

La traduction de la parole est un thème de recherche récent, car il combine deux problèmes scientifiques complexes : la reconnaissance de la parole et la traduction automatique. Mais on imagine sans mal les applications potentielles : systèmes de réservation multilingues, aide au tourisme, indexation cross-lingue de contenus multimédias, assistant pour l'échange d'informations et la négociation, etc. Cette thèse a porté sur la traduction automatique et plus particulièrement sur la traduction de la parole reconnue automatiquement. La tâche retenue est la traduction des discours des députés européens aux sessions plénières du parlement européen, entre l'anglais et l'espagnol.

Nos recherches ont débuté par la conception d'un décodeur pour le modèle « IBM-4 », un modèle statistique performant à base de mots. Ce décodeur a été entièrement développé au cours de cette thèse. Au milieu de l'année 2006, l'avènement de Moses, un décodeur par groupes de mots libre et à l'état de l'art, nous a donné l'opportunité de poursuivre nos recherches avec un autre modèle de traduction. Nous avons envisagé une collaboration entre les deux décodeurs, mais elle n'a malheureusement pas produit l'amélioration espérée.

Dans nos expériences avec les deux décodeurs, le modèle de langage quadrigramme neuronal, développé originellement pour la reconnaissance de la parole, s'est avéré très performant dans les deux sens de traduction, amenant des améliorations sensibles pour toutes les mesures automatiques. Les systèmes de traduction mis en œuvre dans cette thèse ont été très compétitifs à la dernière évaluation TC-STAR, en février 2007.

Au cœur des systèmes de traduction par groupes de mots comme Moses, se trouve la table de traduction, sorte de « dictionnaire bilingue ». Les scores qu'elle contient sont le résultat de choix heuristiques. Nous avons proposé un algorithme inspiré de celui du Perceptron pour modifier de façon discriminante ces scores en observant les erreurs de traduction sur un ensemble de développement. Un gain substantiel a été observé dans un sens de traduction mais n'a pas été confirmé sur une autre tâche. Nous pensons que ces résultats contrastés pourraient être dus à un défaut de cohérence, ou de *lissage*, entre les scores de la table de traduction.

L'amélioration de l'interaction entre la reconnaissance de la parole et la traduction présente plusieurs aspects. À notre connaissance, nous avons été les premiers à mesurer l'impact du taux de mots erronés de la reconnaissance sur les performances de la traduction, et d'évaluer séparément les impacts respectifs du modèle de langage source et du modèle acoustique.

Un autre aspect est la prise en compte de l'ambiguïté de la sortie de la reconnaissance automatique, c'est-à-dire les mots entre lesquels le système de reconnaissance « hésite ». D'après notre propre expérience et les articles parus à ce sujet, toutes les méthodes imposent leurs compromis respectifs, que l'on traduise le treillis de mots produit par la reconnaissance, un réseau de confusion ou une liste de  $n$  meilleures hypothèses.

Nous nous sommes ensuite placés dans le cadre de la traduction d'un flux de mots produit par un système de reconnaissance « inconnu ». Plusieurs traitements spécifiques à la parole sont utiles, pour gérer les mots répétés et autres disfluences. Nous avons constaté l'importance de transformer les données à traduire pour les faire ressembler aux données d'entraînement du système. Mais de façon surprenante, en matière de divergence entre données d'entraînement et données de test, nos expériences ont montré que la *punctuation* était au moins aussi importante que les mots. Nous avons proposé un algorithme d'insertion de ponctuations dont le seul critère est la *quantité* de ponctuations insérées. Cet algorithme a permis d'améliorer très nettement le score BLEU, mesurant la qualité de traduction.

Enfin, nous avons modifié le système de reconnaissance de manière à lui faire insérer ou supprimer plus de mots. Bien que le score BLEU puisse légèrement bénéficier d'un taux d'insertion plus élevé, le taux de mots erronés WER semble le bon critère à minimiser par la reconnaissance pour obtenir les meilleures performances de traduction de la parole.