Approches statistiques discriminantes pour l'interprétation sémantique multilingue de la parole

Bassam Jabaian[1], Fabrice Lefèvre[1], Laurent Besacier[2]
(1) LIA, Université d'Avignon et des Pays de Vaucluse, Avignon, France
{bassam.jabaian,fabrice.lefevre}@univ-avignon.fr
(2) LIG, Université Joseph Fourrier, Grenoble, France laurent.besacier@imag.fr

RÉSUMÉ

Les approches statistiques sont maintenant très répandues dans les différentes applications du traitement automatique de la langue et le choix d'une approche particulière dépend généralement de la tâche visée. Dans le cadre de l'interprétation sémantique multilingue, cet article présente une comparaison entre les méthodes utilisées pour la traduction automatique et celles utilisées pour la compréhension de la parole. Cette comparaison permet de proposer une approche unifiée afin de réaliser un décodage conjoint qui à la fois traduit une phrase et lui attribue ses étiquettes sémantiques. Ce décodage est obtenu par une approche à base de transducteurs à états finis qui permet de composer un graphe de traduction avec un graphe de compréhension. Cette représentation peut être généralisée pour permettre des transmissions d'informations riches entre les composants d'un système d'interaction vocale homme-machine.

ABSTRACT

Discriminative statistical approaches for multilingual speech understanding

Statistical approaches are now widespread in the various applications of natural language processing and the elicitation of an approach usually depends on the targeted task. This paper presents a comparison between the methods used for machine translation and speech understanding. This comparison allows to propose a unified approach to perform a joint decoding which translates a sentence and assign semantic tags to the translation at the same time. This decoding is achieved through a finite-state transducer approach which allows to compose a translation graph with an understanding graph. This representation can be generalized to allow the rich transmission of information between the components of a human-machine vocal interface.

MOTS-CLÉS : compréhension multilingue, système de dialogue, CRF, graphes d'hypothèses.

KEYWORDS: multilingual understanding, dialogue system, CRF, hypothesis graphs.

1 Introduction

Aujourd'hui, les approches statistiques sont très utilisées pour toutes les applications du traitement automatique de la langue (reconnaissance de la parole, traduction automatique,

analyse syntaxique, étiquetage sémantique...). La performance d'une approche particulière dépend énormément de la tâche à laquelle elle est appliquée. Et, selon les tâches, les approches permettant les meilleures performances ne sont pas toujours les mêmes.

Par exemple, pour une tâche de compréhension de la parole (*Spoken Language Understanding*, SLU), assimilable à un étiquetage séquentiel en concepts, les champs aléatoires conditionnels (*Conditional Random Fields*, CRF) (Lafferty *et al.*, 2001) utilisés dans leur version chaîne linéaire sont les plus performants (Hahn *et al.*, 2010). Alors que pour la traduction automatique, ce sont les modèles de traduction log-linéaires à base de segments sous-phrastiques (*Log-linear Phrase-Based Statistical Machine Translation*, LLPB-SMT) (Koehn *et al.*, 2003), qui sont le plus souvent utilisés.

Cependant, malgré les différences entre les approches statistiques, celles-ci présentent des points communs et les frontières entre les unes et les autres ont tendance à s'estomper. On voit, par exemple, des travaux autour de l'utilisation d'approches discriminantes de type CRF pour la traduction automatique (Och et Ney, 2002; Liang *et al.*, 2006; Lavergne *et al.*, 2011), tandis que les approches de traduction à base de segments, sont aussi utilisées dans d'autres tâches du traitement automatique de la langue, comme la conversion graphème-phonèmes (Rama *et al.*, 2009) ou le décodage de Part-Of-Speech (Gascó i Mora et Sánchez Peiró, 2007).

Dans cet article nous comparons les approches CRF-SLU et LLPB-SMT pour les tâches de compréhension et de traduction. Pour cela nous proposons d'utiliser et d'optimiser une approche LLPB-SMT pour la compréhension de la parole, et par ailleurs d'intégrer des modèles à base de CRF à un module de traduction automatique. Cette étude nous permet de mettre en avant les spécificités de chaque tâche et d'évaluer les performances des approches respectives sur ces tâches.

D'autre part, nous avons montré dans un travail précédent (Jabaian *et al.*, 2010, 2011) que l'utilisation de la traduction automatique constitue une solution efficace pour la portabilité multilingue d'un module de compréhension d'une langue vers une autre. Cette portabilité peut être obtenue en cascadant un module de traduction avec un module de compréhension (pour traduire les entrées d'un utilisateur vers une langue pour laquelle nous disposons d'un système de compréhension).

Dans certains cas, la meilleure hypothèse de traduction n'est pas l'hypothèse pour laquelle le système de compréhension génère la meilleure hypothèse (souvent pour des raisons liées à l'ordre des mots). Et donc la sélection préalable de la meilleure traduction n'optimise pas forcément le système lorsqu'on se place selon un scénario de compréhension multilingue.

Nous nous basons sur la comparaison réalisée entre les deux tâches afin de pouvoir proposer un modèle qui pourra gérer la traduction et la compréhension d'une manière similaire permettant un décodage conjoint entre les modules. Ce décodage conjoint permettra de sélectionner des traductions en tenant compte des hypothèses d'étiquetage sémantique. Dans cet esprit, nous ne cherchons plus la meilleure traduction possible mais la traduction qui sera étiquetée sémantiquement de la meilleure manière possible.

Nos expériences sont basées sur le corpus de dialogue français MEDIA sur lequel nous apprenons un système de compréhension du français. Dans le but de pouvoir utiliser ce système pour étiqueter des entrées en italien, nous apprenons un système de traduction de l'italien vers français, qui sera utilisé ensuite lors des tests pour traduire les entrées italiennes vers le français afin de les fournir en entrée du système de compréhension.

Cet article est organisé de la manière suivante : la section 2 présente l'utilisation d'une approche de traduction automatique pour la compréhension de la parole. La section 3 décrit l'utilisation des CRF pour la traduction automatique. Notre proposition pour un décodage conjoint entre la compréhension et la traduction est présentée dans la section 4. Enfin la section 5 présente l'étude expérimentale et les résultats.

2 Méthode de traduction pour la compréhension

Le problème de la compréhension d'un énoncé utilisateur peut être vu comme un problème de traduction de la séquence de mots qui forme cet énoncé (langue source) vers une séquence de concepts (langue cible). (Macherey *et al.*, 2001, 2009) ont montré que les approches de la traduction automatique statistique peuvent être utilisées avec un certain succès pour une tâche de compréhension de la parole. Cette approche part du principe que les séquences de concepts sont les traductions des séquences de mots initiales.

Malgré l'apparente similitude entre les tâches de compréhension et de traduction, la compréhension a ses spécificités qui doivent être prises en considération afin de pouvoir améliorer les performances obtenues par une approche de traduction comme LLPB-SMT.

Les différences entre une tâche de traduction classique (d'une langue naturelle vers une autre) et l'utilisation de la traduction pour la compréhension (traduction d'une langue vers des étiquettes sémantiques) peuvent être résumées comme suit :

– la sémantique d'une phrase respecte l'ordre dans lequel les mots sont émis contrairement à une tâche de traduction où les mots traduits peuvent avoir un ordre différent de l'ordre des mots de la phrase source selon le couple de langues considérées ;
– dans une tâche de traduction, un mot source peut n'être aligné à aucun mot cible (fertilité = 0), alors que pour la compréhension chaque mot doit être aligné à un concept, sachant que les mots qui ne contribuent pas au sens de la phrase sont étiquetés par un concept spécifique `NULL` ;
– enfin, les mesures d'évaluation sont différentes entre les deux tâches (BLEU (Papineni *et al.*, 2002) pour la traduction vs. CER pour la compréhension) et donc les outils utilisés pour l'optimisation des systèmes de traduction doivent être adaptés pour optimiser le score CER.

En suivant l'hypothèse que la sémantique d'une phrase respecte l'ordre dans lequel les mots sont émis, nous proposons d'imposer une contrainte de monotonie pendant la traduction (décodage monotone), qui oblige le décodeur à respecter, en fonction de l'ordre des mots initiaux, l'ordre des concepts générés.

Une difficulté majeure du processus de traduction automatique est l'alignement d'un mot de la langue source avec le mot correspondant dans le langue cible. Vu que les corpus utilisés pour apprendre des systèmes de traduction sont des corpus alignés au niveau des phrases, une étape d'alignement automatique est nécessaire pour obtenir l'alignement en mots. Cependant, la plupart des corpus de compréhension sont étiquetés (alignés) au niveau des segments conceptuels et donc l'utilisation de ces informations d'alignement peut être avantageuse pour aider le processus d'alignement.

Pour cela nous proposons d'utiliser les corpus en format BIO (Begin Inside Outside) (Ramshaw

et Marcus, 1995). Ce format garanti que chaque mot de la phrase source est aligné à son concept correspondant et donc aucun alignement automatique supplémentaire n'est requis. De cette façon, l'extraction de la table de segments est obtenue à partir d'un corpus avec un alignement parfait (non bruité).

Vu que nous cherchons à évaluer les hypothèses générées par cette approche du point de vue de la compréhension (la mesure d'évaluation du système de compréhension étant le CER et non pas le score BLEU) nous proposons de modifier l'algorithme MERT (Och, 2003) afin d'optimiser le CER directement.

3 Méthode de compréhension pour la traduction

Dans cette approche, le problème de la traduction d'une phrase est considéré comme un problème d'étiquetage de la séquence de mots source, avec comme étiquettes possibles les mots de la langue cible. L'apprentissage d'un étiqueteur fondé sur une approche CRF pour une tâche de traduction nécessite un corpus annoté (traduit) au niveau des mots. L'application des modèles IBM (Brown *et al.*, 1993) permet d'obtenir automatiquement des alignements en mots à partir d'un corpus bilingue aligné au niveau des phrases.

Comme pour la compréhension, où plusieurs mots peuvent être associés à un seul concept, plusieurs mots source peuvent être alignés avec un seul mot cible. Pour gérer cela, la proposition la plus simple est d'appliquer la même méthode utilisée pour la compréhension : le passage au format BIO. Ainsi la séquence française "je voudrais" qui est alignée au mot italien "vorrei" sera représentée comme : <je, B_vorrei> <voudrais, I_vorrei>.

La difficulté principale pour apprendre des modèles CRF pour la traduction est liée au nombre élevé d'étiquettes (correspondant à la taille du vocabulaire de la langue cible). (Riedmiller et Braun, 1993) ont proposé d'utiliser l'algorithme RPROP pour l'optimisation des paramètres de modèles lorsqu'il s'agit d'un modèle avec un nombre important de paramètres. Cet algorithme réduit le besoin en mémoire par rapport à d'autres algorithmes d'optimisation (Turian *et al.*, 2006).

Un autre défaut important de l'utilisation des CRF pour la traduction est qu'ils ne prennent pas en compte le réordonnancement des mots et que le modèle de langage cible limité par la complexité algorithmique lors du décodage. Afin d'obtenir un système de traduction efficace à base de CRF, (Lavergne *et al.*, 2011) ont proposé un modèle fondé sur des transducteurs à états finis qui composent les différents étapes du processus de traduction. Nous l'appellerons CRFPB-SMT car il intégre aussi un mécanisme pour la modélisation d'une table de traduction par segments sous-phrastiques (appelés tuples dans ce contexte).

Le décodeur proposé pour ce modèle est une composition de transducteurs à états finis pondérés (*Weighted Finite State Transducer*, WFST) qui met en ouvre les fonctionnalités standards des WFST, disponibles dans des bibliothèques logicielles comme OpenFST (Allauzen *et al.*, 2007). Essentiellement, le décodeur de traduction est une composition de transducteurs qui représentent les étapes suivantes : le réordonnancement et la segmentation de la phrase source selon les tuples de mots, l'application du modèle de traduction (mise en correspondance des parties source et cible des tuples) avec une valuation des hypothèses à base de CRF et, enfin, la composition avec un modèle de langage dans la langue cible. (Kumar et Byrne, 2003)

a proposé une architecture assez similaire qui utilise un modèle ATTM (*Alignment Template Translation Models*) au lieu des CRF comme modèle de traduction.

Cette architecture permet de voir la traduction d'une phrase comme une composition ($\circ$) de transducteurs dans l'ordre suivant :

$$\lambda_{traduction} = \lambda_S \circ \lambda_R \circ \lambda_T \circ \lambda_F \circ \lambda_L$$

sachant que λ_S est l'accepteur de la phrase source, λ_R représente un modèle de réordonnancement, λ_T est un dictionnaire de tuples, qui associe des séquences de la langue source avec leurs traductions possibles en se basant sur l'inventaire des tuples lors de l'apprentissage, λ_F est une fonction d'extraction de motifs (*feature matcher*), qui permet d'attribuer des scores de probabilité aux tuples en les comparant aux motifs des fonctions caractéristiques du modèle CRF et λ_L est un modèle de langage de la langue cible.

4 Décodage conjoint pour la traduction et la compréhension, application à la compréhension multilingue

Notre étude des relations entre les différentes approches est réalisée avec l'objectif de pouvoir les combiner du mieux possible pour la portabilité multilingue d'un système de compréhension.

Dans des travaux précédents (Jabaian *et al.*, 2010), nous avons montré que la meilleure méthode pour porter un système de compréhension existant vers une nouvelle langue est aussi la plus simple : traduire les énoncés utilisateurs de la nouvelle langue vers la langue du système existant et ensuite faire étiqueter les énoncés (traduits) par ce système.

Notre proposition est basée sur une cascade d'un système de traduction (LLPB-SMT) et d'un système de compréhension (CRF-SLU). La meilleure hypothèse générée par le système de traduction constitue l'entrée du système de compréhension. Cependant, d'autres hypothèses de traductions peuvent différer (même sensiblement, par exemple dans l'ordre des mots) et ces variantes peuvent être mieux interprétées par l'étiqueteur sémantique. Donc la sélection a priori de la meilleure traduction n'optimise pas forcément le comportement du système global.

Pour faire face à ce problème nous proposons d'effectuer un décodage conjoint entre la traduction et la compréhension. Ce décodage conjoint aura l'avantage de pouvoir optimiser la sélection de la traduction en prenant compte des étiquettes qui peuvent être attribuées aux différentes traductions possibles.

La proposition d'utiliser l'approche CRFPB-SMT utilisant des transducteurs pour la traduction de graphes d'hypothèses peut être appliquée la compréhension. Donc un système de compréhension $\lambda_{comprehension}$ peut être obtenu de la même manière que proposé dans la section 3. Cette représentation nous permet alors d'obtenir un graphe de compréhension similaire à celui obtenu pour la traduction. Vu que le vocabulaire des sorties du graphe de traduction est le même que celui de l'entrée du graphe de compréhension, ces deux graphes peuvent être composés facilement en utilisant la fonction de composition pour donner un graphe permettant le décodage conjoint :

$$\lambda_{conjoint} = \lambda_{traduction} \circ \lambda_{comprehension}$$

Cette composition prend une phrase de la langue cible en entrée et attribue une séquence de concept à cette phrase en passant par un étiqueteur disponible dans la langue source. Elle nous permet d'obtenir un décodage conjoint entre la traduction et la compréhension dans la mesure où les probabilités des deux modèles sont prises en compte. Un tel décodage ne cherche pas à optimiser la traduction en soi, mais à optimiser le choix d'une traduction qui donnera une meilleure compréhension automatique.

Le transducteur $\lambda_{conjoint}$ peut être généralisé pour permettre de composer un graphe de reconnaissance de la parole avec un graphe de compréhension dans le cadre d'un système de dialogue. Dans un tel cas des procédures d'élagage devront être prises en compte afin d'assurer que les opérations de composition puissent être réalisées selon les contraintes classiques (temps de calcul et espace mémoire machine disponible).

4.1 Travaux connexes

Ce problème rejoint, dans son esprit, le problème classique de la cascade des composants d'un système d'interaction vocal homme-machine. Dans une architecture standard, le système de reconnaissance de la parole transmet sa meilleure hypothèse de transcription au système de compréhension. Vu que cette hypothèse est bruitée, elle n'est pas forcément l'hypothèse que le système de compréhension pourra étiqueter le mieux.

Plusieurs travaux ont proposé un décodage conjoint entre la reconnaissance et la compréhension de la parole pour prendre en compte les n-meilleures hypothèses de reconnaissance lors de l'étiquetage sémantique. Ces premiers travaux (Tür *et al.*, 2002; Servan *et al.*, 2006; Hakkani-Tür *et al.*, 2006) ont proposé d'utiliser un réseau de confusion entre les différentes sorties de reconnaissance pour obtenir un graphe d'hypothèses. Le système de compréhension dans ces propositions a été représenté par un WFST, dont les poids sont obtenus pas maximum de vraisemblance sur les données d'apprentissage. Et le décodage conjoint est obtenu par la composition du graphe de reconnaissance avec le graphe de compréhension.

Les résultats positifs obtenus par ces propositions ont encouragé d'autres travaux dans la même ligne. Vu que les modèles les plus performants dans la littérature sont les CRF, (Anoop Deoras et Hakkani-Tur, 2012) a proposé d'utiliser des modèles CRF au lieu des WFST pour l'étape de compréhension.

Dans la lignée de ces travaux, notre proposition cherche à obtenir un décodage conjoint pour la traduction et la compréhension. Les deux systèmes étant de natures différentes, leur combinaison et leur optimisation conjointe sont rendues délicates, d'où l'intérêt d'uniformiser les systèmes pour les deux tâches.

5 Expériences et résultats

Toutes nos expériences utilisent le corpus de dialogue français MEDIA. Le corpus MEDIA décrit dans (Bonneau-Maynard *et al.*, 2005) couvre un domaine lié aux réservations d'hôtel et aux informations touristiques. Ce corpus est annoté avec 99 étiquettes qui représentent la sémantique du domaine.

Le corpus est constitué de 1257 dialogues regroupés en 3 parties : un ensemble d'apprentissage (environ 13k phrases), un ensemble de développement (environ 1,3k phrases) et un ensemble d'évaluation (environ 3,5k phrases). Un sous-ensemble de données d'apprentissage (environ 5,6k phrases), de même que les ensembles de tests et de développement sont manuellement traduits en italien.

Un système de type LLPB-SMT est utilisé pour apprendre un système de compréhension du français sur le corpus MEDIA, et le sous-ensemble traduit de ce corpus est utilisé comme corpus parallèle pour apprendre un modèle de traduction à base de CRF. Ensuite l'approche CRFPB-SMT à base de transducteurs est évaluée séparément pour la traduction et la compréhension avant d'être utilisée dans le cadre d'un décodage conjoint traduction/compréhension.

Le taux d'erreur en concepts (*Concept Error Rate*, CER) est le critère d'évaluation retenu pour évaluer la tâche de compréhension. Le CER est l'équivalent du taux d'erreur en mots (WER), et peut être défini comme le rapport de la somme des concepts omis, insérés et substitués sur le nombre de concepts dans la référence. D'autre part le score BLEU (Papineni *et al.*, 2002) qui se base sur des comptes de n-grammes communs entre hypothèse et référence est retenu pour évaluer la tâche de traduction.

5.1 Evaluation des systèmes de traduction à base de segments pour une tâche de compréhension

La boîte à outils MOSES (Koehn *et al.*, 2007) a été utilisée pour apprendre un modèle LLPB-SMT pour la compréhension du français. Nos premières tentatives ont clairement montré des performances inférieures à celles d'un modèle CRF-SLU de référence (CER 23,2% après réglage des paramètres avec MERT pour le LLPB-SMT à comparer aux 12,9% pour CRF-SLU [1].).

Les améliorations progressives du modèle proposées dans la section 2 sont évaluées dans le tableau 1. L'utilisation de la contrainte de monotonie durant le décodage permet une réduction de 0,5% absolu. Convertir les données selon le formalisme BIO avant la phase d'apprentissage réduit le CER de façon significative de 2,4%. Enfin, optimiser le score CER à la place du score BLEU réduit le CER de 0,4% supplémentaire. Enfin, l'ajout d'une liste de villes à l'ensemble d'apprentissage avant réapprentissage du modèle LLPB-SMT répond au problème du traitement des mots hors-vocabulaire et permet une réduction finale de 0,8%.

Les résultats montrent qu'en dépit de réglages fins de l'approche LLPB-SMT, les approches à base de CRF obtiennent toujours les meilleures performances pour une tâche de compréhension (CER de 12,9% pour CRF-SLU vs. 18,3% pour LLPB-SMT).

Une analyse rapide du type d'erreur montre que les méthodes utilisant des CRF ont un haut niveau de suppressions comparativement aux autres types d'erreurs, tandis que la méthode LLPB-SMT présente un meilleur compromis entre les erreurs de suppression et d'insertion, et ce bien qu'elle aboutisse à un CER plus élevé. Un nombre important d'erreurs causées par le modèle LLPB-SMT pour la compréhension est dû à une mauvaise segmentation (le plus souvent une sur-segmentation) des phrases. Cette caractéristique des modèles LLPB-SMT mène à une distribution équilibrée d'erreurs entre les omissions, les insertions et les substitutions, alors que pour CRF-SLU un grand nombre d'erreurs venait des omissions.

1. Se référer à (Jabaian *et al.*, 2011) pour plus de détails sur le modèle CRF-SLU

Modèle	Sub	Om	Ins	CER
Initial	5,4	4,1	14,6	24,1
+MERT (BLEU)	5,6	8,4	9,2	23,2
+Décodage monotone	6,2	7,8	8,7	22,7
+Format BIO	5,7	9,3	5,3	20,3
MERT (CER)	5,3	9,2	4,6	19,1
Traitement de mots HV	5,8	7,4	5,1	**18,3**

TABLE 1: Les améliorations itératives du modèle LLPB-SMT pour la compréhension du français (CER%).

5.2 Evaluation des étiqueteurs sémantiques pour une tâche de traduction automatique

Afin de pouvoir évaluer notre proposition d'utiliser une approche CRF-SLU pour la traduction nous utilisons la partie traduite manuellement (du français vers l'italien) du corpus MEDIA comme corpus parallèle pour apprendre le modèle de traduction. L'outil GIZA++ (disponible avec MOSES) a été utilisé pour apprendre automatiquement un alignement mot à mot entre les corpus des deux langues et l'outil Wapiti (Lavergne *et al.*, 2010) a été utilisé pour apprendre les paramètres des modèles CRF.

Dans un premier temps, nous cherchons à apprendre un modèle CRF-SLU pour la traduction, en utilisant l'algorithme RPROP comme proposé dans la section 3. Des fonctions caractéristiques de type 4-grammes symétriques sur les observations et bi-grammes sur les étiquettes sont utilisées pour apprendre ce modèle. Les performances obtenues sont présentées dans le tableau 2. Les résultats montrent que la performance du modèle CRF-SLU (BLEU de 42,5) est significativement moins bonne que la performance obtenue par la méthode LLPB-SMT classique utilisant MOSES avec des paramètres de base (47,2) [2].

Afin d'avoir une comparaison juste entre les deux méthodes, nous cherchons à évaluer l'approche LLPB-SMT dans les mêmes conditions que l'approche CRF-SLU. La méthode LLPB-SMT utilise un modèle de réordonnancement alors que CRF-SLU, dédié à l'étiquetage séquentiel, ne comprend pas un tel modèle. Pour cela nous rajoutons une contrainte de monotonie dans le décodage pour l'approche LLPB-SMT empêchant tout réordonnancement. Il est aussi important de mentionner que l'approche LLPB-SMT utilise un modèle de langage pour sélectionner la meilleure traduction. Les performances du modèle LLPB-SMT de référence sont obtenues en utilisant un modèle de langage tri-grammes (utilisé généralement dans les systèmes de traduction). Cependant la complexité algorithmique de l'approche CRF-SLU ne permet pas d'utiliser un tel modèle de langage sur les étiquettes.

Afin d'évaluer les approches CRF-SLU et LLPB-SMT dans les mêmes conditions, et vu qu'on ne peut pas augmenter la taille des fonctions caractéristiques du modèle CRF, nous proposons de dégrader l'approche LLPB-SMT et de réévaluer sa performance en utilisant un modèle de langage de type bi-grammes.

Par ailleurs, en observant les sorties du modèle CRF-SLU, nous remarquons que les mots

2. Se référer à (Jabaian *et al.*, 2011) pour plus de détails sur le modèle LLPB-SMT.

	CRF-SLU	LLPB-SMT
référence	42,5	47,2
décodage monotone	42,5	46,3
bi-grammes	42,5	46,0
traitement de mots HV	**43,5**	**46,0**

TABLE 2: Comparaison entre les modèles LLPB-SMT et CRF-SLU pour la traduction de l'italien vers le français (BLEU %).

inconnus (hors-vocabulaire) dans le test ont été traduits par d'autres mots du corpus cible selon le contexte général de la phrase, contrairement à l'approche LLPB-SMT qui a tendance à projeter les mots hors-vocabulaire tels qu'ils sont dans la phrase traduite. Ces mots, étant dans la plupart des cas des noms de ville ou de lieux, leur traduction ne change pas d'une langue à l'autre, et donc leur projection dans la sortie traduite est avantageuse pour les modèles LLPB-SMT. Pour cela nous proposons un pré-traitement des mots inconnus dans la phrase source permettant de les récupérer en sortie dans l'approche CRF-SLU.

Les résultats présentés dans le tableau 2 montrent que le décodage monotone dégrade la performance du modèle LLPB-SMT de 0,91% absolu. L'utilisation d'un modèle de langage bi-grammes augmente la perte de 0,3% supplémentaire. Le traitement des mots hors-vocabulaire permet au modèle CRF-SLU de récupérer 1,0% de score BLEU par rapport au modèle CRF-SLU de référence. On remarque que malgré la dégradation du modèle LLPB-SMT et les améliorations du modèle CRF-SLU, la performance de ce dernier reste inférieure à celle du modèle LLPB-SMT (43,5% pour les CRF vs. 46,0% pour LLPB-SMT).

5.3 Evaluation des systèmes à base de transducteurs CRFPB-SMT pour la traduction et la compréhension

Un modèle de traduction CRFPB-SMT à base de transducteurs valués par des CRF pour la traduction a été construit comme décrit dans la section 3. Ce modèle a été construit à partir de l'outil n-code (Crego *et al.*, 2011), implémenté pour apprendre des modèles de traduction à base de n-grammes (Mariño *et al.*, 2006).

Cet outil utilise la bibliothèque OpenFst (Allauzen *et al.*, 2007) pour construire un graphe de traduction qui est la composition de plusieurs transducteurs. La différence entre le modèle implémenté par cet outil et le modèle qu'on cherche à développer réside dans les poids du modèle de traduction. Nous adaptons donc cet outil pour interroger les paramètres d'un modèle CRF afin d'estimer les probabilités de traduction et ensuite nous appliquons une normalisation des scores de probabilité obtenus par ce modèle sur les différents chemins du graphe (comme cela a été proposé dans (Lavergne *et al.*, 2011)).

Dans n-code le modèle de réordonnancement, proposé par (Crego et Mariño, 2006), est fondé sur une approche à base de règles apprises automatiquement sur les données d'entrainement. Cette approche nécessite un étiquetage grammatical des phrases source et un alignement au niveau des mots entre les phrases source et les phrases cible pour apprendre le modèle λ_R. Nous avons utiliser les outils TreeTagger (Schmid, 1994) pour obtenir l'étiquetage grammatical

Modèle	Langue	BLEU
LLPB-SMT	IT –> FR	47,2
CRF-SLU		43,5
CRFPB-SMT		44,1

TABLE 3: Comparaison entre les différentes approches (LLPB-SMT, CRF-SLU, CRFPB-SMT) pour la traduction de l'italien vers le français.

et GIZA++ pour l'alignement en mots. Le modèle de langage utilisé dans nos expériences est un modèle tri-grammes appris sur la partie cible de notre corpus d'apprentissage à l'aide de l'outil SRILM (Stolcke, 2002).

Le tableau 3 présente une comparaison entre trois modèles : le modèle CRFPB-SMT, le modèle LLPB-SMT (de référence) et le modèle CRF-SLU de base (présenté dans la section précédente). Les résultats présentés dans ce tableau montrent que l'approche CRFPB-SMT à base de transducteurs donne des performances inférieures mais comparables à celles obtenues par l'approche LLPB-SMT.

Malgré une dégradation de 3,1 points absolu, ces performances restent assez élevées en valeur pour une tâche de traduction (malgré un ensemble d'apprentissage de taille réduite), ce qui s'explique dans notre contexte par le vocabulaire limité du domaine. Cette différence de performance est comparable à celle observée par le LIMSI (Lavergne *et al.*, 2010) (en ne considérant que l'utilisation des paramètres de base).

D'autre part les résultats montrent que l'utilisation de graphes d'hypothèses dans CRFPB-SMT est doublement avantageuse par rapport à l'utilisation d'une approche CRF simple ; en plus du fait qu'elle permette de traiter des graphes en entrées, cette approche permet d'emblée d'augmenter la performance du système d'environ 1 point absolu.

Le mécanisme utilisé pour obtenir des graphe de traduction peut être utilisé d'une manière similaire pour la compréhension. Dans un premier temps, le graphe d'hypothèse de concepts est obtenu en composant tous les modèles $\lambda_S \circ \lambda_R \circ \lambda_T \circ \lambda_F \circ \lambda_L$ comme cela a été proposé pour la traduction. Cette approche donne un CER de 15,3%, bien moins bon que l'approche CRF-SLU de base (12,9%).

Afin de prendre les spécificités de la compréhension (qui ne comprend pas de modèle de réordonnancement, ni de modèle de langage cible final), nous proposons d'obtenir le graphe de sorties en combinant uniquement les modèles $\lambda_S \circ \lambda_F$. Cela nous a permis d'augmenter la performance de cette approche de 2,2% absolu (15,3% vs 13,1%) permettant de retrouver quasiment les mêmes performances qu'avec CRF-SLU (13,1% vs 12.9%). Une comparaison entre les performances des différentes versions est donnée dans le tableau 4. Par la suite, CRFPB-SMT simplifié est utilisé pour toutes les expériences de compréhension.

5.4 Décodage conjoint dans un scénario de compréhension multilingue

Un décodage conjoint pour la traduction et la compréhension a été appliqué comme nous l'avons proposé dans la section 4. Ce décodage consiste à transmettre le graphe de traduction

Modèle	Sub	Del	Ins	CER
CRF-SLU	3,1	8,1	1,8	12,9
CRFPB-SMT (complet) ($\lambda_S \circ \lambda_R \circ \lambda_T \circ \lambda_F \circ \lambda_L$)	4,2	8,8	2,3	15,3
CRFPB-SMT (simplifié) ($\lambda_S \circ \lambda_T \circ \lambda_F$)	3,5	7,6	2,0	13,1

TABLE 4: Evaluation des approches basées sur les CRF pour la compréhension du français.

en entrée du module de compréhension (incluant les scores pondérés relatifs à la traduction) et ensuite récupérer en sortie un graphe de compréhension qui intègre les scores de traduction et de compréhension. Ce décodage permettra d'étiqueter des phrases en italien en combinant un système de traduction italien vers français et un système de compréhension du français

Pour cela nous avons adapté l'accepteur du modèle de compréhension du français (donné dans la dernière ligne du tableau 4 décrit dans 5.3) pour prendre des graphes en entrée (au lieu d'une hypothèse unique). Ce transducteur génère un graphe valué de compréhension qui prend en compte les scores de traduction.

Au moment du décodage les deux scores (traduction et compréhension) sont pris en considération. Dans un premier temps nous proposons que le score final pour chaque chemin du graphe soit l'addition simple du score de traduction et du score de compréhension sur ce chemin [3]. Le meilleur chemin est ensuite sélectionné parmi l'ensemble des chemins possibles dans le graphe. Ce chemin représente donc un décodage conjoint entre la traduction et la compréhension (marginalisation de la variabilité aléatoire liée à la traduction intermédiaire).

Afin de pouvoir se positionner par rapport à l'état de l'art, nous proposons de réaliser le décodage conjoint selon deux modes : le système de traduction utilisé est un modèle LLPB-SMT (en utilisant la boîte à outils MOSES) dans le premier et un CRFPB-SMT (comme décrit dans 5.3) dans le second. Dans les deux cas les performances du décodage conjoint sont comparées avec ou sans prise en compte du graphe d'hypothèses complet. Dans un premier cas, le meilleur chemin (1-best) du graphe de traduction est fourni en entrée du système de compréhension. Dans un second cas, l'oracle du graphe de traduction est donné en entrée au module de compréhension. Les scores oracle représentent une évaluation fondée sur le chemin du graphe qui se rapproche le plus de la référence de la traduction. Il est alors possible de mesurer l'impact de la qualité de la traduction sur les performances de compréhension.

Le résultat de cette comparaison est donné dans la tableau 5. Nous avons aussi calculé les scores oracle (pour la traduction et la compréhension) sur les sorties des différents couplages de modules, et nous avons calculé le score BLEU sur la traduction sélectionnée par le décodage conjoint (dernière colonne du tableau 5).

La première ligne de ce tableau constitue la combinaison de référence (sans l'utilisation de graphe) dans laquelle la sortie de MOSES est donnée en entrée d'un modèle CRF. Les résultats montrent que le graphe de traduction permet d'améliorer la performance du système par rapport au système de 1-meilleure traduction (CER 19,7% vs. 19,9% pour LLPB-SMT et 21,3 vs. 21,7 pour CRF). L'utilisation d'un graphe de traduction donne aussi des meilleurs performances par rapport à la combinaison avec son oracle (CER 19,7% vs. 19,8% pour

3. Une expérience préliminaire pour mesurer l'impact de la pondération des scores est présentée dans (Jabaian, 2012).

Traduction			Compréhension (CRF)		
Modèle	**Sortie**	**BLEU/Oracle**	**Entrée**	**CER/Oracle**	**BLEU**
LLPB-SMT	1-best	47,2/47,2	1-best	19,9/19,9	47,2
	graphe	46,9/47,9	1-best(graphe)	19,9/19,4	46,9
	graphe	46,9/47,9	oracle(graphe)	19,8/19,3	47,9
	graphe	46,9/47,9	graphe	**19,7/19,1**	46,3
CRFPB-SMT	graphe	44,1/44,9	1-best(graphe)	21,7/21,1	44,1
	graphe	44,1/44,9	oracle(graphe)	21,6/21,1	44,9
	graphe	44,1/44,9	graphe	**21,3/20,6**	43,9

TABLE 5: Evaluation des différents configurations de compréhension multilingue françcais-italien, variant selon le type d'information transmise entre les 2 étapes (1-best, oracle ou graphe).

LLPB-SMT et 21,3 vs. 21,6 pour CRF). La différence entre la performance obtenue par le décodage conjoint en utilisant un modèle LLPB-SMT pour la traduction et celle obtenue en utilisant un modèle CRF (CER 19,7,8% vs. 21,3%) peut être expliquée par la différence entre la performance de ces deux modèles (BLEU 46,9% vs 44,1%).

Il est important de mentionner que seuls les couplages prenant des graphes en entrée de la compréhension permettent de sélectionner la traduction en fonction de l'étiquetage qui lui sera appliqué. Dans les autres cas la sélection de la traduction se fait indépendamment. On remarque que le score BLEU de la traduction sélectionnée par le décodage conjoint est moins bon que celui par la meilleure traduction (46,3 vs. 47,2 pour LLPB-SMT et 43,9 vs. 44,1 pour CRF) malgré le fait que le premier est plus performant en CER. Cela montre l'intérêt de la méthode conjointe à base de graphes qui permet de sélectionner la traduction qui pourra être étiquetée de la meilleure façon possible.

Les scores oracle montrent que la meilleure hypothèse sélectionnée lors du décodage n'est pas forcément la plus proche de la référence parmi les hypothèses du graphe. Cependant, ce résultat est encourageant du fait que la performance du système peut être encore améliorée en ajustant les poids des modèles vu que des meilleures hypothèses se trouvent dans le graphe. Un décodage optimal permettra d'améliorer le CER de 0,6% absolu pour un décodage en composant avec un graphe LLPB-SMT pour la traduction (19,7% vs 19,1%) et de 0,7% absolu pour un décodage en composant avec un graphe CRF pour la traduction (21,3% vs 20,6%).

6 Conclusion

Dans cet article nous avons évalué et comparé des approches statistique à la fois pour la compréhension de la parole et pour la traduction automatique. Nous avons observé que l'approche discriminante CRF reste la meilleure approche pour la compréhension de la parole, malgré toutes les adaptations de l'approche LLPB-SMT pour la tâche. Une approche de type CRF pour la traduction a plusieurs limites et les performances de cette approche peuvent être améliorées par un modèle à base de transducteurs permettant l'intégration de traitements adaptés (réordonnancement, segmentation, modèle de langage cible).

Nous avons alors pu proposer et évaluer une approche de décodage conjoint entre la traduction et la compréhension dans le contexte d'un système de compréhension de la parole multilingue. Nous avons montré qu'avec un tel décodage nous pouvons obtenir de bonnes performances tout en proposant un système homogène sur les deux tâches sous-jacentes.

Dans le contexte d'un système de dialogue homme-machine complet un décodage conjoint entre la reconnaissance de la parole et la traduction pourra être ajouté. Dans ce cas un graphe de reconnaissance devra être composé avec un graphe de traduction. Cette composition permettra au système de reconnaissance de transmettre des informations plus riches au système de compréhension et le système de compréhension transmettra à son tour des informations riches au gestionnaire de dialogue ce qui influencera positivement la performance globale du système.

Références

ALLAUZEN, C., RILEY, M., SCHALKWYK, J., SKUT, W. et MOHRI, M. (2007). OpenFst : A general and efficient weighted finite-state transducer library. *In CIAA*.

ANOOP DEORAS, G. T. et HAKKANI-TUR, D. (2012). Joint decoding for speech recognition and semantic tagging. *In INTERSPEECH*.

BONNEAU-MAYNARD, H., ROSSET, S., AYACHE, C., KUHN, A. et MOSTEFA, D. (2005). Semantic annotation of the french media dialog corpus. *In EUROSPEECH*.

BROWN, P. F., PIETRA, S. D., PIETRA, V. J. D. et MERCER, R. L. (1993). The mathematics of statistical machine translation : Parameter estimation. *Computational Linguistics*, 19(2):263–311.

CREGO, J. M. et MARIÑO, J. B. (2006). Improving statistical mt by coupling reordering and decoding. *Machine Translation*, 20(3):199–215.

CREGO, J. M., YVON, F. et MARIÑO, J. B. (2011). Ncode : an open source bilingual n-gram smt toolkit. *The Prague Bulletin of Mathematical Linguistics*, 96:49–58.

GASCÓ I MORA, G. et SÁNCHEZ PEIRÓ, J. (2007). Part-of-speech tagging based on machine translation techniques. *Pattern Recognition and Image Analysis*, pages 257–264.

HAHN, S., DINARELLI, M., RAYMOND, C., LEFÈVRE, F., LEHNEN, P., DE MORI, R., MOSCHITTI, A., NEY, H. et RICCARDI, G. (2010). Comparing stochastic approaches to spoken language understanding in multiple languages. *IEEE Transactions in Audio, Speech and Language Processing*, 19(6):1569–1583.

HAKKANI-TÜR, D. Z., B., F., RICCARDI, G. et TÜR, G. (2006). Beyond asr 1-best : Using word confusion networks in spoken language understanding. *Computer Speech and Language*, pages 495–514.

JABAIAN, B. (2012). *Systèmes de compréhension et de traduction de la parole : vers une approche unifiée dans le cadre de la portabilité multilingue des systèmes de dialogue*. Thèse de doctorat, CERI - Université d'Avignon, Avignon.

JABAIAN, B., BESACIER, L. et LEFÈVRE, F. (2010). Investigating multiple approaches for slu portability to a new language. *In INTERSPEECH*.

JABAIAN, B., BESACIER, L. et LEFÈVRE, F. (2011). Comparaison et combinaison d'approches pour la portabilité vers une nouvelle langue d'un système de compréhension de l'oral. *In TALN*.

KOEHN, P., HOANG, H., BIRCH, A., CALLISON-BURCH, C., FEDERICO, M., BERTOLDI, N., COWAN, B., SHEN, W., MORAN, C., ZENS, R. *et al.* (2007). Moses : Open source toolkit for statistical machine translation. *In ACL*.

KOEHN, P., OCH, F. et MARCU, D. (2003). Statistical phrase-based translation. *In HLT-NAACL*.

KUMAR, S. et BYRNE, W. (2003). A weighted finite state transducer implementation of the alignment template model for statistical machine translation. *In HLT-NAACL*.

LAFFERTY, J., MCCALLUM, A. et PEREIRA, F. (2001). Conditional random fields : Probabilistic models for segmenting and labeling sequence data. *In ICML*.

LAVERGNE, T., CAPPÉ, O. et YVON, F. (2010). Practical very large scale CRFs. *In ACL*.

LAVERGNE, T., CREGO, J. M., ALLAUZEN, A. et YVON, F. (2011). From n-gram-based to crf-based translation models. *In WSMT*.

LIANG, P., TASKAR, B. et KLEIN, D. (2006). Alignment by agreement. *In HLT-NAACL*.

MACHEREY, K., BENDER, O. et NEY, H. (2009). Application of statistical machine translation approaches to spoken language understanding. *In IEEE ICASSP*.

MACHEREY, K., OCH, F. J. et NEY, H. (2001). Natural language understanding using statistical machine translation. *In INTERSPEECH*.

MARIÑO, J. B., BANCHS, R. E., CREGO, J. M., de GISPERT, A., LAMBERT, P., FONOLLOSA, J. A. R. et COSTA-JUSSÀ, M. R. (2006). N-gram-based machine translation. *Computational Linguistic*, 32(4):527–549.

OCH, F. (2003). Minimum error rate training in statistical machine translation. *In ACL*.

OCH, F. J. et NEY, H. (2002). Discriminative training and maximum entropy models for statistical machine translation. *In ACL*.

PAPINENI, K., ROUKOS, S., WARD, T. et ZHU, W. (2002). Bleu : a method for automatic evaluation of machine translation. *In ACL*.

RAMA, T., SINGH, A. et KOLACHINA, S. (2009). Modeling letter-to-phoneme conversion as a phrase based statistical machine translation problem with minimum error rate training. *In HLT-NAACL*.

RAMSHAW, L. et MARCUS, M. (1995). Text chunking using transformation-based learning. *In The Workshop on Very Large Corpora*.

RIEDMILLER, M. et BRAUN, H. (1993). A direct adaptive method for faster backpropagation learning : The RPROP algorithm. *In ICNN*.

SCHMID, H. (1994). Probabilistic part-of-speech tagging using decision trees. *In NMLP*.

SERVAN, C., RAYMOND, C., B., F. et NOCERA, P. (2006). Conceptual decoding from word lattices : application to the spoken dialogue corpus MEDIA. *In INTERSPEECH*.

STOLCKE, A. (2002). Srilm-an extensible language modeling toolkit. *In ICASSP*.

TÜR, G., WRIGHT, J. H., GORIN, A. L., RICCARDI, G. et HAKKANI-TÜR, D. Z. (2002). Improving spoken language understanding using word confusion networks. *In INTERSPEECH*.

TURIAN, J. P., WELLINGTON, B. et MELAMED, I. D. (2006). Scalable discriminative learning for natural language parsing and translation. *In NIPS*.

