



Décodage hybride dans les SRAP pour l'indexation automatique des documents multimédias

Mohamed BOUAZIZ

Sous l'encadrement de :
M. Antoine LAURENT

10 Juillet 2013

INTRODUCTION

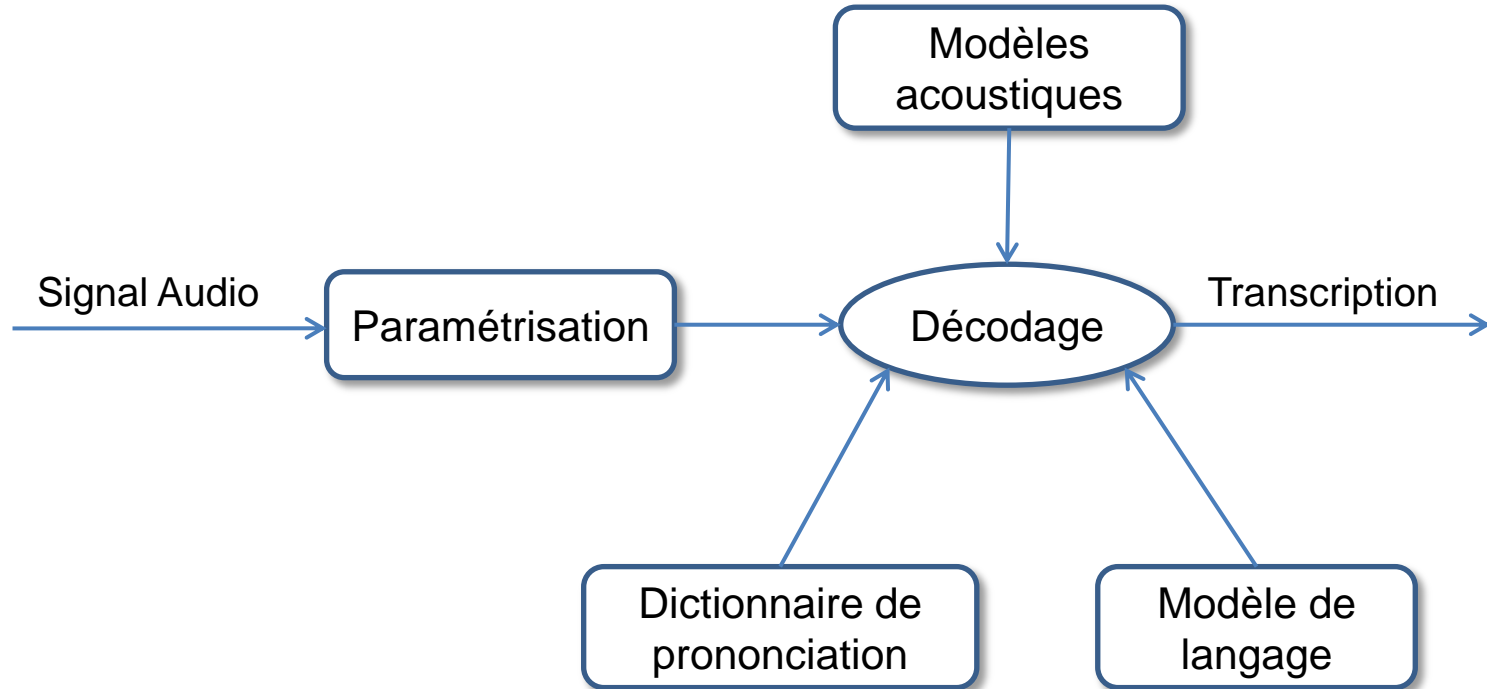
- Information audio : une interface d'échange
- Progrès considérables des Systèmes de Reconnaissance Automatique de la Parole (SRAP)
 - ✓ Taux d'erreur : 10%

Utilisations :

- Entrée vocale (dictée vocale, applications GPS, etc.)
- Source de données ou de connaissances

INTRODUCTION

SRAP à vocabulaire fixe :



INTRODUCTION

Défis :

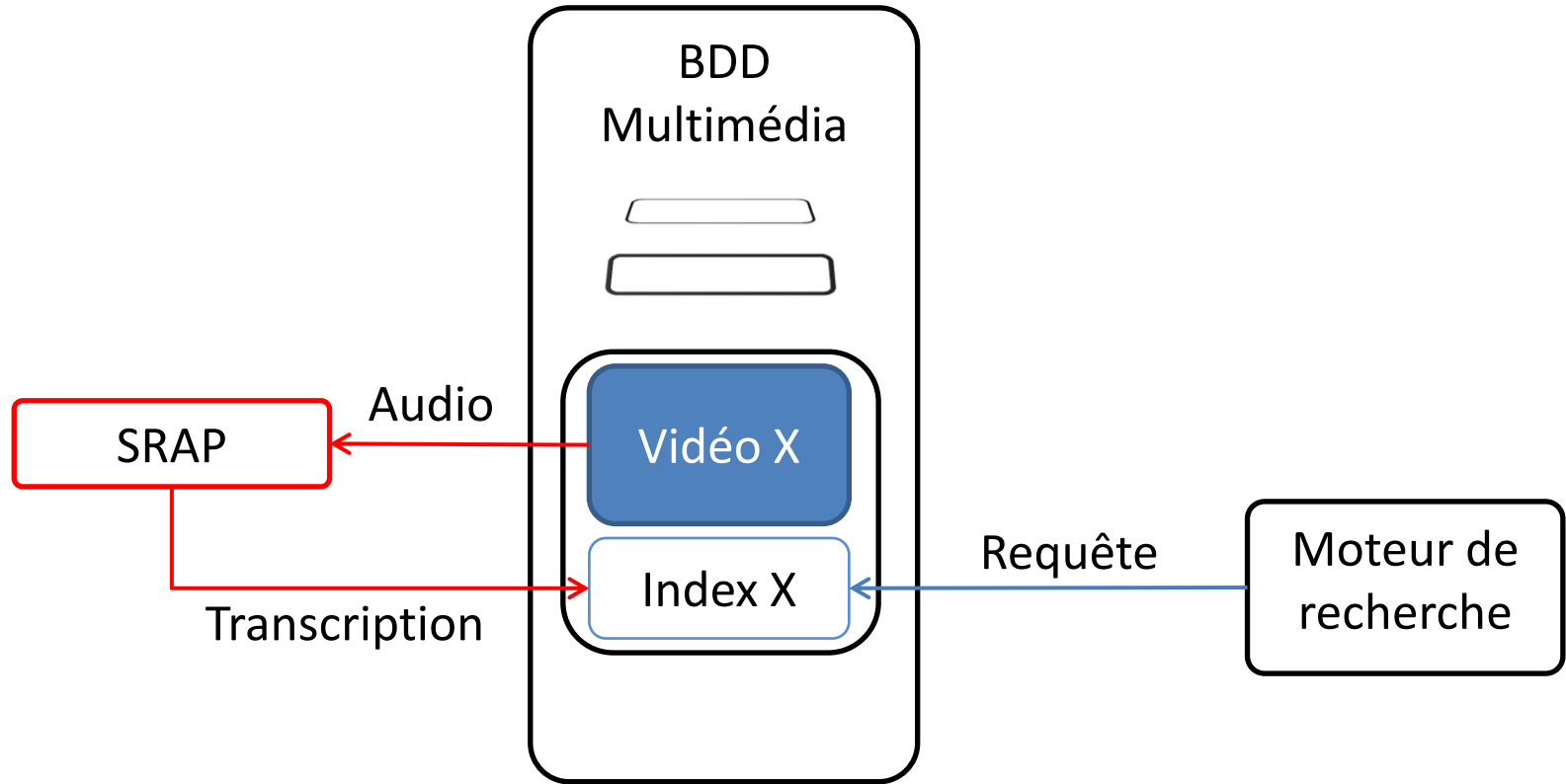
- Variabilité des conditions d'enregistrement
- Diversité des langues et des accents
- Mots « hors-vocabulaire »

INTRODUCTION

Impact des mots hors-vocabulaire (OOVs) :

- Détérioration de la performance du système
- Effet sur la reconnaissance des mots voisins [Dufour 2008]
- Influence sur les traitements potentiels :
 - Traduction de la Parole
 - Indexation Automatique

INTRODUCTION



INTRODUCTION

Objectifs :

- Adapter le SRAP du LIUM à l'indexation automatique des documents multimédias
- Remédier au problème des OOVs
 - ➔ Utilisation du « décodage hybride »

DÉCODAGE HYBRIDE

Unité sous-lexicale (SW) :

	<i>sympathiquement</i>
Morphèmes	<i>sympath + -ique + -ment</i>
Syllabes	<i>sym + pa + thi + que + ment</i>
Phonèmes	<i>ss un pp aa tt ii kk (ee) mm an</i>

- Etat de l'art

- Approche proposée

- Expérimentations

DÉCODAGE HYBRIDE

Langues :

Langues Flexionnelles	Langues Analytiques
Allemand Slovène Arabe Russe Anglais	Chinois

- **Français** : Pas de travaux

MOTS HORS-VOCABULAIRE

[Bisani et al. 2005]

- Décomposition en « Graphonèmes »
- Insertion des SW dans le Modèle de Langage (ML)
 - ➔ ML Hybride
- Langue : Anglais

Vocab	% OOVs	% WER
9k (5k W + 4k SW)	10	21,43 (30% ↘)

*W : mot

✓ Réduction de l'effet des OOVs sur le voisinage

MOTS HORS-VOCABULAIRE

[Shaik et al. 2011]

- SW : (**morphèmes ou syllabes**) + graphonèmes
- Langue : Allemand

Vocab	% OOVs	% Reco. OOVs	% WER
300k (5k W – 295k SW)	2,3	40	27.0 (5% ↘)

MOTS HORS-VOCABULAIRE

[Gelas et al. 2012]

- SW : **morphèmes** ou syllabes
- Langue : Amharique (peu dotée)

Vocab	% OOVs	% Reco. OOVs	% WER
65k	9	75	7,9 (49% ↘)

INDEXATION AUTOMATIQUE DES DOCUMENTS MULTIMÉDIAS (IADM)

SpeechBot [Thong 2002] :

- Moteur de recherche :
 - Données audio
 - SRAP
- Satisfaction des requêtes : 77,5%

Extension : Décodage Hybride [Logan et al. 2002] :

- SW : phonèmes + particules
- Précision + Rappel : ↗
- Fausses Alarmes : ↗

• Etat de l'art

• Approche proposée

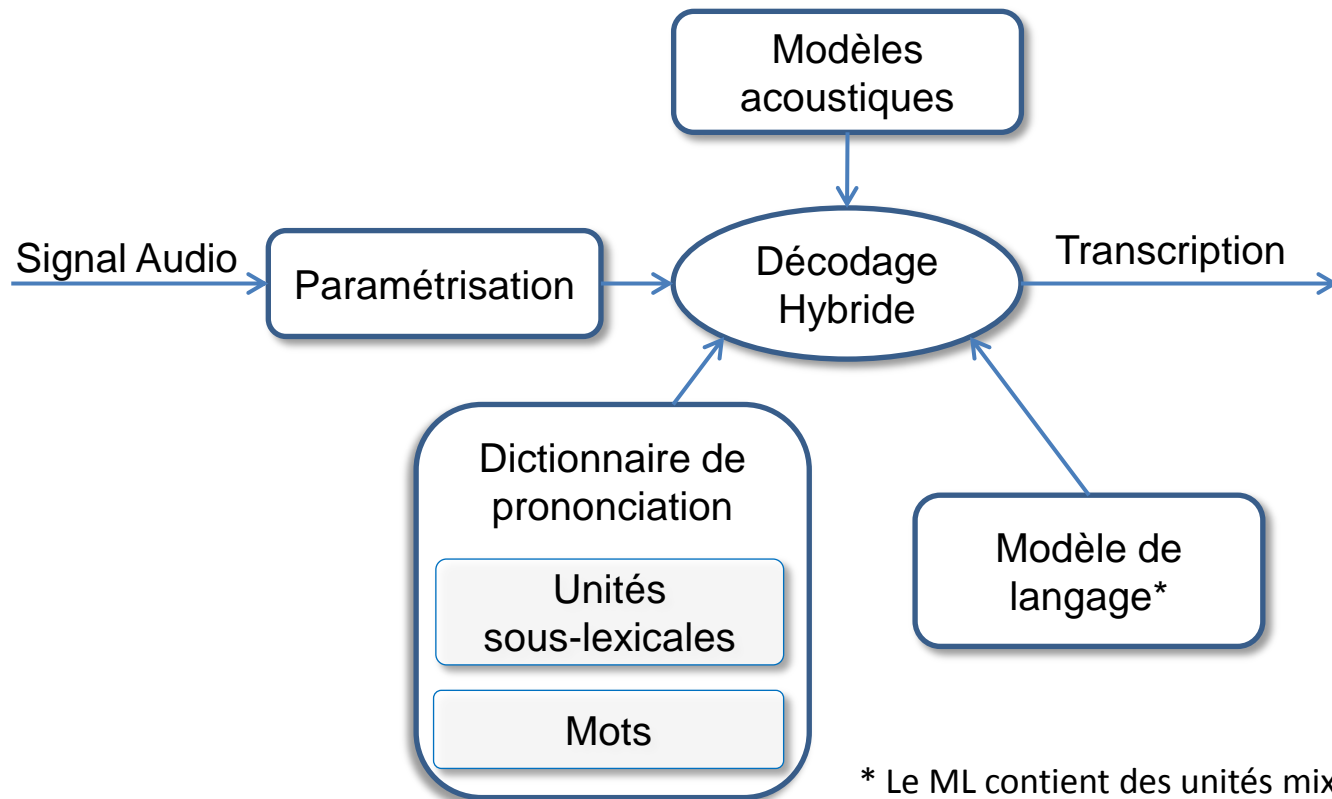
• Expérimentations

IADM + Vocabulaire Ouvert [Allauzen et al. 2005] :

- Extraction des nouveaux mots à partir des métadonnées
- Enrichissement dynamique du vocabulaire
- Pas d'adaptation du ML

% OOVs \searrow	% Reco. des EN introduites
30	84

SRAP À DÉCODAGE HYBRIDE



* Le ML contient des unités mixtes

- Etat de l'art
- Approche proposée
- Expérimentations

CHOIX DES OOVs

Sources des OOVs :

- Etat de l'art
 - **Approche proposée**
 - Expérimentations
- Taille du vocabulaire du système insuffisante
 - Morphologie riche
 - Domaine inadapté
 - Évolutivité du vocabulaire à traiter

l'IADM et les OOVs :

- Émergence continue de nouveaux mots
- Entités nommées : noms propres...
- Importance pour les moteurs de recherche

CHOIX DE L'UNITÉ SOUS-LEXICALE

Morphèmes :

sympath + -ique + -ment

- langues à morphologie riche
- Variabilité grammaticale
- Mot : lemme + terminaisons

Phonèmes :

ss un pp aa tt ii kk (ee) mm an

- Nombre fini pour chaque langue
- Reconstruction des graphèmes
- Modèle de langage : manque de contraintes

• Etat de l'art

• Approche proposée

• Expérimentations

CHOIX DE L'UNITÉ SOUS-LEXICALE

Syllabes :

- Susceptible de retrouver des mots
- Un nombre suffisant de contraintes

bouaziz

X bouazizi

✓ **bou + a + ziz**

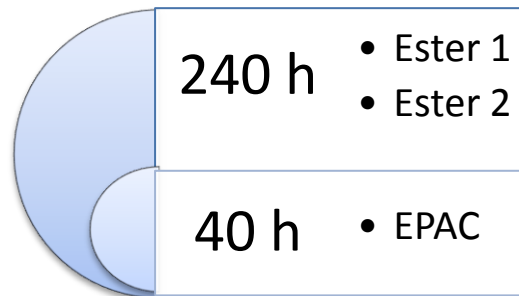
X au [bout à ziz]anie

SRAP DE BASE (LIUM'08)

Apprentissage acoustique :

- Paramétrisation PLP : 39 paramètres

- Données d'apprentissage :



- Modèles dépendant du :
 - Type du canal (Studio, Téléphone)
 - Genre du locuteur (Masculin, Féminin)

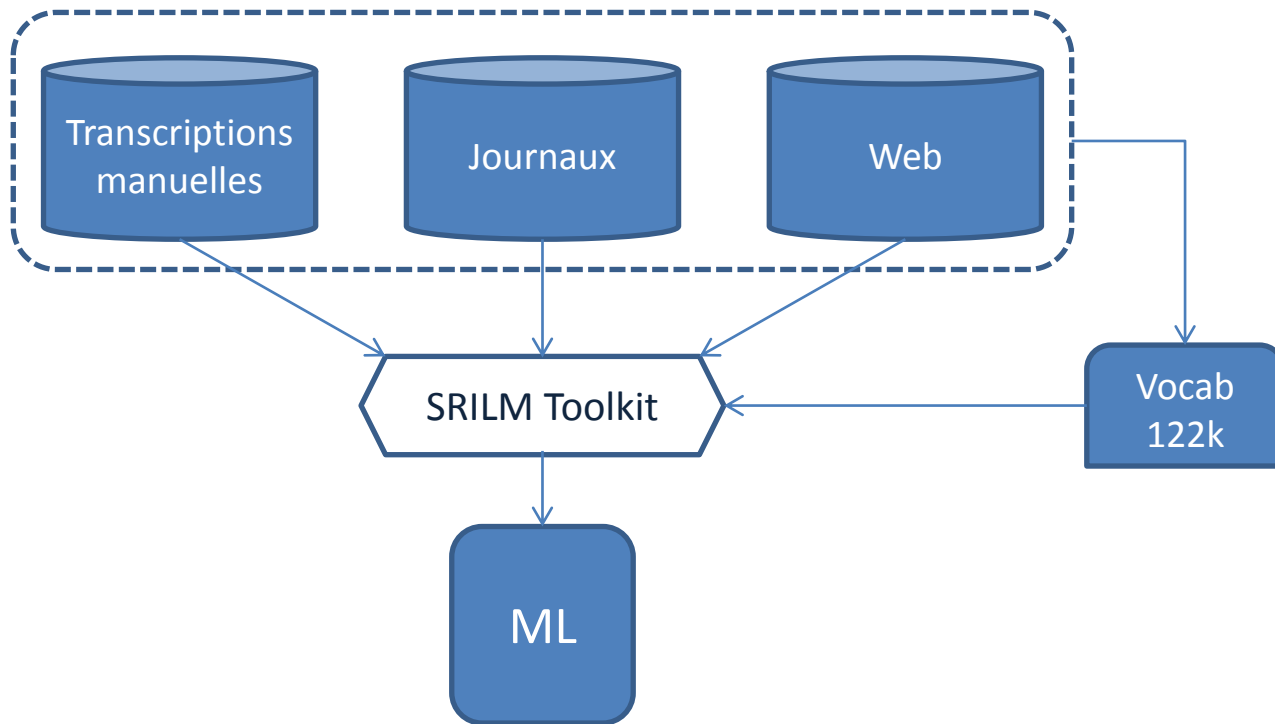
- Etat de l'art

- Approche proposée

- Expérimentations

SRAP DE BASE (LIUM'08)

Apprentissage linguistique :



- Etat de l'art
- Approche proposée
- Expérimentations

SRAP DE BASE (LIUM'08)

Dictionnaire phonétique :

- Phonétisation :
 - Dictionnaire : BDLEX
 - Outil : LIA_PHON

Phonèmes	Fillers
35	5

SRAP DE BASE (LIUM'08)

Segmentation en locuteurs :

- Critère d'Information Bayésien (BIC)
- Meilleur système (ESTER 2)

Décodage :

- Sphinx v3.7
- 5 passes
- Transformation : CMLLR

• Etat de l'art

• Approche proposée

• Expérimentations

MISE EN ŒUVRE

Données :

- Etat de l'art
 - Approche proposée
 - Expérimentations
- Projet REPERE : 60h
 - Annotation : noms de personnes

Chaîne	Emissions	Durée *
LCP	5	1h 45
BFM	2	1h 15

* 3h de corpus

MISE EN ŒUVRE

Données :

Phase	Volume
Apprentissage (R_train)	42 h
Développement (R_dev)	9 h
Test (R_test)	9 h

- Etat de l'art
- Approche proposée
- Expérimentations

MISE EN ŒUVRE

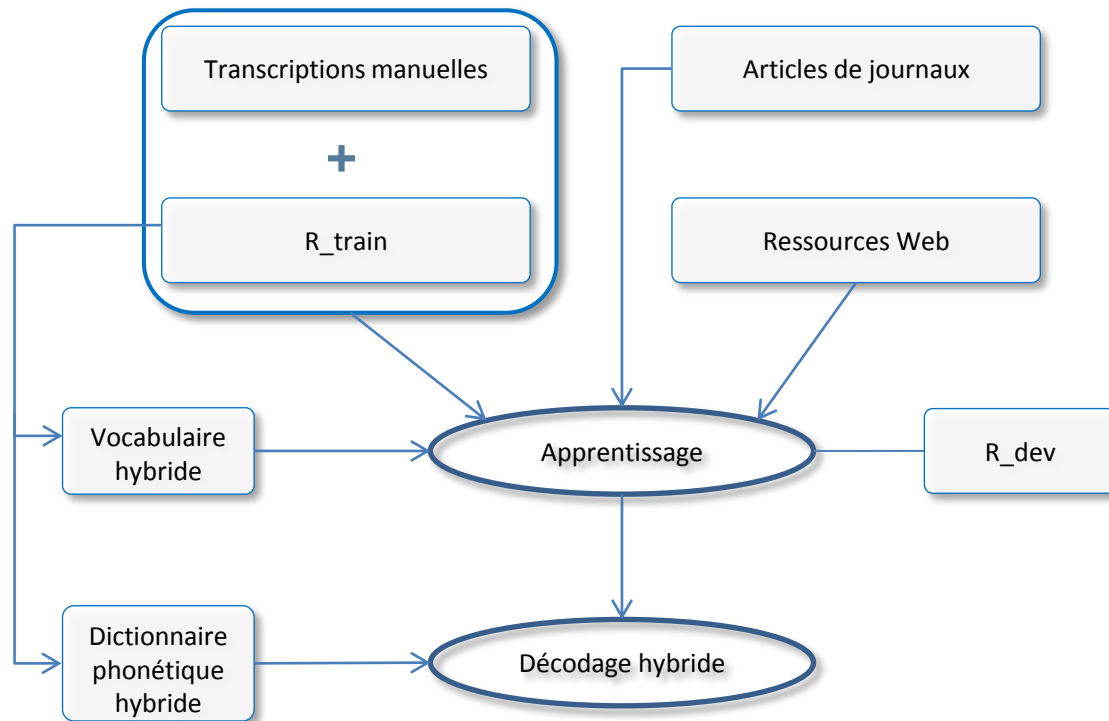
Apprentissage :

- Syllabation : LIA_PHON
 - Sans étiquette : de + la + ge
 - Avec étiquette : **de***SyllEtiqu* + **la***SyllEtiqu* + **ge***SyllEtiqu*

Noms propres	Occurrences	Syllabes
1225	7947	1074

MISE EN ŒUVRE

Apprentissage :



- Etat de l'art
- Approche proposée
- Expérimentations

MISE EN ŒUVRE

Apprentissage :

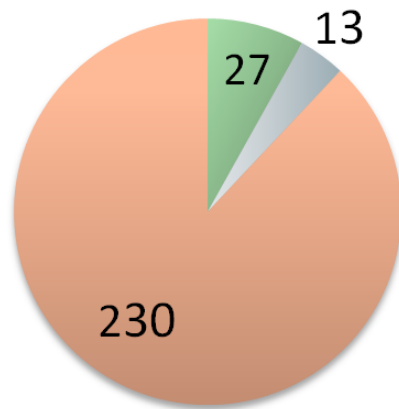
- Etat de l'art
- Approche proposée
- Expérimentations

	Transcriptions + R_train		Articles de journaux		Web	
	3g	4g	3g	4g	3g	4g
ML : syllabes non étiquetées (S1)	0,57	0,55	0,42	0,44	0,01	0,01
ML : syllabes étiquetées (S2)	0,51	0,48	0,48	0,49	0,01	0,01

ÉVALUATION

Métriques d'évaluation :

- Noms propres (R_test) :



- Originellement OOV
- R_train (découpés)
- Artificiellement OOV

ÉVALUATION

Métriques d'évaluation :

- Moteur de recherche :
 - Pas de détection des entités nommées
➔ Découpage de la requête
- Découpage des noms propres
 - Référence (stm)
 - Hypothèse (ctm)
- Sclite (NIST)

• Etat de l'art

• Approche proposée

• Expérimentations

ÉVALUATION

Résultats :

Système	% Reco. de noms propres HV	% Reco. de syll. dans les noms propres HV	SER
Système de base	0%	-	18,70%
S1 (non étiqueté)	31,39%	42,82%	19,90%
S2 (étiqueté)	31,10%	42,95%	24,90%

- Etat de l'art
- Approche proposée
- Expérimentations

CONCLUSION & PERSPECTIVES

- ✓ Premier essai pour la langue française
 - Reconnaissance des OOVs
 - Décodage hybride
- Prise en compte des autres catégories
- Exploration d'autres stratégies de décodage

Merci...