



**HAL**  
open science

# Corpus oral pour les mesures de confiance pour la Traduction Automatique de la Parole

Marwa Hadj Salah

► **To cite this version:**

Marwa Hadj Salah. Corpus oral pour les mesures de confiance pour la Traduction Automatique de la Parole. Sciences de l'Homme et Société. 2014. dumas-01023835

**HAL Id: dumas-01023835**

**<https://dumas.ccsd.cnrs.fr/dumas-01023835v1>**

Submitted on 15 Jul 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# **Corpus Oral pour les Mesures de Confiance pour la Traduction Automatique de la Parole**

**Nom:** Hadj Salah  
**Prénom :** Marwa

UFR LLASIC

---

Mémoire de Master 2 Recherche

Spécialité : Industries de la Langue

Parcours : TALEP

Sous la direction de M. Laurent Besacier et M. Benjamin Lecouteux

Année universitaire 2013-2014

## ***Remerciements***

*Je tiens à exprimer ma immense gratitude à mes deux encadrants M. Laurent Besacier et M. Benjamin Lecouteux pour m'avoir guidé et pour le soutien qu'ils m'ont apporté tout au long de la réalisation de ce mémoire de recherche. Ils ont su me faire profiter de leurs profondes connaissances et leurs nombreuses compétences dans le domaine du traitement automatique de la langue.*

*Qu'il me soit permis de leurs exprimer mon plus profond respect et mes vifs remerciements.*

*Je tiens également à remercier très sincèrement les membres de jury (M.George Antoniadis et M.Didier Schwab) qui m'ont fait l'honneur de bien vouloir accepter d'évaluer ce travail.*

*Je profite de cette occasion pour exprimer mon respect à tous mes enseignants, pour leurs nombreuses aides et leurs continuels encouragements durant ce Master.*

*A toute personne qui a participé de près ou de loin dans l'accomplissement de ce modeste travail.*

**MARWA**

# SOMMAIRE

Liste des tableaux .....	6
Liste des figures .....	6
Résumé .....	7
Abstract .....	7
Introduction générale .....	8
<b>Chapitre 1: Présentation du stage</b> .....	10
<b>Chapitre 2: Etat de l'art</b> .....	12
1. Introduction.....	13
2. Reconnaissance automatique de la parole.....	14
2.1. Principe de base.....	14
2.2. Extraction des paramètres.....	15
2.3. Modélisation acoustique.....	15
2.3.1. Algorithme d'expectation-maximisation (EM).....	16
2.3.2. Dictionnaire phonétisé .....	16
2.3.3. Alignement des phonèmes avec le signal... ..	16
2.3.4 Adaptation... ..	16
2.3.5 Evaluation des systèmes de RAP.....	17
2.4 Modèle de langage.....	17
3. Traduction automatique statistique.....	18
3.1 Principe de base.....	18
3.2 Modèle de traduction.....	18
3.2.1 Notion d'alignement.....	19
3.3 Décodeur.....	19
4. Traduction automatique de la parole.....	19
4.1 Spécificités de la traduction de la parole.....	20
4.1.1 Disfluences et registres de langue.....	20
4.1.2 Ponctuation, casse et segmentation.....	21
5. Corpus de post-éditions initialement disponible pour notre étude.....	21
6. Etat de l'art sur l'enregistrement du corpus.....	22

7. Revue de l'état de l'art sur les mesures de confiance .....	22
7.1 Mesures de confiance en TA.....	22
7.2 Mesures de confiance en RAP.....	23
8. Plan du travail pour le mémoire.....	24
<b>Chapitre 3: Enregistrement du Corpus oral.....</b>	<b>25</b>
1. Enregistrement : tâche et conditions d'enregistrement .....	26
1.1 Fiche signalétique.....	27
2. Conclusion.....	27
<b>Chapitre 4: Adaptation du système de reconnaissance de la parole.....</b>	<b>28</b>
1. Système de reconnaissance de la parole KALDI.....	29
1.1 Architecture .....	30
1.1.1 Modèle de Langage.....	30
1.1.2 Modèle Acoustique .....	30
2. Expérimentations .....	31
2.1 Corpus de test .....	31
2.2 Problèmes de normalisation liés à la Traduction de la parole.....	32
2.2.1 Cohérence dans la forme des données.....	32
2.3 Adaptation du modèle du langage et Ajout du vocabulaire .....	35
2.3.1 Corpus d'adaptation.....	35
2.3.2 Ajout du nouveau vocabulaire et phonétisation.....	35
2.3.3 Construction d'un nouveau Modèle de langage.....	35
2.3.4 Interpolation avec l'ancien modèle de langage .....	36
2.4 Récapitulatif des Résultats de RAP obtenus.....	36
3. Conclusion.....	36
<b>Chapitre 5 : Génération de nouvelles traductions et des étiquettes .....</b>	<b>37</b>
1 Système de Traduction .....	38
1.1 Architecture.....	38
1.1.1 Modèle de langage.....	39
1.1.2 Modèle de traduction.....	39
1.1.3 Décodage.....	39
2 Génération des étiquettes avec TERp-A .....	39
2.1 Translation Edit Rate (TER) .....	39
2.2 TERp.....	39
2.3 Le score BLEU .....	40

2.4 Résultats expérimentaux obtenus.....	40
3 Conclusion.....	42
<b>Conclusion et perspectives.....</b>	<b>43</b>
Conclusion .....	44
Perspectives.....	45
<b>ANNEXES.....</b>	<b>46</b>
Annexe 1: Script Perl réalisés.....	47
Annexe 2 : Extrait du corpus référence nettoyé .....	47
Annexe 3 : corpus de référence version 1 préparé pour KALDI.....	48
Annexe 4 : corpus de référence version 2 préparé pour KALDI.....	48
Annexe 5 : Exemple de phonétisation avec LIA-PHON. ....	49
Annexe 6 : comparaison entre les sorties de RAP sans et avec application du script de correspondance .....	49
Annexe 7 : Exemples de sorties de TERp-A.....	50
<b>Bibliographie .....</b>	<b>51</b>

## LISTE DES TABLEAUX

<b>Tableau 1 :</b> Locuteurs participants à la tâche de l'enregistrement .....	26
<b>Tableau 2 :</b> Fiche signalétique .....	27
<b>Tableau 3 :</b> Taille du nouveau lexique.....	35
<b>Tableau 4 :</b> L'ensemble de résultats de RAP obtenus .....	36
<b>Tableau 5:</b> Récapitulatif de tous les résultats obtenus .....	41

## LISTE DES FIGURES

<b>Figure 1 :</b> Schéma de la traduction automatique de la parole .....	20
<b>Figure 2 :</b> Plan de travail pour le mémoire.....	24
<b>Figure 3 :</b> Reconnaissance automatique de la parole.....	29
<b>Figure4 :</b> Vue simplifiée des divers composants de KALDI [17].....	30
<b>Figure 5 :</b> Extrait du corpus référence nettoyé .....	32
<b>Figure 6 :</b> La sortie de RAP avant et après application du script de correspondance .....	34
<b>Figure 7:</b> Système de traduction automatique .....	38
<b>Figure 8 :</b> Schéma décrivant les trois contributions réalisées .....	45

**Mots clés :** Traduction automatique de la parole, Mesures de confiance, Reconnaissance automatique de la parole, Traduction automatique statistique, Corpus enrichi.

## Résumé

La traduction automatique de la parole combine deux domaines connexes, à savoir la reconnaissance automatique de la parole et la traduction. En vue d'améliorer la sortie d'un système de traduction et réduire, ainsi, le taux d'erreur, il est évident qu'il faut faire appel à des mesures de confiance qui sont capables de donner une estimation sur l'exactitude de l'hypothèse. Toutefois, jusqu'à présent, il n'y a aucun corpus de systèmes susceptible de prédire les mesures de confiance pour la traduction automatique de la parole. Pour combler ce vide, nous avons d'une part enrichi un corpus existant à l'aide d'enregistrements oraux et d'autre part effectué des expérimentations avec l'ensemble de nos données. Dans le présent mémoire nous avons suivie une méthode pour classer les étiquettes générées telles que B : (Bon) ou M : (Mauvais). Nous avons obtenu de bons résultats pour les mots considérés comme bons (65,49%) et encourageants pour les mots jugés mauvais (34,51%).

**Keywords:** The automatic speech translation, confidence Measures, automatic speech recognition, statistical Machine translation, Enriched corpora.

## Abstract

The automatic speech translation combines two related domains, i.e. the automatic speech recognition and Machine translation. In order to improve the output of a translation system; and, hence, minimize the error rate, it's necessary to use confidence measures which can efficiently evaluate the accuracy of the hypothesis. However, there has been no corpus with such measures, so far. Therefore, to fill in this gap, we have first added oral recordings to an existing corpus and later, experimented with the whole of our new data. In this paper, we describe among other things strategy we adopted to classify the obtained labels such G: (Good) or B: (Bad). We obtained good results with good words (65,49%) and encouraging ones with words rated as bad (34.51%).

## Introduction générale

Ce mémoire de recherche s'inscrit au terme de la deuxième année de Master Sciences du langage, spécialité Industries de la langue (IDL) de l'Université Stendhal Grenoble 3. L'organisme accueillant est l'équipe GETALP (Groupe d'Étude en Traduction Automatique/Traitement Automatisé des Langues et de la Parole) du laboratoire d'informatique de Grenoble (LIG). C'est un groupe pluridisciplinaire de recherche comportant des informaticiens, des linguistes, des phonéticiens, des traducteurs et traiteurs de signaux oraux ... dont le but est : « *d'aborder tous les aspects théoriques, méthodologiques et pratiques de la communication et du traitement de l'information multilingue (écrite ou orale)* » [21].

Ce mémoire a été co-encadré par M. Laurent BESACIER, professeur à l'Université Joseph Fourier – Grenoble et M. Benjamin LECOUTEUX, Maître de conférences à l'Université Pierre-Mendès-France.

La traduction automatique de la parole constitue un champ de recherche riche et très important, laissant envisager diverses applications : la traduction d'un journal télévisé, la simplification de communications commerciales ou diplomatiques, l'usage dans un cadre touristique en vue de pouvoir s'informer rapidement dans un pays dont la langue est étrangère, ou encore dans un cadre militaire pour faciliter l'échange avec la population locale etc.

Evidemment, la traduction automatique de la parole (TAP) combine deux domaines larges et complémentaires que sont : la reconnaissance automatique de la parole et la traduction automatique. Autrement dit, il s'agit tout d'abord de reconnaître la parole prononcée par un locuteur dans une langue cible et de la transcrire automatiquement sous forme d'un texte, ensuite de traduire à l'aide d'un système de traduction statistique la sortie du système de reconnaissance de la parole dans une langue cible.

Par ailleurs, il est très intéressant d'évaluer la qualité de sortie d'un système de traduction de la parole, en faisant appel à des mesures de confiance qui cherchent à prédire le score de confiance que nous pouvons accorder aux hypothèses produites. Notre travail se concentre sur la traduction automatique de la parole (parole lue pour l'instant). **Plus précisément nous nous intéressons au problème original des mesures de confiance pour**

**la traduction de parole. Ce problème n'a pas encore été abordé dans la littérature car il n'existe pas à ce jour de corpus oral permettant une telle étude. Le but de ce mémoire de Master-R consiste à combler ce manque.**

Ainsi, ce manuscrit est organisé comme suit :

- Le premier chapitre propose un état de l'art et une description du contexte de notre travail de recherche, en présentant les deux domaines fondateurs de la traduction de la parole, et nous expliquons en détail la notion de mesures de confiance en TAP.
- Le deuxième chapitre décrit brièvement la tâche et les conditions d'enregistrement de notre corpus oral ainsi que le matériel utilisé.
- Le troisième chapitre présente tout d'abord, le système de reconnaissance de la parole que nous avons utilisé, ensuite nous décrivons de façon approfondie les expérimentations réalisées et les résultats obtenus.
- Le dernier chapitre s'attache à montrer le système de traduction automatique utilisé pour traduire les hypothèses de RAP, ainsi que l'outil pour la génération des étiquettes qui seront nécessaires pour construire, plus tard, des systèmes d'estimation de mesures de confiance pour la TAP.

# **Chapitre 1 :**

## **Présentation du stage**

## **Présentation de l'organisme d'accueil :**

Le Laboratoire d'Informatique de Grenoble (LIG) [20] est créé en 2007 d'envergure et qui est partenaire de : CNRS, Grenoble INP, INRIA Grenoble, l'Université Joseph Fourier, l'Université Pierre-Mendès-France, et de Université Stendhal. Le LIG regroupe environ 500 chercheurs, enseignants chercheurs, doctorants ainsi que personnels en support à la recherche. De plus, le laboratoire d'informatique de Grenoble est composé de 22 équipes de recherche, dont les travaux réalisés se basent principalement sur les axes thématiques de recherche suivantes : (Génie des logiciels et des systèmes d'information, méthodes formelles et méthodes de langages, systèmes interactifs et cognitifs, systèmes répartis et Réseaux, Traitement des données et de connaissances etc.)

## **Présentation de l'équipe GETALP du LIG :**

Le GETALP (Groupe d'Étude en Traduction Automatique/Traitement Automatisé des Langues et de la Parole) est une équipe du Laboratoire d'Informatique de Grenoble (LIG), abouti à l'union de groupe de chercheurs en traitement automatique de l'écrit et de la parole.

L'épistémologie du travail de GETALP se base principalement sur : « *des allers-retours continus entre collectes de données, investigations fondamentales, développement de systèmes opérationnels, applications et évaluations expérimentales* » [site Getalp]

Citant les points saisisants du GETALP comme l'approche écologique, la démarche agnostique, la multidisciplinarité, assister l'humain dans des situations de communication et d'interaction, le développement des outils et des ressources tels que les plateformes web de développement utilisant la collaboration de bases lexicales multilingues, la construction des corpus oraux et écrits nécessaires pour le traitement des langues peu dotées.

Par ailleurs, les axes de recherche de l'équipe GETALP sont fondés sur : la traduction et transcription automatiques de la parole, la traduction assistée par ordinateur, traitement des langues peu dotées, Génie logiciel du multilinguisme, collecte des ressources lexicales multilingues etc.

# **Chapitre 2 :**

# **Etat de l'art**

## 1. Introduction

Toutes les technologies modernes du traitement automatique du langage Naturel (TALN) commettent des erreurs ; ainsi pour réduire le taux d'erreur, il est nécessaire d'utiliser des méthodes en vue d'améliorer le fonctionnement et la sortie du système en faisant appel à des mesures de confiance qui fournissent une estimation sur l'exactitude de l'hypothèse.

En effet, les mesures de confiance sont utilisées dans diverses applications du TAL notamment dans la reconnaissance automatique de la parole (RAP) ou (ASR) en *anglais* «*automatic speech recognition*» et dans la traduction automatique (TA) ou (MT) en *anglais* «*Machine Translation*».

La reconnaissance automatique de la parole est une technique informatique qui permet à une machine d'interpréter une suite de mots prononcés par un locuteur humain et de transcrire la parole en texte. Toutefois, un tel système ne doit pas se tromper en faisant des confusions sur les mots utilisés et par conséquent produire un texte erroné. Cela va de soi, les mêmes mots doivent être exactement reconnus.

Il y a bien de publications concernant les mesures de confiance en reconnaissance automatique de la parole. Mais, ce qu'il faut signaler à cet égard c'est que tout le monde admet qu'il s'agit d'un problème assez délicat qui demeure jusqu'à présent sans solution parfaite.

Evidemment, en RAP, les mesures de confiance cherchent à prédire le score de confiance que nous pouvons associer au résultat (au niveau du phonème, au niveau du mot ou bien au niveau de la phrase) produit par le moteur de reconnaissance. Par conséquent, ce score va autoriser l'acceptation ou le rejet de l'hypothèse.

Par ailleurs, les systèmes de traduction automatique statistiques (TAS) ou (SMT en *anglais*) ont réalisé des progrès importants vers la production des traductions acceptables pour l'utilisateur. Toutefois, malgré l'amélioration continue des performances de ces systèmes informatiques, il n'y a toujours pas des résultats de traduction qui sont totalement satisfaisants. Alors pour bien traduire un texte  $t_1$  rédigé dans une langue source vers un texte  $t_2$  rédigé dans une langue cible, il est intéressant d'appliquer des mesures de confiance sur les hypothèses de traduction.

En effet en traduction automatique, les mesures de confiance sont utilisées généralement au niveau des mots ou encore au niveau des phrases pour indiquer à l'utilisateur (lecteur, post-éditeur, traducteur) les éventuelles erreurs produites par le système.

La traduction automatique de la parole (TAP) nécessite la réunion des deux domaines de la reconnaissance automatique de la parole et la traduction automatique statistique à la fois. Dans ce cas, il faut tenir compte des difficultés inhérentes à chacun de ces deux volets. Ainsi, afin d'obtenir des résultats satisfaisants, un système de traduction de la parole doit s'auto-évaluer (accepter, corriger ou supprimer une hypothèse de traduction), ce qui nécessite l'intégration des mesures de confiance qui tentent d'indiquer dans quelle proportion une décision prise par le système est fiable.

**Nous nous intéressons dans ce projet, aux mesures de confiance pour la traduction automatique de la parole. Ainsi pour accomplir cette tâche, nous allons tout d'abord enrichir un corpus initial par des enregistrements oraux, puis se concentrer sur l'analyse des erreurs et aux mesures de confiance pour la TAP.**

Dans la première partie de ce chapitre nous allons présenter les deux domaines fondateurs de la traduction automatique de la parole que sont : la reconnaissance automatique de la parole et la traduction automatique statistique, ensuite nous décrivons les spécificités inhérentes à la traduction de la parole. Après avoir défini le domaine de TAP, nous présentons le corpus de post-éditions initial que nous allons enrichir par des enregistrements oraux. Dans un deuxième temps, nous passons à une description bien détaillée concernant les mesures de confiance en RAP et en TA. Enfin nous présentons les deux parties majeures à aborder dans notre projet de recherche.

## **2. Reconnaissance automatique de la parole**

### **2.1 Principe de base**

La reconnaissance automatique de la parole repose dans la majorité des systèmes actuels sur une approche probabiliste [Jelinek 01]. Un système de RAP a pour but d'associer une suite de mots ou phonèmes à une suite de vecteurs acoustiques. Par conséquent, ce système recherche la séquence de mots la plus vraisemblable  $\hat{W} = w_1 w_2 \dots w_n$  parmi toutes les suites de mots  $W$  possibles sachant l'observation acoustique  $X = x_1 x_2 \dots x_m$ .

Mathématiquement, la séquence de mots  $\hat{W}$  devra alors maximiser l'équation suivante :

$$\hat{W} = \arg \max_W P(W|X)$$

Par application du théorème de Bayes, on obtient la formule suivante :

$$\hat{W} = \arg \max_W \frac{P(X|W)P(W)}{P(X)}$$

Comme, la suite de vecteurs acoustiques  $X$  n'influençant pas  $\arg \max W$ , l'équation peut se simplifier en :

$$\hat{W} = \arg \max_W P(X|W)P(W)$$

Deux types de modèles probabilistes sont utilisés dans la reconnaissance automatique de la parole en vue de la recherche de la séquence de mots la plus vraisemblable : le modèle acoustique qui fournit la valeur de  $P(X|W)$ , c'est la probabilité d'observer la suite de vecteurs acoustiques  $X$  lorsque la séquence de mots  $W$  est prononcée, et le modèle de langage qui fournit la valeur de  $P(W)$ , il associe une probabilité à toute séquence de mots et participe aux choix des candidats pour la suite de la phrase à reconnaître.

## 2.2 Extraction des paramètres

Evidemment, le signal de parole ne peut pas être directement utilisé. Il est indispensable de passer par l'étape d'extraction de ses paramètres afin d'identifier les informations pertinentes du signal. Il faut alors découper le signal de parole en trames, puis un vecteur de paramètres acoustiques est extrait pour chaque trame. Les paramètres acoustiques les plus utilisés de nos jours dans les systèmes de RAP sont les coefficients cepstraux MFCC « Mel-scale Frequency Cepstral Coefficients » [Davis 1980] ainsi que la méthode PLP « Perceptual Linear Prediction » [Hermansky 1990]. Cette extraction permet d'obtenir une suite de vecteurs acoustiques  $X = x_1 x_2 \dots x_m$ .

## 2.3 Modélisation acoustique

Les modèles acoustiques utilisés en reconnaissance automatique de la parole reposent principalement sur les modèles de Markov cachés MCC ou HMM en anglais « Hidden Markov Models » [Jelinek 1976, Rabiner 1989].

En effet, les MCC sont des automates probabilistes à états finis qui permettent de déterminer la probabilité d'émission d'une suite d'observations.

Pour un système de reconnaissance automatique de la parole, les émissions désignent les vecteurs de caractéristiques du signal audio composés en général des coefficients MFCC.

Les MMC respectent la propriété de Markov d'ordre 1 : l'état courant ne dépend que de l'état précédent et non pas des états antérieurs.

### **2.3.1 Algorithme d'expectation-maximisation (EM)**

L'apprentissage des modèles acoustiques conduit à estimer les paramètres des chaînes de Markov ainsi que des densités d'observations accordées aux états, cela se fait généralement à l'aide de l'algorithme d'Expectation-Maximisation (EM) [Dempster 1977] ayant pour objectif la recherche du maximum de vraisemblances dans les paramètres des modèles probabilistes *a posteriori* quand le modèle dépend de variables latentes non observables.

### **2.3.2 Dictionnaire phonétisé**

L'apprentissage des modèles acoustiques ne peut pas être réalisé sans l'utilisation d'un dictionnaire phonétisé. Étant donné qu'un système de RAP est basé sur les phonèmes, il est indispensable d'associer une suite de phonèmes propre pour chaque mot du dictionnaire.

En général, l'approche exploitée est la construction du dictionnaire de façon automatique n l'améliorant des fois par l'approche manuelle, comme par exemple via le système proposé par [Béchet 2001] basé sur des règles de phonétisation qui transforment de manière automatique les graphèmes en phonèmes.

Ainsi, ce dictionnaire doit regrouper tous les mots qui seront utilisés lors de l'apprentissage du modèle acoustique ainsi que leur phonétisation.

### **2.3.3 Alignement des phonèmes avec le signal**

Pour estimer les paramètres des modèles acoustiques, il est important de passer par une phase d'alignement entre les phonèmes et le signal. Cet alignement permettra l'association des vecteurs acoustiques aux états du MMC à l'aide de l'algorithme de Viterbi [Viterbi 1967] ou de l'algorithme forward-backward [Baum 1972].

Ainsi, afin d'obtenir des modèles acoustiques performants, la phonétisation de chaque transcription doit être celle la plus pertinente.

### **2.3.4 Adaptation**

L'adaptation des modèles acoustiques est importante en vue de gérer de la meilleure façon possible plusieurs contraintes comme : la variabilité des locuteurs, les différents conditions d'enregistrement du signal de parole, etc.

Par conséquent, cette adaptation rend le système de reconnaissance plus robuste et plus performant face à des locuteurs inconnus manquants dans les données d'apprentissages.

Plusieurs techniques d'adaptation des modèles acoustiques sont couramment utilisées : « Maximum Likelihood Linear Regression » (MLLR) [Leggetter 1995], « Constrained

Maximum Likelihood Linear Regression » (CMLLR) [Digalakis 1995] et «Maximum A Posteriori » (MAP) [Gauvain 1994].

### 2.3.5 Evaluation des systèmes de RAP

Les systèmes de reconnaissance automatique de la parole sont évalués de façon systématique au niveau de taux d'erreur par mot.

Le WER (Word Error Rate – Taux d'erreurs de mots) est la mesure classique la plus utilisée pour l'évaluation des systèmes de RAP, elle est obtenue en calculant la distance minimale entre la transcription du système et la référence.

En effet, le WER cherche l'alignement qui peut minimiser la distance et prend en compte les erreurs suivantes :

- **Substitutions** : mot mal reconnu dans la transcription du système de RAP par rapport à la référence.
- **Insertions** : mot supplémentaire dans la transcription du système de RAP par rapport à la référence.
- **Délétions** : mot non reconnu dans la transcription du système de RAP par rapport à la référence.

⇒  $Distance = (Substitutions + Insertions + Délétions) / \text{Nombre de mots de la référence}$ .

Ainsi le WER s'exprime par l'expression suivante :

$$WER = (Substitutions + Insertions + Délétions) / \text{Nombre de mots de la référence} * 100$$

## 2.4 Modèle de langage

Dans un système de reconnaissance automatique de la parole ou un système de traduction automatique, le modèle de langage exige des contraintes linguistiques pour guider le décodage acoustique en choisissant la séquence de mots la plus vraisemblable. Il permet alors d'accorder une probabilité à toute séquence de mots et de contribuer aux choix des candidats pour la suite de la phrase à reconnaître en RAP. [2]

Il existe plusieurs modèles de langage comme le modèle n-gramme, le modèle Trigger, le modèle cache etc. Néanmoins, le modèle n-gramme est couramment le plus utilisé dans les systèmes de RAP et TA.

Etant données : une séquence de mots  $W$  de longueur  $k$ ,  $h_i = w_1, \dots, w_{i-1}$  représentant l'historique d'un mot  $w_i$  et la probabilité du modèle de langage qui est représenté par  $P(W)$  en RAP.

Dans le modèle n-gramme, l'historique d'un mot est constitué des  $n-1$  mots antérieurs, où  $n$  est l'ordre du modèle qui prend en général la valeur 3 ou 4.

Ainsi, la probabilité  $P(W)$  est exprimée par :

$$P(W) \simeq P(w_1) \prod_{i=2}^k P(w_i | w_{i-1}, \dots, w_1)$$

### 3. Traduction automatique statistique

#### 3.1 Principe de base

La traduction automatique statistique repose sur l'apprentissage de modèles statistiques à partir des corpus parallèles qui désignent un ensemble de textes accompagnés de leurs traductions alignés au niveau des phrases.

De manière formelle, il s'agit de traduire une phrase source  $f = f_1 \dots f_j$  rédigée dans une langue source vers une phrase  $e = e_1 \dots e_l$  rédigée dans une langue cible.

En effet, la TAS utilise un modèle de traduction noté  $P(e|f)$ , qui permet de déterminer la probabilité qu'une phrase cible  $e$  soit la traduction d'une phrase source  $f$ , ainsi qu'un modèle de langage noté  $P(e)$  qui indique tout simplement la probabilité que  $e$  soit une phrase.

Comme en RAP, formellement, la traduction automatique statistique applique la formule de Bayes :

$$\hat{e} = \arg \max_e P(e|f) = \frac{P(f|e)P(e)}{P(f)}$$

Toutefois, la probabilité  $P(f)$  associée à la phrase source étant connu a priori, par suite l'équation peut se simplifier en :

$$\hat{e} = \arg \max_e P(f|e)P(e)$$

#### 3.2 Modèle de traduction

La tâche de traduction automatique statistique présuppose la présence d'un corpus parallèle aligné au niveau des phrases. Le modèle de traduction permet de retrouver les traductions les plus vraisemblables d'un groupe de mots, alors il peut être défini comme des tables de traduction associant pour chaque groupe de mots leurs correspondants dans la langue cible.

En effet, il y'a deux types de modèles de traduction : un modèle à base des mots, comme les cinq modèles statistiques IBM proposés par [Brown et al.] permettant de calculer la probabilité  $P(f|e)$  grâce à des paramètres estimés sur un corpus bilingue, le modèle HMM proposé par [Vogel et al. 96] en améliorant le modèle IBM 2 etc. De plus, il y'a le modèle à

base des séquences prenant comme unité de traduction une suite de mots en vue de simplifier le processus de traduction, il consiste alors à segmenter la phrase source en séquences qui seront traduites dans une langue cible, puis de réordonner ces traductions à l'aide d'un modèle de distorsion afin de produire la traduction finale.

### 3.2.1 Notion d'alignement

Pour estimer les probabilités de traduction, il est indispensable d'aligner les mots de la langue cible sur ceux de la langue source, c'est-à-dire d'indiquer pour chaque mot de la phrase source son correspondant existant dans la phrase cible.

Comme le but d'un modèle de traduction est de donner la traduction la plus vraisemblable en se basant sur un ou plusieurs corpus parallèles, le modèle de langage repose sur des données monolingues en langue cible et son rôle est le même qu'en reconnaissance automatique de la parole, c'est de confirmer que la traduction produite soit celle la plus probable.

Toutefois, un système de TAS est composé non seulement d'un modèle de traduction et d'un modèle de langage, mais aussi d'un décodeur permettant le calcul de l'argument maximum *argmax*.

### 3.3 Décodeur

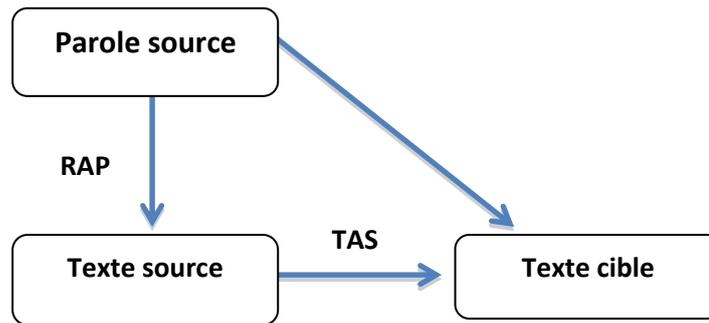
Le but du décodeur est de rechercher la meilleure hypothèse cible  $\hat{e}$  parmi toutes celles possibles en se basant sur les probabilités de deux modèles de traduction et de langage et en exploitant l'algorithme *Espérance-maximisation*.

## 4. Traduction automatique de la parole

La traduction automatique de la parole représente aujourd'hui un champ de recherche passionnant qui vise à transcrire puis traduire un flux de parole prononcé dans une langue source vers une langue cible.

En effet, la traduction automatique de la parole combine deux domaines connexes que sont la reconnaissance automatique de la parole et la traduction automatique statistique.

Parmi les projets et les prototypes réalisés au cours de ces dernières années en TAP, nous pouvons citer le projet *Verbmobil* [Wahlster 2000] dont le but était de développer un système mobile de traduction de la parole spontanée, ainsi que le projet européen TC-STAR qui avait pour objectif l'amélioration des performances de la traduction parole à parole.



**Figure 1** : Schéma de la traduction automatique de la parole

## 4.1 Spécificités de la traduction de la parole

### 4.1.1 Disfluences et registres de langue

A l'oral, la voix d'un journaliste qui présente une émission de télé ou la voix de « Simone Héroult » pour la SNCF la plus connue en France, présentent moins de disfluences qu'une interview d'un passant dans la rue ou encore qu'une conversation ou un dialogue spontané.

Le registre de langue ou aussi le niveau de langue, à l'oral comme à l'écrit peut être défini comme un mode d'expression adapté à une situation de communication. Il se caractérise par le choix du vocabulaire, respect ou non respect des règles grammaticales, utilisation d'abréviation etc.

En français, on distingue trois registres principaux : le registre soutenu, utilisé notamment à l'écrit dans les textes littéraires ou encore dans des communications officielles, reconnu par l'utilisation d'un vocabulaire plus riche ou des règles de grammaires respectées. De plus il y a le registre courant qui est employé à l'oral comme à l'écrit avec des personnes que l'on ne connaît pas intimement, il se caractérise par l'utilisation d'un vocabulaire simple et par le respect des règles de grammaire et syntaxe. Le troisième niveau est le registre familier, utilisé principalement à l'oral entre des amis et des proches, comme il se caractérise par l'emploi d'un vocabulaire relâché, l'utilisation des abréviations et des règles de grammaire parfois incorrectes.

En outre, la parole spontanée comporte plusieurs disfluences qui se produisent spontanément comme les répétitions, les hésitations, les pauses pleines, les faux départs, lapsus etc.

La plupart de ces phénomènes propres à l'oral sont transcrites par un système de reconnaissance de la parole.

Par conséquent, un système de traduction de la parole doit considérer toutes ces marques qui font de la transcription de la parole un type d'écrit très spécifique [Déchelotte 2007].

#### **4.1.2 Ponctuation, casse et segmentation**

La traduction automatique de la parole se diffère de la traduction automatique de l'écrit au niveau des problèmes et des difficultés rencontrées.

En effet, dans le cas de la parole continue, il n'y a pas de frontières entre les mots, par suite la transcription automatique va produire comme résultat une séquence de mots, sans segmentation claire en phrases, sans aucune ponctuation et aucune casse, ce qui rend la délimitation des phrases incorrecte.

En addition, un système de reconnaissance de la parole ne tient pas compte des informations prosodiques qui se trouvent dans le signal de parole, ni de la ponctuation et de la casse.

Bref, pour traduire de la parole d'une langue source vers une langue cible, il est nécessaire de récupérer la ponctuation qui manque et de rétablir la casse d'une manière correcte.

### **5. Corpus de post-éditions initialement disponible pour notre étude**

La traduction automatique statistique se base sur un corpus parallèle, alors pour produire des traductions valables, un système de TAS nécessite des données de haute qualité pour être efficacement entraînées, améliorées et évaluées. Néanmoins, le développement d'un ensemble de données de valeur est une tâche relativement coûteuse.

Dans le cadre de notre projet nous allons utiliser et enrichir le corpus de [3], c'est une large base des données de 10.881 hypothèses de traductions SMT (Français/Anglais, basé sur les syntagmes) post-éditées ou corrigées à la main, un annotateur bilingue vérifie la qualité de l'hypothèse de traduction, puis corrige les résultats.

Ces post-éditions ont été obtenues à l'aide d' « Amazon's Mechanical Turk », en essayant de respecter certaines règles éthiques.

Evidemment, la méthode de collection des données nécessite des indications claires et une vérification stricte pour parvenir à la qualité.

Chaque post-édition a été évaluée comme étant une traduction de la phrase source « FR » et classée suivant sa qualité (parfait, assez bon ou partiellement erronée)

Une analyse partielle des données collectées a montré la haute qualité des corrections, avec plus de 87% des post-éditions qui améliorent les hypothèses et plus de 94% des post-éditions effectuées par la plateforme en ligne « crowdsourcing », qui sont de qualité professionnelle.

*Potet et al.* ont également fait post-editer 1500 traductions de référence gold-standard (de corpus parallèles, bilingues, professionnels) et ils ont remarqué que 72% de ces traductions nécessitaient des rectifications au cours de la post-édition. Puis, ils ont calculé une mesure de proximité entre les différentes sortes de traductions et ils ont noté que les traductions de référence sont presque aussi dissimilaires des hypothèses de TAS que des hypothèses post-éditées.

En addition, l'analyse des résultats montre que même les correcteurs non qualifiés y compris ceux dont la langue maternelle n'est pas la langue cible, sont capables de produire des traductions de valeur, dans un contexte normal. [3]

## **6. Etat de l'art sur l'enregistrement de corpus**

Dans ce mémoire nous allons enrichir le corpus parallèle présenté précédemment [3] avec des enregistrements oraux de l'ensemble des données exprimées dans la langue source « Français ».

L'enregistrement de corpus est une opération technique et matérielle qui consiste à collecter des données audio, et qui doit être effectuée dans des conditions expérimentales (lieu d'enregistrement, enregistrement numérique etc.).

Ainsi, avant de commencer l'enregistrement, il faut choisir les locuteurs, le type des données, ainsi que l'outil à utiliser pour effectuer cette tâche.

En outre, afin d'obtenir une transcription acceptable, il est indispensable de gérer les spécificités de la parole. En effet, l'intervention de n'importe quelle personne est éventuelle dans le fichier audio à transcrire : le signal de parole peut varier en fonction du locuteur: son sexe (Homme/Femme), son âge (Adulte/Enfant), sa langue, son accent, son débit de parole, ou encore son état physiologique et psychologique (stress, maladie, affects/émotions etc.)

Parmi les erreurs commises par les locuteurs et produisant une perturbation du signal de parole: il y a le non-respect des tours de parole où l'intervention des deux locuteurs au même instant.

## **7. Revue de l'état l'art sur les mesures de confiance**

### **7.1 Mesures de confiance en TA**

Une fonctionnalité intéressante dans le contexte de la post-édition est l'aptitude d'un système de traduction automatique à s'auto-évaluer, c'est-à-dire à fournir, outre les

traductions automatiques, des éléments d'estimation quantitative qui s'apercevront de la confiance qu'a le système dans son résultat de sortie.

En traduction automatique, les mesures de confiance (MC) portent sur les phrases ou encore sur les mots.

Au niveau des phrases : étant données une phrase source ainsi que son hypothèse de traduction, le système associe un score de confiance pour l'hypothèse caractérisant sa qualité, en vue de l'accepter, la corriger ou la supprimer.

Egalement, l'estimation des mesures de confiance au niveau des mots consiste à prédire leur justesse dans la phrase cible générée par un système de traduction automatique en jugeant chaque mot de l'hypothèse s'il est correct ou incorrect, et par conséquent estimer la fiabilité d'une sortie de traduction afin de vérifier et post-éditer les segments mal traduits.

Diverses approches ont été proposées pour faire face au problème des mesures de confiance au niveau des mots, et la plupart d'elles convergent sur deux questions majeures : quels paramètres et quelles combinaisons sont efficaces? Et puis quel classifieur conviendrait le mieux pour entraîner les ensembles de paramètres? [5]

Dans l'un des premiers ouvrages ; et d'ailleurs l'un des plus connus dans ce domaine, (*Blatz et al., 2003*) rassemblent un grand nombre de paramètres, en employant des réseaux de neurones et des réseaux naïfs Bayésiens. Parmi ces paramètres, nous trouvons : la liste des *N*-meilleures hypothèses ou « *N-best lists based features* » ou les probabilités a posteriori « *Word Posterior Probability* » (*WPP*), proposées dans (*Ueffing et al., 2003*). [5]

Par ailleurs, (*Luong Ngoc Quang*) a étudié l'impact sur le calcul des mesures de confiance de divers types de paramètres dans son travail réalisé en **2012** : paramètres lexicaux et syntaxiques tirées de la sortie du système de traduction, en présentant la méthode qui permet de labelliser automatiquement les corpus utilisés ; Comme il a intégré les divers paramètres et a proposé une classification approprié des mots à l'aide du classifieur à base de champs aléatoires conditionnels. [5].

## **7.2 Mesures de confiance en RAP**

En reconnaissance automatique de la parole, les MC tentent à prédire le score de confiance que nous pouvons accorder à la sortie du moteur de reconnaissance, que ce soit pour un phonème, un mot ou une phrase.

La contribution des mesures de confiance permet ainsi de faire ressortir les mots mal reconnus ou encore les mots hors-vocabulaire. Par conséquent, ils peuvent être utilisés pour identifier

les erreurs de reconnaissance de la parole et distinguer les parties de transcription pertinentes de celles qui ne le sont pas.

Il y'a bien de travaux réalisés concernant les mesures de confiance en RAP, parmi les quelles, nous citons la publication de (**Lecouteux et al., 2009**) sur « *la détection des mots hors vocabulaire par combinaison des mesures de confiance de haut et bas niveaux* » dans le cadre des systèmes de RAP continu grand vocabulaire. Ils ont proposé une méthode inspirée par les mesures de confiance consistant à analyser les hypothèses du système dans le but d'y révéler automatiquement les erreurs liées aux mots hors vocabulaires. [7].

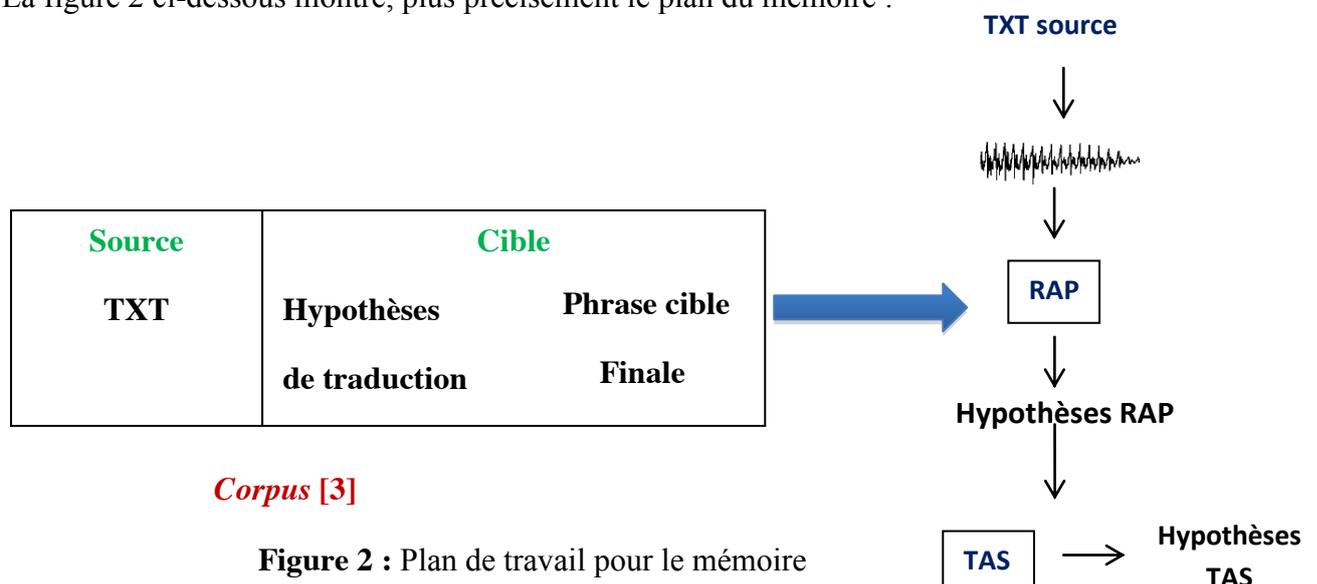
D'une autre part, (**Christian Raymond**) a présenté une approche d'estimation de la qualité des hypothèses de reconnaissance en se basant sur des mesures de confiance. Ces MC permettent d'identifier où les sorties ont divers taux d'erreur et sont notamment appropriées pour établir une stratégie d'acceptation ou rejet au niveau du module de dialogue.

D'autres travaux dans le cadre des mesures de confiance exploitent des informations directement tirées du processus de décodage comme : le nombre de mots en compétition à la fin d'un nœud, les vraisemblances normalisées, etc. [7].

## 8. Plan du travail pour le mémoire

Dans ce mémoire de recherche, nous abordons deux thèmes majeurs : nous nous intéressons en premier lieu à l'enrichissement du corpus [3] par des enregistrements audio ; Ensuite, en deuxième lieu nous allons nous concentrer aux mesures de confiance pour la traduction automatique de la parole, autrement dit, nous allons analyser les erreurs et estimer les mesures de confiance des hypothèses fournies par un système de reconnaissance automatique de la parole, puis des hypothèses produites par un système de traduction automatique.

La figure 2 ci-dessous montre, plus précisément le plan du mémoire :



**Figure 2 : Plan de travail pour le mémoire**

# **Chapitre 3:**

## **Enregistrement du Corpus oral**

L'enregistrement d'un corpus de parole est une tâche importante qui exige beaucoup de temps et de concentration, pour aboutir à un résultat satisfaisant.

Ce chapitre a pour objectif de présenter la tâche et les conditions d'enregistrement, ainsi que l'ensemble d'énoncés à faire prononcer par des locuteurs francophones natifs, en vue de constituer un corpus oral pertinent et bien structuré. Nous verrons alors, comment nous avons organisé notre travail pour atteindre ce but.

## 1. Enregistrement : tâche et conditions d'enregistrement

Evidemment, la première partie de notre projet vise à collecter des enregistrements oraux des 881 dernières phrases, fois 3, rédigées dans la langue source « FR » du Corpus [3], pour la Traduction automatique de la parole, notant que parmi les 21 300 mots du corpus, nous trouvons aussi des mots en Allemand, en Espagnol et en Italien, car le corpus comporte des données journalistiques de type « news » représentant des sujets de politique, sciences ou économie, extraits de différents sites web journalistiques tels que : (*Project syndicate, libération et le Figaro*) dont le contenu est publié en différentes langues.

En effet, nous avons segmenté l'ensemble de données en paquets de 176 phrases, et seulement le dernier paquet comporte 177 phrases ; Notant que chaque paquet est prononcé par 3 locuteurs. Ainsi, notre corpus inclut 2643 enregistrements (3 signaux pour chacune des 881 phrases) produits par 15 locuteurs francophones natifs, comme il est présenté dans le tableau ci-dessous :

<i>Locuteurs</i>	<i>Genre</i>	<i>Age</i>	<i>Nb de phrases</i>	<i>Locuteurs par paquet</i>
L01	M	29	176	<b>HYPs 1</b> : 1 <sup>er</sup> paquet de 881 phrases
L02	M	41	176	
L03	F	40	176	
L04	F	20	176	
L05	M	33	177	
L06	F	23	176	<b>HYPs 2</b> : 2 <sup>ème</sup> paquet de 881 phrases
L07	F	21	176	
L08	M	25	176	
L09	F	24	177	
L10	M	22	176	
L11	F	21	176	<b>HYPs 3</b> : 3 <sup>ème</sup> paquet de 881 phrases
L12	M	27	176	
L13	F	20	176	
L14	F	23	176	
L15	F	41	177	

**Tableau 1** : Locuteurs participants à la tâche de l'enregistrement

Par ailleurs, chaque enregistrement dure entre une heure et demi et deux heures, sachant que nos phrases comportent en moyenne 25 mots et que nous estimions expérimentalement qu'il faut 10 secondes environ pour prononcer chacune d'elles.

L'enregistrement des locuteurs a été fait en conditions réelles dans une chambre sourde au sein de l'appartement DOMUS du LIG. La chaîne d'acquisition est constituée d'un microphone, d'un modèle TASCAM-DR-100 relié à l'entrée de la carte son d'un ordinateur, ainsi que d'un logiciel d'enregistrement des phrases lues, qui a été développé par l'équipe GETALP du LIG.

Par conséquent, notre corpus comprend 2643 fichiers audio enregistrés au format Wav accompagnés de leurs fichiers de transcription orthographique.

## 1.1 Fiche signalétique

<b>Corpus</b>	
<b>Version</b>	<i>Corpus parallèle [3] enrichi par des enregistrements oraux</i>
<b>Type</b>	<i>Parole lue</i>
<b>Locuteurs</b>	<i>15 Adultes (9 Femmes &amp; 6 Hommes), âgés entre 20 et 41 ans</i>
<b>Enregistrement</b>	<i>Dans des conditions réelles</i>
<b>Durée</b>	<i>5h48</i>
<b>Contenu</b>	<i>Corpus audio &amp; Transcriptions orthographiques</i>
<b>Lieu de l'enregistrement</b>	<i>DOMUS - LIG</i>

**Tableau 2** : Fiche signalétique

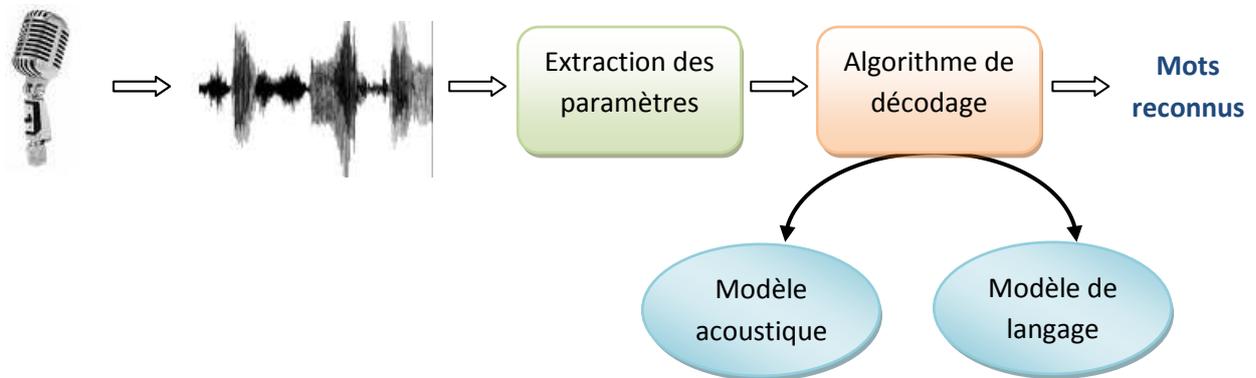
## 2. Conclusion

Dans ce chapitre, nous avons décrit brièvement la tâche d'enregistrement de notre corpus oral, en indiquant toutes les dispositions prise en compte, les conditions ainsi que le lieu d'enregistrement. Cela va de soi, nos données sont prêtes à être traitées par le système de reconnaissance de la parole.

**Chapitre 4:**

**Adaptation du système de  
reconnaissance de la parole**

Après avoir terminé l'enregistrement de notre corpus oral, nous passons à l'étape de reconnaissance automatique de la parole, afin de décoder ou de transcrire la parole sous forme de texte, ce qui est illustré dans la figure ci-dessous :



**Figure 3 :** Reconnaissance automatique de la parole

Au cours de ce chapitre, nous allons présenter dans un premier temps le système de RAP que nous avons utilisé ainsi que son architecture globale ; Ensuite nous détaillerons nos expérimentations et les traitements proposés en vue d'adapter le système de reconnaissance de la parole, tout en indiquant les résultats obtenus.

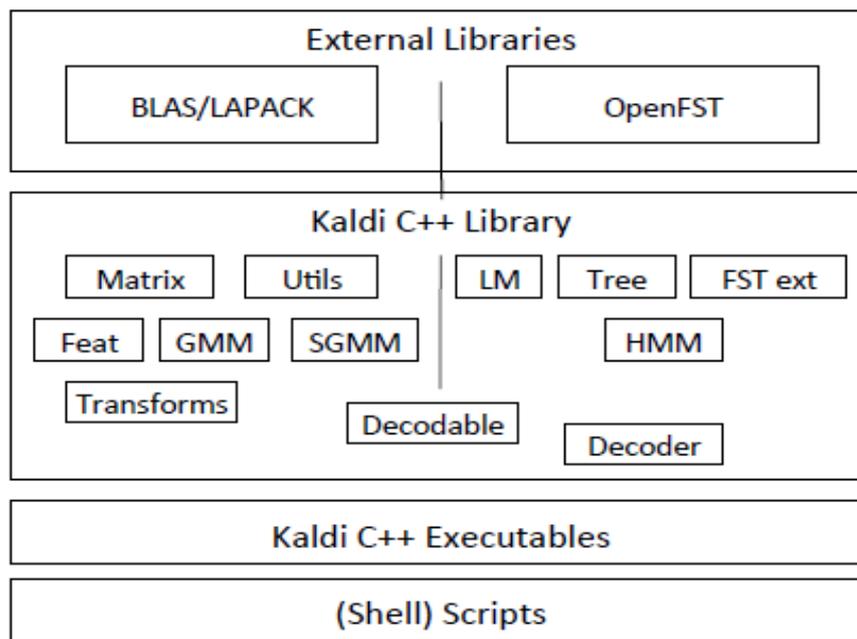
## 1. Système de reconnaissance de la parole KALDI

Le système de reconnaissance de la parole KALDI [Povey *et al.*,2011] est une boîte à outils open-source, gratuite, pour la recherche de l'identification de la parole. En effet, KALDI fournit un système d'identification de la parole basé sur des transducteurs à états finis (en utilisant le OpenFst, qui est disponible gratuitement) avec documentation détaillées et des scripts pour élaborer des systèmes d'identification complets. KALDI s'écrit en C++, et la bibliothèque centrale soutient la modélisation de tailles de contexte phonétique arbitraires, modélisation acoustique avec *subspace Gaussian mixture models (SGMM)* ainsi que *standard Gaussianmixture models*, de même que toutes les transformations courantes aussi bien affines que linaires.

KALDI est disponible sous Licence Apache v2.0, qui est hautement non-restrictive, par conséquent accessible à un grand nombre d'utilisateurs.

## 1.1 Architecture

Ci-dessous une vue simplifiée des divers composants de KALDI. Les modules de la bibliothèque peuvent être classés en deux groupes : ceux qui dépendent des bibliothèques d'algèbre linéaire et ceux qui dépendent du OpenFst. La classe décodable sert d'intermédiaire/relie entre les deux (moitiés). Les modules au bas du schéma dépendent de ceux en haut.



**Figure4** : vue simplifiée des divers composants de KALDI [17]

### 1.1.1 Modèle de Langage

Le modèle de langage sert à déterminer les probabilités des diverses suites de mots qui peuvent apparaître dans le résultat de décodage. En effet, le modèle de langage utilisé est trigramme avec environ 55K mots, et a été construit par [18] en exploitant l'ensemble des données du corpus *ESTER* ainsi que les textes extraits du corpus français *Gigaword*.

### 1.1.2 Modèle Acoustique

Le GMM et le Sous-espace GMM (SGMM) modélisent tous les deux les probabilités de chaque état HMM (Modèle de Markov caché) avec un modèle de mélanges gaussiens, mais dans l'approche SGMM les moyennes de gaussiennes et les poids de mélanges gaussiens sont

générés à partir du sous-espace phonétique et du sous-espace représentant les locuteurs avec l'ensemble des projections de poids[18].

## 2. Expérimentations

Etant donné qu'un système de RAP interprète un signal de parole, il fournit en sortie des transcriptions (une suite de mots) en minuscules, sans caractères accentués ni signes de ponctuation. Par conséquent, afin d'obtenir des résultats de reconnaissance de la parole pertinents et acceptables, il est nécessaire de passer par l'étape de normalisation des données de test utilisées.

Nous allons présenter dans cette section, les dispositions prises en compte en vue d'atteindre ce but.

### 2.1 Corpus de test

Après avoir extrait les 881 dernières phrases en langue source du corpus [3], celles-ci sont traitées par un script de normalisation que nous avons créé, en vue de nettoyer les références pour l'évaluation de taux d'erreur de reconnaissance du système de RAP. Ainsi, nous avons appliqué les modifications suivantes :

- corriger les fautes de frappe et quelques fautes d'orthographe manuellement.
- Supprimer les doubles espaces.
- Convertir tout le texte en minuscule.
- Eliminer les signes de ponctuation et certains caractères spéciaux.
- Appliquer la commande « iconv » pour convertir les caractères accentués en bdlex (ex. é → e1, è → e2 etc.)

Toutefois, dans un premier temps, nous avons évalué la qualité de la reconnaissance du système en termes de WER, en faisant le minimum de changement sur les transcriptions de référence. De ce fait, nous avons laissé les nombres, les mots composés (ex. week-end, ex-parti...), les abréviations (ex. Km, m<sup>2</sup> ...) ainsi que les caractères spéciaux (ex. %, ...) tels qu'ils sont.

La figure ci-dessous montre une partie du corpus avant et après normalisation :

nous comprenons le sens des protestations mais n' acceptons pas le vandalisme a exprime1 le delpute1 cdu du parlement reigional peter beuth  
les groupes parlementaires de l' opposition spd les verts et la gauche die linke ne partageaient pas cette opinion  
le csu prend le pretexte de quelques murs peinturlure1s pour criminaliser les protestations a dit janine wissler du parti la gauche linke  
au lieu de re1flelchir aux arguments et revendications des etudiants on les elvacue  
le parti la gauche a delpose1 une motion qui somme le ministre de l' intelrieur de remettre un rapport sur les conditions de l' intervention de la police  
la nouvelle silicon-valley

**Figure 5 :** Extrait du corpus référence nettoyé

En outre, nous avons utilisé dans notre première expérimentation : le modèle du langage décrit précédemment, le modèle acoustique *SGMM*, mais tout en conservant le lexique de KALDI.

D'après les résultats de sortie de RAP, le taux d'erreur par mot est relativement élevé « *WER = 29,05%* », sachant que la perplexité du corpus par rapport au modèle du langage est élevé aussi « *PPL = 218* ».

## **2.2 Problèmes de normalisation liés à la Traduction de la parole**

Afin d'obtenir des résultats de traduction de la parole pertinents et raisonnables, il est nécessaire de posséder un jeu de données, cohérent et approprié à l'application sur la reconnaissance automatique de la parole et encore sur la traduction automatique statistique.

Toutefois, comme il indiqué dans le premier chapitre, le corpus parallèle [3] contient : le texte source, les hypothèses de traduction ainsi que les post-éditions en langue cible ; Autrement dit, nous avons des contraintes à respecter, car les post-éditions avec lesquelles nous allons après comparer nos résultats de sortie de traduction, sont déjà établies. De ce fait, nous devons adapter le système de RAP avec ces contraintes.

Dans ce cas, il est nécessaire d'établir un rapport d'harmonie et d'organisation logique entre les diverses données utilisées, et à la fois de rapprocher les hypothèses de RAP à traduire, du format des post-éditions rédigées dans la langue cible «EN».

### **2.2.1 Cohérence dans la forme des données**

- *Usage des chiffres*

En reconnaissance automatique de la parole, les chiffres et les nombres s'écrivent systématiquement en toute lettre, car ils sont prononcés, mais, en traduction automatique, il

est préférable et convenu, d'utiliser les chiffres pour exprimer une année, un montant, une surface, des pourcentages etc.

- ***La casse***

Contrairement à la reconnaissance automatique de la parole, la casse est présente en traduction automatique.

- ***La ponctuation***

La ponctuation est assurément absente en reconnaissance de la parole, mais elle peut être présente en traduction automatique, dans le but d'organiser le texte et de le rendre plus compréhensible.

- ***Les abréviations et les mots composés***

En reconnaissance automatique de la parole, les abréviations (ex. km, M., Mme ...) ne font pas réellement sens, car nous traitons de la parole. En revanche, il est possible d'utiliser les abréviations ainsi que les acronymes en traduction automatique.

Par ailleurs, les tirets des mots composés (ex. week-end, ex-parti ...) sont fréquemment éliminés en RAP, en vue de conserver une taille raisonnable pour le lexique du système ; ce qui n'est pas le cas pour la TA.

- ***Les caractères spéciaux***

Tout comme les abréviations, les caractères spéciaux tels que le symbole de pourcentage « % », le symbole degré Celsius « °C », le symbole euro « € » etc., sont écrits en toute lettre en reconnaissance de la parole ; Toutefois en traduction automatique, ces caractères spéciaux sont présents.

## **Traitements proposés :**

Etant donné que les hypothèses de sortie du système de RAP comportent que des lettres et des caractères en format ***Bdelx***, et que nos transcriptions de références ne doivent pas être trop modifiées, dans ce cas nous proposons de post-éditer les hypothèses de RAP avant de calculer le score, en vue de les assimiler aux références.

Autrement dit, nous avons appliqué un script de correspondance que nous avons créé avec le langage Perl, sur la sortie de RAP pour atteindre les objectifs suivants :

- Transformer l'écriture des nombres qui est en toutes lettres, vers une écriture en chiffres. (ex. deux milles sept → 2007, douze virgule six → 12,6 ...)
- Rajouter les tirets pour les mots composés (week end → week-end, nord américain → nord-américain ...)
- Utiliser les abréviations (ex. madame → mme, note de la rédaction → ndlr ...)
- Utiliser certains caractères spéciaux (ex. pourcent → %, fois → × ...)

La figure ci-dessous illustre la différence entre les hypothèses de RAP telles qu'elles étaient produits par KALDI, et après application du script de correspondance :

- L14P4\_P4-2\_36 le holding doit rester a2 **cent pourcents** une entreprise familiale  
 - L14P4\_P4-2\_37 hambourg a2 long terme l' e1diteur aubert bourdin fera sa socie1te1 entre les mains de ses enfants  
 - L14P4\_P4-2\_38 les **deux** enfants e1viteront a2 **vingt sept** ans et sans fermement de1cide1 a2 inte1grer la socie1te1 id aubert bourdin **soixante neuf** ans au spiegel  
 - L14P4\_P4-2\_39 tous les de1tails ont e1te1 re1gle1s a2 l' avance  
 - L14P4\_P4-2\_40 la question de savoir s' ils seront actif dans l' entreprise se posera par la suite  
 - L14P\_P4-2\_41 il faudra voir e1galement s' ils en sont capables  
 - L14P4\_P4-2\_42 on ne peut pas imposer ses enfants a2 **sept milles cinq cents** collaborateurs sans savoir s' ils ont la capacite1 de s' en occuper

- L14P4\_P4-2\_36 le holding doit rester a2 **100 %** une entreprise familiale  
 - L14P4\_P4-2\_37 hambourg a2 long terme l' e1diteur aubert bourdin fera sa socie1te1 entre les mains de ses enfants  
 - L14P4\_P4-2\_38 les **2** enfants e1viteront a2 **27** ans et sans fermement de1cide1 a2 inte1grer la socie1te1 id aubert bourdin **69** ans au spiegel  
 - L14P4\_P4-2\_39 tous les de1tails ont e1te1 re1gle1s a2 l' avance  
 - L14P4\_P4-2\_40 la question de savoir s' ils seront actif dans l' entreprise se posera par la suite  
 - L14P4\_P4-2\_41 il faudra voir e1galement s' ils en sont capables  
 - L14P4\_P4-2\_42 on ne peut pas imposer ses enfants a2 **7500** collaborateurs sans savoir s' ils ont la capacite1 de s' en occuper

**Figure 6 :** La sortie de RAP avant et après application du script de correspondance

Ainsi, le taux d'erreur par mot a diminué de **29,05%** à **26,6%** ce qui est raisonnable, car les données de test utilisées contiennent pas mal de mots étrangers, et ceux-ci se répètent plusieurs fois dans le corpus.

## 2.3 Adaptation du modèle du langage et Ajout du vocabulaire

La troisième stratégie proposée, est celle la plus employée en vue de diminuer le taux d'erreur de mot du système de reconnaissance de la parole ; En effet, elle consiste à adapter le modèle de langage utilisé.

### 2.3.1 Corpus d'adaptation

Nous avons extrait les 10.000 phrases rédigées dans la langue source « FR » du corpus [3] dans le but d'utiliser ces données pour ajouter un nouveau lexique au système KALDI Français, et par suite adapter le modèle de langage.

Evidemment, il est important de passer par l'étape de normalisation du corpus avant de l'exploiter dans notre travail.

### 2.3.2 Ajout du nouveau vocabulaire et phonétisation

Pour faire une mise à jour du lexique, nous avons tokeniser tout d'abord le corpus d'adaptation en mots, ensuite nous avons fait la différence entre ces derniers et le vocabulaire français du système KALDI.

Après avoir extrait le nouveau vocabulaire, nous avons passé à l'étape de phonétisation à l'aide de l'outil **LIA-PHON** [Bechet, 2001], en vue d'associer chaque nouveau mot avec la ou les phonétisations correspondantes; Finalement nous avons fait l'union entre l'ancien et le nouveau lexique.

Le tableau ci-dessous nous montre la taille de notre nouveau vocabulaire de décodage :

<i>Ancien Lexique</i>	<b>347 783 mots</b>
<i>Union entre Ancien Lexique &amp; Nouveau Lexique</i>	<b>436 433 mots</b>

**Tableau 3** : Taille du nouveau lexique

### 2.3.3 Construction d'un nouveau Modèle de langage

Nous avons construit un modèle de langage à trigrammes incluant le nouveau vocabulaire, à l'aide de la boîte à outils **SRILM** [Stolcke, 2002], dans le but de l'interpoler

avec le modèle de langage générique. En effet, la perplexité du modèle entraîné mesurée sur les transcriptions de référence de notre corpus oral est de 400.

### 2.3.4 Interpolation avec l'ancien modèle de langage

L'interpolation de modèles de langage consiste à combiner linéairement les deux modèles thématiques afin de construire un nouveau modèle trigramme de la langue cible. En effet, cette approche peut donner de meilleurs résultats que l'entraînement d'un seul modèle sur les données disponibles, comme elle peut diminuer le taux de reconnaissance du système. De ce fait, nous avons interpolé notre modèle de langage créé avec celui du système de référence, en attribuant le poids d'interpolation 0,5.

### 2.4 Récapitulatif des Résultats de RAP obtenus

	<i>Taux d'erreur par mot (%WER)</i>	<i>WER après application du script de correspondance</i>	<i>PPL : Perplexité</i>	<i>Taux de mots hors vocabulaire (%MHV)</i>
<i>Expérience 2</i>	29,05 %	26,6 %	218	10,05%
<i>Expérience 3</i>	29,2 %	26,9 %	225	9,11%

**Tableau 4 :** L'ensemble de résultats de RAP obtenus

## 3. Conclusion

Pour qu'un système de reconnaissance automatique de la parole puisse interpréter de manière fiable et robuste une suite de mots prononcés, il faut tenir compte des différents types d'erreurs qu'il peut commettre. Ainsi dans nos expériences, nous avons cherché à identifier précisément les principales sources d'erreurs comme les problèmes de normalisation, et puis les mots hors vocabulaire qui n'appartiennent pas au lexique du moteur de reconnaissance ; Mais évidemment, les mots étrangers (en Allemand, en Espagnol et en Italien) n'ont pas été phonétisé correctement par le phonétiseur Français.

En outre, même si nous avons adapté le modèle de langage et ajouté 88 650 nouveaux mots au lexique, le **WER** a augmenté de **26,6%** à **26,9%**. Cela va de soi, nous allons utiliser dans les prochaines expérimentations de TA, les sorties de RAP produites lors de la deuxième expérience.

**Chapitre 5 :**  
**Génération de nouvelles traductions**  
**et des étiquettes**

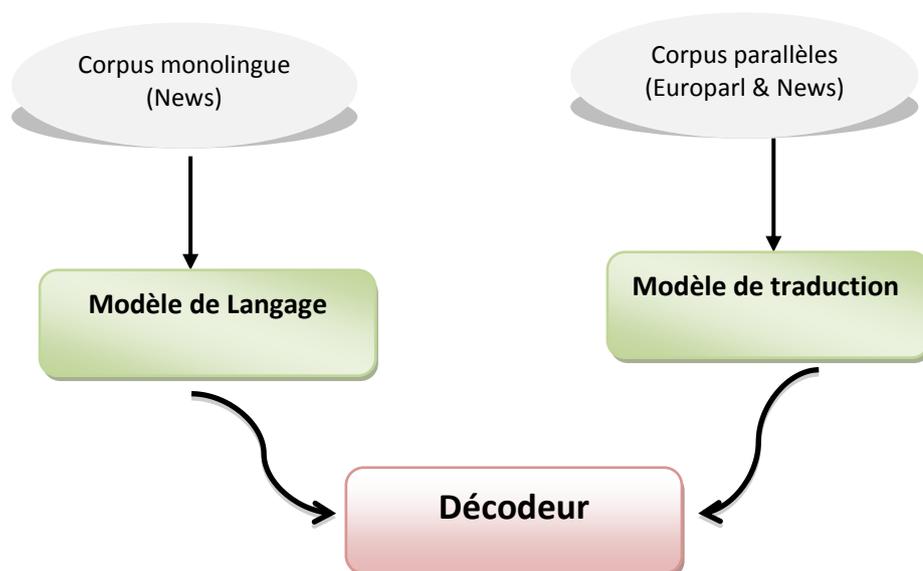
Après avoir obtenu les hypothèses de sortie en Français du système de reconnaissance de la parole, nous allons les mettre en entrée d'un système de traduction automatique en vue de les traduire vers l'anglais et générer des étiquettes qui seraient notamment utiles pour les mesures de confiance du système.

Dans ce dernier chapitre, nous nous intéressons dans un premier temps à l'architecture globale du système de TA utilisé, puis dans un second temps, nous allons évaluer les résultats produits en les comparant avec les post-éditions du corpus [3] et nous poursuivons avec la génération des étiquettes. Enfin nous présentons les différents résultats obtenus tout au long de nos expérimentations.

## 1. Système de Traduction

Nous allons utiliser le système MOSES-LIG [19] pour traduire nos données. Ainsi, les trois composants principaux sont : un modèle de langage, un modèle de traduction et un décodeur.

### 1.1 Architecture



**Figure 7 :** Système de traduction automatique

### **1.1.1 Modèle de langage**

Le modèle de la langue cible est un modèle de langage n-gram standard formé en utilisant le SRI Language Modeling Toolkit [Stoche, 2002] sur le corpus monolingue *News*. La technique de lissage appliquée est la *Kneser-Ney* modifiée avec interpolation [19].

### **1.1.2 Modèle de traduction**

Le modèle de traduction est entraîné en utilisant les deux corpus parallèles *Europarl* et *News*. Tout d'abord, les mots du corpus ont été alignés, les paires de termes en langue source et leurs équivalents en langue cible, ont été extraits du corpus d'entraînement bilingue dont les mots sont alignés, à l'aide des scripts fournis avec le décodeur Moses [Koehn et al. 2007]. Le résultat est une table de traduction contenant toutes les expressions alignées [19].

### **1.1.3 Décodage**

Pour le décodage, le système utilise une combinaison log-linéaire du modèle de la langue cible et treize modèles de traductions tirées de la table de traduction. Etant donné que le système peut être réglé au préalable, en ajustant des poids de combinaisons log-linéaires sur un corpus de développement, les méthodes utilisées sont celles de Minimum Error Rate Training (MERT) de [Och, 2013] [19].

## **2. Génération des étiquettes avec TERp-A**

### **2.1 Translation Edit Rate (TER)**

Le taux d'erreur d'éditions ou en anglais « Translation Edit Rate » (TER) est une progression du score WER qui est utilisé fréquemment pour l'évaluation des systèmes de RAP. Toutefois, le TER ajoute l'opération de déplacement d'un ensemble de mots contigus vers la gauche ou la droite, aux trois autres opérations d'éditions (Insertion, suppression et substitution). Notant que, tout déplacement est compté comme une seule opération sans tenir compte du nombre de mots décalés ou de l'amplitude de décalage.

### **2.2 TERp**

TERp est un développement de TER qui a non seulement hérité du succès de cette évolution métrique et outil d'alignement, mais qui a aussi éliminé ses défauts étant donné les opérations linguistiques d'édition telles que les associations radicaux, de synonymes et

substitution de locutions, évidemment, y compris les opérations conventionnelles (correspondance exacte, Insertion, Deletion , substitutions et décalage).

Evidemment, ces ajouts nous évitent de classer le mot d'hypothèse en tant qu'Insertion ou Substitution s'il l'a la même racine ou fait parti de la même liste de synonyme représentée par *WordNet*, ou s'il s'agit de la paraphrase d'un mot dans la référence. [Luong, 2012]

### 2.3 Le score BLEU

Le score BLEU en anglais « Bilingual Evaluation Understudy » a été proposé par [Papineni et al., 2002]. C'est une métrique qui mesure la similarité entre la traduction candidate et celle de référence en utilisant une précision n-gramme (pourcentage de vrais positifs) modifiée.

En effet, le score BLEU varie de 0 à 1, comme il peut être aussi représenté sous forme de pourcentage, et il est encore plus meilleur lorsqu'il est plus grand.

### 2.4 Résultats expérimentaux obtenus

Nous avons effectué trois expérimentations de TA pour les références et les hypothèses de sortie du système de RAP, en vue de retenir finalement les meilleurs résultats de traduction. Par suite, nous avons utilisé les données suivantes :

REF : représente les transcriptions de référence utilisées en RAP

HYPs1 : ce sont les hypothèses de sortie de la RAP du 1<sup>er</sup> paquet contenant 881 phrases.

HYPs2 : ce sont les hypothèses de sortie de la RAP du 2<sup>ème</sup> paquet de 881 phrases.

HYPs3 : ce sont les hypothèses de sortie de la RAP du 3<sup>ème</sup> paquet de 881 phrases.

MOY : représente la moyenne des 3 hypothèses (HYPs1 + HYPs2 + HYPs3), c'est la sortie du système de RAP contenant 2643 phrases.

De ce fait, nous avons traduit toutes les hypothèses de sortie de RAP ainsi que la REF à l'aide du système *baseline* du *LIG*, dans le but de calculer le TER, le Terpa et le score BLEU pour chacune de celles-ci. Toutefois, dans un premier temps nous avons déterminé ces valeurs en utilisant les données textuelles telles qu'elles sont ; Autrement dit les post-éditions (PE) du Corpus [3] comportent des signes de ponctuations, tandis que les hypothèses de RAP non. Ensuite, dans un deuxième temps nous avons enlevé automatiquement les ponctuations du

fichier PE pour voir le gain ; et après dans la troisième expérimentation nous avons calculé le TER, le Terpa ainsi que le BLEU, tout en conservant les ponctuations pour les PE, mais en rajoutant ces derniers pour les hypothèses de sortie de RAP à l'aide du script de [L.Besacier].

Par conséquent, tous résultats expérimentaux sont illustrés dans le tableau suivant :

Paquets	RAP (%WER)	TA (PE avec ponct)			TA (PE sans ponct)			TA (HYP RAP & PE avec ponct)		
		TER	Terpa	Bleu	TER	Terpa	Bleu	TER	Terpa	Bleu
<i>REF</i>		41%	49%	36,36%	39%	41%	36,57%	42%	49%	36,06%
<i>HYPs 1</i>	27,3%	53%	63%	26,71%	52%	57%	26,79%	50%	56%	30,11%
<i>HYPs 2</i>	26,3%	53%	62%	27,07%	52%	56%	26,95%	49%	55%	30,56%
<i>HYPs 3</i>	26,3%	52%	62%	27,36%	51%	55%	27,38%	49%	55%	31,03%
<i>MOY</i>	26,6%	52%	62%	27,04%	52%	56%	27,04%	49%	55%	30,57%

**Tableau 5:** récapitulatif de tous les résultats obtenus

Comme il est indiqué dans le tableau ci-dessus, les meilleurs résultats sont ceux de la troisième expérience : le TER de la moyenne des hypothèses est égale à 49%, le TERp-A 55% ainsi que le score BLEU 30,57%.

Nous allons présenter dans ce qui suit les étiquettes générées par TERp-A pour une hypothèse de traduction et une référence de TA (post-édition), de la troisième expérience :

Chaque mot ou une phrase de l'hypothèse est aligné respectivement à un mot ou une phrase de la référence avec une opération d'édition : (I : Insertion, S : Substitution, T : correspondances radicales, Y: correspondances synonyme, P: substitutions de phrases et D : Délétion), ainsi nous avons suivi la méthode de [Luong, 2012] pour classer les étiquettes générées comme G : Good (Bon) ou B : Bad (mauvais). Par suite, nous n'avons pas considéré les D car cela signifie que le mot apparaît que dans la référence, Cependant, l'absence d'une édition indique une correspondance exacte, alors nous avons remplacé tous les espaces par (E : Exact).

Après avoir obtenus les 6 catégories (E, T, Y, S, P, I) nous les simplifiant en une classe binaire : E, T, Y sont classés comme (Bon) et les S, P, I sont classés comme (Mauvais).

Enfin, nous comparons le nombre de termes trouvés avec le nombre de mots des hypothèses. Ainsi, nous avons obtenu 65,49% (Bon) et 34,51% (Mauvais).

### ***3. Conclusion***

Dans ce chapitre, nous avons présenté au début le système de traduction automatique statistique que nous avons utilisé ainsi que son architecture globale. Ensuite nous avons détaillé les étapes effectuées pour la génération des étiquettes en utilisant l'outil TERp-A. Enfin nous avons appliqué la méthode de *[luong, 2012]* pour calculer le pourcentage des mots considérés comme (Bon) et ceux considérés comme (mauvais), tout en indiquant les résultats obtenus.

# **Conclusion et perspectives**

## Conclusion

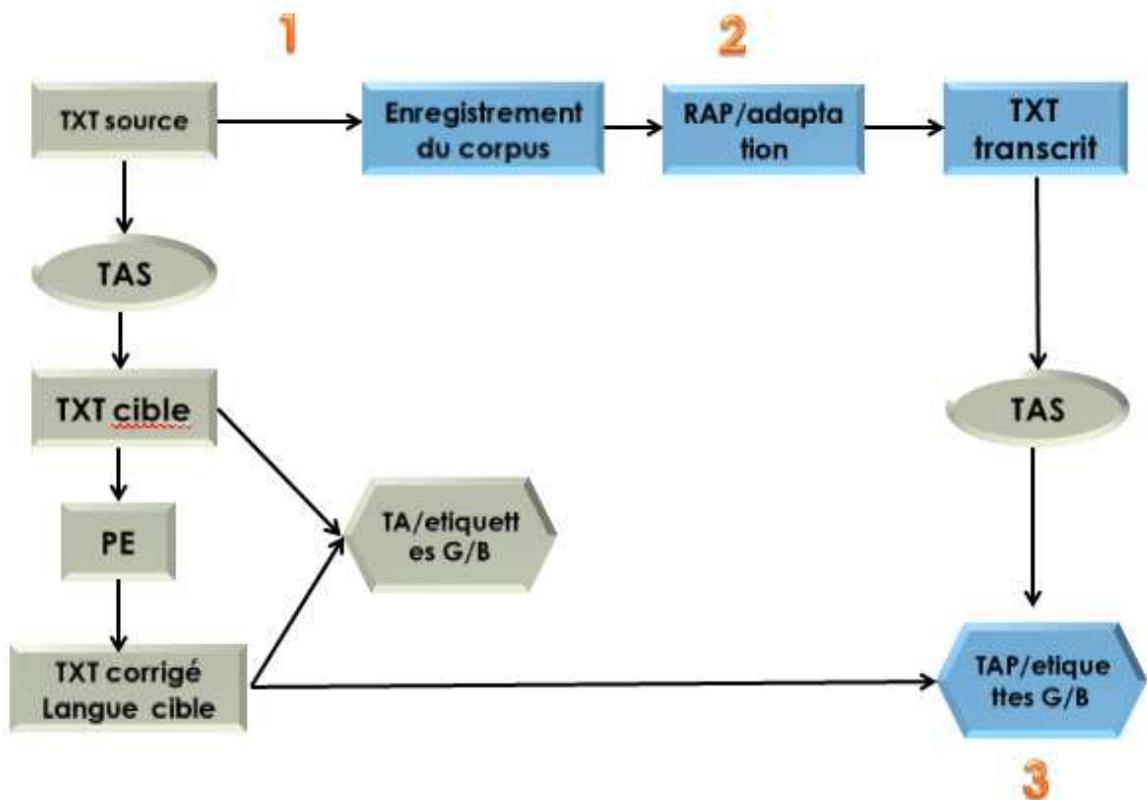
Dans le premier chapitre de ce manuscrit, nous avons d'abord décrit les mesures de confiance et leur nécessité dans le but de l'amélioration du fonctionnement et la sortie d'un système de TAP. De plus, nous avons exposé les deux domaines fondateurs de la traduction de la parole que sont la RAP et la TA en tenant compte des difficultés inhérentes à chacun de ces deux volets. D'autre part, nous avons présenté le corpus de post-éditions initial à enrichir par des enregistrements oraux. Dans un second temps, nous avons décrit de façon bien détaillée les mesures de confiance utilisées en reconnaissance automatique de la parole et en Traduction automatique de l'écrit.

Les autres chapitres correspondaient à notre contribution, à savoir, construire le premier corpus pour l'évaluation de mesures de confiance en traduction de la parole. Pour cela, nous avons enrichi le corpus écrit initial de post-éditions du LIG [3].

Dans le second chapitre de ce manuscrit, nous avons décrit la tâche et les conditions d'enregistrement de notre corpus oral, en indiquant toutes les dispositions prises en compte pour constituer un corpus de parole approprié.

Dans le troisième chapitre nous avons présenté le système de reconnaissance de la parole utilisé pour transcrire les enregistrements oraux, et nous avons détaillé les expérimentations réalisées dans le but d'adapter le système de RAP. Nous avons obtenu à la fin des résultats de sortie de RAP acceptables évalués en terme de taux d'erreur par mot.

Dans la quatrième partie de ce manuscrit, nous avons présenté le système de traduction automatique statistique que nous avons utilisé pour traduire les hypothèses de sortie du système de RAP. Ensuite, nous avons évalué les sorties de TAS en les comparant avec les post-éditions existantes du Corpus exploité. Enfin, nous avons suivie une méthode pour générer les étiquettes et classer les mots considérés comme bien traduits (B) et ceux considérés comme mal traduits (M). Ces étiquettes et le corpus oral associé constituent l'apport de ce mémoire.



**Figure 8 :** Schéma décrivant les trois contributions réalisées

## Perspectives

A partir des contributions détaillées et le centre de recherche décrit en introduction, quelques perspectives peuvent être envisagées.

Premièrement, il est important d'implémenter des systèmes de mesures de confiance pour la traduction automatique de la parole pour l'auto-évaluation du système de TAP. Autrement dit, de donner outre les sorties de transcriptions, des éléments d'estimation quantitatives qui s'apercevront de la confiance qu'a le système dans son résultat de sortie, dans le but de la post-édition. En effet les mesures de confiance peuvent porter sur les mots ou encore sur les phrases : étant données une phrase source ainsi que son hypothèse de traduction, le système va associer un score de confiance pour l'hypothèse pour caractériser sa qualité, en vue de l'accepter, la corriger ou la supprimer.

Par ailleurs, Pour améliorer la qualité de reconnaissance de la parole du système, il est envisageable d'utiliser un autre phonétiseur automatique pour l'ensemble des mots étrangers (mots en Allemand, Espagnol et Italien) existants dans le corpus [3].

# **ANNEXES**

## Annexe 1: Script Perl réalisés

<b>correspondance.pl</b>	Faire les correspondances entre les chiffres écrits en toutes lettres et les nombres écrits en chiffres. Faire les correspondances entre les mots composés comme (a-t-il, week-end). Faire les correspondance pour les abréviations.
<b>normalize.pl</b>	Nettoyer et normaliser les références .
<b>Decoupe.pl</b>	Découper un corpus en paquets
<b>Calc-Label.pl</b>	Calculer les pourcentages des mots considérés comme Bon et ceux considérés comme Mauvais
<b>extract-new-Voc.pl</b>	Extraire le nouveau vocabulaire
<b>CreatTransCorrigee.pl</b>	Affecter une transcription corrigée pour chaque fichier WAV.
<b>addBalise.pl</b>	Ajouter des balises afin de respecter le bon format pour le système KALDI.

## Annexe 2: Extrait du corpus référence nettoyé

les chirurgiens de los angeles ont dit qu' ils etaient outrels a  
delclare1 m camus

j' ai entendu parler d' une possible proceldure judiciaire a2 venir a-t-  
il dit

jean-claude camus a estimel que johnny hallyday eltait parti beaucoup  
trop vite pour los angeles avec l' autorisation du chirurgien parisien  
j' ai eu ce meldecin en personne en ligne qui m' a assure1 qu' il n' y  
avait aucun proble2me pour ce voyage et cela m' avait stupelfait eltant  
donnel que j' ai eu moi-me3me deux hernies discales a-t-il ajoutel  
peu aparavant sur rtl il avait indique1 que le chanteur eltait  
apparemment parti sans drain sans rien apre2s son opelration parisienne  
le dr stelphane delajoux est tre2s connu dans le milieu artistique  
il avait opelrel l' actrice marie trintignant a2 vilnius quelques jours  
avant son delce2s dans une clinique de neuilly-sur-seine ou2 il  
exerc5ait a2 l' elpoque

en 2007 il avait elgalement opelrel l' actrice charlotte gainsbourg pour  
relsorber un helmatome a2 la suite d' une helmorragie celrelbrale

### Annexe 3 : corpus de référence version 1 préparé pour KALDI

<p>&lt;s&gt; me olivier metzner l' avocat de l' ordre des me1decins l' a accuse1 d' e3tre un homme sans scrupules poursuivi re1gulie2rement devant le conseil de l' ordre et qui est loin d' e3tre recommandable &lt;/s&gt; (L01P1_P1-0_10)</p>
<p>&lt;s&gt; apre2s plusieurs jours de silence le dr ste1phane delajoux s' est exprime1 vendredi par le biais de son avocat me david koubbi &lt;/s&gt; (L01P1_P1-0_11)</p>
<p>&lt;s&gt; celui-ci a indique1 que l' intervention s' e1tait parfaitement bien de1roule1e et que les examens post-opelratoires e1taient normaux &lt;/s&gt; (L01P1_P1-0_12)</p>
<p>&lt;s&gt; l' ope1ration n' e1tait pas he1morrhagique et ne ne1cessitait donc pas la pose d' un drain a-t-il ajoute1 &lt;/s&gt; (L01P1_P1-0_13)</p>
<p>&lt;s&gt; en tout e1tat de cause la tourne1e du chanteur dont les ennuis de sante1 se multiplient depuis cet e1tel semble entre parenthe2ses &lt;/s&gt; (L01P1_P1-0_14)</p>
<p>&lt;s&gt; de toute e1vidence une reprise de la tourne1e le 8 janvier a2 amiens ndlr comme pre1vu e5a semble tre2s court a delclare1 m camus &lt;/s&gt; (L01P1_P1-0_15)</p>
<p>&lt;s&gt; l' association udf hausse le ton et somme le nouveau centre de ne plus utiliser son sigle &lt;/s&gt; (L01P1_P1-0_16)</p>

### Annexe 4 : corpus de référence version 2 préparé pour KALDI

L01P1\_P1-0\_01 les chirurgiens de los angeles ont dit qu' ils e1taient outre1s a delclare1 m camus

L01P1\_P1-0\_02 j' ai entendu parler d' une possible proce1dure judiciaire a2 venir a-t-il dit

L01P1\_P1-0\_03 jean-claude camus a estime1 que johnny hallyday e1tait parti beaucoup trop vite pour los angeles avec l' autorisation du chirurgien parisien

L01P1\_P1-0\_04 j' ai eu ce me1decin en personne en ligne qui m' a assure1 qu' il n' y avait aucun proble2me pour ce voyage et cela m' avait stupel1fait e1tant donne1 que j' ai eu moi-me3me deux hernies discales a-t-il ajoute1

L01P1\_P1-0\_05 peu auparavant sur rtl il avait indique1 que le chanteur e1tait apparemment parti sans drain sans rien apre2s son ope1ration parisienne

L01P1\_P1-0\_06 le dr ste1phane delajoux est tre2s connu dans le milieu artistique

## Annexe 5 : Exemple de phonétisation avec LIA-PHON

abondent	aa bb on dd an tt
abondent	aa bb on dd an tt
abondent	aa bb on dd an
abondent	aa bb on dd an
abonnements	aa bb oo nn eu mm an zz
abonnements	aa bb oo nn eu mm an zz
abonnements	aa bb oo nn eu mm an
abonnements	aa bb oo nn eu mm an
abonnements	aa bb oo nn mm an zz
abonnements	aa bb oo nn mm an zz
abonnements	aa bb oo nn mm an
abonnements	aa bb oo nn mm a

## Annexe 6 : comparaison entre les sorties de RAP sans et avec application du script de correspondance

L11P1\_P1-0\_02 j' ai entendu parler un possibles proce1dures judiciaires a2 venir **a t il** dit

L11P1\_P1-0\_02 j' ai entendu parler un possibles proce1dures judiciaires a2 venir **a-t-il** dit

L07P2\_P2-2\_31 toutefois commence a2 pre1sent a2 la re1alisation de stuttgart **vingt et un** a2 co3te1 du nouveau trace1 de la ligne inter city perd un

L07P2\_P2-2\_31 toutefois commence a2 pre1sent a2 la re1alisation de stuttgart **21** a2 co3te1 du nouveau trace1 de la ligne inter-city perd un

L07P2\_P2-2\_44 des experts pensent que **dix** a2 **quinze milliards** sont plus re1aliste

L07P2\_P2-2\_44 des experts pensent que **10** a2 **15** milliards sont plus re1aliste

L07P2\_P2-3\_04 depuis **mille neuf cents quatre vingt sept** cette ligne est construite au ralenti

L07P2\_P2-3\_04 depuis **1987** cette ligne est construite au ralenti

L11P1\_P1-2\_16 mais il a fallu encore plus de **six heures** d' une re1union commenc5ait a2 onze heures pour que tout le monde on est d' accord et qu' a2 **dix huit heures** les textes soient sydney

L11P1\_P1-2\_16 mais il a fallu encore plus de **6 h 00** d' une re1union commenc5ait a2 11 h 00 pour que tout le monde on est d' accord et qu' a2 **18 h 00** les textes soient sydney

## Annexe 7 : Exemples de sorties de TERp-A

**Sentence ID:** [1]

**Original Reference:** the surgeons of los angeles said they were outraged , mr camus said .

**Original Hypothesis:** surgeons in los angeles , it said they were also told me to camus .

**Reference:** the surgeons of los angeles said they were outraged , mr camus said .

**Hypothesis:** surgeons in los angeles , it said they were also told me to camus .

**Hypothesis After Shift:** surgeons in los angeles , it said they were also me to camus told .

**Alignment:**(DESEEIIEESSEYE)

**Sentence ID:** [2]

**Original Reference:** the surgeons of los angeles said they were outraged , mr camus said .

**Original Hypothesis:** it is a surgeon in los angeles , it said they were also said to camus .

**Reference:** the surgeons of los angeles said they were outraged , mr camus said .

**Hypothesis:** it is a surgeon in los angeles , it said they were also said to camus .

**Hypothesis After Shift:** it is a surgeon in los angeles , it said they were also to camus said .

**Alignment:**(IISTSEEIIEEDSSEE)

**Sentence ID:** [3]

**Original Reference:** the surgeons of los angeles said they were outraged , mr camus said .

**Original Hypothesis:** surgeons in los angeles , it said these and also said to camus .

**Reference:** the surgeons of los angeles said they were outraged , mr camus said .

**Hypothesis:** surgeons in los angeles , it said these and also said to camus .

**Hypothesis After Shift:** surgeons in los angeles said , it these and also to camus said .

**Alignment:**(DESEEEDDESIIIEE)

**Sentence ID:** [4]

**Original Reference:** i have heard about a possible legal action to come , he said .

**Original Hypothesis:** i have heard about a possible legal procedures to come , a-t-il said .

**Reference:** i have heard about a possible legal action to come , he said .

**Hypothesis:** i have heard about a possible legal procedures to come , a-t-il said .

**Hypothesis After Shift:** i have heard about a possible legal procedures to come , a-t-il said .

**Alignment:**(EEEEEEPEESE)

## *Bibliographie*

---

- [1] Anthony Rousseau, « **La traduction automatique de la parole** », Thèse- pour l'obtention du doctorat de l'Université du Maine, (07/12/2012).
- [2] Laurent Besacier, François Portet « **Cours Robotique de la parole** », [http://membres-liglab.imag.fr/portet/resources/ApplisTAP\\_4.pdf](http://membres-liglab.imag.fr/portet/resources/ApplisTAP_4.pdf)
- [3] Marion Potet, Emmanuelle, Esperança-rodier, Laurent Besacier, Hervé Blanchon. « **Collection of a Large Database of French-English SMT Output Corrections**» LREC (8th International Conference on Language Resources and Evaluation, Istanbul (Turkey), may 2012.
- [4] Ngoc-Quang Luong, Benjamin Lecouteux, Laurent Besacier. « **LIG System for WMT13 QE Task: Investigating the Usefulness of Features in Word Confidence Estimation for MT**». Proceedings of the 8th Workshop on Statistical Machine Translation, Sofia, Bulgaria, aug 2013.
- [5] Luong Ngoc Quang. «**Integrating Lexical, Syntactic, and System-based Features to improve Word Confidence Estimation in SMT** », Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 3: RECITAL, pages 43–56, Grenoble, 4 au 8 juin 2012.
- [6] Ngoc-Quang Luong, Laurent Besacier, Benjamin Lecouteux. « **Word Confidence Estimation and its Integration in Sentence Quality Estimation for Machine Translation.** » Proceedings of the fifth international conference on knowledge and systems engineering (KSE), Hanoi, Vietnam, oct 2013.
- [7] Lecouteux Benjamin and Linares Georges and Favre benoit. «**Combined low level and high level features for out-of-vocabulary word detection**», International conference of the Speech Communication Association ISCA InterSpeech'09 Tokyo Japan, 1187-1190, 2009.
- [8] Benjamin Lecouteux, « **Reconnaissance automatique de la parole guidée par des transcriptions a priori** », Thèse- Université d'Avignon - École Doctorale 166 I2S «Mathématiques et Informatique» Laboratoire d'Informatique (EA 4128), (05/12/2008).

- [9] Richard Dufour, « **Transcription automatique de la parole spontanée** », Thèse – pour l’obtention du doctorat à l’Université du Maine, (01/12/2010).
- [10] Julie Mauclair, « **Mesures de confiance en traitement automatique de la parole et applications**», Thèse – pour l’obtention du doctorat à l’Université du Maine. (05/12/2006).
- [11] Elisabeth Delais- Roussarie, « **Linguistique de Corpus**», [Lien pour le cours](#) .
- [12] François Yvon et Guillaume Wisniewski, « **Mesures de confiance en traduction automatique** », Proposition de thèse en collaboration avec Lingua Custodia, Draft - Mai 2013, (07/06/2013).
- [13] Christian Raymond, « **Mesures de confiance pour la reconnaissance de la parole dans des applications de dialogue homme-machine**», *Majestic* , Octobre, 2003, Marseille, France.
- [14] Laurent Besacier, «**Cours Communication langagière**», (Master IDL), <http://www-clips.imag.fr/geod/User/laurent.besacier/CoursCL/3.4.pdf>
- [15] Daniel Déchelotte, « **Traduction automatique de la parole par méthodes statistiques** », Thèse- pour l’obtention du doctorat de l’Université Paris-Sud 11, (17/12/2007).
- [16] Matthew Snover, Nitin Madani, Bonnie Dorr, Richard Schwartz. « **TERp System Description** » In MetricsMATR workshop at AMTA (2008).
- [17] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Luk’as Burget, Ondřej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motl’rcek, Yanmin Qian, Petr Schwarz, Jan Silovsk’, Georg Stemmer, Karel Vesel’ « **The Kaldi Speech Recognition Toolkit** » , **conference paper**, Hilton Waikoloa Village, Big Island, Hawaii, US , 2011.
- [18] Michel Vacher, Benjamin Lecouteu, François Portet, « **Multichannel Automatic Recognition of Voice Command in a Multi-Room Smart Home : an Experiment involving Seniors and Users with Visual Impairment** », Interspeech 2014, Singapore, Singapore. Proceedings of Interspeech 2014, pp. 1-5.
- [19] Marion Potet, Laurent Besacier and Hervé Blanchon, «**The LIG machine translation system for WMT 2010**», Uppsala, Sweden. July 15-16, 2010. vol. 1/1: pp. 167-172.
- [20] Laboratoire d’Informatique de Grenoble [www.liglab.fr](http://www.liglab.fr)
- [21] équipe GETALP du Laboratoire d’Informatique de Grenoble [getalp.imag.fr](http://getalp.imag.fr)

**DECLARATION**

- Ce travail est le fruit d'un travail personnel et constitue un document original.
  - Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
  - Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

Nom : **HADJ SALAH**

Prenom : **MARWA**

Date : **25/06/2014**

Signature :

