

## Alignement automatique pour la compréhension littérale de l’oral par approche segmentale

Stéphane Huet et Fabrice Lefèvre  
Université d’Avignon, LIA-CERI, France  
{stephane.huet,fabrice.lefevre}@univ-avignon.fr

**Résumé.** Les approches statistiques les plus performantes actuellement pour la compréhension automatique du langage naturel nécessitent une annotation segmentale des données d’entraînement. Nous étudions dans cet article une alternative permettant d’obtenir de façon non-supervisée un alignement segmental d’unités conceptuelles sur les mots. L’impact de l’alignement automatique sur les performances du système de compréhension est évalué sur une tâche de dialogue oral.

**Abstract.** Most recent efficient statistical approaches for language understanding require a segmental annotation of the training data. In this paper we study an alternative that obtains a segmental alignment of conceptual units with words in an unsupervised way. The impact of the automatic alignment on the understanding system performance is evaluated on a spoken dialogue task.

**Mots-clés :** Alignement non-supervisé, compréhension de la parole.

**Keywords:** Unsupervised alignment, spoken language understanding.

### 1 Introduction

Une des toutes premières étapes pour construire un système de compréhension de l’oral pour les systèmes de dialogue est l’extraction de concepts littéraux à partir d’une séquence de mots issue d’un système de reconnaissance de la parole. Pour résoudre ce problème d’étiquetage en concepts, un certain nombre de techniques sont disponibles. Ces techniques reposent sur des modèles classiques maintenant, qui peuvent être génératifs ou discriminants, parmi lesquels on peut citer : les modèles de Markov cachés, les transducteurs à états finis, les modèles de Markov à entropie maximale, les machines à vecteurs supports, les réseaux bayésiens dynamiques (*Dynamic Bayesian Networks*, DBN) ou encore les champs de Markov conditionnels (*Conditional Markov Random Fields*, CRF (Lafferty *et al.*, 2001)). Dans (Hahn *et al.*, 2010), il est montré que les CRF permettent d’obtenir les meilleures performances sur la tâche MEDIA (Bonneau Maynard *et al.*, 2008) en français, mais aussi sur deux corpus comparables en italien et en polonais. De même, la robustesse des CRF a pu être montrée en observant ses résultats sur la compréhension de transcriptions manuelles et automatiques.

Dans beaucoup d’approches, l’interprétation littérale se contente d’une relation lexique-concept ; c’est ainsi le cas du système PHOENIX (Ward, 1991) basé sur la détection de mots-clefs. L’approche segmentale fait une analyse plus fine en considérant la phrase comme une séquence de segments lors de son interprétation. Elle permet alors de relier correctement les différents niveaux d’analyse : lexicaux, syntaxiques et sémantiques. Toutefois, afin de simplifier la mise en œuvre, les segments ont été définis spécifiquement pour l’annotation conceptuelle et n’ont pas de relation imposée avec les unités syntaxiques (*chunks*, groupes syntaxiques...). Une autre raison est que l’objectif étant d’utiliser le module d’interprétation au sein de systèmes de dialogue oral, les données qui sont ici traitées sont fortement bruitées (langage naturel très spontané et agrammatical, erreurs dues à la reconnaissance automatique de la parole), ce qui perturbe fortement les analyseurs syntaxiques.

L’approche segmentale présente aussi l’intérêt de pouvoir découpler la détection d’une unité conceptuelle de l’estimation de sa valeur. La valeur correspond à la normalisation de la forme de surface associée au concept ; par exemple si au concept `temps-depart` est associé le segment « pas avant 11h », sa valeur est « matin ». De même pour « entre 8h et 12h » ou « dans la matinée ». L’estimation de la valeur nécessite donc un ancrage des concepts sur les mots de la phrase. Il est alors possible de traiter le problème de normalisation à partir de règles sous formes

d'expressions régulières ou à l'aide de modèles de langages spécifiques à chaque concept (ce qui permet alors une approche intégrée de l'interprétation (Lefèvre, 2007)). Dans le cas d'approches globales (c-à-d non segmentales), la détection des valeurs doit être associée aux classes conceptuelles à reconnaître, comme dans (Mairesse *et al.*, 2009). Le processus est alourdi considérablement et n'est plus envisageable que lorsque le nombre de valeurs possibles est limité.

L'inconvénient majeur de l'approche segmentale est bien sûr son coût : associer des étiquettes conceptuelles à la transcription d'un dialogue est une tâche déjà complexe et sa complexité est augmentée par la délimitation précise du support (segment lexical) correspondant à chaque unité. La campagne d'évaluation des systèmes de compréhension automatique MEDIA a été la première occasion de réaliser et de rendre disponible un corpus de taille conséquente possédant une annotation segmentale ; toutefois la difficulté reste entière à chaque fois qu'il faut développer un corpus pour une nouvelle tâche.

Nous proposons dans cette étude un procédé permettant de réduire l'effort nécessaire à la production de corpus d'entraînement pour obtenir des modèles d'annotation conceptuelle segmentale. En faisant l'hypothèse que les unités conceptuelles associées à une phrase ont été détectées automatiquement ou fournies par un expert, nous étudions comment les associer à leur support lexical dans la phrase sans connaissances *a priori*. Dans ce cadre, des techniques d'alignement de la traduction automatiques sont utilisées pour obtenir de manière non-supervisée une annotation conceptuelle segmentale. Si l'idée d'appliquer des méthodes issues de la traduction automatique pour construire des systèmes de compréhension n'est pas nouvelle (Hahn *et al.*, 2010), l'apport principal de notre étude est de montrer l'intérêt des méthodes d'alignement pour le problème de l'alignement segmental.

Nous présentons dans l'article les adaptations nécessaires à l'application de techniques d'alignement dans ce nouveau contexte. Elles sont peu nombreuses afin de garder la plus grande généralité à l'approche et aussi de pouvoir bénéficier d'outils logiciels déjà disponibles. Nous évaluons la qualité des alignements par l'approche non-supervisée par rapport aux annotations de référence dans deux situations d'intérêt : l'une où l'ordre des concepts dans la phrase est connu *a priori* et l'autre où ce n'est pas le cas. Finalement, la véritable évaluation de l'approche consiste à utiliser des alignements produits pour entraîner des systèmes de compréhension à base de CRF afin de mesurer leur impact sur les performances finales du système.

Après un rappel des principes du décodage conceptuel segmental dans la section 2, l'alignement automatique de corpus parallèles sera présenté dans la section 3 avec les particularités liées à l'alignement de concepts sémantiques. La présentation des expériences et les commentaires des résultats seront donnés en section 4.

## 2 Décodage conceptuel segmental

Si l'interprétation littérale peut être vue comme une traduction de la langue naturelle vers l'ensemble des séquences d'étiquettes sémantiques, alors les méthodes et les modèles de traduction peuvent être utilisés. Considérant le fait que le nombre de concepts est en général bien plus faible que la taille du vocabulaire de mots, ce type particulier de traduction peut aussi être considéré simplement comme un problème de classification dans lequel les constituants conceptuels représentent les classes à reconnaître. Il est donc possible de réaliser l'interprétation par le biais de méthodes et modèles de classification. Les approches discriminantes modélisent la distribution de probabilités conditionnelles d'une séquence de constituants sémantiques (concepts)  $c_1 \dots c_n$  considérant une séquence de mots  $w_1 \dots w_T : P(c_1^n | w_1^T)$ . Dans l'approche générative, c'est la probabilité jointe  $P(c_1^n, w_1^T)$  qui sera modélisée, permettant de calculer des inférences pour la prédiction de données ou l'apprentissage des paramètres.

Les modèles génératifs (de type modèles de Markov cachés) ont été introduits en premier pour traiter le problème de compréhension par des approches probabilistes (Levin & Pieraccini, 1995). Des variantes ont été proposées plus récemment offrant plus de degrés de liberté dans la modélisation (par exemple (He & Young, 2005; Lefèvre, 2007)). Depuis, les modèles log-linéaires ont assez clairement montré leur supériorité sur des tâches d'étiquetage séquentiel (Hahn *et al.*, 2010). Plusieurs variantes existent, se distinguant par les hypothèses d'indépendance conditionnelle entre les variables et par l'étape de normalisation. Les CRF (Lafferty *et al.*, 2001) représentent des chaînes linéaires de variables aléatoires indépendantes, toutes conditionnées sur la séquence entière étiquetée, et la normalisation est globale à la séquence.

Certaines approches génératives comme les DBN permettent l'inférence dans des modèles multi-niveaux (Lefèvre, 2007) et prennent donc en compte intrinsèquement la segmentation. Pour les modèles ne permettant pas la



FIGURE 1 – Exemple d'alignement des mots avec leurs concepts sémantiques.

représentation à plusieurs niveaux, comme les CRF, il convient de représenter la notion de segment directement au niveau des étiquettes à classer. Le formalisme BIO est utilisé : B est ajouté aux étiquettes débutant un segment, I à l'intérieur d'un segment et O aux segments hors-domaine (si ceux-ci ne sont pas déjà traités par une étiquette spécifique). Dans le cas de la figure 1, la séquence de concepts devient : B-cmd-tache I-cmd-tache I-cmd-tache B-null I-null B-loc-ville I-loc-ville I-loc-ville I-loc-ville I-loc-ville I-loc-ville B-tps-date I-tps-date B-tps-date I-tps-date I-tps-date.

### 3 Alignement de concepts sémantiques

L'alignement automatique est une des problématiques majeures du domaine de la traduction automatique. Les alignements mot-à-mot sont ainsi utilisés pour construire des tables de segments qui sont au cœur de nombreux systèmes de traduction statistiques actuels (Koehn *et al.*, 2007). L'alignement en traduction consiste à trouver quels mots correspondent dans deux phrases qui sont la traduction l'une de l'autre. Ce processus est confronté à plusieurs difficultés : certains mots ne sont associés à aucun mot dans la traduction ; d'autres au contraire sont traduits par plusieurs mots ; les règles syntaxiques peuvent enfin différer suivant les langues, les mots en relation pouvant ainsi être à des positions très différentes d'une langue à une autre.

Plusieurs modèles statistiques ont été proposés pour aligner deux phrases (Brown *et al.*, 1993). Un de leurs grands intérêts est qu'ils sont construits à partir de corpus parallèles alignés au niveau des phrases, sans aucune annotation manuelle au niveau des mots. Formellement, à partir d'une phrase  $S = s_1 \dots s_m$  exprimée dans une langue source et de sa traduction  $T = t_1 \dots t_n$ , un alignement  $A = a_1 \dots a_m$  de type IBM revient à connecter chaque mot de  $S$  à un mot de  $T$  ( $a_j \in \{1, \dots, n\}$ ) ou au mot vide ( $a_j = 0$ ), ce dernier rendant compte des mots cibles non traduits. Les modèles statistiques IBM permettent d'évaluer la probabilité de traduction de  $S$  vers  $T$  en estimant  $P(S, A|T)$ . Le meilleur alignement  $\hat{A}$  peut être déterminé à partir de ce critère à l'aide d'un algorithme de Viterbi :  $\hat{A} = \operatorname{argmax}_A P(S, A|T)$ .

Les différents modèles IBM diffèrent suivant leur niveau de complexité. IBM1 fait des hypothèses fortes quant à l'indépendance entre les alignements et se limite à l'évaluation des probabilités de transfert  $P(s_i|t_j)$ . Le modèle HMM (Vogel *et al.*, 1996), qui est une amélioration du modèle IBM2, prend en compte un nouveau paramètre  $P(a_j|a_{j-1}, n)$  qui suppose une dépendance d'ordre 1 entre les variables d'alignement. Les modèles suivants (IBM3 à IBM5) introduisent la notion de distorsion, qui mesure la probabilité de réordonnement des mots de  $T$  par rapport à la position des mots  $S$  avec lesquels ils sont alignés, et celle de fertilité, qui détermine le nombre moyen de mots sources qui sont alignés dans une même phrase avec un mot cible donné. Afin d'améliorer la qualité des alignements, les modèles IBM sont généralement appliqués dans les deux directions de traduction. On opère ensuite une intersection entre les deux alignements ainsi obtenus, en étendant les alignements aux frontières des groupes de mots déjà connectés suivant des heuristiques (Och *et al.*, 1999).

Si l'on dispose d'une méthode capable de repérer automatiquement les concepts contenus dans un tour de parole, l'annotation segmentale peut être obtenue au moyen d'un alignement entre les mots du tour du parole  $S = w_1^T$  et les concepts  $T = c_1^n$  qui ont été détectés (Fig. 1). Les concepts générés suivent idéalement le même ordre que les mots avec lesquels ils doivent être alignés. Un cadre plus réaliste consiste toutefois à considérer que les concepts sont produits sous forme de sacs plutôt que de séquences ordonnées.

Les méthodes statistiques conçues pour la traduction automatique sont pertinentes dans notre cadre applicatif, en considérant que la langue cible est celle des concepts. Il existe toutefois des différences notables entre les deux domaines d'application. Tout d'abord, chaque mot ne peut être au plus aligné qu'à un concept, alors qu'un concept est aligné à au moins un mot. En conséquence, la fertilité des mots ne peut être que 1 dans le cas d'un alignement des concepts vers les mots et celle des concepts est supérieure ou égale à 1 dans la direction opposée. Par construction des modèles IBM, l'alignement des concepts vers les mots ne permet d'aligner qu'un mot au plus

par concept, ce qui empêche d’avoir un nombre d’alignements satisfaisants pour cette direction.

Une autre différence notable avec la traduction réside dans l’absence du mot vide dans le cas de l’alignement avec les concepts sémantiques, chaque mot de la séquence à aligner devant être aligné à au moins un concept. Le rôle sémantique `null` attribué aux mots qui ne sont pas associés à un concept pertinent dans le cadre de la compréhension joue un rôle similaire au mot vide. Toutefois, des expériences préliminaires sur l’alignement ont montré qu’il était préférable d’introduire l’étiquette `null` dans la liste des concepts pour aider les modèles d’alignement à reconnaître que certains mots ne sont pas associés à des concepts pertinents.

Enfin, une dernière différence concerne la notion de séquentialité. En effet, alors que l’ordre des mots d’une langue n’est pas aléatoire et suit une certaine logique propre aux règles syntaxiques de la langue, cela n’est pas le cas lorsque l’on doit aligner une séquence de mots avec un sac de concepts. Or, les modèles HMM et IBM2 à IBM5 ont des paramètres qui supposent l’influence de la position du mot source correspondant ou celles des traductions des mots cibles adjacents sur la position d’un mot cible. Le caractère aléatoire des positions des étiquettes conceptuelles est ainsi de nature à perturber ces types de modèles, contrairement à IBM1.

## 4 Expériences et résultats

### 4.1 Cadre expérimental

Les méthodes employées sont évaluées sur le corpus MEDIA (Bonneau Maynard *et al.*, 2008). Les données se composent de dialogues homme-machine collectées à l’aide d’une procédure de magicien d’Oz dans le domaine de la négociation pour des services touristiques. Ce corpus, réalisé dans le cadre d’une tâche réaliste, est annoté par 145 concepts sémantiques différents et est constitué de données audio, accompagnées de leur transcription automatique et de leur référence. Le corpus est divisé en trois parties : un ensemble d’apprentissage (12 k tours de parole), un ensemble de développement (1,2 k) et un ensemble de test (3 k).

Les expériences menées sur les méthodes d’alignement ont été réalisées sur le corpus de développement au moyen de l’outil MGIZA++ (Gao & Vogel, 2008), une version multithreadée de GIZA++ qui présente l’avantage d’offrir des paramètres<sup>1</sup> permettant d’appliquer sur les corpus de développement et de test des modèles d’alignement précédemment appris. L’étape de décodage conceptuel a été évaluée quant à elle sur le corpus de test. Différentes configurations ont été testées : versions manuelle ou automatique de la transcription, prise en compte ou non des valeurs pour le calcul des erreurs. Plusieurs types d’ordonnement des concepts à aligner ont en outre été considérés : une idéale laissant telles quelles les séquences de concepts de la référence et deux autres plus réalistes triant les concepts dans l’ordre alphabétique ou de manière aléatoire, simulant ainsi des sacs de concepts.

L’ordre de grandeur du temps mis pour la réalisation de ces expériences est de quelques minutes pour l’alignement automatique des 12 k phrases, de quelques heures pour l’apprentissage des modèles CRF<sup>2</sup> et de quelques secondes pour le décodage du test.

### 4.2 Résultats expérimentaux

La qualité de l’alignement est estimée à l’aide de l’AER (*Alignment Error Rate*), une métrique souvent employée en traduction automatique (Och & Ney, 2000). Si  $H$  représente les alignements suggérés par la méthode automatique et  $R$  les alignements de la référence, l’AER est calculée par la relation<sup>3</sup> :

$$AER = 1 - \frac{2 \times |H \cap R|}{|H| + |R|} \quad (1)$$

Comme les alignements sont ensuite utilisés pour étiqueter et comme les positions des concepts à aligner ne correspondent pas dans toutes les versions du corpus à celui de la référence, nous avons considéré qu’un alignement était un couple  $(w_i, c_j)$  plutôt que  $(i, j)$ .

1. `previousa, previsoust, previousn...`

2. Wapiti a été utilisé dans nos expériences, <http://wapiti.limsi.fr/>.

3. L’égalité est simplifiée par rapport à celle donnée habituellement du fait qu’ici tous les alignements sont considérés comme sûrs dans la référence.

Le tableau 1 présente les résultats d'alignement mesurés sur le corpus de développement selon le mode choisi pour ordonner les concepts et selon la direction considérée d'alignement. Les trois premières lignes montrent le résultats obtenus à partir de la chaîne d'itérations des modèles IBM<sup>4</sup> utilisée pour construire les modèles de traduction par MOSES, le système à l'état de l'art le plus utilisé actuellement (Koehn *et al.*, 2007). Comme attendu, l'AER mesuré avec des modèles d'alignement dans le sens concept  $\rightarrow$  mot (deuxième ligne), qui ne peuvent associer qu'un mot maximum à chaque concept, est bien supérieur à celui obtenu avec des modèles construits dans le sens opposé (première ligne). L'utilisation de l'heuristique par défaut de MOSES (*grow-diag-final*) montre toutefois que la symétrisation des alignements obtenus dans les deux directions conduit à un gain en terme d'AER (troisième ligne). Les modèles IBM1, contrairement aux autres modèles, ne prennent pas en compte l'ordre de succession des mots sources et cibles dans leurs calculs de probabilités d'alignement, ce qui les rend intéressants pour traiter des sacs de concepts. Les résultats obtenus en appliquant des modèles IBM1 puis en symétrisant les alignements dans les deux sens (quatrième ligne) montrent que ces modèles conduisent finalement à des performances inférieures à IBM4 et même à HMM (dernière ligne), que ce soit pour des concepts triés selon les ordres alphabétique ou aléatoire (2 dernières colonnes).

	Séquentiel	Alphabétique	Aléatoire
mot $\rightarrow$ concept IBM4	14,4	29,2	28,6
concept $\rightarrow$ mot IBM4	40,9	51,6	49,0
symétrisé IBM4	<b>12,8</b>	<b>27,3</b>	<b>25,7</b>
symétrisé IBM1	28,2	33,2	33,1
symétrisé HMM	14,8	29,9	28,7

TABLE 1 – AER (%) sur le corpus de développement en variant les modèles d'alignement et leur direction.

Les résultats précédents montrent que les performances mesurées lorsque l'on considère les séquences de concepts de la référence sont fortement dégradées quand on est face à des sacs de concepts. Afin de limiter les perturbations des modèles IBM qui apprennent des paramètres sur la séquentialité, nous réordonnons les séquences après un premier alignement  $A_1$  produit par le modèle IBM4 symétrisé. Deux stratégies ont alors été considérées pour déterminer la nouvelle position de chaque concept  $c_j$  : l'une réalisant une simple moyenne des positions des mots  $w_i$  avec lesquels il est aligné suivant  $A_1$  (Tab. 2, deuxième colonne), l'autre obtenue en pondérant chaque position par les probabilité de transfert  $P(w_i|c_j)$  et  $P(c_j|w_i)$  déterminées par IBM4 (troisième colonne). L'utilisation des modèles d'alignement appris sur le corpus d'apprentissage ainsi réordonné montre une amélioration significative de l'AER, la diminution de l'AER étant plus importante encore en prenant en compte les probabilités de transfert. Cette phase de réordonnement peut être réitérée tant que les performances continuent à être améliorées en utilisant à l'itération  $i$  des corpus d'apprentissage et de développement réordonnés suivant les derniers modèles d'alignement  $A_i$  obtenus. En procédant ainsi jusqu'à l'itération 3 pour l'ordre alphabétique et jusqu'à l'itération 7 dans le cas aléatoire, l'AER atteint une valeur inférieure à 20% (dernière colonne). Il est à noter que le tri aléatoire conduit à de meilleurs résultats que l'ordre alphabétique du fait que les modèles IBM4 apprennent des probabilités de distorsion qui sont davantage biaisées lorsque l'on effectue un tri alphabétique, en voyant apparaître plus souvent les mêmes séquences de concepts alors que celles-ci ne sont pas celles de la référence.

	Initial	Réordonnement 1 <sup>ère</sup> itération		Réordonnement dernière itération
		simple	avec probabilités de transfert	avec probabilités de transfert
Alphabétique	27,3	22,2	21,0	<b>19,4</b>
Aléatoire	25,7	21,9	20,2	<b>18,5</b>

TABLE 2 – AER (%) sur le corpus de développement en variant la stratégie de réordonnement des concepts.

Pour la tâche de compréhension, des CRF sont entraînés à partir de corpus d'entraînement dont l'étiquetage est réalisé soit par des experts, soit par des méthodes automatiques d'alignement. Le critère de performance utilisé pour évaluer la compréhension est le taux d'erreur en concepts (*Concept Error Rate*, CER). Il s'obtient en faisant le ratio de la somme des concepts de l'hypothèse substitués, insérés ou omis et du nombre total de concepts présents dans l'annotation manuelle de référence, calculés après un alignement de Levenshtein entre les deux séquences

4. 5 itérations d'IBM1, 5 itérations d'HMM, 3 itérations d'IBM3 puis 3 itérations d'IBM4.

de concepts hypothèse et référence. Le concept `null` n'est pas pris en compte lors du calcul du score. À partir d'un système à l'état de l'art, les dégradations dues aux différentes conditions d'alignement sont rapportées dans le Tableau 3. On retiendra qu'en tenant compte des valeurs, l'augmentation du CER est au plus de 8,0 % (17,6 % à 25,6 %), que l'apport de l'ordre la ramène à 3,7 % (17,6 % à 21,3 %) et enfin qu'avec la transcription automatique la dégradation liée aux alignements automatiques est moins grande (resp. 5,8 et 2,0 %).

Erreurs en Concept (Concept+valeur)	Manuel	Séquentiel	Alphabétique	Aléatoire
Transcription manuelle	13,9 (17,6)	17,7 (21,3)	22,6 (26,4)	22,0 (25,6)
Transcription automatique (WER 31 %)	24,7 (29,8)	27,1 (31,8)	31,5 (36,4)	30,6 (35,6)

TABLE 3 – CER (%) du décodage conceptuel en variant la méthode d'alignement des données d'entraînement.

## Conclusion

Dans cette étude nous proposons une approche non-supervisée au problème de l'alignement des unités conceptuelles pour la compréhension automatique du langage naturel. La qualité de l'alignement obtenu, déjà bonne dans le cas général (< 20 % d'erreurs sur les associations mot-concept), est améliorée par la connaissance de l'ordre des unités à aligner (< 15 %). Lorsque l'on utilise des CRF appris sur des corpus où les mots sont alignés automatiquement avec des sacs de concepts, l'impact mesuré sur les erreurs d'annotation, de l'ordre de 8 %, est réduit à 6 % quand le texte analysé est transcrit automatiquement. Aussi nous pensons que le rapport coût/performance est plutôt favorable à la méthode proposée.

## Références

- BONNEAU MAYNARD H., DENIS A., BÉCHET F., DEVILLERS L., LEFÈVRE F., QUIGNARD M., ROSSET S. & VILLANEUVE J. (2008). MEDIA : évaluation de la compréhension dans les systèmes de dialogue. In *L'évaluation des technologies de traitement de la langue, les campagnes Technolangue*, p. 209–232. Hermès, Lavoisier.
- BROWN P. F., DELLA PIETRA S. A., DELLA PIETRA V. J. & MERCER R. L. (1993). The mathematics of statistical machine translation : Parameter estimation. *Computational Linguistics*, **19**(2), 263–311.
- GAO Q. & VOGEL S. (2008). Parallel implementations of word alignment tool. In *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, p. 49–57, Columbus, OH, USA.
- HAHN S., DINARELLI M., RAYMOND C., LEFÈVRE F., LEHEN P., DE MORI R., MOSCHITTI A., NEY H. & RICCARDI G. (2010). Comparing stochastic approaches to spoken language understanding in multiple languages. *IEEE Transactions on Audio, Speech and Language Processing (TASLP)*, **PP**, 1–15.
- HE Y. & YOUNG S. (2005). Spoken language understanding using the hidden vector state model. *Speech Communication*, **48**(3–4), 262–275.
- KOEHN P., HOANG H., BIRCH A., CALLISON-BURCH C., FEDERICO M., BERTOLDI N., COWAN B., SHEN W., MORAN C., ZENS R., DYER C., BOJAR O., CONSTANTIN A. & HERBST E. (2007). Moses : Open source toolkit for statistical machine translation. In *Proceedings of ACL, Companion Volume*, p. 177–180, Prague, Czech Republic.
- LAFFERTY J., MCCALLUM A. & PEREIRA F. (2001). Conditional random fields : Probabilistic models for segmenting and labeling sequence data. In *Proceedings of ICML*, p. 282–289, Williamstown, MA, USA.
- LEFÈVRE F. (2007). Dynamic bayesian networks and discriminative classifiers for multi-stage semantic interpretation. In *Proceedings of ICASSP*, Honolulu, Hawaii.
- LEVIN E. & PIERACCINI R. (1995). Concept-based spontaneous speech understanding system. In *Proceedings of Eurospeech*, p. 555–558, Madrid, Spain.
- MAIRESSE F., GAŠIĆ M., JURČIČEK F., KEIZER S., THOMSON B., YU K. & YOUNG S. (2009). Spoken language understanding from unaligned data using discriminative classification models. In *Proceedings of ICASSP*, Taipei, Taiwan.
- OCH F. J. & NEY H. (2000). A comparison of alignment models for statistical machine translation. In *Proceedings of Coling*, volume 2, p. 1086–1090, Saarbrücken, Germany.
- OCH F. J., TILLMANN C. & NEY H. (1999). Improved alignment models for statistical machine translation. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, p. 20–28, College Park, MD, USA.
- VOGEL S., NEY H. & TILLMANN C. (1996). HMM-based word alignment in statistical translation. In *Proceedings of Coling*, volume 2, p. 836–841, Copenhagen, Denmark.
- WARD W. (1991). Understanding Spontaneous Speech. In *Proceedings of ICASSP*, p. 365–368, Toronto, Canada.