

# Fusion d'informations pour la reconnaissance automatique d'adresses postales dans le cadre de la théorie des fonctions de croyance

David Mercier<sup>1,2</sup>

<sup>1</sup>Laboratoire HeuDiaSyC, Université de Technologie de Compiègne

<sup>2</sup>SOLYSTIC, Gentilly

Sous la direction de Thierry Denœux<sup>1</sup> et Marie-Hélène Masson<sup>1</sup>

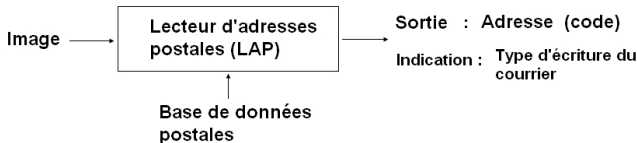
Encadrement Solystic Geneviève Cron<sup>2</sup>

Mercredi 6 Décembre 2006



# Introduction

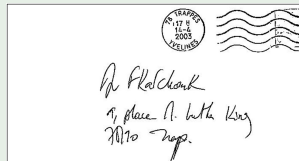
- L'entreprise **Solystic** conçoit et commercialise des **lecteurs d'adresses postales (LAP)** en charge de la reconnaissance d'adresses postales :



## Un envoi dactylographié



## Un envoi manuscrit



# Introduction

- Plusieurs étapes sont nécessaires pour passer d'une image à une adresse postale.
  - Chacune implique :
    - différents algorithmes ;
    - différentes chaînes de décision.
- ⇒ La reconnaissance automatique d'adresses postales est un problème difficile.
- Une voie d'amélioration réside dans la fusion ou combinaison d'informations issues de plusieurs LAP.
- ⇒ Ce travail de thèse a consisté à proposer des solutions à ce problème de fusion.

# Introduction

- Plusieurs étapes sont nécessaires pour passer d'une image à une adresse postale.
- Chacune implique :
  - différents algorithmes ;
  - différentes chaînes de décision.

⇒ La reconnaissance automatique d'adresses postales est un problème difficile.

- Une **voie d'amélioration** réside dans la **fusion** ou **combinaison** d'informations issues de plusieurs LAP.

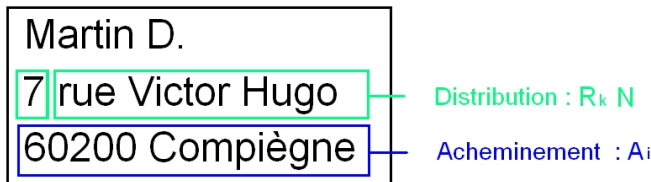
⇒ Ce travail de thèse a consisté à proposer des solutions à ce problème de fusion.



# Caractéristiques de ce problème de fusion

## ⇒ Granularité des sorties

- Une adresse postale se décompose en plusieurs niveaux.



- Au regard de cet envoi, un LAP peut fournir :

| Sortie                                      | Niveau        |
|---------------------------------------------|---------------|
| une adresse complète : « $A_i R_k N$ »      | distribution  |
| une adresse partielle : « $A_i R_k Rejet$ » | rue           |
| une adresse partielle : « $A_i Rejet$ »     | acheminement  |
| une adresse partielle : « Rejet »           | rejet complet |

# Caractéristiques de ce problème de fusion

## ⇒ Informations statistiques

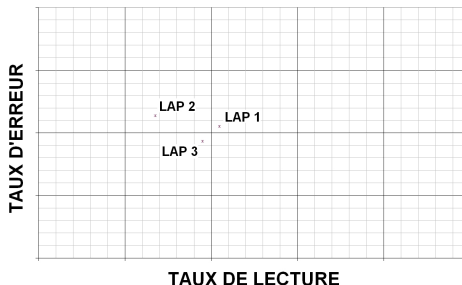
À partir de données étiquetées (ensemble d'apprentissage), on peut calculer à chaque niveau :

- **taux de lecture**  $L = \frac{\text{nombre de bonnes reconnaissances}}{\text{nombre soumis}}$  ;
- **taux d'erreur**  $E = \frac{\text{nombre d'erreurs}}{\text{nombre soumis}}$  ;
- **taux de rejet**  $R = \frac{\text{nombre de rejets effectués}}{\text{nombre soumis}}$  .

On a  $L + E + R = 1$  .

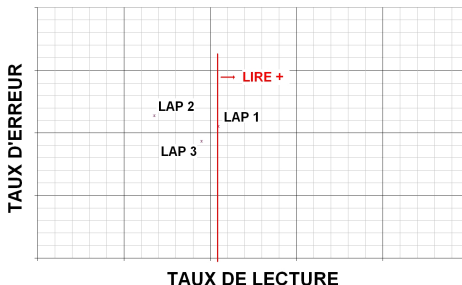
# Objectif

- Améliorer les performances de reconnaissance d'adresses en intégrant des informations issues de plusieurs LAP.
- Plus précisément :



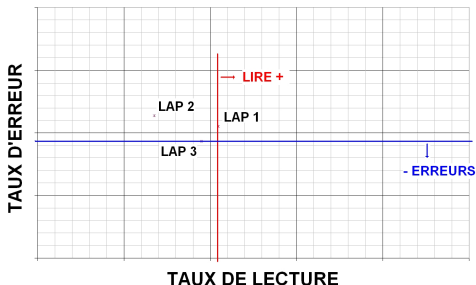
# Objectif

- Améliorer les performances de reconnaissance d'adresses en intégrant des informations issues de plusieurs LAP.
- Plus précisément :



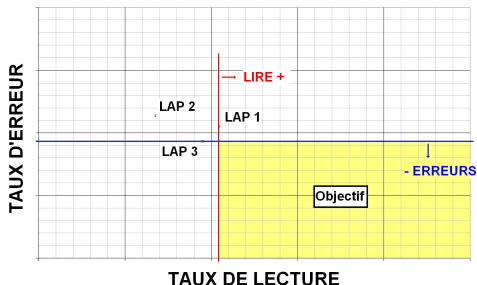
# Objectif

- **Améliorer les performances** de reconnaissance d'adresses en intégrant des informations issues de plusieurs LAP.
- Plus précisément :



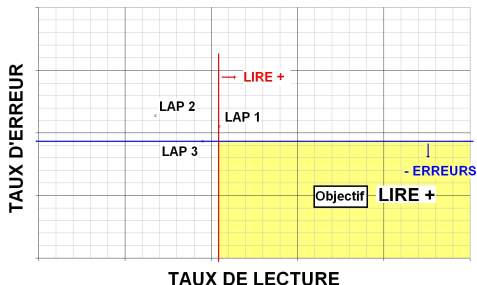
# Objectif

- Améliorer les performances de reconnaissance d'adresses en intégrant des informations issues de plusieurs LAP.
- Plus précisément :



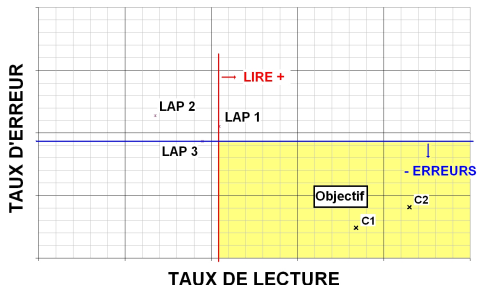
# Objectif

- Améliorer les performances de reconnaissance d'adresses en intégrant des informations issues de plusieurs LAP.
- Plus précisément :



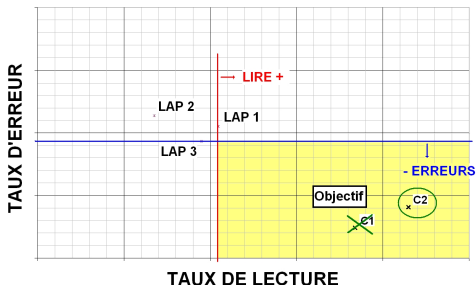
# Objectif

- Améliorer les performances de reconnaissance d'adresses en intégrant des informations issues de plusieurs LAP.
- Plus précisément :



# Objectif

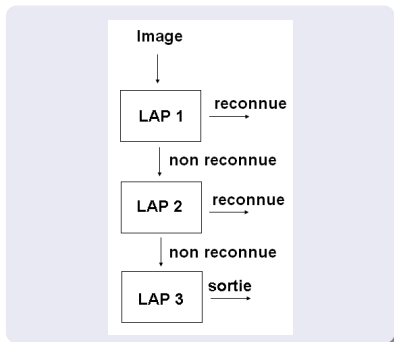
- Améliorer les performances de reconnaissance d'adresses en intégrant des informations issues de plusieurs LAP.
- Plus précisément :



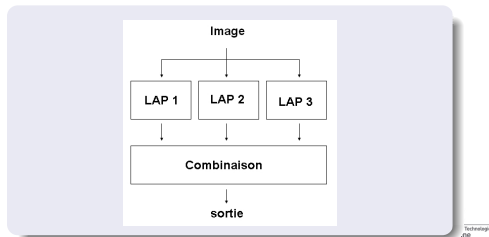
# Approches existantes

## Deux grands schémas

- Le schéma en cascade.



- le schéma en parallèle.



# Approches existantes

## Des combinaisons à base d'heuristiques

- À l'intérieur de ces schémas : essentiellement des combinaisons à base de règles :
  - Brevets ELSAG, Siemens, 2005.
  - Première approche Solystic.
- Ces règles essaient de prendre compte à la fois :
  - 1 la **redondance des informations**,
  - 2 et des **statistiques** de performance des LAP.

⇒ Peu de travaux ont été réalisés en fusion d'adresses postales.

# Approches existantes

## Des combinaisons à base d'heuristiques

- À l'intérieur de ces schémas : essentiellement des combinaisons à base de règles :
  - Brevets ELSAG, Siemens, 2005.
  - Première approche Solystic.
- Ces règles essaient de prendre compte à la fois :
  - ① la **redondance des informations**,
  - ② et des **statistiques** de performance des LAP.

⇒ Peu de travaux ont été réalisés en fusion d'adresses postales.

# Notre approche

⇒ Modélisation de cette fusion dans le cadre du **modèle des croyances transférables**.

- Flexibilité dans la modélisation des granularités et statistiques disponibles.
- Ce modèle a déjà fait ces preuves en fusion d'informations.

# Plan

- 1 Introduction
- 2 Le Modèle des Croyances Transférables (MCT)
- 3 Mécanismes de correction d'une fonction de croyance (MC)
- 4 Modélisation du problème postal dans le cadre du MCT
- 5 Améliorations du modèle
- 6 Conclusion et perspectives

# Plan

- 1 Introduction
- 2 Le Modèle des Croyances Transférables (MCT)
- 3 Mécanismes de correction d'une fonction de croyance (MC)
- 4 Modélisation du problème postal dans le cadre du MCT
- 5 Améliorations du modèle
- 6 Conclusion et perspectives

# Qu'est ce que le MCT ?

- Le MCT est un modèle de représentation et de manipulation de données imparfaites (incertaines et imprécises).
- Il trouve ses origines dans les travaux de **Dempster (1966-1968)** et **Shafer (1976, « A Mathematical Theory of Evidence »)**.
- **Smets (dès 1978)**, développement du **Modèle des croyances transférables**
  - axiomatisation des travaux de Shafer,
  - interprétation subjective et non-probabiliste des **fonctions de croyance**.

# Le Modèle des croyances transférables

Deux niveaux sont distingués dans la modélisation du raisonnement d'un agent rationnel.

## 1 Le niveau crédal

- **Partie statique** : représentation des informations ;
- **Partie dynamique** : combinaison, manipulation des informations (raisonnement).

## 2 Le niveau décisionnel

- Lorsqu'une décision est demandée et que toutes les informations ont été synthétisées : **prise de décision**.

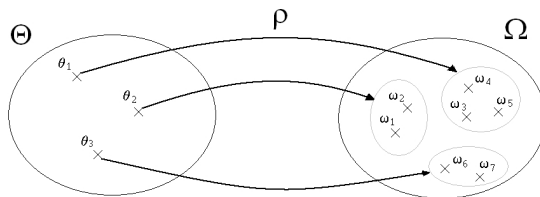
# Fonction de Masse (FM)

## Définitions

- $\Omega = \{\omega_1, \dots, \omega_K\}$  : ensemble fini de réponses à une certaine **question**  $Q$ .  $\Omega$  est appelé **cadre discernement**.
- Toute information relative à  $Q$ , détenue par un agent rationnel  $Ag$ , étant donné un ensemble de connaissances  $BC$ , peut être représentée par une **fonction de masse**  $m$  :  $2^\Omega \rightarrow [0, 1]$ , vérifiant :  $\sum_{A \subseteq \Omega} m(A) = 1$ .
- **Notation complète** :  $m_S^\Omega[BC]$ .
- $m(A)$  : part de croyance allouée au sous-ensemble  $A$  de  $\Omega$ .
- $A \subseteq \Omega$  t.q  $m(A) > 0$  : **élément focal** de  $m$ .
- $m_A(A) = 1, A \subseteq \Omega$ , **fonction de masse catégorique**.
- $m_\Omega =$  **fonction de masse vide** : **ignorance totale**.

# Grossissements et raffinements

- La granularité du cadre de discernement est toujours déterminé par **convention**, chaque élément du cadre pouvant toujours être divisé en différentes alternatives.
- Ainsi les concepts de **grossissement** et **raffinement** sont définis.
- $\Theta$  et  $\Omega$  deux cadres de discernement.  $\Theta$  est un **grossissement** de  $\Omega$ , et  $\Omega$  est un **raffinement** de  $\Theta$  ssi :



# Manipulation des informations

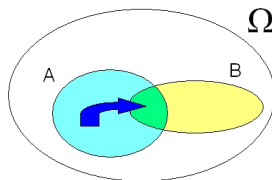
## Règle de Dempster

Deux fonctions de masse  $m_1$  et  $m_2$  issues de sources fiables et distinctes peuvent être combinées par la **règle de Dempster** :

$$m_1 \odot m_2(A) = \sum_{B \cap C = A} m_1(B)m_2(C), \quad \forall A \subseteq \Omega.$$

## Conditionnement

$m$  une FM,  $B \subseteq \Omega$ , la **fonction de masse conditionnelle**  $m[B]$  est définie par :  $m[B] = m \odot m_B$ .



# Principe de minimum d'information

« Choisir la fonction de masse la moins informative parmi l'ensemble des fonctions de masse compatibles avec les informations disponibles ».

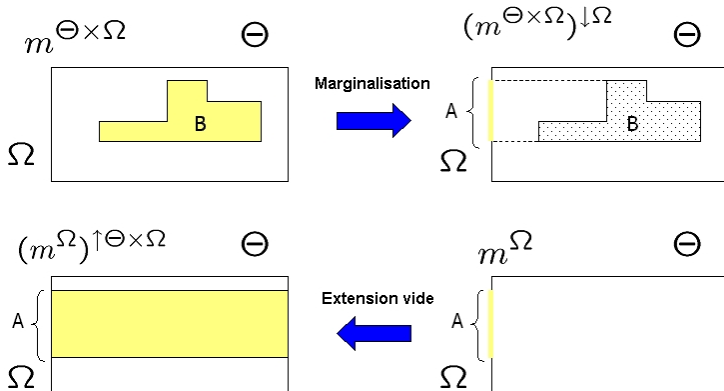
- Ce principe reflète une sorte de conservatisme, de précaution, dans le transfert des masses qui doivent être allouées aux sous-ensembles les plus larges possibles.

Applications :

- extension vide
- déconditionnement.

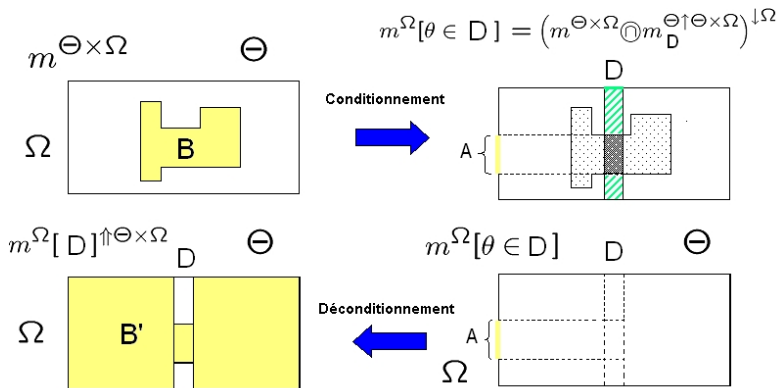
# Marginalisation et extension vide

Dans le cas d'un espace produit



# Conditionnement et déconditionnement

Dans le cas d'un espace produit



# Prise de décision

## Principes généraux

- Au moment de prendre une décision il faut préciser :
  - 1 l'ensemble des décisions  $\mathcal{D}$  pouvant être prises ;
  - 2 et, les états de la nature considérés qui sont regroupés dans le **cadre de pari**  $\Gamma$ .
- **Principes de rationalité** (Savage, DeGroot) : choisir la décision  $d$  minimisant le risque espéré

$$\rho(d) = \sum_{\gamma \in \Gamma} c(d, \gamma) P^{\Gamma}(\{\gamma\}), \quad \text{où}$$

- $P^{\Gamma} : 2^{\Gamma} \rightarrow [0, 1]$  : une mesure de probabilité
- $c : \mathcal{D} \times \Gamma \rightarrow \mathbb{R}$  : une fonction de coût.

# Prise de décision

## MCT

Dans le MCT, ces principes sont acceptés :

- $\Gamma$  : une composition de grossissements ou raffinements de  $\Omega$ ,
- on effectue la **transformation pignistique** :  
 $m^\Omega \rightarrow m^\Gamma \rightarrow \text{probabilité}$

$$BetP(\{\gamma\}) = \sum_{\{A \subseteq \Gamma, \gamma \in A\}} \frac{m(A)}{|A|} \times \frac{1}{1 - m(\emptyset)}.$$

# Prise de décision

Exemple :  $m(\Gamma) = m(\{\text{faire\_mammo}, \neg\text{faire\_mammo}\}) = 1 \implies$  équiprobabilité



# Plan

- 1 Introduction
- 2 Le Modèle des Croyances Transférables (MCT)
- 3 Mécanismes de correction d'une fonction de croyance (MC)**
- 4 Modélisation du problème postal dans le cadre du MCT
- 5 Améliorations du modèle
- 6 Conclusion et perspectives

# Introduction

- La prise en compte de la fiabilité d'une information  $m_S$  est classiquement réalisée par l'opération d'affaiblissement (Shafer, 1976) définie par :

$${}^{\alpha}m_S = (1 - \alpha) m_S + \alpha m_{\Omega} ,$$

où  $\alpha \in [0, 1]$  : taux d'affaiblissement de  $m_S$ .

- Néanmoins :

- la fiabilité de la source est représentée par un unique nombre dans cette opération ; or, une connaissance plus fine pourrait être disponible ; en particulier, cette fiabilité peut dépendre de la réponse à la question  $Q$   
 ⇒ introduction de l'affaiblissement contextuel
- cet opération permet seulement d'affaiblir une source, par exemple, ne pourrait-on pas la renforcer ? ⇒ introduction de mécanismes généraux de correction

# Introduction

- La prise en compte de la fiabilité d'une information  $m_S$  est classiquement réalisée par l'opération d'affaiblissement (Shafer, 1976) définie par :

$${}^{\alpha}m_S = (1 - \alpha) m_S + \alpha m_{\Omega} ,$$

où  $\alpha \in [0, 1]$  : taux d'affaiblissement de  $m_S$ .

- Néanmoins :

- la fiabilité de la source est représentée par un unique nombre dans cette opération ; or, une connaissance plus fine pourrait être disponible ; en particulier, cette fiabilité peut dépendre de la réponse à la question  $Q$

⇒ introduction de l'affaiblissement contextuel

- cet opération permet seulement d'affaiblir une source, par exemple, ne pourrait-on pas la renforcer ? ⇒

introduction de mécanismes généraux de correction



# Introduction

- La prise en compte de la fiabilité d'une information  $m_S$  est classiquement réalisée par l'opération d'affaiblissement (Shafer, 1976) définie par :

$${}^{\alpha}m_S = (1 - \alpha) m_S + \alpha m_{\Omega} ,$$

où  $\alpha \in [0, 1]$  : taux d'affaiblissement de  $m_S$ .

- Néanmoins :
  - la fiabilité de la source est représentée par un unique nombre dans cette opération ; or, une connaissance plus fine pourrait être disponible ; en particulier, cette fiabilité peut dépendre de la réponse à la question  $Q$   
 $\Rightarrow$  introduction de l'affaiblissement contextuel
  - cet opération permet seulement d'affaiblir une source, par exemple, ne pourrait-on pas la renforcer ?  $\Rightarrow$  introduction de mécanismes généraux de correction

# Introduction

- La prise en compte de la fiabilité d'une information  $m_S$  est classiquement réalisée par l'opération d'affaiblissement (Shafer, 1976) définie par :

$${}^{\alpha}m_S = (1 - \alpha) m_S + \alpha m_{\Omega} ,$$

où  $\alpha \in [0, 1]$  : taux d'affaiblissement de  $m_S$ .

- Néanmoins :
  - la fiabilité de la source est représentée par un unique nombre dans cette opération ; or, une connaissance plus fine pourrait être disponible ; en particulier, cette fiabilité peut dépendre de la réponse à la question  $Q$   
 $\Rightarrow$  introduction de l'affaiblissement contextuel
  - cet opération permet seulement d'affaiblir une source ; par exemple, ne pourrait-on pas la renforcer ?  
 $\Rightarrow$  introduction de mécanismes généraux de correction

# Introduction

- La prise en compte de la fiabilité d'une information  $m_S$  est classiquement réalisée par l'opération d'affaiblissement (Shafer, 1976) définie par :

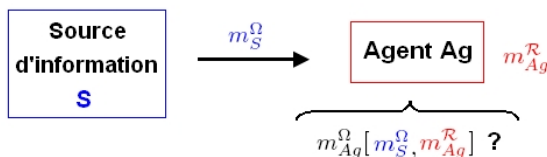
$${}^{\alpha}m_S = (1 - \alpha) m_S + \alpha m_{\Omega} ,$$

où  $\alpha \in [0, 1]$  : taux d'affaiblissement de  $m_S$ .

- Néanmoins :
  - la fiabilité de la source est représentée par un unique nombre dans cette opération ; or, une connaissance plus fine pourrait être disponible ; en particulier, cette fiabilité peut dépendre de la réponse à la question  $Q$   
 $\Rightarrow$  introduction de l'affaiblissement contextuel
  - cet opération permet seulement d'affaiblir une source ; par exemple, ne pourrait-on pas la renforcer ?  
 $\Rightarrow$  introduction de mécanismes généraux de correction

# Modèle général d'une correction

Les modèles de ces deux mécanismes de correction dérivent du modèle suivant :



- L'agent *Ag* cherche la réponse à la question *Q*.
- Il reçoit une information  $m_S^\Omega$  issue d'une source *S*.
- Par ailleurs, *Ag* possède une information  $m_{Ag}^R$  concernant l'état de fiabilité de la source.
- L'agent *Ag* doit alors **corriger** l'information issue de *S* à partir de sa connaissance de la fiabilité de *S*.

# Modèle de l'affaiblissement contextuel

- La source peut être dans deux états : **fiable (F)** ou **non fiable (NF)**,  $\mathcal{R} = \{F, NF\}$ .
- Si  $S$  est fiable (F) :  $m_{Ag}^\Omega[F] = m_S^\Omega$ , si  $S$  est non fiable (NF) :  $m_{Ag}^\Omega[NF](\Omega) = 1$ .
- La connaissance de l'agent sur la fiabilité de la source est décrite conditionnellement aux éléments d'un grossissement  $\Theta = \{\theta_1, \dots, \theta_L\}$  de  $\Omega$ , c'est-à-dire par  $L$  fonctions de masse conditionnelles  $m_{Ag}^{\mathcal{R}}[\theta_k]$ ,  $k = 1, \dots, L$ , de la forme :

$$\begin{cases} m_{Ag}^{\mathcal{R}}[\theta_k](\{F\}) &= 1 - \alpha_k, \\ m_{Ag}^{\mathcal{R}}[\theta_k](\mathcal{R}) &= \alpha_k. \end{cases}$$

# Modèle de l'affaiblissement contextuel

- La source peut être dans deux états : **fiable (F)** ou **non fiable (NF)**,  $\mathcal{R} = \{F, NF\}$ .
- Si  $S$  est fiable (F) :  $m_{Ag}^\Omega[F] = m_S^\Omega$ , si  $S$  est non fiable (NF) :  $m_{Ag}^\Omega[NF](\Omega) = 1$ .
- La connaissance de l'agent sur la fiabilité de la source est décrite conditionnellement aux éléments d'un grossissement  $\Theta = \{\theta_1, \dots, \theta_L\}$  de  $\Omega$ , c'est-à-dire par  **$L$  fonctions de masse conditionnelles  $m_{Ag}^{\mathcal{R}}[\theta_k]$ ,  $k = 1, \dots, L$ , de la forme :**

$$\begin{cases} m_{Ag}^{\mathcal{R}}[\theta_k](\{F\}) &= 1 - \alpha_k, \\ m_{Ag}^{\mathcal{R}}[\theta_k](\mathcal{R}) &= \alpha_k. \end{cases}$$

# Affaiblissement contextuel

Les informations disponibles peuvent alors être combinées de la manière suivante :

$$m_{Ag}^{\Omega} [m_S^{\Omega}, m_{Ag}^{\mathcal{R}}[\theta_1], \dots, m_{Ag}^{\mathcal{R}}[\theta_L]] = (m_{Ag}^{\Omega}[F]^{\uparrow\Omega \times \mathcal{R}} \odot m_{Ag}^{\mathcal{R}}[\theta_1]^{\uparrow\Omega \times \mathcal{R}} \odot \dots \odot m_{Ag}^{\mathcal{R}}[\theta_L]^{\uparrow\Omega \times \mathcal{R}})^{\downarrow\Omega}. \quad (1)$$

Le résultat obtenu est donné par :

$${}^{\alpha}m(A) = \sum_{B \subseteq A} {}^{\alpha}G(A, B)m(B), \quad \forall A \subseteq \Omega,$$

avec  $\alpha = (\alpha_1, \dots, \alpha_L)$  et :

$${}^{\alpha}G(A, B) = \begin{cases} \prod_{\cup \theta_k = (A \setminus B)^*} \alpha_k \prod_{\cup \theta_{\ell} = \bar{A}^*} 1 - \alpha_{\ell} & \text{si } \exists C = \cup \theta_i, B \cup C = A \\ 0 & \text{sinon.} \end{cases}$$

Si  $\Theta = \{\Omega\}$ , le résultat est identique à l'affaiblissement classique.

# Affaiblissement contextuel

Les informations disponibles peuvent alors être combinées de la manière suivante :

$$m_{Ag}^{\Omega} [m_S^{\Omega}, m_{Ag}^{\mathcal{R}}[\theta_1], \dots, m_{Ag}^{\mathcal{R}}[\theta_L]] = (m_{Ag}^{\Omega}[F]^{\uparrow\Omega \times \mathcal{R}} \odot m_{Ag}^{\mathcal{R}}[\theta_1]^{\uparrow\Omega \times \mathcal{R}} \odot \dots \odot m_{Ag}^{\mathcal{R}}[\theta_L]^{\uparrow\Omega \times \mathcal{R}})^{\downarrow\Omega}. \quad (1)$$

Le résultat obtenu est donné par :

$${}^{\alpha}m(A) = \sum_{B \subseteq A} {}^{\alpha}G(A, B)m(B), \quad \forall A \subseteq \Omega,$$

avec  $\alpha = (\alpha_1, \dots, \alpha_L)$  et :

$${}^{\alpha}G(A, B) = \begin{cases} \prod_{\cup \theta_k = (A \setminus B)^*} \alpha_k \prod_{\cup \theta_{\ell} = \bar{A}^*} 1 - \alpha_{\ell} & \text{si } \exists C = \cup \theta_i, B \cup C = A \\ 0 & \text{sinon.} \end{cases}$$

Si  $\Theta = \{\Omega\}$ , le résultat est identique à l'affaiblissement classique.

# Modèle de mécanismes généraux de correction

- La source peut être dans certains nombres d'états  $R_1, \dots, R_N$ ,  $\mathcal{R} = \{R_1, \dots, R_N\}$ .
- L'interprétation de ces états est donnée par une transformation  $m_i$  de  $m$  :
  - si la source est dans l'état  $R_i$ ,  $i \in 1, \dots, N$  :  $m_{Ag}^\Omega[R_i] = m_i$ .
- La connaissance de l'agent sur la fiabilité de la source est décrite par la fonction de masse  $m_{Ag}^\mathcal{R}$  définie par  $m_{Ag}^\mathcal{R}(\{R_i\}) = \gamma_i \forall i \in 1, \dots, N$ , avec  $\sum_{i=1}^N \gamma_i = 1$ .
- Les informations disponibles peuvent alors être combinées de la manière suivante :

$$m_{Ag}^\Omega[m_S^\Omega, m_{Ag}^\mathcal{R}] = \left( \bigodot_{i=1}^N m_{Ag}^\Omega[R_i]^{\uparrow\Omega \times \mathcal{R}} \bigodot m_{Ag}^{\mathcal{R} \uparrow \Omega \times \mathcal{R}} \right)^{\downarrow\Omega}$$

# Mécanismes généraux de correction

Le résultat dépend seulement des  $m_i$  et des  $\gamma_i$ , et il vérifie :

$$m_{Ag}^{\Omega} = \gamma m = \sum_{i=1}^N \gamma_i m_i ,$$

où  $\gamma$  désigne le vecteur composé des  $\gamma_i$ .

## Exemple

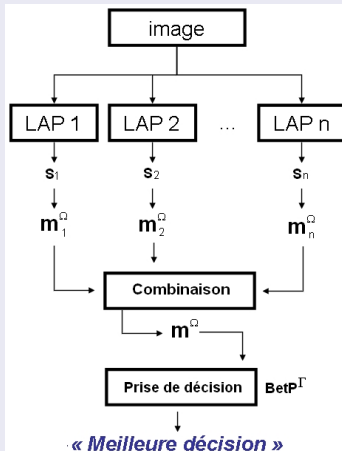
Un mécanisme de correction permettant de **renforcer** ou d'**affaiblir** une fonction de croyance est donné par :

$$\gamma m = \gamma_1 m_{\Omega} + \gamma_2 m + \gamma_3 {}^{tr}m , \quad \text{avec } {}^{tr}m = \begin{cases} \frac{m(A)}{1-m(\Omega)} & \forall A \subset \Omega , \\ 0 & \text{sinon .} \end{cases}$$

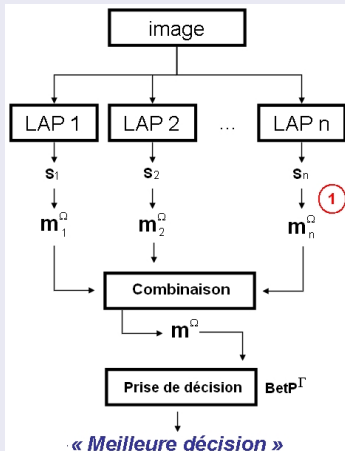
# Plan

- 1 Introduction
- 2 Le Modèle des Croyances Transférables (MCT)
- 3 Mécanismes de correction d'une fonction de croyance (MC)
- 4 Modélisation du problème postal dans le cadre du MCT**
- 5 Améliorations du modèle
- 6 Conclusion et perspectives

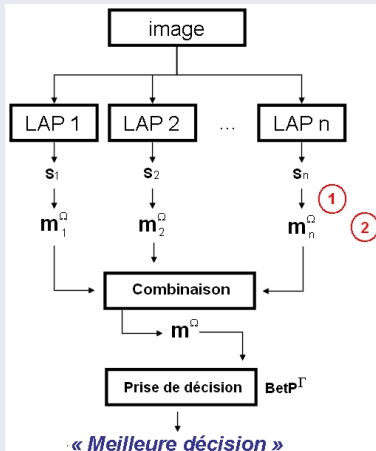
# Un modèle de fusion de décisions postales



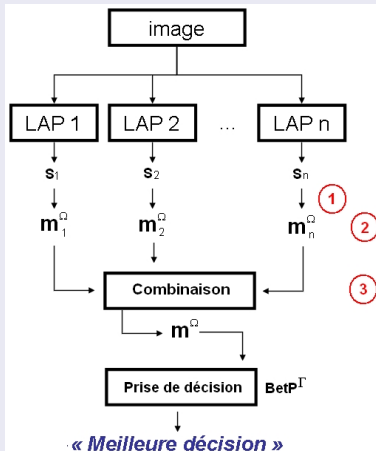
# Un modèle de fusion de décisions postales



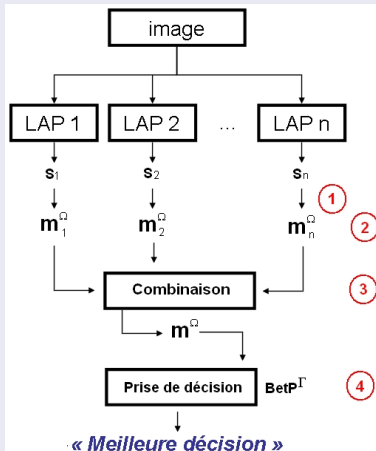
# Un modèle de fusion de décisions postales



# Un modèle de fusion de décisions postales



# Un modèle de fusion de décisions postales



## 1. Choix du cadre de discernement

## Choix du cadre de discernement

- **Question Q** : « À quelle adresse le courrier en cours d'analyse doit-il être envoyé ? »
- La base de données postales contient l'ensemble des adresses postales du pays traité.
- Est-ce que toutes les réponses à **Q** sont dans cette base de données ?
- Non, l'adresse d'un courrier postal peut aussi être invalide.

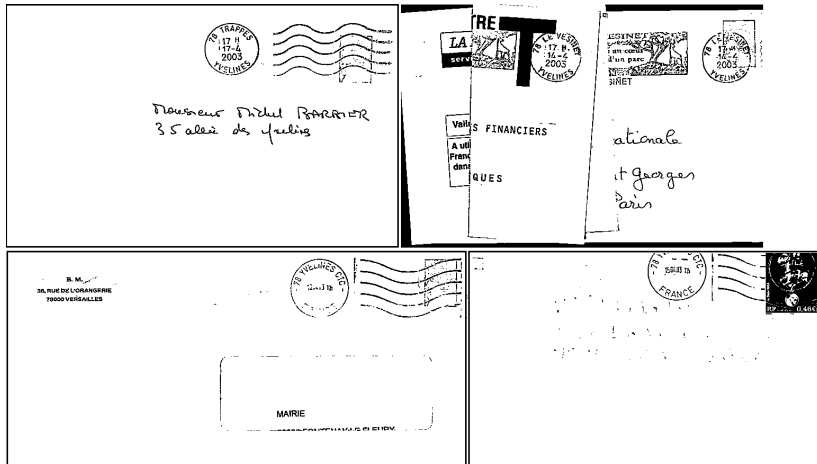
## 1. Choix du cadre de discernement

## Choix du cadre de discernement

- **Question Q** : « À quelle adresse le courrier en cours d'analyse doit-il être envoyé ? »
- La base de données postales contient l'ensemble des adresses postales du pays traité.
- Est-ce que toutes les réponses à **Q** sont dans cette base de données ?
- Non, l'adresse d'un courrier postal peut aussi être invalide.

## 1. Choix du cadre de discernement

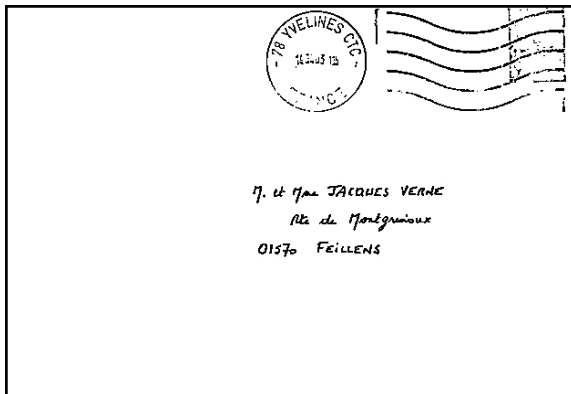
## Exemples (« lettres invalides »)



## 1. Choix du cadre de discernement

## Exemple (une adresse invalide partielle)

- La vérité de l'envoi suivant est de la forme «  $A_i R_j$  numéro de voie invalide ».



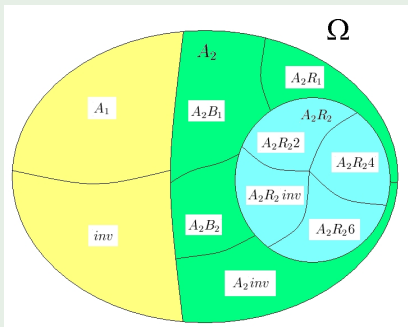
## 1. Choix du cadre de discernement

## Choix du cadre de discernement

Le **cadre de discernement**  $\Omega$  choisi est composé :

- ① des adresses de la base de données postales,
- ② et des adresses invalides (totalement et partiellement).

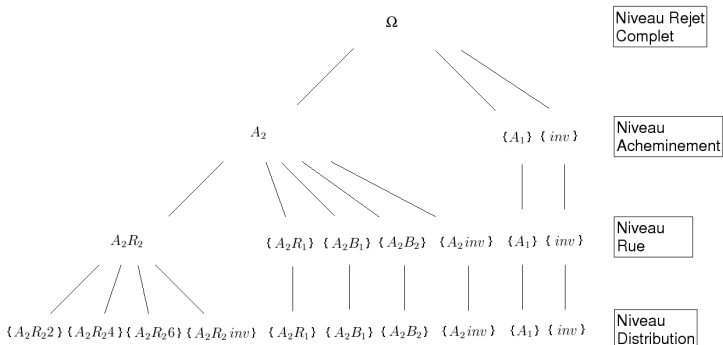
## Exemple sur une base de données réduite



## 1. Choix du cadre de discernement

# Organisation hiérarchique des sorties des LAP

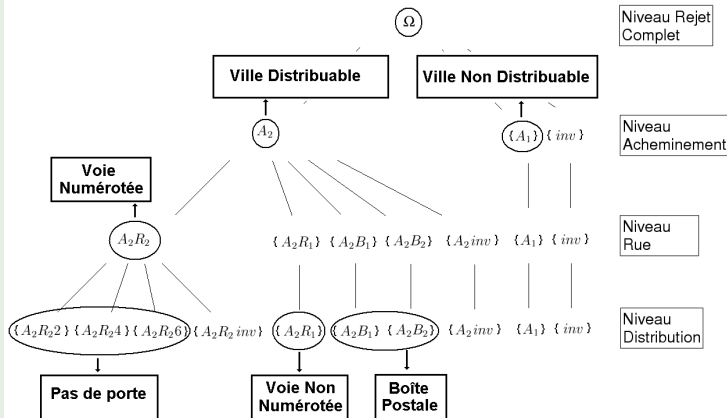
## Exemple sur une base de données réduite



## 1. Choix du cadre de discernement

## Ensemble des sorties possibles des LAP

## Exemple sur une base de données réduite



## 2. Construction des fonctions de masse

## Conversion de la sortie d'un LAP en une FM

- Un ensemble d'apprentissage composé de courriers étiquetés est disponible.
- Idée : allocation de la FM par l'utilisation de la matrice de confusion associée au LAP.**

⇒ Nous nous sommes inspirés de l'affectation de Xu et al. (1992), méthode déjà employée en combinaison de classifieurs sans organisation hiérarchique des sorties.

Affectation de  $m$  dans le cas  $s = \{\omega_k\}$ 

$$\{\omega_k\} \mapsto L = \frac{\sum_{k=1}^K n_{kk}}{n}$$

$$\Omega \setminus \{\omega_k\} \mapsto E = \frac{\sum_{k=1}^K \sum_{\ell=1; \ell \neq k}^K n_{k\ell}}{n}$$

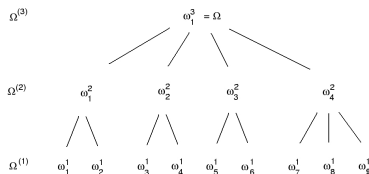
$$\Omega \mapsto R = \sum_{\ell=1}^K n_{(K+1)\ell} / n$$

| Vérité<br>Décision | $\omega_1$   | ...      | $\omega_K$   |
|--------------------|--------------|----------|--------------|
| $\omega_1$         | $n_{11}$     | ...      | $n_{1K}$     |
| $\vdots$           | $\vdots$     | $\ddots$ | $\vdots$     |
| $\omega_K$         | $n_{K1}$     | ...      | $n_{KK}$     |
| $Rejet = \Omega$   | $n_{(K+1)1}$ | ...      | $n_{(K+1)K}$ |

## 2. Construction des fonctions de masse

## Conversion de la sortie d'un LAP en une FM

⇒ Nous avons enrichi cette affectation grâce à la hiérarchie.



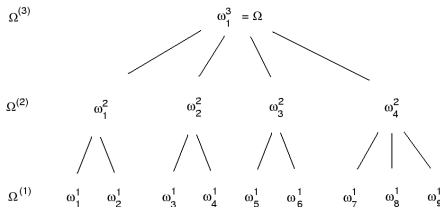
## Un exemple de matrice de confusion

| Vérité                                          | $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_4$ | $\omega_5$ | $\omega_6$ | $\omega_7$ | $\omega_8$ | $\omega_9$ |
|-------------------------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Décisions                                       |            |            |            |            |            |            |            |            |            |
| Niveau 1                                        |            |            |            |            |            |            |            |            |            |
| $\omega_1^1 = \{\omega_1\}$                     | 34         | 1          | 1          | 0          | 0          | 0          | 0          | 0          | 0          |
| $\omega_2^1 = \{\omega_2\}$                     | 2          | 20         | 0          | 0          | 0          | 0          | 1          | 0          | 0          |
| $\omega_3^1 = \{\omega_3\}$                     | 0          | 1          | 24         | 2          | 0          | 0          | 0          | 0          | 0          |
| $\omega_4^1 = \{\omega_4\}$                     | 1          | 0          | 3          | 24         | 0          | 0          | 0          | 0          | 0          |
| $\omega_5^1 = \{\omega_5\}$                     | 0          | 0          | 0          | 1          | 35         | 0          | 0          | 1          | 0          |
| $\omega_6^1 = \{\omega_6\}$                     | 0          | 0          | 0          | 1          | 3          | 15         | 0          | 0          | 1          |
| $\omega_7^1 = \{\omega_7\}$                     | 0          | 2          | 3          | 0          | 4          | 1          | 20         | 10         | 1          |
| $\omega_8^1 = \{\omega_8\}$                     | 0          | 1          | 0          | 0          | 3          | 2          | 5          | 35         | 1          |
| $\omega_9^1 = \{\omega_9\}$                     | 0          | 1          | 0          | 3          | 2          | 0          | 2          | 3          | 31         |
| Niveau 2                                        |            |            |            |            |            |            |            |            |            |
| $\omega_1^2 = \{\omega_1, \omega_2\}$           | 10         | 12         | 0          | 0          | 1          | 0          | 1          | 0          | 0          |
| $\omega_2^2 = \{\omega_3, \omega_4\}$           | 0          | 1          | 20         | 9          | 0          | 0          | 0          | 2          | 0          |
| $\omega_3^2 = \{\omega_5, \omega_6\}$           | 0          | 0          | 0          | 0          | 12         | 24         | 0          | 0          | 0          |
| $\omega_4^2 = \{\omega_7, \omega_8, \omega_9\}$ | 0          | 2          | 4          | 6          | 0          | 3          | 24         | 12         | 15         |
| Niveau 3                                        |            |            |            |            |            |            |            |            |            |
| $\omega_1^3 = \Omega$                           | 1          | 2          | 0          | 4          | 1          | 2          | 1          | 0          | 2          |

## 2. Construction des fonctions de masse

## Affectation hiérarchique

- Remarque :  
Plusieurs degrés d'erreur.
- Une sortie de niveau  $p$  est **correcte à un niveau  $q \geq p$** .



## Exemple

- Si Vérité =  $\omega_1^1$  :
  - $\omega_2^1$  est une sortie de niveau 1 correcte au niveau 2.
  - $\omega_3^1$  est une sortie de niveau 1 correcte au niveau 3.
  - $\omega_1^1$  est une sortie de niveau 1 correcte.

# Affectation hiérarchique

## Notons

- $P$  : nombre de niveau de la hiérarchie.
- $n^{p,q}$  : le nombre de sortie de niveau  $p$  correctes au niveau  $q$  où  $q \in [p, P]$ .
- $n^p = \sum_{q=p}^P n^{p,q}$  : nombre total d'objets classés au niveau  $p$ .
- $u$  : associe à une sortie son parent.
- **Affectation hiérarchique** :
  - Si la sortie  $s$  du LAP est de niveau  $p$  :

$$m(u^{q-p}(s)) = \frac{n^{p,q}}{n^p}, \quad \forall q \in [p, P].$$

## 2. Construction des fonctions de masse

## Affectation hiérarchique

## Exemple

- Si  $s = \omega_3^1$ .
- $m(\omega_3^1) = m(\{\omega_3\})$   
 $= n^{1,1}/n^1 = 245/314$   
 $= 0.78,$
- $m(u(\omega_3^1))$   
 $= m(\{\omega_3, \omega_4\})$   
 $= n^{1,2}/n^1 = 39/314$   
 $= 0.12,$
- $m(u^2(\omega_3^1)) = m(\Omega)$   
 $= n^{1,3}/n^1 = 30/314$   
 $= 0.10.$

| Vérité                                          | $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_4$ | $\omega_5$ | $\omega_6$ | $\omega_7$ | $\omega_8$ | $\omega_9$ |
|-------------------------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Décisions                                       |            |            |            |            |            |            |            |            |            |
| Niveau 1                                        |            |            |            |            |            |            |            |            |            |
| $\omega_1^1 = \{\omega_1\}$                     | 34         | 1          | 1          | 0          | 0          | 0          | 0          | 0          | 0          |
| $\omega_2^1 = \{\omega_2\}$                     | 2          | 23         | 0          | 0          | 0          | 0          | 1          | 0          | 0          |
| $\omega_3^1 = \{\omega_3\}$                     | 0          | 1          | 24         | 2          | 0          | 0          | 0          | 0          | 0          |
| $\omega_4^1 = \{\omega_4\}$                     | 1          | 0          | 3          | 28         | 0          | 0          | 0          | 0          | 0          |
| $\omega_5^1 = \{\omega_5\}$                     | 0          | 0          | 0          | 1          | 35         | 0          | 0          | 1          | 0          |
| $\omega_6^1 = \{\omega_6\}$                     | 0          | 0          | 0          | 1          | 3          | 15         | 0          | 0          | 1          |
| $\omega_7^1 = \{\omega_7\}$                     | 0          | 2          | 3          | 0          | 4          | 1          | 20         | 10         | 1          |
| $\omega_8^1 = \{\omega_8\}$                     | 0          | 1          | 0          | 0          | 3          | 2          | 5          | 35         | 1          |
| $\omega_9^1 = \{\omega_9\}$                     | 0          | 1          | 0          | 3          | 2          | 0          | 8          | 3          | 31         |
| Niveau 2                                        |            |            |            |            |            |            |            |            |            |
| $\omega_1^2 = \{\omega_1, \omega_2\}$           | 10         | 12         | 0          | 0          | 1          | 0          | 1          | 0          | 0          |
| $\omega_2^2 = \{\omega_3, \omega_4\}$           | 0          | 1          | 20         | 9          | 0          | 0          | 0          | 2          | 0          |
| $\omega_3^2 = \{\omega_5, \omega_6\}$           | 0          | 0          | 0          | 0          | 12         | 24         | 0          | 0          | 0          |
| $\omega_4^2 = \{\omega_7, \omega_8, \omega_9\}$ | 0          | 2          | 4          | 6          | 0          | 3          | 24         | 12         | 15         |
| Niveau 3                                        |            |            |            |            |            |            |            |            |            |
| $\omega_1^3 = \Omega$                           | 1          | 2          | 0          | 4          | 1          | 2          | 1          | 0          | 2          |

## 3. Combinaison

## Application de la règle de Dempster

Exemple : trois  $LAP_i$  ayant fourni les sorties  $s_i$  suivantes :

| $s_1 = \{A_2 R_1 17\}$                                                                                                 | $s_2 = \{A_1\}$                                                     | $s_3 = A_2$                                                                                  |
|------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| $\begin{cases} m_1(\{A_2 R_1 17\}) = 0.88 \\ m_1(A_2 R_1) = 0.07 \\ m_1(A_2) = 0.03 \\ m_1(\Omega) = 0.02 \end{cases}$ | $\begin{cases} m_2(\{A_1\}) = 0.9 \\ m_2(\Omega) = 0.1 \end{cases}$ | $\begin{cases} m_3(A_2) = 0.95 \\ m_3(\{A_2 inv\}) = 0.04 \\ m_3(\Omega) = 0.01 \end{cases}$ |

Le calcul de  $m = m_1 \odot m_2 \odot m_3$  donne :

|                     |            |                |            |
|---------------------|------------|----------------|------------|
| $m(\{A_2 R_1 17\})$ | $= 0.0845$ | $m(\{A_1\})$   | $= 0.0002$ |
| $m(A_2 R_1)$        | $= 0.0067$ | $m(\Omega)$    | $= 0.0000$ |
| $m(\{A_2 inv\})$    | $= 0.0002$ | $m(\emptyset)$ | $= 0.9036$ |
| $m(A_2)$            | $= 0.0048$ |                |            |

La masse importante portée par le vide reflète le conflit important entre les décisions fournies.

## 3. Combinaison

## Application de la règle de Dempster

Exemple : trois  $LAP_i$  ayant fourni les sorties  $s_i$  suivantes :

| $s_1 = \{A_2 R_1 17\}$                                                                                                 | $s_2 = \{A_1\}$                                                     | $s_3 = A_2$                                                                                  |
|------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| $\begin{cases} m_1(\{A_2 R_1 17\}) = 0.88 \\ m_1(A_2 R_1) = 0.07 \\ m_1(A_2) = 0.03 \\ m_1(\Omega) = 0.02 \end{cases}$ | $\begin{cases} m_2(\{A_1\}) = 0.9 \\ m_2(\Omega) = 0.1 \end{cases}$ | $\begin{cases} m_3(A_2) = 0.95 \\ m_3(\{A_2 inv\}) = 0.04 \\ m_3(\Omega) = 0.01 \end{cases}$ |

Le calcul de  $m = m_1 \odot m_2 \odot m_3$  donne :

|                     |     |        |                |     |        |
|---------------------|-----|--------|----------------|-----|--------|
| $m(\{A_2 R_1 17\})$ | $=$ | 0.0845 | $m(\{A_1\})$   | $=$ | 0.0002 |
| $m(A_2 R_1)$        | $=$ | 0.0067 | $m(\Omega)$    | $=$ | 0.0000 |
| $m(\{A_2 inv\})$    | $=$ | 0.0002 | $m(\emptyset)$ | $=$ | 0.9036 |
| $m(A_2)$            | $=$ | 0.0048 |                |     |        |

La masse importante portée par le vide reflète le conflit important entre les décisions fournies.

## 4. Prise de décision

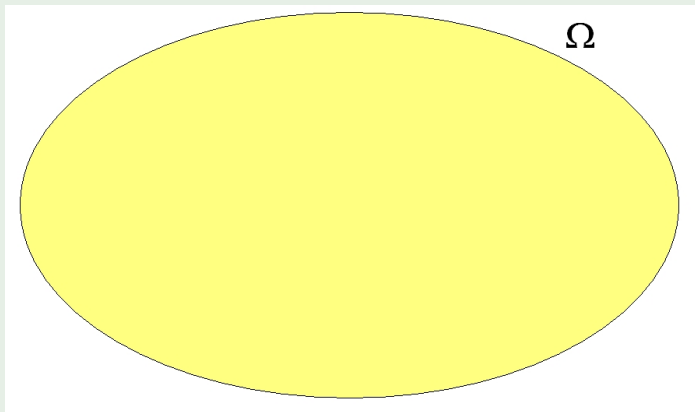
## Choix du cadre de pari

- L'ensemble  $\mathcal{D}$  des décisions pouvant être prises par la combinaison est identique à l'ensemble des sorties possibles des LAP.
- La cardinalité des sous-ensembles de  $\Omega$  n'est pas disponible.
- Ainsi on choisit comme **Cadre de Pari** :  
« l'ensemble le plus grossier contenant les sorties des LAP avec les sorties parentes et les adresses invalides qui leur sont associées ».

## 4. Prise de décision

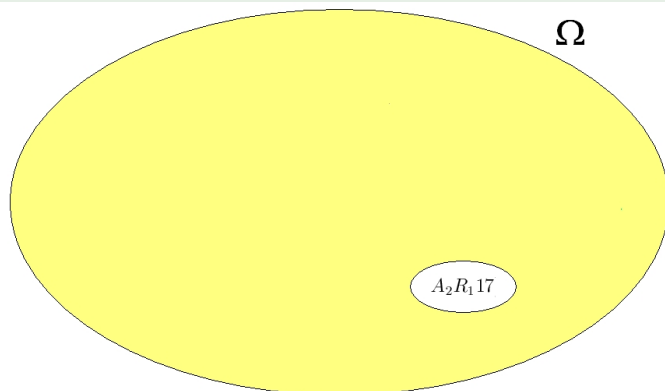
## Construction du cadre de pari

## Exemple



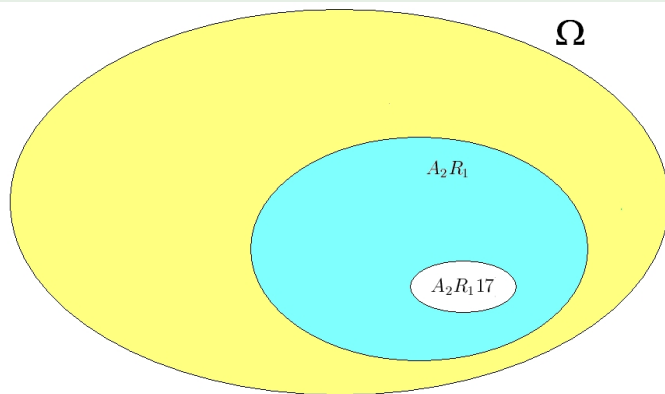
## 4. Prise de décision

## Construction du cadre de pari

Exemple (sortie 1 :  $s_1 = \{A_2R_117\}$ )

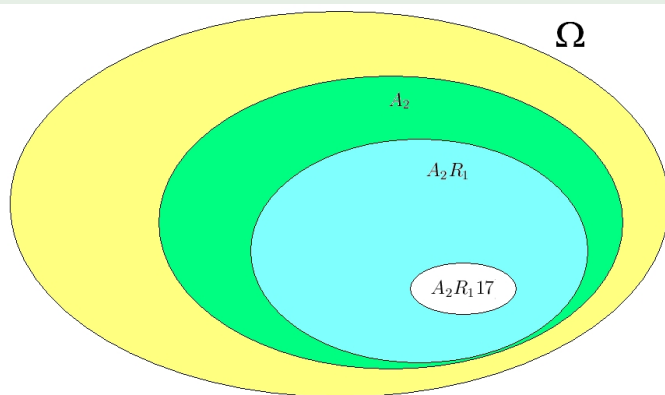
## 4. Prise de décision

## Construction du cadre de pari

Exemple (sortie 1 :  $s_1 = \{A_2R_117\}$ )

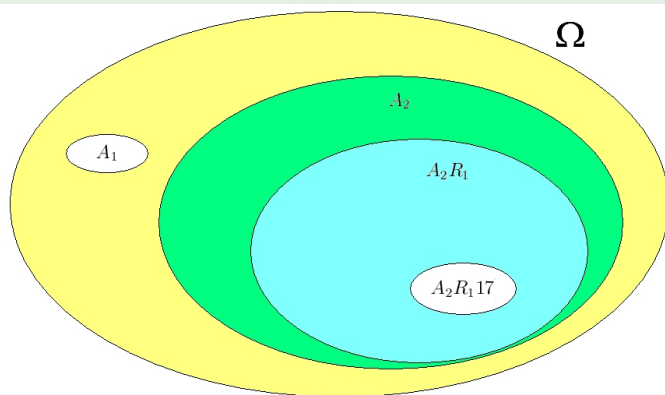
## 4. Prise de décision

## Construction du cadre de pari

Exemple (sortie 1 :  $s_1 = \{A_2R_117\}$ )

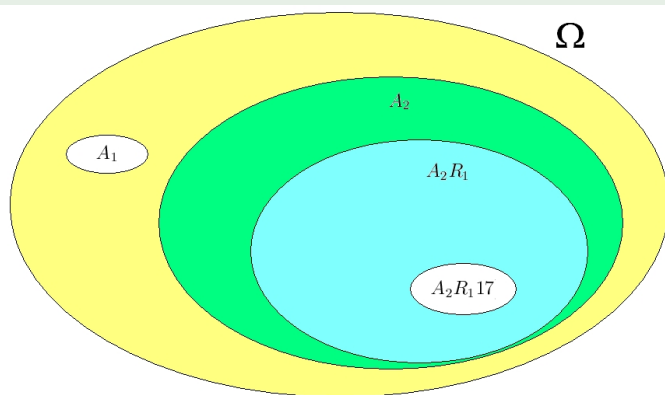
## 4. Prise de décision

## Construction du cadre de pari

Exemple (sortie 2 :  $s_2 = \{A_1\}$ )

## 4. Prise de décision

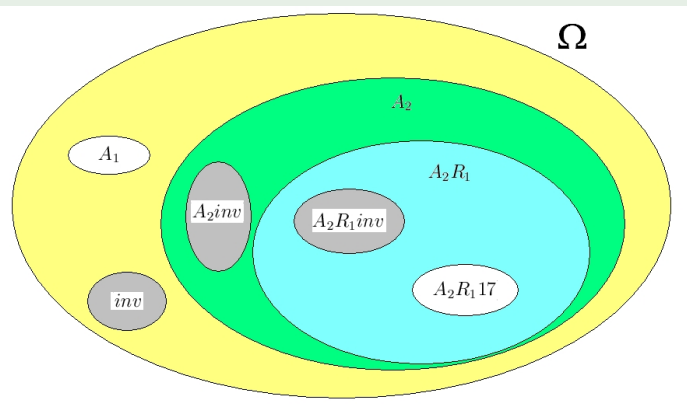
## Construction du cadre de pari

Exemple (sortie 3 :  $s_3 = A_2$ )

## 4. Prise de décision

## Construction du cadre de pari

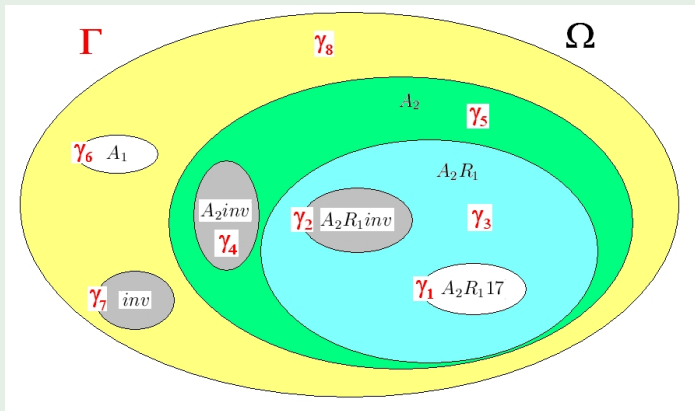
## Exemple (prise en compte des invalides)



#### 4. Prise de décision

# Construction du cadre de pari

### Exemple (cadre de pari finalement retenu)



## 4. Prise de décision

## Choix des coûts

Les différents coûts de décision peuvent être regroupés en :

- coût de bonne réponse, choisi égal à 0 ;
- **coûts de rejet** de niveaux distribution, rue et acheminement (resp. notés  $C_{R_{dis}}$ ,  $C_{R_{rue}}$ ,  $C_{R_{ach}}$ ) ;
- **coûts d'erreur** de niveaux distribution, rue et acheminement (resp. notés  $C_{E_{dis}}$ ,  $C_{E_{rue}}$ ,  $C_{E_{ach}}$ ).

Il existe un ordre suivant le sens commun entre ces coûts :

$$0 \leq C_{R_{dis}} \leq C_{R_{rue}} \leq C_{R_{ach}} \leq C_{E_{dis}} \leq C_{E_{rue}} \leq C_{E_{ach}}.$$

Ces coûts peuvent être déterminés à partir d'un ensemble d'apprentissage de manière à obtenir un comportement attendu de la combinaison.

En pratique, on fixe les coûts d'erreur, et on fait varier les coûts de rejet pour obtenir différents points de fonctionnement de la combinaison.

## 4. Prise de décision

## Choix des coûts

Les différents coûts de décision peuvent être regroupés en :

- coût de bonne réponse, choisi égal à 0 ;
- **coûts de rejet** de niveaux distribution, rue et acheminement (resp. notés  $C_{R_{dis}}$ ,  $C_{R_{rue}}$ ,  $C_{R_{ach}}$ ) ;
- **coûts d'erreur** de niveaux distribution, rue et acheminement (resp. notés  $C_{E_{dis}}$ ,  $C_{E_{rue}}$ ,  $C_{E_{ach}}$ ).

Il existe un ordre suivant le sens commun entre ces coûts :

$$0 \leq C_{R_{dis}} \leq C_{R_{rue}} \leq C_{R_{ach}} \leq C_{E_{dis}} \leq C_{E_{rue}} \leq C_{E_{ach}}.$$

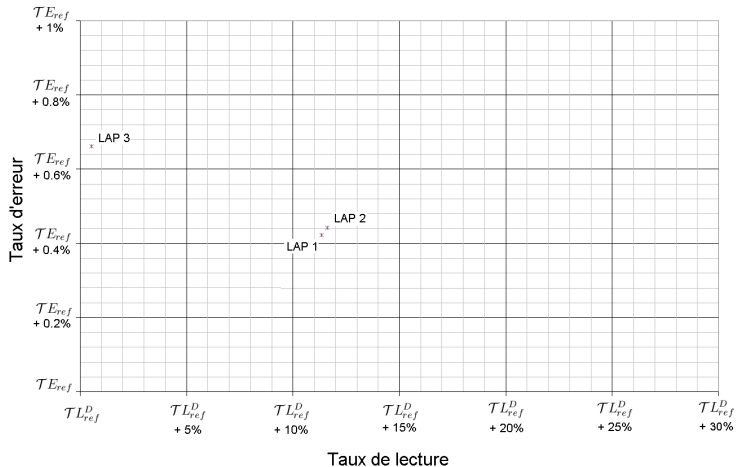
Ces coûts peuvent être déterminés à partir d'un ensemble d'apprentissage de manière à obtenir un comportement attendu de la combinaison.

**En pratique**, on fixe les coûts d'erreur, et on fait varier les coûts de rejet pour obtenir différents points de fonctionnement de la combinaison.

## Résultats expérimentaux

Performances au **niveau distribution**

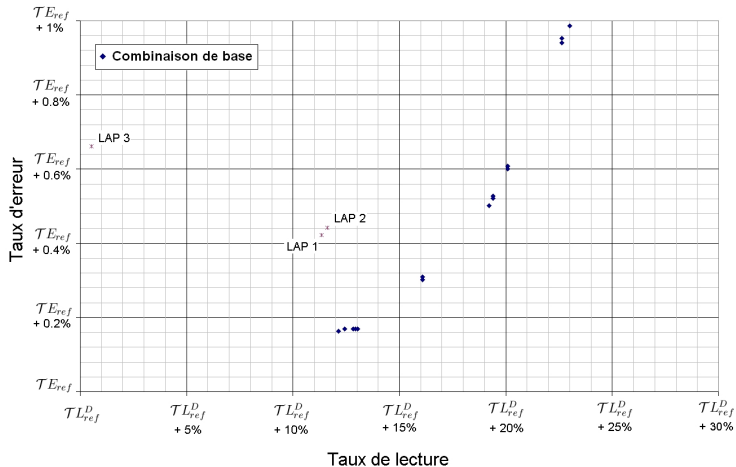
Sur l'ensemble d'apprentissage



## Résultats expérimentaux

Performances au **niveau distribution**

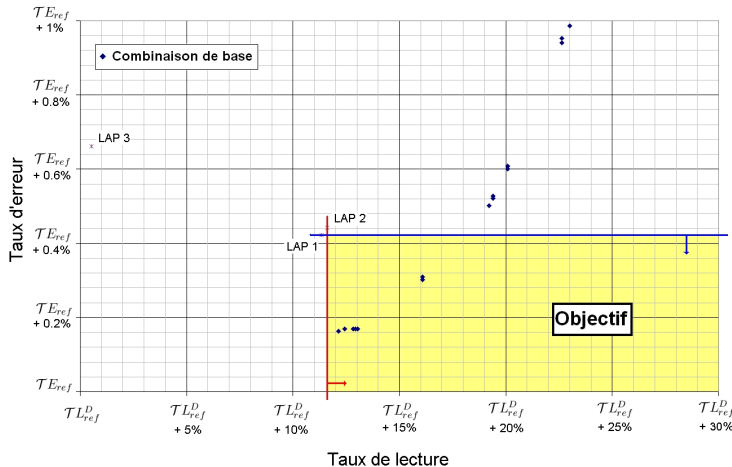
Sur l'ensemble d'apprentissage



## Résultats expérimentaux

Performances au **niveau distribution**

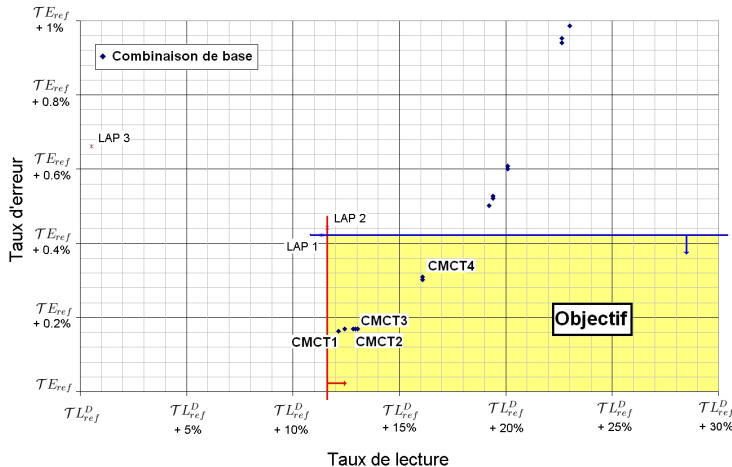
Sur l'ensemble d'apprentissage



## Résultats expérimentaux

Performances au **niveau distribution**

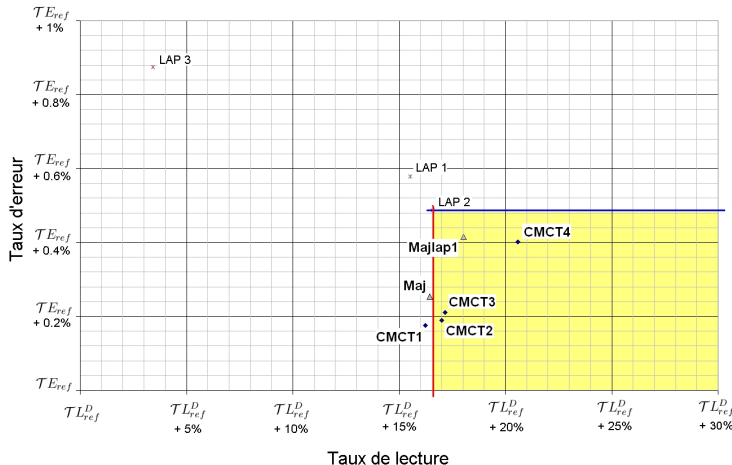
Sur l'ensemble d'apprentissage



## Résultats expérimentaux

## Performances au niveau distribution

Sur l'ensemble de test



# Plan

- 1 Introduction
- 2 Le Modèle des Croyances Transférables (MCT)
- 3 Mécanismes de correction d'une fonction de croyance (MC)
- 4 Modélisation du problème postal dans le cadre du MCT
- 5 Améliorations du modèle**
- 6 Conclusion et perspectives

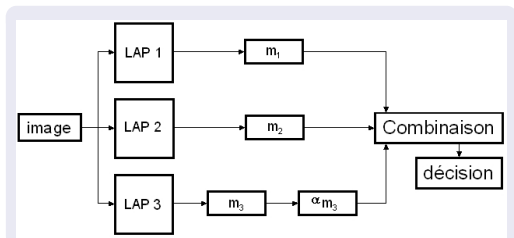
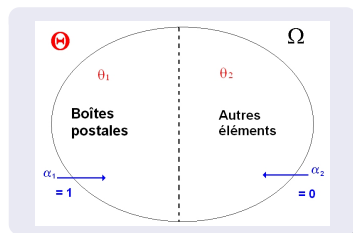
# Prise en compte de nouvelles informations

Nous avons travaillé sur la prise en compte des informations suivantes :

- 1 le LAP 3 ne reconnaît pas les boîtes postales ;
- 2 les LAP 1 et 2 fournissent en plus de leurs sorties un score de confiance ;
- 3 les LAP 1 et 2 ont une dépendance algorithmique.

# Prise en compte de l'information concernant le LAP 3

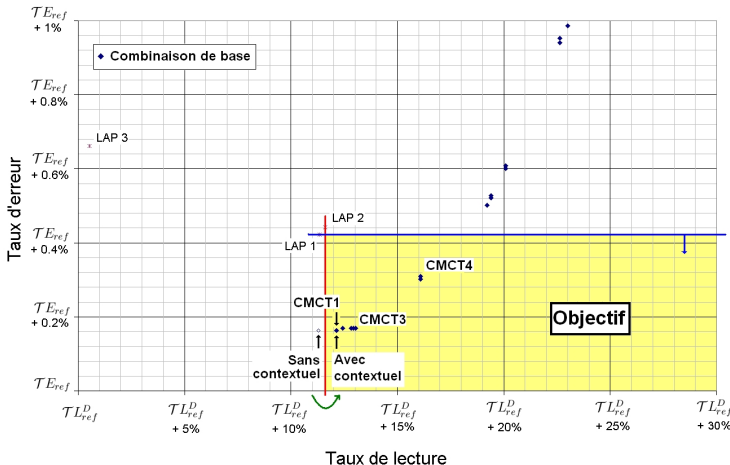
- Correction de l'information fournie par le LAP 3 par un **affaiblissement contextuel** basé sur le grossissement  $\ominus$  suivant, de vecteur d'affaiblissement  $\alpha = (\alpha_1 = 1, \alpha_2 = 0)$ .



Non reconnaissance des boîtes postales de la part du LAP 3

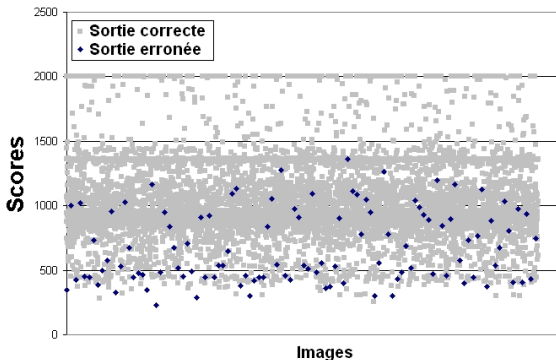
# Performances au niveau distribution

Sur l'ensemble d'apprentissage



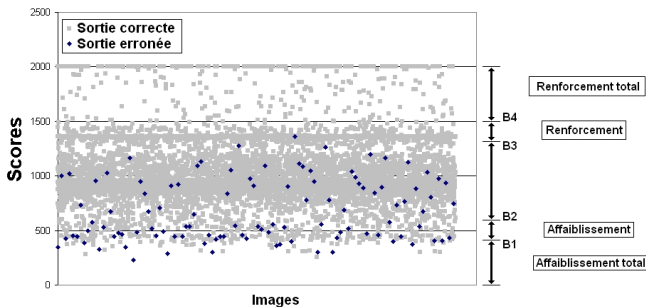
# Information fournie par ces scores

- Afin d'observer la réelle information fournie par le score de confiance associé à une sortie d'un LAP, les sorties correctes et erronées du *LAP1* sont représentées sur la figure suivante en fonction du score les accompagnant.



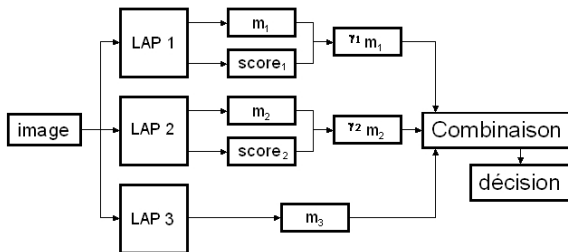
# Information fournie par ces scores

- Les observations précédentes sont valables aussi pour le *LAP2*.
- On voudrait alors **corriger** l'information fournie par chacun des LAP 1 et 2 en fonction de ce score



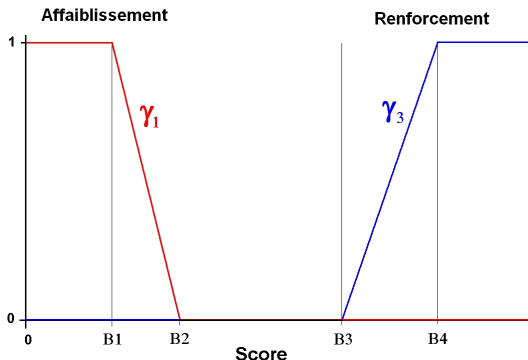
# Utilisation d'un mécanisme général de correction

- Correction de l'information fournie par chacun des LAP 1 et 2 par le **mécanisme général de correction**  $\gamma m = \gamma_1 m_\Omega + \gamma_2 m + \gamma_3 {}^tr m$  qui permet de renforcer ou d'affaiblir une FM.



# Calcul des paramètres de correction

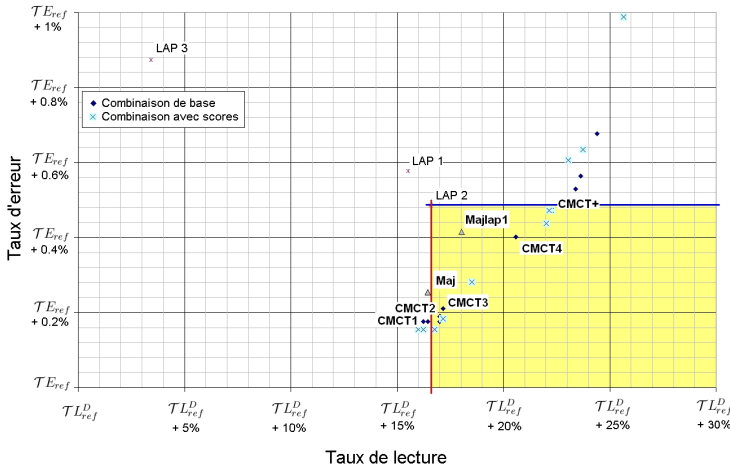
- Pour chaque image, les paramètres ( $\gamma_1, \gamma_2, \gamma_3$ ) du mécanisme  $\gamma m = \gamma_1 m_\Omega + \gamma_2 m + \gamma_3^{tr} m$  sont calculés en fonction du score associé à la sortie du LAP de la manière suivante ( $\gamma_1 + \gamma_2 + \gamma_3 = 1$ ).



Prise en compte des scores de confiance

# Performances au niveau distribution

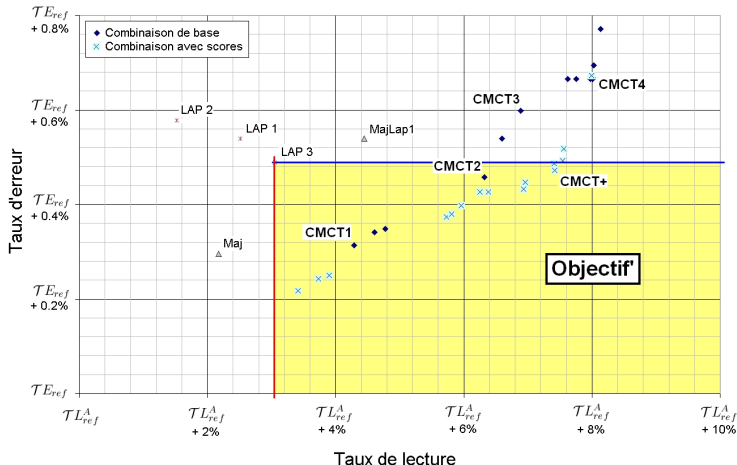
Sur l'ensemble de test



Prise en compte des scores de confiance

# Performances au niveau acheminement

Sur l'ensemble de test



# Plan

- 1 Introduction
- 2 Le Modèle des Croyances Transférables (MCT)
- 3 Mécanismes de correction d'une fonction de croyance (MC)
- 4 Modélisation du problème postal dans le cadre du MCT
- 5 Améliorations du modèle
- 6 Conclusion et perspectives

# Conclusion

- Une modélisation d'une fusion postale a été réalisée dans le cadre du modèle des croyances transférables (**flexibilité, modularité, performance**).
- À partir de **nouveaux mécanismes de correction** de fonction de croyance, nous avons montré comment améliorer encore les performances.
- Ce modèle a été implémenté et testé. Il **fonctionne actuellement** en France, en Belgique et à Jersey. Il est à l'étude pour la Norvège et le Portugal.
- L'ensemble de ces travaux fait l'objet d'une demande de brevet déposée à l'INPI le 17 mars 2006.

# Quelques perspectives

- Continuer l'étude de l'apport du score. Apprentissage automatique des paramètres de correction.
- Utiliser des informations de la base de données, notamment dans les grandes agglomérations.
- Étudier la prise en compte :
  - de plusieurs scores ;
  - de plusieurs sorties pour un LAP ;
  - de la dépendance des LAP.

Fin... *Merci de votre attention*

