_{1,2}(\mathbf{Emmanuel\ Cugny})$	$0,6 \times 0,85 = \mathbf{0,51}$
$m_{1,2}(\mathbf{Jacques\ Chirac})$	$0,4 \times 0,15 = \mathbf{0,06}$
$m_{1,2}(\Omega)$	$0,4 \times 0,85 = \mathbf{0,34}$
$m_{1,2}(\emptyset)$	$\mathbf{0,09}$

$$\Sigma = 1$$



# Problème d'assignation

Locu 3 (H)

**Emmanuel Cugny : 0,51**  
**Jacques Chirac : 0,06**

...

Locu 1 (F)

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

...

Locu 4 (H)

**Emmanuel Cugny : 0,4**  
**Joel Collado : 0,2**

# Algorithme de Kuhn Munkres

- Problème classique en recherche opérationnelle :  $N$  travailleurs pour  $M$  tâches
- Algorithme de Kuhn-Munkres : maximise une fonction de coût entre les locuteurs et les noms



# Matrice d'assignation

	E. Cugny	J. Chirac	M. Bayeu	B. Dagot	J. Collado
LOCUI	0	0	0,89	0,2	0
LOCU3	0,51	0,06	0	0	0
LOCU4	0,4	0	0	0	0,2



# Évaluation



# Corpus

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



# Métriques d'évaluation

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

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

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

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



# Transcriptions manuelles

Systeme	Corpus	Tx. Erreur Durée	Tx. Erreur Nombre
Base (LIUM 06)	Dev	26,64 %	37,40%
Proposé (Milesin)	Dev	11,44 %	12,43 %
Proposé (Milesin)	Test	22,85 %	28,23 %



# Transcriptions automatiques

Corpus	Tx. Err. Dur.
Développement	69,43 %
Test	61,23 %

Systeme du LIUM pour ESTER2

WER : 17,83 %

DER : 10 %



# Conclusion



# Conclusion

- Système basé sur une analyse conjointe signal/texte disponible en OpenSource
- Système de décision : taux d'erreur divisé par deux
- Utilisation de transcriptions entièrement automatiques



# Perspectives

- Passage à l'automatique : analyse des erreurs
- Travail sur les noms propres : modèles n-grammes à classes et phonétisation automatique
- Adaptation des systèmes à la tâche
- Exploitation d'autres informations (sous-titres, modèles acoustiques, ...)



Merci de votre attention



# Influence de la transcription

Transcription	En durée			En nb de locuteurs
	Rappel	Précision	ErrDur	ErrLoc
LIUM	41,74 %	85,44 %	54,79 %	61,35 %
LIMSI	56,31 %	86,17 %	41,24 %	51,79 %

# Influence de la transcription

Transcription	En durée			En nb de locuteurs
	Rappel	Précision	ErrDur	ErrLoc
LIUM	41,74 %	85,44 %	54,79 %	61,35 %
LIMSI	56,31 %	86,17 %	41,24 %	51,79 %



# Vers des transcriptions automatiques

		En durée			En nb de Locuteur
Trans.	Seg/Class.	R	P	ErrDur	ErrLoc
Corpus de développement					
M	M	88,35 %	96,46 %	10,95 %	11,46 %
M	A	59,66 %	79,47 %	38,35 %	- %
A	M	41,74 %	85,44 %	54,79 %	61,35 %
A	A	25,15 %	65,99 %	69,85 %	-
Corpus de test					
M	M	78,00 %	95,68 %	21,95 %	25,81 %
M	A	34,41 %	64,62 %	65,59 %	-
A	M	54,66 %	89,03 %	45,34 %	49,19 %
A	A	42,50 %	85,48 %	57,50 %	-

# Listes de locuteurs

	En durée			En nombre
	R	P	ErrDur	ErrLoc
<i>Corpus de Développement</i>				
Liste globale	87,85%	95,24%	11,44%	12,43 %
Liste par enregistrement	88,35 %	96,46 %	10,95 %	11,46 %
Sans liste	72,86 %	73,39 %	29,27 %	30,22 %
<i>Corpus de Test</i>				
Liste globale	77,10%	92,97%	22,85%	28,23%
Liste par enregistrement	78,00 %	95,68 %	21,95 %	25,81 %
Sans liste	61,77%	64,38 %	38,16 %	35,48 %



FIN

# 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.*

# Matrice d'assignation

	E. Cugny	J. Chirac	M. Bayeu	B. Dagot	J. Collado
LOCUI	0	0	1,52	0,2	0
LOCU3	0,51	0,06	0	0	0
LOCU4	0,4	0	0	0	0,2



# Métriques

- Précision / Rappel

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

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

- Taux d'erreur

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

- En durée
- En nombre de locuteurs

# Transcriptions manuelles et entités nommées

Systeme	Tx. Erreur Durée	Tx. Erreur Nombre
Nemesis	17,77 %	21,58 %
LIA_NE	11,44 %	12,43 %



# Publications

- Speaker identification using belief functions - *IPMU 2010 (IEEE), Dortmund*
- Analyse conjointe du signal sonore et de sa transcription pour l'identification nommée du locuteur - *Revue TAL*
- 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*