



Projection Interlingue d'Étiquettes pour l'Annotation Sémantique Non Supervisée

Othman Zennaki, Nasredine Semmar, Laurent Besacier

► **To cite this version:**

Othman Zennaki, Nasredine Semmar, Laurent Besacier. Projection Interlingue d'Étiquettes pour l'Annotation Sémantique Non Supervisée. TALN 2016, Jul 2016, Paris, France. Actes de la conférence conjointe JEP-TALN-RECITAL. <hal-01350117>

HAL Id: hal-01350117

<https://hal.archives-ouvertes.fr/hal-01350117>

Submitted on 29 Jul 2016

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.

Projection Interlingue d'Étiquettes pour l'Annotation Sémantique Non Supervisée

Othman Zennaki^{1,2} Nasredine Semmar¹ Laurent Besacier²

(1) CEA, LIST, Laboratoire Vision et Ingénierie de Contenus, F-91191, Gif-sur-Yvette, France

(2) Laboratoire d'Informatique de Grenoble, Univ. Grenoble-Alpes, Grenoble, France

othman.zennaki, nasredine.semmar@cea.fr, laurent.besacier@imag.fr

RÉSUMÉ

Nos travaux portent sur la construction rapide d'outils d'analyse linguistique pour des langues peu dotées en ressources. Dans une précédente contribution, nous avons proposé une méthode pour la construction automatique d'un analyseur morpho-syntaxique via une projection interlingue d'annotations linguistiques à partir de corpus parallèles (méthode fondée sur les réseaux de neurones récurrents). Nous présentons, dans cet article, une amélioration de notre modèle neuronal, avec la prise en compte d'informations linguistiques externes pour un annotateur plus complexe. En particulier, nous proposons d'intégrer des annotations morpho-syntaxiques dans notre architecture neuronale pour l'apprentissage non supervisé d'annotateurs sémantiques multilingues à gros grain (annotation en *SuperSenses*). Nous montrons la validité de notre méthode et sa généralité sur l'italien et le français et étudions aussi l'impact de la qualité du corpus parallèle sur notre approche (généré par traduction manuelle ou automatique). Nos expériences portent sur la projection d'annotations de l'anglais vers le français et l'italien.

ABSTRACT

Cross-lingual Annotation Projection for Unsupervised Semantic Tagging.

This work focuses on the development of linguistic analysis tools for resource-poor languages. In a previous study, we proposed a method based on cross-language projection of linguistic annotations from parallel corpora to automatically induce a morpho-syntactic analyzer. Our approach was based on Recurrent Neural Networks (RNNs). In this paper, we present an improvement of our neural model. We investigate the inclusion of external information (POS tags) in the neural network to train a multilingual SuperSenses Tagger. We demonstrate the validity and genericity of our method by using parallel corpora (obtained by manual or automatic translation). Our experiments are conducted for cross-lingual annotation projection from English to French and Italian.

MOTS-CLÉS : Multilinguisme, transfert crosslingue, annotation sémantique, réseaux de neurones récurrents.

KEYWORDS: Multilingualism, cross-lingual transfer, semantic tagging, recurrent neural networks.

1 Introduction

En vue de minimiser le besoin en ressources annotées, plusieurs études se sont intéressées au développement d'outils du TAL (Traitement Automatique de la Langue) basés sur des méthodes non supervisées et semi-supervisées (Collins & Singer, 1999; Goldberg, 2010). Par exemple, de très bonnes performances sont obtenues par les systèmes utilisant des approches fondées sur *la*

projection interlingue d'annotations (Yarowsky *et al.*, 2001). L'idée générale consiste à utiliser des corpus parallèles (préalablement alignés automatiquement au niveau des mots) pour transférer des annotations d'une langue *source* (richement dotée) vers une autre langue *cible* (faiblement dotée). Ce type d'approche permet d'exploiter, pour la construction automatique de ressources pour des langues peu dotées, les ressources disponibles pour des langues bien dotées ainsi que les annotations (ou systèmes) déjà disponibles pour ces mêmes langues.

Nos travaux se situent dans ce cadre. Nous explorons la possibilité d'induire des ressources et des outils multilingues adéquats à moindre coût pour l'analyse des langues faiblement dotées. Nous avons, par le passé, proposé une méthode utilisant des réseaux de neurones récurrents (on utilise l'acronyme RNN en anglais - pour *Recurrent Neural Networks*), en vue de construire un annotateur morpho-syntaxique multilingue (Zennaki *et al.*, 2015b,a). Nous avons évalué les performances de notre méthode sur cette tâche, dans les cadres non supervisé et semi-supervisé. Dans le cas non supervisé, notre méthode combinée avec une méthode fondée sur l'utilisation d'alignements automatiques source-cible au niveau des mots, avait donné des résultats très encourageants (au niveau des meilleures techniques de l'état de l'art). Dans le cas semi-supervisé (exploitation d'un petit corpus annoté manuellement pour l'adaptation du modèle non supervisé initial sur la langue cible) l'utilisation des RNN s'était révélée intéressante sur des scénarios d'adaptation avec peu de données en langue cible.

Contribution Nous proposons dans cet article une amélioration de notre approche (basée sur les réseaux de neurones récurrents) avec la prise en compte d'informations linguistiques de bas niveau pour la construction d'annotateurs linguistiques plus complexes. Ainsi, nous démontrons que l'intégration des annotations morpho-syntaxiques dans notre modèle neuronal est utile pour la construction d'annotateurs sémantiques à gros grain (*SuperSenses*) multilingues. Cette tâche d'annotation en *SuperSenses* prend une importance grandissante dans plusieurs applications du TAL. Cependant, à notre connaissance, il n'y a pas de travaux sur cette tâche pour le français. La réalisation d'un tel système pour le français est donc un autre contribution de cet article.

Méthodologie Notre méthode ne nécessite qu'un corpus parallèle (ou multi-parallèle) entre une langue source bien dotée (riche en corpus annotés en sens) et une ou plusieurs langues cibles moins bien dotées. Un corpus parallèle peut être obtenu soit par construction manuelle (corpus non bruité) ou en utilisant un système de traduction automatique (corpus bruité). Nous étudions l'impact de la qualité du corpus parallèle sur notre approche en utilisant des traductions manuelles ou automatiques du corpus *SemCor* (Miller *et al.*, 1993) (corpus anglais annoté en sens, utilisé notamment pour la tâche de désambiguïsation lexicale multilingue de Semeval 2013) vers l'italien et le français. Ceci nous permettra aussi d'évaluer la généralité de notre méthode.

Plan Après un bref état de l'art donné dans la section 2, nous présentons la tâche d'annotation à gros grain (*SuperSenses Tagging*) dans la partie 3. Notre modèle est décrit dans la partie 4 et son évaluation est présentée dans la partie 5. Enfin, la partie 6 conclut notre étude et présente quelques perspectives.

2 État de l'art

Plusieurs études ont utilisé la projection interlingue pour le transfert d'annotations linguistiques à partir de langues bien dotées (*source*) vers les langues faiblement dotées (*cible*) en vue de créer

des outils d'analyse linguistique pour les langues cibles. La projection interlingue d'annotations a été introduite par (Yarowsky *et al.*, 2001), en utilisant un corpus parallèle pour adapter des outils monolingues (analyseurs morphologiques, analyseurs morpho-syntaxiques et analyseurs syntaxiques de surface) à de nouvelles langues. Cette méthode a été depuis utilisée avec succès dans plusieurs autres travaux.

On peut citer par exemple : l'apprentissage d'analyseurs morpho-syntaxiques de bonne qualité par projection (Das & Petrov, 2011; Duong *et al.*, 2013; Wisniewski *et al.*, 2014), le transfert d'entités nommées (Kim *et al.*, 2012), d'annotations syntaxiques (Jiang *et al.*, 2011), la projection d'annotations en sens réalisée par (Bentivogli *et al.*, 2004), l'annotation en rôles sémantiques sur l'allemand par projection interlingue à partir de la paire de langues anglais-allemand (Padó & Lapata, 2005), dont la généralité a plus spécifiquement été évaluée dans (Pado & Pitel, 2007). Pour la compréhension automatique de la parole, (Jabaian *et al.*, 2013) utilisent cette technique pour transférer un système du français vers l'italien et l'arabe avec un faible coût de supervision.

Souvent, le transfert entre les langues repose sur l'existence de corpus parallèles et l'utilisation d'alignements au niveau des mots entre langue source et langue cible. Plusieurs outils permettent d'obtenir automatiquement de tels alignements, dont GIZA++ (Och & Ney, 2000).

Cependant, les performances des algorithmes d'alignement au niveau des mots ne sont pas toujours satisfaisantes (Allauzen & Wisniewski, 2010) et cette étape a un impact significatif sur la performance de la projection d'annotations linguistiques (Fraser & Marcu, 2007). Pour cette raison, notre approche utilise un corpus parallèle aligné au niveau des phrases seulement et n'applique aucun pré-traitement du type *alignement automatique en mots* qui peut être source d'erreurs et de bruit.

À l'instar de notre approche, plusieurs auteurs ont proposé d'utiliser des représentations interlingues/multilingues plutôt que des alignements bruités. En particulier, de nombreuses contributions se sont appuyées sur les travaux de (Mikolov *et al.*, 2013) sur l'apprentissage de vecteurs de mots (ou *word embeddings*) et leur extension à des représentations bilingues (ou multilingues) (Al-Rfou *et al.*, 2013; Luong *et al.*, 2015). À partir de ces représentations, le transfert d'annotations en langue source est déduit pour les mots ou fragments cibles les plus proches dans l'espace commun aux deux langues. Afin d'apprendre de telles représentations, plusieurs ressources peuvent être utilisées telles que des lexiques bilingues (Gouws & Søggaard, 2015) et des corpus parallèles (Gouws *et al.*, 2015; Zennaki *et al.*, 2015b,a). L'utilisation de représentations interlingues a par ailleurs donné lieu à des résultats satisfaisants sur plusieurs tâches du TAL. Par exemple, l'annotation morpho-syntaxique multilingue (Zennaki *et al.*, 2015b,a; Gouws & Søggaard, 2015), l'annotation interlingue en SuperSenses (Gouws & Søggaard, 2015), la reconnaissance interlingue d'entités nommées (Täckström *et al.*, 2012), la classification interlingue de documents (Gouws *et al.*, 2015), l'annotation syntaxique interlingue (Xiao & Guo, 2015), et l'annotation en rôles sémantiques interlingue (Titov & Klementiev, 2012).

Notre approche de transfert interlingue est par ailleurs décrite dans (Zennaki *et al.*, 2015b,a). Le travail présenté ici est différent puisqu'il applique notre méthode à une autre tâche mais va aussi plus loin en essayant de prendre en compte des étiquettes morpho-syntaxiques pour la construction d'un annotateur sémantique à gros grain. Il s'inscrit dans la lignée des études précédemment décrites puisque nous essayons aussi d'apprendre, via le RNN, un espace de représentation multilingue commun à plusieurs langues. En entrée du réseau, notre représentation commune est fondée simplement sur la cooccurrence des mots sources et cibles dans un corpus parallèle (empreinte distributionnelle). Nous utilisons cette représentation pour apprendre (à partir d'une annotation en *SuperSenses* côté source) un annotateur sémantique à gros grain pour des textes en langue cible.

Concluons cet état de l'art en mentionnant que plusieurs travaux ont étudié la possibilité d'utiliser les réseaux de neurones dans le TAL (Bengio *et al.*, 2003; Collobert *et al.*, 2011; Henderson, 2004; Mikolov *et al.*, 2010; Federici & Pirrelli, 1993). (Federici & Pirrelli, 1993) font partie des premiers à avoir développé des annotateurs morpho-syntaxiques basés sur les réseaux de neurones, (Bengio *et al.*, 2003) et (Mikolov *et al.*, 2010) ont utilisé les réseaux de neurones pour construire des modèles de langages. (Collobert *et al.*, 2011) ont employé les techniques de réseaux de neurones profonds pour l'apprentissage multitâche qui comprend : l'annotation morpho-syntaxique, la reconnaissance d'entités nommées et l'annotation sémantique. (Henderson, 2004) a aussi proposé des méthodes d'apprentissage d'analyseurs syntaxiques fondées sur les réseaux de neurones. Nous présentons dans la partie suivante la tâche d'annotation en *SuperSenses* ainsi qu'un état de l'art des travaux traitant de cette tâche.

3 Annotation en SuperSenses

L'annotation en SuperSenses (SuperSenses Tagging – SST –) est une tâche du TAL qui consiste à annoter chaque unité du texte, avec un jeu d'étiquettes sémantiques générales définies par les catégories lexicographiques de WordNet (*SuperSenses*). Elle peut être vue comme une tâche à cheval entre la reconnaissance d'entités nommées (REN) et la désambiguïation lexicale (Word Sense Disambiguation – WSD) : tout en étant une extension de la REN, elle est une simplification de la WSD.

Reconnaissance d'entités nommées Issue de la recherche d'information, la reconnaissance d'entités nommées (REN) consiste à rechercher des objets textuels (i.e. un mot, ou un groupe de mots) catégorisables dans des classes telles que noms de personnes, noms d'organisations ou d'entreprises, noms de lieux, quantités, distances, valeurs, dates, etc. La REN fait sans doute partie des tâches les plus étudiées du TAL et apparaît en effet comme fondamentale pour diverses applications telles que l'analyse de contenu, la recherche et l'extraction d'information, les systèmes de question-réponse, le résumé automatique, etc. (Ehrmann, 2008). Néanmoins, cette tâche reste limitée à la reconnaissance de catégories générales en omettant des unités informatives (catégories plus fines) potentiellement importantes dans certains contextes applicatifs.

Désambiguïation lexicale Le processus de désambiguïation lexicale consiste à sélectionner les sens corrects d'instances contextualisées de mots ambigus, parmi l'ensemble de leurs sens possibles (ou sens candidats). La REN peut prendre part à un processus de désambiguïation lexicale en tant qu'information sémantique. Concrètement, les entités nommées sont utilisées comme « filtre » au niveau des restrictions de sélection (ou sous-catégorisation sémantique) des sens d'une unité lexicale parmi plusieurs dizaines de milliers de sens très spécifiques. Ces sens sont inventoriés dans des dictionnaires, incluant les entités nommées. Un tel degré de granularité (très fin), rend la distinction entre les sens trop subtile pour être capturée automatiquement de façon robuste. De plus, l'apprentissage automatique de modèles nécessite des données manuellement annotées en sens, difficiles à obtenir.

WordNet Le *Princeton WordNet* (Fellbaum, 1998), est probablement la ressource la plus utilisée pour la désambiguïation lexicale (Word Sense Disambiguation – WSD). Il est organisé autour de la notion d'ensemble de synonymes (synsets) décrits par une partie du discours (nom, verbe, adjectif, adverbe), une définition et leurs liens (hyperonyme, hyponyme, antonyme, etc.). Chaque sens d'un item lexical (entrée) correspond à un synset. La version courante du Princeton WordNet (version 3.0) comprend 155 287 items lexicaux pour un total de 117 659 synsets.

Annotation sémantique à gros grain (*SuperSenses Tagging*) La complexité à modéliser et à traiter l’ambiguïté lexicale ainsi que les limitations de la REN ont fait émerger des tâches d’annotation sémantique lexicale (sens lexical) à gros grain. Ce type de tâche présente plusieurs avantages. Par exemple, l’intérêt d’utiliser les annotations en sens à gros grain, afin de lever l’ambiguïté lexicale des mots, a été souligné dès l’apparition des premiers réseaux lexico-sémantiques (Peters *et al.*, 1998). (Kohomban & Lee, 2005) ont proposé d’utiliser les catégories lexicographiques du *Princeton WordNet* comme amorce à l’annotation avec des sens plus spécifiques (à grain fin).

L’annotation en *SuperSenses* (*SuperSenses Tagging* – SST –) est l’une de ces tâches alternatives, la tâche d’annotation en *SuperSenses* a été introduite par (Ciaramita & Johnson, 2003), comme une étape de désambiguïsation intermédiaire pour la réalisation de la WSD. Ces mêmes auteurs ont proposé d’utiliser un perceptron structuré entraîné et évalué sur le SemCor Corpus (Miller *et al.*, 1993). (Ciaramita & Johnson, 2003) ont appelé « *SuperSenses* » les 41 catégories lexicographiques du *Princeton WordNet* (Fellbaum, 1998). Les catégories WordNet sont vues comme la principale base pour la construction d’un ensemble de catégories plus exhaustives en vue de mieux annoter les concepts occurant dans une phrase. Un autre avantage de l’utilisation de ces catégories est leur universalité (elles sont communes à plusieurs langues), ce qui permet leur utilisation pour des tâches telles que la traduction automatique ou la recherche et l’extraction d’informations multilingues. L’annotation en *SuperSenses* a été aussi utilisée comme première étape pour plusieurs tâches, telles que la désambiguïsation en sens (Ye & Baldwin, 2007), et le réordonnement des hypothèses d’un parser (Collins & Koo, 2005).

Les étiquettes *SuperSenses* Le jeu d’étiquettes *SuperSenses* original comprend au total 41 sens, répartis en deux catégories : 26 étiquettes pour représenter les sens des noms et 15 autres pour représenter les sens des verbes, plus une étiquette unique (catch-all) pour les autres unités (adjectifs, adverbes, etc.). (Ciaramita & Johnson, 2003) présentent ces étiquettes en détail.

4 Méthode proposée

Étant donné un corpus parallèle/multi-parallèle (Anglais-Italien et Anglais-Français dans notre cas) entre une langue source bien dotée et une ou plusieurs langues cibles moins bien dotées, notre but est de créer, en utilisant des approches faisant appel à la projection interlingue d’annotations, des ressources et des outils multilingues pour l’analyse des langues cibles. Notre méthode est basée sur l’utilisation des réseaux de neurones récurrents (RNN) associés à une représentation interlingue des mots (des langues source et cibles).

Dans ce travail qui fait suite à nos travaux sur la construction d’annotateurs morpho-syntaxiques multilingues (Zennaki *et al.*, 2015b,a), nous proposons une amélioration de notre approche avec la prise en compte d’annotations externes par le RNN (voir figure 2), et aussi sa validation sur une nouvelle tâche plus complexe : l’annotation sémantique à gros grain multilingue (*SuperSenses Tagging*).

Avant de décrire notre annotateur en *SuperSenses* multilingue basé sur les réseaux de neurones récurrents (on utilise l’acronyme **RNN-SST** en anglais - pour **Recurrent Neural Networks based SuperSense Tagger**), nous décrivons tout d’abord l’approche par projection simple à laquelle nous allons en partie nous comparer (et qui sera aussi combinée — au cours des expériences qui vont suivre — avec la méthode que nous proposons).

4.1 Annotateur en *SuperSenses* non supervisé par Projection Simple

Notre approche pour construire un annotateur en *SuperSenses* non supervisé par projection simple (décrit par l'algorithme 1) est très proche de celle introduite par (Yarowsky *et al.*, 2001). Ces auteurs utilisent l'alignement automatique en mots (obtenu à partir d'un corpus parallèle) pour projeter les annotations de la langue *source* vers la langue *cible*, en vue de construire des ressources et outils linguistiques pour la langue *cible*.

Nous avons utilisé comme corpus multi-parallèle, pour nos expériences, le *MultiSemCor* (voir 5.1). Le côté *source* correspond au *SemCor* Anglais (Miller *et al.*, 1993), c'est un sous-ensemble du Corpus de *Brown* (Kucera & Francis, 1979) annoté avec les synsets du *Princeton WordNet* (Fellbaum, 1998).

L'algorithme 1 de projection simple est décrit dans l'encadré ci-dessous :

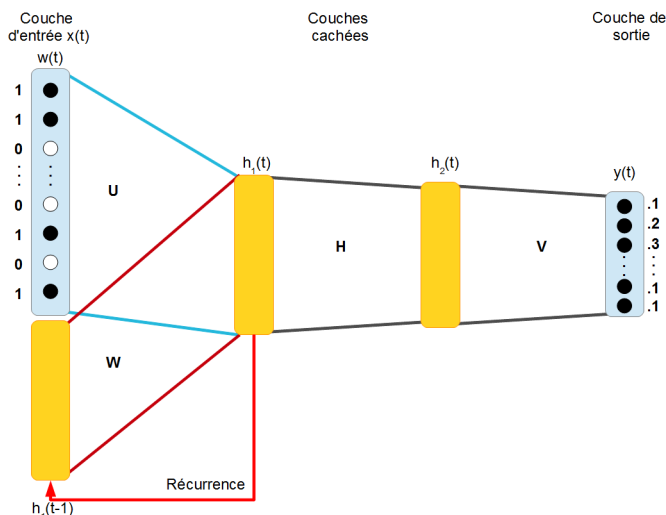
Algorithme 1 : Annotateur en *SuperSenses* par projection simple

- 1 : Annoter le SemCor Anglais en *SuperSenses*, en utilisant les correspondances *synsets* \Leftrightarrow *SuperSenses*.
 - 2 : Aligner automatiquement le corpus parallèle *MultiSemCor* en utilisant GIZA++ ou un autre outil d'alignement en mots.
 - 3 : Projeter les annotations directement pour les alignements 1-1.
 - 4 : Pour les correspondances N-1, projeter l'annotation du mot se trouvant à la position N/2 arrondi à l'entier supérieur.
 - 5 : Annoter les mots non-alignés avec l'étiquette la plus fréquente qui leur est associée dans le corpus.
 - 6 : Apprendre un annotateur en *SuperSenses* à partir de la partie cible du corpus désormais annotée (par exemple, dans notre cas, nous utilisons TNT tagger (Brants, 2000)).
-

4.2 Annotateur en *SuperSenses* fondé sur une architecture RNN (RNN-SST)

Dans cette section, nous décrivons l'adaptation de notre approche proposée dans nos travaux antérieurs sur l'annotation morpho-syntaxique multilingue (Zennaki *et al.*, 2015b,a), pour la construction d'un annotateur multilingue en *SuperSenses* basé sur les RNNs (RNN-SST), avant de présenter dans la section suivante trois améliorations de l'architecture du RNN-SST, qui permettent la prise en compte d'annotations externes (morpho-syntaxiques dans le cas présent) pour l'annotation en *SuperSenses* (RNN-SST-POS). Notre approche nécessite la disponibilité d'un *MultiSemCor* (voir 5.1), obtenu soit par traduction automatique soit par traduction manuelle du *SemCor* Anglais.

Concernant le choix du réseau neuronal, il existe deux architectures majeures des réseaux de neurones : les réseaux de neurones Feed-forward (Bengio *et al.*, 2003) et les réseaux de neurones récurrents (Elman, 1990; Mikolov *et al.*, 2010). (Sundermeyer *et al.*, 2013) ont montré que les modèles de langue statistiques basés sur une architecture récurrente présentent de meilleures performances que les modèles basés sur une architecture Feed-forward. Cela vient du fait que les réseaux de neurones récurrents utilisent un contexte de taille non limitée, contrairement aux réseaux Feed-forward dont la topologie limite la taille du contexte pris en compte. Cette propriété a motivé notre choix d'utiliser, dans nos expériences, un réseau de neurones de type récurrent (Elman, 1990).

FIGURE 1 – Annotateur multilingue en *SuperSenses* fondé sur un RNN

4.2.1 Description du modèle

L'architecture du RNN-SST est présentée à titre d'illustration dans la figure 1. Nous utilisons l'architecture du RNN ayant donné les meilleurs résultats lors de nos travaux précédents, ainsi, le RNN-SST est composé d'une succession de quatre couches de neurones : une couche d'entrée au temps t notée $x(t)$, une première couche cachée $h_1(t)$ (aussi appelée couche de contexte), une deuxième couche cachée $h_2(t)$, et une couche de sortie $y(t)$. Chaque neurone de la couche d'entrée est relié à tous les neurones de la couche cachée par les matrices des poids U et W . Les neurones des deux couches cachées sont connectés entre eux par la matrice des poids H . La matrice des poids V connecte tout neurone de la deuxième couche cachée à chaque neurone de la couche de sortie.

Dans notre modèle, la couche d'entrée est formée par la concaténation de la représentation vectorielle $w(t)$ du mot courant, et de la première couche cachée au temps précédent $h_1(t-1)$, ce qui confère au réseau neuronal son aspect récurrent. La première étape de notre modèle est donc d'associer à chaque mot w (appartenant aux vocabulaires des langues source et cible) une représentation vectorielle spécifique.

On souhaite construire un espace de représentation multilingue pour les mots (en langue source et cible), dans lequel un mot source et sa traduction cible possèdent des représentations vectorielles proches. Ceci permettrait alors d'utiliser l'annotateur en *SuperSenses* de type RNN (appris initialement sur le côté source) pour annoter un texte en langue cible.

Généralement, un mot source et sa traduction cible cooccurrent dans un nombre important de bi-phrases du corpus parallèle, par conséquent leurs empreintes distributionnelles sont proches. Notre espace de représentation interlingue est construit en associant à chaque mot (source, cible) son empreinte distributionnelle V_{wi} , $i = 1, \dots, N$, où N est le nombre de bi-phrases dans le corpus parallèle. Si w apparaît dans la $i^{\text{ème}}$ bi-phrase alors $V_{wi} = 1$; par conséquent, les neurones de la couche d'entrée représentant le mot courant w , sont mis à 0 sauf ceux correspondants aux bi-phrases contenant le mot w , qui sont mis à 1.

La couche de sortie de notre modèle comporte 42 neurones, correspondant aux 41 étiquettes *SuperSenses* plus une étiquette vide. Les valeurs en sortie sont calculées comme suit :

- Propagation en avant de l'entrée $x(t)$ pour calculer la valeur de la première couche cachée :

$$h_1(t) = f(U(t)w(t) + W(t)h_1(t - 1)) \quad (1)$$

- Calculer la valeur de la deuxième couche cachée :

$$h_2(t) = f(H(t)h_1(t)) \quad (2)$$

- Calculer la sortie du réseau (prédiction du modèle) $y(t)$:

$$y(t) = g(V(t)h_2(t)) \quad (3)$$

où $f(z)$ est une fonction d'activation de type *sigmoïde* et $g(z)$ est une fonction de type *softmax*.

On utilise la fonction *softmax* sur la couche de sortie afin d'obtenir des scores assimilables à des probabilités, le mot w en entrée du réseau est annoté par l'étiquette la plus probable en sortie du réseau.

4.2.2 Construction du modèle

Notre modèle (RNN-SST) a été entraîné sur le corpus d'apprentissage (côté source du corpus parallèle MultiSemCor annoté en *SuperSenses* — SemCor Anglais —) par descente de gradient stochastique basée sur la **R**étro-**P**ropagation (**RP**) du gradient de l'erreur (Rumelhart *et al.*, 1985) et la **R**étro-**P**ropagation du gradient de l'erreur à **T**ravers le **T**emps (**RPTT**) (Rumelhart *et al.*, 1985). L'apprentissage de notre modèle neuronal est un processus itératif sur le corpus d'apprentissage (coté *source* annoté en *SuperSenses* du corpus parallèle). Après chaque itération (époque), l'erreur d'annotation du modèle est estimée sur un corpus de validation. La convergence du modèle est obtenue si le taux d'erreur d'annotation ne diminue plus d'une époque à une autre. Généralement le modèle converge en 5 à 10 époques.

Comme présenté précédemment, le modèle neuronal (annotateur en *SuperSenses* basé sur les RNN) ainsi appris est utilisé pour annoter les textes *cible* en *SuperSenses*, via l'utilisation de notre espace de représentation multilingue. Il est important de noter que dans le cas d'un corpus parallèle multilingue, le même modèle neuronal pourra annoter en *SuperSenses* toutes les langues *cibles* sans être re-entraîné. On dispose donc d'un véritable annotateur multilingue.

4.3 Ajout d'informations externes (RNN-SST-POS)

Dans la section précédente, nous avons défini la façon dont notre modèle neuronal est construit pour l'annotateur en *SuperSenses*. Nous avons néanmoins réalisé une modification spécifique de ce modèle avec la prise en compte des traits morpho-syntaxiques pour une meilleure adaptation de notre modèle à la tâche d'annotation en *SuperSenses*. L'idée sous-jacente à cette modification est d'utiliser les annotations morpho-syntaxiques (annotations basiques) pour lever une partie de l'ambiguïté relative à l'annotation en *SuperSenses*. Nous pensons que l'intégration des annotations morpho-syntaxiques devrait permettre d'améliorer les performances de notre modèle neuronal.

Nous proposons trois architectures neuronales pour l'intégration des annotations morpho-syntaxiques à différents niveaux de représentation. Comme spécifié dans la figure 2, l'intégration peut se faire soit au niveau de la couche d'entrée soit au niveau de la première couche cachée ou bien au niveau de la deuxième couche cachée. Dans ces trois modèles, l'annotation morpho-syntaxique du mot en cours

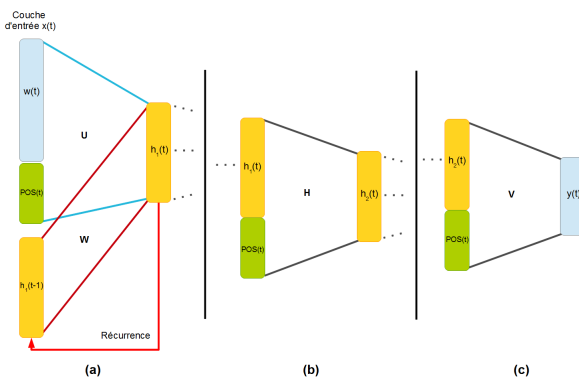


FIGURE 2 – Architectures RNN-SST-POS avec prise en compte de l’annotation morpho-syntactique (POS) du mot en cours $w(t)$ sur trois niveaux ; de gauche à droite : (a) au niveau de la couche d’entrée, (b) au niveau de la première couche cachée, (c) au niveau de la deuxième couche cachée.

$w(t)$ est représentée par un vecteur ($POS(t)$) de même taille que l’ensemble des annotations morpho-syntactiques considérées (nous utilisons le jeu de 12 étiquettes morpho-syntactiques universelles défini dans (Petrov *et al.*, 2012)), toutes les composantes du vecteur sont égales à zéro, sauf la composante correspondant à l’indice de l’annotation morpho-syntactique du mot en cours qui est égale à 1.

5 Mise en œuvre de notre approche

Pour tester notre approche dans un cadre multilingue, nous l’évaluons sur l’italien et le français considérés comme langues cibles. La seule source de données utilisée dans les deux cas est le corpus parallèle multilingue MultiSemCor (MSC), résultant des traductions du SemCor Anglais vers l’italien (manuellement et automatiquement) et vers le français (automatiquement).

5.1 Corpus

SemCor Le SemCor (Miller *et al.*, 1993)¹ est un sous-ensemble du Corpus de Brown (Kucera & Francis, 1979). Sur les 700 000 mots de ce dernier, environ 230 000 sont annotés avec des *synsets* du Princeton *WordNet*.

MultiSemCor Nous disposons du MultiSemCor Anglais/Italien (MSC-IT-1) construit par traduction manuelle du SemCor Anglais vers l’italien (Bentivogli & Pianta, 2005)². Afin d’étudier l’influence de la qualité du corpus parallèle (traduction manuelle/automatique) sur notre système, nous traduisons aussi automatiquement le SemCor Anglais vers l’italien pour obtenir un Mutli-SemCor Anglais/Italien (MSC-IT-2) construit automatiquement. Ne disposant pas de MultiSemcor Anglais/Français construit manuellement, nous utilisons deux traductions du SemCor vers le français provenant de deux systèmes de traduction distincts : - MultiSemcor Anglais/Français (MSC-FR-1)³ obtenu avec le système de traduction statistique anglais-français basé sur la boîte à outils *Moses* (Hoang & Koehn, 2008) mis au point par le Laboratoire d’Informatique de Grenoble (Besacier *et al.*, 2012) ; - MultiSemcor Anglais/Français (MSC-FR-2) obtenu avec le système de traduction statistique anglais-français en ligne (Google Traduction⁴).

1. <http://web.eecs.umich.edu/~mihalcea/downloads.html#semcor>

2. <http://multisemcor.fbk.eu/>

3. <http://getalp.imag.fr/static/wsd/TALN2015/ressources/frenchSemcor3.0.zip>

4. <https://translate.google.com/>

Corpus d'apprentissage Puisque le SemCor est annoté avec les *synsets* de Princeton *WordNet* et que l'apprentissage de nos modèles se fait sur des *SuperSenses*, nous avons réalisé une conversion des annotations du SemCor des *synsets* *WordNet* vers les *SuperSenses*. Nous avons ensuite appris nos systèmes sur les différentes versions du MultiSemCor (décrites précédemment), avec coté source le SemCor annoté en *SuperSenses*.

Corpus d'évaluation Pour estimer les performances de nos modèles et pouvoir les comparer avec des modèles existants, nous utilisons le corpus d'évaluation de la tâche 12 (désambiguïsation lexicale multilingue) de Semeval 2013 (Navigli *et al.*, 2013), qui est un corpus d'évaluation traduit en 5 langues (anglais, français, allemand, italien, espagnol) pour lequel nous utilisons les textes italien et français. Cependant, ces textes de la campagne d'évaluation Semeval 2013 (tâche 12) ont été annotés en sens issus de *BabelNet* (Navigli & Ponzetto, 2012), nous avons donc tout d'abord réalisé une conversion des sens *BabelNet* vers les *synsets* *WordNet*, puis les *synsets* *WordNet* ont été convertis vers les *SuperSenses*.

5.2 Systèmes évalués

Les objectifs de nos expérimentations sont les suivants : d'une part nous souhaitons valider notre modèle neuronal sur une nouvelle tâche (l'annotation en *SuperSenses* RNN-SST) ; d'autre part, nous souhaitons évaluer l'apport de la prise en compte des annotations morpho-syntaxiques dans notre modèle neuronal pour l'annotation en *SuperSenses* (RNN-SST-POS) ; enfin, nous voulons évaluer l'influence de la qualité des corpus parallèles (traduction manuelle ou automatique) sur les performances de nos modèles neuronaux (RNN-SST et RNN-SST-POS) et du modèle basé sur la projection interlingue d'annotations.

En résumé :

- nous avons construit quatre annotateurs en *SuperSenses* basés sur la projection interlingue (Projection Simple), méthode que nous avons présentée dans la section 4.1, en utilisant les corpus MultiSemCor (MSC-IT-1, MSC-IT-2, MSC-Fr-1, et MSC-FR-2) décrits dans la section 5.1.
- nous avons aussi utilisé les corpus MultiSemCor pour construire des espaces de représentations multilingues anglais-italien-français.
- nous avons utilisé le SemCor anglais annoté en *SuperSenses* pour l'apprentissage de notre modèle neuronal multilingue RNN-SST, via nos espaces de représentation multilingues.
- nous avons annoté le SemCor anglais en utilisant *TreeTagger* (analyseur morpho-syntaxique proposé par (Schmid, 1995)), puis avons utilisé le SemCor ainsi annoté, pour l'apprentissage de trois modèles neuronaux RNN-SST-POS, correspondant aux trois architectures proposées dans la section 4.3 :
 - RNN-SST-POS-In : ajout de l'annotation morpho-syntaxique au niveau de la couche d'entrée.
 - RNN-SST-POS-H1 : ajout de l'annotation morpho-syntaxique au niveau de la première couche cachée.
 - RNN-SST-POS-H2 : ajout de l'annotation morpho-syntaxique au niveau de la deuxième couche cachée.

5.3 Résultats et discussion

Comme mentionné précédemment, nos modèles présentés ci-dessus ont été évalués sur les corpus d'évaluations de Semeval 2013 (tâche 12 - désambiguïsation lexicale - WSD), ce qui nous permet une

comparaison avec les systèmes ayant participé à la tâche 12 de Semeval 2013, mais cette comparaison est indirecte du fait que ces systèmes n’ont pas été construits spécialement pour l’annotation en *SuperSenses* mais pour la désambiguïsation lexicale en utilisant les sens BabelNet (nous avons donc opéré une conversion des sens BabelNet vers les *SuperSenses*). Nous avons identifié deux systèmes provenant de cette tâche, se rapprochant le plus de notre cadre expérimentale :

- **MFS Semeval 2013** : Système de référence fourni par SemEval 2013 pour la tâche de WSD, ce système utilise l’heuristique du sens WordNet le plus fréquent (MFC - *Most Frequent Sense*) et constitue une *baseline* solide (car elle fait appel aux sens issus d’une base lexicale multilingue, BabelNet, tandis que nos méthodes tentent d’induire ces sens à partir de l’anglais).
- **GETALP** : Système non-supervisé pour la WSD proposé par (Schwab *et al.*, 2012) basé sur un algorithme à colonies de fourmis.

Nous précisons également que deux autres systèmes ont participé à cette tâche, le système DAEBAK ! (Navigli & Lapata, 2010) et le système UMCC-DLSI (Gutiérrez Vázquez *et al.*, 2011). Contrairement à notre système et aux systèmes GETALP et MFS Semeval 2013, les systèmes DAEBAK ! et UMCC-DLSI bénéficiaient de ressources externes et de la richesse des informations de BabelNet. Notre cadre expérimental non supervisé se rapprochant plus des systèmes GETALP et MFS Semeval 2013, nous pensons qu’il est plus judicieux de nous comparer aux performances de ces deux systèmes⁵.

Le modèle BARISTA (*Bilingual Adaptive Reshuffling with Individual Stochastic Alternatives*) (Gouws & Søggaard, 2015) est un modèle fondé sur l’utilisation d’une liste de paires de mots (mot source-traduction cible ou bien mot source-mot cible portant la même information linguistique) pivots pour la construction d’une représentation distribuée bilingue. Cette représentation est apprise en utilisant un modèle neuronal sur un corpus bilingue dans lequel les mots pivots d’une langue sont remplacés aléatoirement par les mots correspondants dans l’autre langue.

Par souci d’exhaustivité, nous avons choisi de reporter, dans notre tableau des résultats, les performances sur l’annotation en *SuperSenses* en danois du modèle BARISTA, malheureusement, nous n’avons pas pu nous évaluer nous même sur le danois (corpus d’évaluation non disponible).

Le tableau 1 présente les performances en taux d’étiquetage correct de nos étiqueteurs en *SuperSenses* neuronaux (avec ou sans prise en compte des annotations morpho-syntaxiques), des étiqueteurs en *SuperSenses* fondés sur une projection interlingue, de la référence MFS Semeval 2013, du système GETALP de Semeval 2013 et du modèle BARISTA. De ces résultats, nous pouvons faire les observations suivantes :

- Nos modèles obtiennent les meilleurs résultats sur l’italien en utilisant le corpus non bruité MSC-IT-1 (traduction manuelle du SemCor vers l’italien).
- La qualité du corpus parallèle influe significativement sur les performances de nos modèles : en utilisant des corpus bruités (traduction automatique du SemCor vers l’italien et le français) les performances de la totalité de nos modèles se dégradent ; cela se vérifie également, à moindre mesure, d’une traduction automatique à l’autre (le corpus MSC-FR-1 obtenu par Google Traduction est moins bruité que le corpus MSC-FR-2 obtenu par le système Moses).
- Les meilleurs performances sont celles des modèles résultant de la combinaison entre projection simple et RNN, ce qui montre la complémentarité de ces deux approches.
- Dans le cas des corpus bruités, l’approche neuronale semble plus robuste que l’approche *baseline* par projection.

5. DAEBAK ! et UMCC-DLSI ont obtenus pour la tâche d’annotation en *SuperSenses* respectivement 68.1% et 72.5% sur l’italien, et 59.8% et 67.6 % sur le français

- Le modèle neuronal le plus performant est RNN-SST-POS-H2, ce qui démontre que la prise en compte des annotations morpho-syntaxiques au niveau de la deuxième couche est la plus pertinente ; une prise en compte tardive des informations morpho-syntaxiques semble donc préférable pour un annotateur sémantique à gros grain neuronal.

Modèle	italien		français	
Modèles Semeval 2013				
MFS Semeval 2013	60.7%		52.4%	
GETALP (Schwab <i>et al.</i> , 2012)	40.2%		34.6%	
Modèles fondés sur la Projection Interlingue				
	MSC-IT-1 trad man.	MSC-IT-2 trad. auto	MSC-FR-1 trad. auto	MSC-FR-2 trad auto.
Projection Simple	61.3%	45.6%	42.6%	44.5%
Modèles Neuronaux				
RNN-SST	59.4%	46.2%	46.2%	47%
RNN-SST-POS-In	61%	47%	46.5%	47.3%
RNN-SST-POS-H1	59.8%	46.5%	46.8%	47.4%
RNN-SST-POS-H2	63.1%	48.7%	47.7%	49.8%
Combinaisons de la Projection Interlingue et des Modèles Neuronaux				
Projection+RNN-SST	62%	46.7%	46.5%	47.4%
Projection+RNN-SST-POS-In	62.9%	47.4%	46.9%	47.7%
Projection+RNN-SST-POS-H1	62.5%	47%	47.1%	48%
Projection+RNN-SST-POS-H2	63.5%	49.2%	48%	50.1%
Modèle Basé sur les Représentations Distribuées Bilingues				
BARISTA (Gouws & Søgaard, 2015)	danois			
	58.3%			

TABLE 1 – Performances en taux d’étiquetage correct (Projection Simple, RNN-SST, RNN-SST-POS, Combinaisons Projection+RNN) - et une comparaison directe avec deux modèles Semeval 2013 et une autre indirecte avec le Modèle BARISTA sur le Danois.

Le tableau 3 illustre un exemple 2 où la prise en compte de l’annotation morpho-syntaxique au niveau de la deuxième couche cachée de notre annotateur neuronal en SupserSenses (RNN-SST-POS-H2) a permis de lever l’ambiguïté lié au mot *conseiller*, annoté par le RNN-SST comme *verb.communication* alors que l’annotation exacte est *noun.person*.

Français	... qui est également conseiller sur le climat pour le Mexique.
RNN-SST	... conseiller/ verb.communication ...
RNN-SST-POS-H2	... conseiller/ noun.person ...

TABLE 2 – Effet de la prise en compte des annotations morpho-syntaxiques.

6 Conclusion

Dans cet article, nous avons présenté une approche utilisant les réseaux de neurones récurrents comme annotateurs sémantiques multilingues à gros grain (*SuperSenses*). Nous avons démontré que la prise en compte des annotations morpho-syntaxiques dans notre architecture neuronale améliore les performances de notre modèle sur l’annotation en *SuperSenses*. Nous avons ainsi montré la faisabilité et la généralité de l’approche sur deux langues cibles : l’italien et le français. Par ailleurs, nous avons aussi pu évaluer l’impact de la qualité du corpus parallèle sur notre approche (corpus obtenu par traductions manuelles ou automatiques). Dans nos futurs travaux, nous envisageons d’utiliser des réseaux de neurones récurrents bidirectionnels pour prendre en compte aussi le contexte droit du mot (actuellement notre RNN n’utilise que le contexte gauche et est donc, d’une certaine façon, désavantagé par rapport à l’approche de projection simple fondée sur l’étiqueteur TNT qui prend en compte les contextes gauche et droit du mot courant). L’utilisation de représentations distribuées bilingues (*bilingual word embeddings*) directement à l’entrée de notre réseau récurrent est également une perspective intéressante.

Références

- AL-RFOU R., PEROZZI B. & SKIENA S. (2013). Polyglot : Distributed word representations for multilingual nlp. *Proceedings of the 17th CoNLL*, p. 183–192.
- ALLAUZEN A. & WISNIEWSKI G. (2010). Modèles discriminants pour l’alignement mot-à-mot. *TAL*, **50**(3).
- BENGIO Y., DUCHARME R., VINCENT P. & JAUVIN C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, **3**, 1137–1155.
- BENTIVOGLI L., FORNER P. & PIANTA E. (2004). Evaluating cross-language annotation transfer in the multiseacor corpus. In *Proceedings of the 20th CoNLL*, p. 364.
- BENTIVOGLI L. & PIANTA E. (2005). Exploiting parallel texts in the creation of multilingual semantically annotated resources : the multiseacor corpus. *Natural Language Engineering*, **11**(03), 247–261.
- BESACIER L., LECOUTEUX B., AZOUZI M. & LUONG N.-Q. (2012). The lig english to french machine translation system for iwslt 2012. In *IWSLT*, p. 102–108.
- BRANTS T. (2000). Tnt : a statistical part-of-speech tagger. In *Applied natural language processing*.
- CIARAMITA M. & JOHNSON M. (2003). Supersense tagging of unknown nouns in wordnet. In *EMNLP*.
- COLLINS M. & KOO T. (2005). Discriminative reranking for natural language parsing. *Computational Linguistics*, **31**(1), 25–70.
- COLLINS M. & SINGER Y. (1999). Unsupervised models for named entity classification. In *EMNLP*.
- COLLOBERT R., WESTON J., BOTTOU L., KARLEN M., KAVUKCUOGLU K. & KUKSA P. (2011). Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*, **12**.
- DAS D. & PETROV S. (2011). Unsupervised part-of-speech tagging with bilingual graph-based projections. In *ACL-HLT*, volume 1, p. 600–609.
- DUONG L., COOK P., BIRD S. & PECINA P. (2013). Simpler unsupervised pos tagging with bilingual projections. In *ACL (2)*, p. 634–639.
- EHRMANN M. (2008). *Les entités nommées, de la linguistique au TAL*. PhD thesis, Univ. Paris 7.
- ELMAN J. L. (1990). Finding structure in time. *Cognitive science*, **14**(2), 179–211.
- FEDERICI S. & PIRRELLI V. (1993). Analogical modelling of text tagging. *unpublished report, Istituto diLinguistica Computazionale, Italy*.
- FELLBAUM C. (1998). *WordNet*. Wiley Online Library.
- FRASER A. & MARCU D. (2007). Measuring word alignment quality for statistical machine translation. *Computational Linguistics*, **33**(3), 293–303.
- GOLDBERG A. B. (2010). *New directions in semi-supervised learning*. PhD thesis, University of Wisconsin.
- GOUWS S., BENGIO Y. & CORRADO G. (2015). Bilbowa : Fast bilingual distributed representations without word alignments. *ICML 2015*.
- GOUWS S. & SØGAARD A. (2015). Simple task-specific bilingual word embeddings. In *NAACL-HLT*.
- GUTIÉRREZ VÁZQUEZ Y., FERNÁNDEZ ORQUÍN A., MONTORO GUIJARRO A., VÁZQUEZ PÉREZ S. *et al.* (2011). Enriching the integration of semantic resources based on wordnet.
- HENDERSON J. (2004). Discriminative training of a neural network statistical parser. In *ACL*, p. 95.
- HOANG H. & KOEHN P. (2008). Design of the mooses decoder for statistical machine translation. In *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, p. 58–65.
- JABAIAI B., BESACIER L. & LEFEVRE F. (2013). Comparison and combination of lightly supervised approaches for language portability of a spoken language understanding system. *IEEE TASLP*, **21**(3), 636–648.
- JIANG W., LIU Q. & LÜ Y. (2011). Relaxed cross-lingual projection of constituent syntax. In *CoNLL*.
- KIM S., TOUTANOVA K. & YU H. (2012). Multilingual named entity recognition using parallel data and metadata from wikipedia. In *Proceedings of the 50th Annual Meeting of the ACL*, volume 1, p. 694–702.

- KOHOMBAN U. S. & LEE W. S. (2005). Learning semantic classes for word sense disambiguation. In *Proceedings of the 43rd Annual Meeting on ACL*, p. 34–41.
- KUCERA H. & FRANCIS W. (1979). A standard corpus of present-day edited american english, for use with digital computers (revised and amplified from 1967 version).
- LUONG T., PHAM H. & MANNING C. D. (2015). Bilingual word representations with monolingual quality in mind. In *Proceedings of the 1st Workshop on VSMNLP*, p. 151–159.
- MIKOLOV T., CHEN K., CORRADO G. & DEAN J. (2013). Efficient Estimation of Word Representations in Vector Space. In *The Workshop Proceedings of ICLR 2013*.
- MIKOLOV T., KARAFIÁT M., BURGET L., CERNOCKÝ J. & KHUDANPUR S. (2010). Recurrent neural network based language model. In *INTERSPEECH 2010*, p. 1045–1048.
- MILLER G. A., LEACOCK C., TENGI R. & BUNKER R. T. (1993). A semantic concordance. In *Proceedings of the workshop on HLT*, p. 303–308.
- NAVIGLI R., JURGENS D. & VANNELLA D. (2013). Semeval-2013 task 12 : Multilingual word sense disambiguation. In *SEM 2013*, volume 2, p. 222–231.
- NAVIGLI R. & LAPATA M. (2010). An experimental study of graph connectivity for unsupervised word sense disambiguation. *Pattern Analysis and Machine Intelligence*, **32**(4), 678–692.
- NAVIGLI R. & PONZETTO S. P. (2012). Babelnet : The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, **193**, 217–250.
- OCH F. J. & NEY H. (2000). Improved statistical alignment models. In *ACL*, p. 440–447.
- PADÓ S. & LAPATA M. (2005). Cross-linguistic projection of role-semantic information. In *HLT and EMNLP*.
- PADO S. & PITEL G. (2007). Annotation précise du français en sémantique de rôles par projection cross-linguistique. *TALN-07*.
- PETERS W., PETERS I. & VOSSEN P. (1998). Automatic sense clustering in eurowordnet.
- PETROV S., DAS D. & McDONALD R. (2012). A universal part-of-speech tagset. In *LREC'12*.
- RUMELHART D. E., HINTON G. E. & WILLIAMS R. J. (1985). *Learning internal representations by error propagation*. Rapport interne, DTIC Document.
- SCHMID H. (1995). Treetagger: a language independent part-of-speech tagger. *Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart*, **43**, 28.
- SCHWAB D., GOULIAN J., TCHECHMEDJIEV A. & BLANCHON H. (2012). Ant colony algorithm for the unsupervised word sense disambiguation of texts : Comparison and evaluation. In *COLING*, p. 2389–2404.
- SUNDERMEYER M., OPARIN I., GAUVAIN J.-L., FREIBERG B., SCHLUTER R. & NEY H. (2013). Comparison of feedforward and recurrent neural network language models. In *ICASSP*, p. 8430–8434 : IEEE.
- TÄCKSTRÖM O., McDONALD R. & USZKOREIT J. (2012). Cross-lingual word clusters for direct transfer of linguistic structure. In *NAACL-HLT*, p. 477–487.
- TITOV I. & KLEMENTIEV A. (2012). Crosslingual induction of semantic roles. In *ACL*, volume 1, p. 647–656.
- WISNIEWSKI G., PÉCHEUX N., GAHBICHE-BRAHAM S. & YVON F. (2014). Cross-lingual part-of-speech tagging through ambiguous learning. In *EMNLP*, volume 14, p. 1779–1785.
- XIAO M. & GUO Y. (2015). Annotation projection-based representation learning for cross-lingual dependency parsing. *CoNLL 2015*, p. 73.
- YAROWSKY D., NGAI G. & WICENTOWSKI R. (2001). Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the first international conference on HLT*, p. 1–8.
- YE P. & BALDWIN T. (2007). Melb-ylb : Preposition sense disambiguation using rich semantic features. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, p. 241–244.
- ZENNAKI O., SEMMAR N. & BESACIER L. (2015a). Unsupervised and lightly supervised part-of-speech tagging using recurrent neural networks. In *PACLIC 29*.
- ZENNAKI O., SEMMAR N. & BESACIER L. (2015b). Utilisation des réseaux de neurones récurrents pour la projection interlingue d'étiquettes morpho-syntaxiques à partir d'un corpus parallèle. In *TALN 2015*.