WoNeF : amélioration, extension et évaluation d'une traduction française automatique de WordNet

Quentin Pradet, Jeanne Baguenier Desormeaux, Gaël de Chalendar, Laurence Danlos

▶ **To cite this version:**

Quentin Pradet, Jeanne Baguenier Desormeaux, Gaël de Chalendar, Laurence Danlos. WoNeF : amélioration, extension et évaluation d'une traduction française automatique de WordNet. TALN 2013 - 20ème conférence du Traitement Automatique du Langage Naturel, Jun 2013, Les Sables d'Olonne, France. pp.76-89. cea-00932340

HAL Id: cea-00932340
https://cea.hal.science/cea-00932340

Submitted on 16 Jan 2014

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

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

WoNeF : amélioration, extension et évaluation d'une traduction française automatique de WordNet

Quentin Pradet[1] Jeanne Baguenier-Desormeaux[1]
Gaël de Chalendar[1] Laurence Danlos[2]
(1) CEA, LIST, Laboratoire Vision et Ingénierie des Contenus,
Gif-sur-Yvette, F-91191, France;
(2) Univ Paris Diderot, Sorbonne Paris Cité, ALPAGE, UMR-I 001 INRIA
{quentin.pradet,gael.de-chalendar}@cea.fr
laurence.danlos@linguist.univ-paris-diderot.fr

RÉSUMÉ

Identifier les sens possibles des mots du vocabulaire est un problème difficile demandant un travail manuel très conséquent. Ce travail a été entrepris pour l'anglais : le résultat est la base de données lexicale WordNet, pour laquelle il n'existe encore que peu d'équivalents dans d'autres langues. Néanmoins, des traductions automatiques de WordNet vers de nombreuses langues cibles existent, notamment pour le français. JAWS est une telle traduction automatique utilisant des dictionnaires et un modèle de langage syntaxique. Nous améliorons cette traduction, la complétons avec les verbes et adjectifs de WordNet, et démontrons la validité de notre approche via une nouvelle évaluation manuelle. En plus de la version principale nommée WoNeF, nous produisons deux versions supplémentaires : une version à haute précision (93% de précision, jusqu'à 97% pour les noms), et une version à haute couverture contenant 109 447 paires (littéral, synset).

ABSTRACT

WoNeF, an improved, extended and evaluated automatic French translation of WordNet

Identifying the various possible meanings of each word of the vocabulary is a difficult problem that requires a lot of manual work. It has been tackled by the WordNet lexical semantics database in English, but there are still few resources available for other languages. Automatic translations of WordNet have been tried to many target languages such as French. JAWS is such an automatic translation of WordNet nouns to French using bilingual dictionaries and a syntactic langage model. We improve the existing translation precision and coverage, complete it with translations of verbs and adjectives and enhance its evaluation method, demonstrating the validity of the approach. In addition to the main result called WoNeF, we produce two additional versions : a high-precision version with 93% precision (up to 97% on nouns) and a high-coverage version with 109,447 (literal, synset) pairs.

MOTS-CLÉS : WordNet, désambiguïsation lexicale, traduction, ressource.

KEYWORDS: WordNet, Word Sense Disambiguation, translation, resource.

© ATALA

1 Introduction

WordNet est une base de données lexicale en développement depuis les années 80 (Fellbaum, 1998). Cette base est organisée autour du concept de **synset** (ensemble de synonymes), chaque synset représentant un sens très précis à l'aide d'une définition et d'un certain nombre de mots que nous nommons littéraux. Ces synsets sont liés par différentes relations sémantiques telles que la méronymie et l'hyponymie. Malgré des défauts reconnus (Boyd-Graber *et al.*, 2006) principalement liés à la granularité trop fine des sens, WordNet reste une ressource extrêmement utile et reproduire ce travail pour d'autres langues serait coûteux et difficile à maintenir. Et malgré quelques problèmes théoriques, (Fellbaum et Vossen, 2007; de Melo et Weikum, 2008) montrent que traduire WordNet en gardant sa structure et ses synsets mène à des ressources linguistiques utiles.

Les traductions automatiques de WordNet emploient une approche dite d'extension (*extend approach*) : la structure de WordNet est préservée et seuls les littéraux sont traduits. Trois techniques principales représentent cette approche dans la littérature. La plus simple utilise des dictionnaires bilingues pour faciliter le travail des lexicographes qui filtrent ensuite manuellement les entrées proposées (Vossen, 1998; Pianta *et al.*, 2002; Tufis *et al.*, 2004). Une deuxième méthode de traduction utilise des corpus parallèles, ce qui évite l'utilisation de dictionnaires qui peuvent entraîner un biais lexicographique. (Dyvik, 2004) représente cette méthode en s'appuyant sur des *back-translations* entre le norvégien et l'anglais, alors que (Sagot et Fišer, 2008) combinent un lexique multilingue et les différents WordNets de BalkaNet comme autant de sources aidant à la désambiguïsation. Enfin, plus récemment, des ressources telles que Wikipédia ou le Wiktionnaire ont été explorées. Grâce aux nombreux liens entre les différentes langues de ces ressources, il est possible de créer de nouveaux wordnets (de Melo et Weikum, 2009; Navigli et Ponzetto, 2010) ou d'améliorer des wordnets existants (Hanoka et Sagot, 2012).

Concernant le français, l'EuroWordNet (Vossen, 1998) est la première traduction française de WordNet. C'est une ressource d'une couverture limitée qui demande des améliorations significatives avant de pouvoir être utilisée (Jacquin *et al.*, 2007), et qui n'est ni libre ni librement accessible. WOLF est une seconde traduction initialement construite à l'aide de corpus parallèles (Sagot et Fišer, 2008) et étendue depuis avec différentes techniques (Apidianaki et Sagot, 2012). WOLF est distribué sous une licence libre compatible avec la LGPL et c'est aujourd'hui le WordNet français standard. Enfin, JAWS (Mouton et de Chalendar, 2010) est une traduction des noms de WordNet développée à l'aide de dictionnaires bilingues et d'un modèle de langue syntaxique.

Nos travaux étendent et améliorent les techniques utilisées dans JAWS et l'évaluent à l'aide d'une adjudication de deux annotateurs. Le résultat de ce travail est WoNeF [1]. Il se décline en trois versions pour répondre à différents besoins. Le WoNeF principal a un F-score de 70.9%, une autre version a une précision de 93.3%, et une dernière contient 109 447 paires (littéral, synset).

L'approche de JAWS consiste à combiner des sélecteurs variés permettant de choisir les traductions adaptées à chaque synset (section 2). Les contributions principales de cet article sont l'amélioration de JAWS et sa complétion en ajoutant les verbes et les adjectifs (section 3) et son évaluation (sections 4 et 5). Cette évaluation se fait à travers une adjudication elle-même validée par la mesure de l'accord inter-annotateur, ce qui montre la validité de l'approche par extension pour traduire WordNet.

1. Ce travail a été en partie financé par le projet ANR ASFALDA ANR-12-CORD-0023.

© ATALA

2 JAWS

2.1 Processus de traduction

(Mouton et de Chalendar, 2010) ont conçu JAWS comme un algorithme faiblement supervisé qui ne demande aucune donnée annotée manuellement. Pour traduire un wordnet source, JAWS s'appuie sur un dictionnaire bilingue et un modèle de langue syntaxique pour le langage cible.

Le dictionnaire bilingue est une concaténation du dictionnaire bilingue SCI-FRAN-EurADic [2] et des liens entre les Wiktionnaires français et anglais [3]. Le modèle de langue syntaxique a été entraîné sur un grand corpus extrait du web (Grefenstette, 2007). Le corpus a été analysé par LIMA (Besançon *et al.*, 2010), une chaîne d'analyse linguistique ici utilisée comme un analyseur syntaxique à base de règles produisant des dépendances syntaxiques fines. Pour une relation donnée r et un mot x, le modèle de langue indique quels sont les 100 premiers mots co-occurrant le plus fréquemment avec x dans la relation r. Avec le mot *avion* et la relation de complément du nom, le mot *billet* modifie le plus *avion* : *billet d'avion* est fréquent dans le corpus. Le modèle de langue ici présenté peut-être visualisé sur `http://www.kalisteo.fr/demo/semanticmap/index.php`.

Grâce aux dictionnaires, JAWS n'a pas besoin de sélectionner les littéraux de chaque synset parmi l'ensemble du vocabulaire mais seulement parmi un petit nombre de candidats (9 en moyenne). Le processus de traduction se fait en trois étapes :

1. Créer un wordnet vide : la structure de WordNet est préservée, mais les synsets eux-mêmes n'ont pas de littéraux associés.
2. Choisir les traductions les plus faciles parmi les candidats des dictionnaires pour commencer à remplir JAWS.
3. Étendre JAWS de manière incrémentale en utilisant le modèle de langue, les relations entre synsets et le JAWS déjà existant.

Sélecteurs initiaux Quatre algorithmes que nous nommons sélecteurs initiaux choisissent des traductions correctes parmi celles qui sont proposées par les dictionnaires. Premièrement, les mots qui apparaissent dans un seul synset ne sont pas ambigüs et il suffit d'ajouter toutes leurs traductions au WordNet français : c'est le sélecteur par monosémie. C'est le cas de *grumpy* : toutes ses traductions sont validées dans le synset où il apparaît. Deuxièmement, le sélecteur par unicité identifie les mots n'ayant qu'une seule traduction et la valident dans tous les synsets où elle est présente. Les cinq synsets contenant *pill* en anglais sont ainsi complétés avec *pilule*. Un troisième sélecteur vise à traduire les mots qui ne sont pas dans le dictionnaire en utilisant directement la traduction anglaise : c'est le sélecteur des transfuges. Un quatrième sélecteur utilise la distance d'édition de Levenshtein : si la distance entre un mot anglais et sa traduction est petite, on peut considérer que c'est le même sens (c'est le cas par exemple pour *portion* ou encore *university*), malgré l'existence de certains faux amis. Ces quatre sélecteurs produisent une première version du WordNet français qui contient assez de traductions pour pouvoir ensuite utiliser le modèle de langue et continuer de compléter les synsets.

2. `http://catalog.elra.info/product_info.php?products_id=666`
3. `http://www.wiktionary.org/`

© ATALA

Expansion de JAWS JAWS étant partiellement rempli, une nouvelle étape d'expansion tire parti des relations entre les synsets de WordNet pour valider de nouvelles traductions. Par exemple, si :

– un synset S1 est méronyme d'un synset S2 dans WordNet,
– il existe un contexte où un littéral dans S1 est méronyme d'un littéral candidat C dans S2,

alors ce littéral est considéré comme correct. La tâche de traduction est ainsi réduite à une tâche de comparaison entre d'une part les relations lexicales entre les synsets de WordNet et d'autre part les relations lexicales entre les lexèmes du français.

Prenoms l'exemple de *quill* qui peut se traduire par *piquant* ou *plume* (Figure 1). Dans WordNet, *quill* est méronyme de *porcupine* qui a déjà été traduit par *porc-épic* par un sélecteur initial. Dans le modèle de langue, *piquant* fait partie des compléments du noms de *porc-épic* mais ce n'est pas le cas de *plume*. Ici, la relation de complément du nom implique la méronymie et c'est donc *piquant* qu'il faut choisir comme la traduction correcte de *quill*. Le modèle de langue a permis la désambiguïsation parmi les deux traductions possibles.

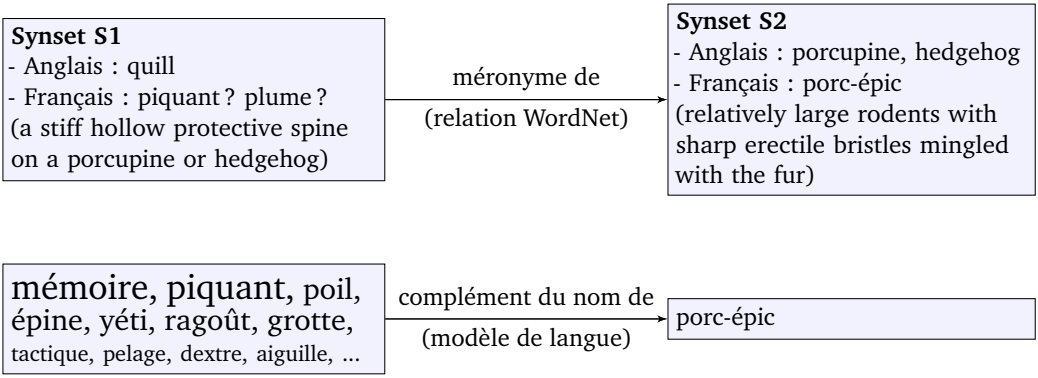

FIGURE 1 – Traduction via la relation de méronymie de partie.

Un problème potentiel avec cette approche est que la relation de complément du nom n'est pas limitée à la méronymie. Par exemple, le mot *mémoire* qui apparaît dans le modèle de langue (Figure 1) vient d'un livre intitulé *Mémoires d'un porc-épic*. Heureusement, *mémoire* n'est pas dans les candidats de *quill* et ne peut pas être choisi comme une traduction. Paradoxalement, le modèle de langue ne peut pas choisir entre deux mots très différents, mais est capable de choisir la traduction correcte d'un mot polysémique. Alors que traduire WordNet automatiquement avec un dictionnaire ou un modèle de langue syntaxique est impossible, combiner les deux ressources permet de résoudre le problème.

Chaque sélecteur suit le même principe que le sélecteur par méronymie de partie et traduit de nouveaux synsets en identifiant les relations entre lexèmes via le modèle de langue syntaxique. La correspondance entre la relation de complément du nom et la relation de méronymie est directe, mais ce n'est pas le cas pour les autres relations : il n'y a par exemple pas de relation syntaxique qui exprime directement la synonymie entre deux lexèmes. Pour ces relations, il est nécessaire d'employer soit des motifs lexicaux (Hearst, 1992) soit des relations syntaxiques de deuxième ordre (Lenci et Benotto, 2012). Ce sont ces dernières, aussi nommées relations paradigmatiques,

© ATALA

que JAWS utilise. Pour la synonymie, si deux mots partagent les mêmes co-occurents dans une relation syntaxique donnée, alors ils peuvent être synonymes dans ce contexte. Pour les noms, les relations syntaxiques qui donnent les meilleurs résultats sont les relations de complément du nom, d'objet du verbe et d'apposition. Concrètement, si deux noms qui modifient les mêmes noms sont les objets des mêmes verbes ou sont apposés aux mêmes noms, alors il est probable qu'ils soient synonymes et si l'un des deux est déjà dans un synset, alors on peut y ajouter le second. Par exemple, *avant-propos* et *préface* partagent les mêmes compléments du noms : *livre, édition, ouvrage*. Le sélecteur par synonymie peut ajouter *avant-propos* une fois que le littéral *préface* est dans JAWS. (Mouton et de Chalendar, 2010; Mouton, 2011) décrivent d'autres sélecteurs exploitant notamment les relations d'hyperonymie et d'hyponymie.

2.2 Limites de JAWS

JAWS souffre d'un certain nombre de limites. Avant tout, il ne contient que des noms, ce qui empêche de l'utiliser dans de nombreuses applications. Ensuite, la façon dont il a été évalué rend difficile tout jugement sur sa qualité. En effet, JAWS a été évalué en le comparant à l'EuroWordNet du français et à WOLF 0.1.4 (qui date de 2008). Ces deux WordNets du français ne sont pas des annotations de références : ils souffrent soit d'une précision limitée soit d'une couverture limitée.

M&C ont décidé de compléter cette évaluation limitée par une évaluation manuelle des littéraux n'existant pas dans WOLF, mais elle n'a été faite que sur 120 paires (littéral, synset). La précision de JAWS est évaluée à 67,1% (Mouton, 2011), ce qui est plus bas que celle de WOLF 0.1.4 et considérablement plus bas que la précision de WOLF 1.0b [4]. Ce score, même bas, est à prendre avec précaution étant donné la taille de l'échantillon de test : l'intervalle de confiance est d'environ 25%. Une autre limite de JAWS est qu'il ne contient qu'une seule et unique ressource qui ne correspond pas à tous les besoins.

À notre connaissance, les traductions automatiques de WordNet actuelles n'existent qu'en une seule version où les auteurs décident eux-mêmes quelle métrique optimiser. Nous fournissons aussi une telle version, mais ajoutons aussi deux ressources qui peuvent servir des besoins différents. Même si notre WoNeF à haute précision est petit, il peut être utilisé comme une annotation de référence et servir pour entraîner un système d'apprentissage. Une ressource à haute couverture peut servir de base à une correction manuelle ou servir pour une intersection à d'autres ressources, ce qui est la raison pour laquelle nous en fournissons une aussi.

3 WoNeF : un JAWS nominal amélioré

Cette section présente les trois améliorations essentielles qui ont étés apportées à JAWS. Un changement non détaillé est celui qui a mené à une meilleure rapidité d'exécution : JAWS se construit en plusieurs heures contre moins d'une minute pour WoNeF, ce qui a facilité les expérimentations.

4. Nous remercions Benoît Sagot pour nous avoir fourni cette version préliminaire de WOLF 1.0.

© ATALA

3.1 Sélecteurs initiaux

Les sélecteurs initiaux de JAWS ne sont pas optimaux. Alors que les sélecteurs par monosémie et par unicité sont conservés, nous avons changé les autres sélecteurs. Premièrement, le sélecteur des transfuges est supprimé : sa précision était très basse, même pour les noms.

Deuxièmement, un nouveau sélecteur considère les traductions candidates provenant de plusieurs mots anglais différents dans un synset donné : c'est le sélecteur par sources multiples. Par exemple, dans le synset *line, railway line, rail line (the road consisting of railroad track and roadbed)*, les littéraux français *ligne de chemin de fer* et *voie* sont des traductions à la fois de *line* et *railway line*, et sont donc choisis comme traductions.

Troisièmement, le sélecteur de la distance de Levenshtein a été amélioré. 28% du vocabulaire anglais est d'origine française (Finkenstaedt *et al.*, 1973), et l'anglicisation a produit des transformations prévisibles. Il est possible d'appliquer ces mêmes transformations aux littéraux candidats français, et seulement alors d'appliquer la distance de Levenshtein. Nous commençons par supprimer les accents, puis appliquons différentes opérations. Par exemple, l'inversion des lettres "r" et "e" prend en compte (*order*/*ordre*) et (*tiger*/*tigre*) [5]. Toutes les transformations ne s'appliquent qu'à la fin des mots : *-que* est transformé en *-k* ou *-c* (*marque* devient *mark*), *-té* vers *-ty* (*extremité* devient *extremity*), etc. Les faux-amis ne sont toujours pas explicitement pris en compte.

3.2 Apprentissage de seuils

Dans JAWS, chaque littéral anglais ne peut avoir qu'une traduction française correspondante. La traduction choisie est celle qui a le meilleur score, indépendamment des scores des traductions moins bien notées. Cela a pour effet de rejeter des candidats valides et d'accepter des candidats erronés. Par exemple, JAWS n'inclut pas *particulier* au synset *(a human being) "there was too much for one person to do"* parce que *personne* est déjà inclus avec un score supérieur.

Dans WoNeF, nous avons donc appris un seuil pour chaque partie du discours et sélecteur. Nous avons d'abord généré les scores pour toutes les paires (littéral, synset) candidates, puis trié ces paires par score. Les 12 399 paires présentes dans l'évaluation manuelle associée à WOLF 1.0b (notre ensemble d'apprentissage) ont été jugées correctes tandis que les paires n'y étant pas ont été jugées erronées. Nous avons ensuite calculé les seuils maximisant la précision et le F-score. Le seuil qui maximise le F-score est utilisé dans les ressources à haut F-score et à haute couverture, tandis que le seuil maximisant la précision est utilisé dans la ressource à haute précision.

Une fois que ces seuils sont définis, les sélecteurs choisissent tous les candidats au-dessus du nouveau seuil, ce qui a deux effets positifs :

– des candidats valides ne sont plus rejetés simplement parce qu'un meilleur candidat est aussi sélectionné, ce qui améliore à la fois le rappel et la couverture.
– les candidats invalides qui étaient jusque-là acceptés sont maintenant rejetés grâce au seuil plus strict : la précision s'en retrouve augmentée.

5. La distance de Damerau-Levenshtein qui prend en compte les inversions n'importe-où dans un mot (Damerau, 1964) a donné de moins bons résultats.

© ATALA

3.3 Vote

Après l'application des différents sélecteurs, notre WordNet est large mais contient des synsets bruités. Comme toutes les traductions automatiques de WordNet, WoNeF doit alors être nettoyé (Sagot et Fišer, 2012). Dans WoNeF, le bruit provient de différents facteurs :

– les sélecteurs essaient d'inférer des informations sémantiques à partir d'une analyse syntaxique sans prendre en compte toute la complexité de l'interface syntaxe-sémantique,
– l'analyseur syntaxique produit lui-même des résultats bruités,
– le modèle de langue syntaxique est produit à partir d'un corpus extrait du web lui-même bruité (texte mal écrit, contenu non textuel, phrases non françaises) et n'est pas une « distribution idéale » (Copestake et Herbelot, 2012),
– les traductions déjà choisies sont considérées comme valides dans les étapes suivantes alors que ce n'est pas toujours le cas.

Pour la ressource haute-précision, il fallait donc un moyen de ne garder que les littéraux pour lesquels les sélecteurs étaient les plus confiants. Étant donné que, contrairement à JAWS, plusieurs sélecteurs peuvent choisir une même traduction (sous-section 3.2), notre solution est simple et efficace : les traductions validées par un bon sélecteur ou par plusieurs sélecteurs moyens sont conservées tandis que les autres sont supprimées. Ce principe de vote est aussi appelé méthode d'ensemble en apprentissage automatique. Les sélecteurs performants varient d'une partie du discours à une autre : le choix est fait sur un ensemble de développement contenant 10% de notre référence.

Cette opération de nettoyage ne conserve que 18% des traductions (de 87 757 paires (littéral, synset) à 15 625) mais la précision grimpe de 68,4% à 93,3%. Cette ressource à haute précision peut être utilisée comme donnée d'entraînement. Un défaut classique des méthodes de vote est de ne choisir que des exemples faciles et peu intéressants, mais la ressource obtenue ici est équilibrée entre les synsets ne contenant que des mots monosémiques et d'autres synsets contenant des mots polysémiques et plus difficiles à désambiguïser (section 5.2).

3.4 Extension aux verbes, adjectifs et adverbes

Les travaux sur JAWS ont commencé par les noms parce qu'ils représentent 70% des synsets dans WordNet. Nous avons continué ce travail sur les autres parties du discours qui sont aussi importantes pour examiner le sens d'un texte donné : verbes, adjectifs et adverbes. Les sélecteurs génériques ont ici été modifiés, mais il s'agira dans le futur d'implémenter des sélecteurs prenant en compte les spécificités des différentes parties du discours dans WordNet.

Verbes Les sélecteurs choisis pour les verbes sont le sélecteur par unicité et par monosémie. En effet, la distance de Levenshtein a donné des résultats médiocres pour les verbes : seuls 25% des verbes choisis par ce sélecteur étaient des traductions correctes. Concernant les sélecteurs syntaxiques, seul le sélecteur par synonymie a donné de bons résultats, alors que le sélecteur par hyponymie avait les performances d'un classifieur aléatoire.

© ATALA

Adjectifs Les adjectifs sont traduits de la même manière que les noms : tout d'abord un nombre limité de sélecteurs initiaux remplit un WordNet vide, puis les sélecteurs syntaxiques complètent cette traduction avec le modèle de langue syntaxique. Tous les sélecteurs initiaux sont ici choisis, et le sélecteur syntaxique choisi est le sélecteur par synonymie. Ils ont donné de bons résultats qui sont présentés dans la section 5.3.

Adverbes Nous n'avons pas d'annotation de référence pour les adverbes, ce qui explique qu'ils ne sont pas inclus dans WoNeF : nous ne pouvons évaluer leur précision. Cependant, la comparaison avec WOLF (section 5.4) montre que les adverbes ont de meilleurs résultats que les autres parties du discours, ce qui laisse penser que c'est une ressource de qualité. C'est une ressource aussi très complémentaire : 87% des adverbes proposés ne sont pas dans WOLF. Une fusion entre WoNeF et WOLF aurait trois fois plus d'adverbes que WOLF seul.

4 WoNeF : un JAWS évalué

4.1 Développement d'une annotation de référence

L'évaluation de JAWS souffre d'un certain nombre de limites (section 2.2). Pour évaluer rigoureusement notre propre traduction de WordNet, nous avons produit une annotation de référence. Pour chaque partie du discours, 300 synsets ont été annotés par deux annotateurs locuteurs natifs du français. Pour chaque traduction candidate fournie par nos dictionnaires, il fallait décider si oui ou non elle appartenait au synset. Puisque les dictionnaires ne proposent pas de candidats pour tous les synsets et que certains synsets n'ont pas de candidat valable, le nombre réel de synsets non vides est inférieur à 300 (section 4.2).

Durant l'annotation manuelle, nous avons rencontré une difficulté importante découlant de la tentative de traduire WordNet dans une autre langue. Dans le cas de l'anglais vers le français, la plupart des difficultés proviennent des verbes et adjectifs figurant dans une collocation. Dans WordNet, ils peuvent être regroupés d'une manière qui fait sens en anglais, mais qui ne se retrouve pas directement dans une autre langue. Par exemple, l'adjectif *pointed* est le seul élément d'un synset défini comme *direct and obvious in meaning or reference ; often unpleasant ; "a pointed critique" ; "a pointed allusion to what was going on" ; "another pointed look in their direction"*. Ces exemples se traduiraient par trois adjectifs différents en français : *une critique dure*, *une allusion claire* et *un regard appuyé*. Il n'existe pas de solution satisfaisante lors de la traduction d'un tel synset : le synset résultant contiendra soit trop soit trop peu de traductions. Nous avons décidé de ne pas traduire ces synsets dans notre annotation manuelle. Ces problèmes de granularité concernent 3% des synsets nominaux, 8% des synsets verbaux et 6% des synsets adjectivaux. Actuellement, WoNeF ne détecte pas de tels synsets.

L'autre difficulté principale découle de traductions manquantes, ce qui peut être considéré comme un défaut de nos ressources. Les sens rares d'un mot sont parfois absents. Par exemple, le sens *to catch* du jeu du chat (ou du loup) et le sens *coat with beaten egg* du verbe *to egg* ne sont pas présents. Aucun de ces sens ne sont dans les synsets les polysémiques (définis à la section 5.2), ce qui confirme que cela ne se produit que pour les sens rares. Pourtant, WoNeF pourrait être amélioré en utilisant des dictionnaires spécifiques pour, par exemple, les espèces (comme dans (Sagot et Fišer, 2008)), les termes médicaux, les entités nommées (en utilisant Wikipedia) et ainsi

 © ATALA

de suite. Un autre exemple est celui des adjectifs de jugement : il n'y a pas de bonne traduction de *weird* en français. Même si la plupart des dictionnaires fournissent *bizarre* comme traduction, on ne retrouve pas dans *bizarre* l'aspect *stupide* du mot *weird* : les deux adjectifs ne sont pas substituables dans tous les contextes, ce qui est un problème si l'on considère que le sens d'un synset doit être conservé par la traduction.

4.2 Accord inter-annotateurs

Malgré les difficultés mentionnées ci-dessus, l'annotation résultante a été validée par la mesure de l'accord inter-annotateurs, qui montre que l'approche par extension pour la création de nouveaux wordnets est valide et peut produire des ressources utiles. Deux annotateurs humains, auteurs de cet article, respectivement linguiste informaticien et informaticien linguiste, ont annoté de façon indépendante les mêmes synsets choisis au hasard pour chaque partie du discours. Ils ont utilisé WordNet pour examiner les synsets voisins, le dictionnaire Merriam-Webster, le TLFi (Pierrel, 2003) et des moteurs de recherche pour attester l'utilisation des divers sens des mots considérés. Après adjudication faite par ces deux annotateurs en confrontant leurs opinions en cas de désaccord, l'annotation de référence a été formée.

	Noms	Verbes	Adjectifs
Kappa de Fleiss	0.715	0.711	0.663
Synsets non-vides	270	222	267
Candidats par synset	6.22	14.50	7.27

TABLE 1 – Accord inter-annotateurs sur l'annotation de référence

La table 1 montre l'accord inter-annotateur évalué par le kappa de Fleiss pour les trois parties du discours annotées. Même s'il s'agit d'une métrique discutée (Powers, 2012), toutes les tables d'évaluation existantes considèrent ces scores comme étant suffisamment élevés pour décrire cet accord inter-annotateurs comme « bon » (Gwet, 2001), ce qui nous permet de dire que notre annotation de référence est de bonne qualité. L'approche par extension pour la traduction de WordNet est elle aussi validée.

5 Résultats

Nous présentons dans cette section les résultats de WoNeF. Nous commençons par décrire les résultats après l'application de l'étape des sélecteurs initiaux seulement puis ceux de la ressource complète. Notre annotation de référence est découpée en deux parties : 10% des littéraux forment l'ensemble de développement utilisé pour choisir les sélecteurs s'appliquant aux différentes versions de WoNeF, tandis que les 90% restant forment l'ensemble de test servant à l'évaluation. Précision et rappel sont calculés sur l'intersection des synsets présents dans WoNeF et l'annotation de référence considérée, que ce soit l'ensemble de test de notre propre adjudication (sections 5.1 à 5.3) ou WOLF (section 5.4). Par exemple, la précision est la fraction des paires (littéral, synset) correctes au sein de l'intersection en question.

© ATALA

5.1 Sélecteurs initiaux

Pour les noms, les verbes et les adjectifs, nous avons calculé l'efficacité de chaque sélecteur initial sur notre ensemble de développement, et utilisé ces données pour déterminer ceux qui doivent être inclus dans la version ayant une haute précision, celle ayant un F-score élevé et celle présentant une grande couverture. Les scores ci-dessous sont calculés sur l'ensemble de test, plus grand et plus représentatif.

	P	R	F1	C
monosémie	71.5	76.6	74.0	54 499
unicité	91.7	63.0	75.3	9 533
sources multiples	64.5	45.0	53.0	27 316
Levenshtein	61.9	29.0	39.3	20 034
haute précision	**93.8**	50.1	65.3	13 867
haut F-score	71.1	**72.7**	**71.9**	82 730
haute couverture	69.0	69.8	69.4	**90 248**

TABLE 2 – Sélecteurs initiaux sur l'ensemble des traductions (noms, verbes et adjectifs). La couverture C est le nombre total de paires (littéral, synset).

La table 2 montre les résultats de cette opération. La couverture donne une idée de la taille des ressources. En fonction des objectifs de chaque ressource, les sélecteurs initiaux choisis seront différents. Différents sélecteurs peuvent choisir plusieurs fois une même traduction, ce qui explique que la somme des couvertures soit supérieure à la couverture de la ressource à haute couverture. Fait intéressant non visible dans la table, le sélecteur le moins efficace pour les verbes est la distance de Levenshtein avec une précision de l'ordre de 25% : les faux amis semblent être plus nombreux pour les verbes.

5.2 Résultats globaux

Nous nous intéressons maintenant aux résultats globaux (Table 3). Ils comprennent l'application des sélecteurs initiaux et des sélecteurs syntaxiques. Le mode de haute précision applique également un vote (section 3.3). Comme pour la table précédente, la couverture C indique le nombre de paires (littéral, synset).

	Tous synsets				Synsets BCS			
	P	R	F1	C	P	R	F1	C
haute précision	**93.3**	51.5	66.4	15 625	**90.4**	36.5	52.0	1 877
haut F-score	68.9	73.0	**70.9**	88 736	56.5	62.8	**59.1**	14 405
haute couverture	60.5	**74.3**	66.7	**109 447**	44.5	**66.9**	53.5	**23 166**

TABLE 3 – Résultats globaux : tous les synsets et synsets BCS.

Dans WordNet, les mots sont majoritairement monosémiques, mais c'est une petite minorité de mots polysémiques qui est la plus représentée dans les textes. C'est justement sur cette minorité que nous souhaitons produire une ressource de qualité. Pour l'évaluer, nous utilisons la liste des

© ATALA

synsets **BCS** (Basic Concept Set) fournie par le projet BalkaNet (Tufis *et al.*, 2004). Cette liste contient les 8 516 synsets lexicalisés dans six traductions différentes de WordNet, et représente les synsets les plus fréquents et ceux qui comportent le plus de mots polysémiques. Les résultats montrent le nombre de synsets BCS pour les ressources à haut F-score et haute couverture. Alors que les ressources à haut F-score et à haute couverture perdent en précision pour les synsets BCS, ce n'est pas le cas pour la ressource à haute précision. En effet, le mécanisme de vote rend la ressource haute-précision très robuste, et ce même pour les synsets BCS.

5.3 Résultats par partie du discours

		P	R	F1	C
haute précision	noms	**96.8**	56.6	71.4	11 294
	verbes	**68.4**	41.9	52.0	1 110
	adjectifs	**90.0**	36.7	52.2	3 221
haut F-score	noms	71.7	73.2	**72.4**	59 213
	JAWS	70.7	68.5	69.6	55 416
	verbes	48.9	**76.6**	**59.6**	9 138
	adjectifs	69.8	71.0	70.4	20 385
haute couverture	noms	61.8	**78.4**	69.1	**70 218**
	verbes	45.4	61.5	52.2	**18 844**
	adjectifs	69.8	**71.9**	**70.8**	**20 385**

TABLE 4 – Résultats par partie du discours. JAWS ne contient que des noms : il est comparé à la ressource nominale à haut F-score.

La table 4 montre les résultats détaillés pour chaque partie du discours. Concernant les noms, le mode de haute précision utilise deux sélecteurs, tous deux fondés sur la relation syntaxique de complément du nom : le sélecteur par méronymie décrit à la section 2.1, et le sélecteur par hyponymie. La ressource de haute précision pour les noms est notre meilleure ressource. La version avec le F-score optimisé a un F-score de 72,4%, ce qui garantit que peu de paires (littéral, synset) sont absentes tout en ayant une précision légèrement supérieure à celle de JAWS.

Les résultats des verbes sont moins élevés. L'explication principale est que les verbes sont en moyenne plus polysémiques dans WordNet et nos dictionnaires que les autres parties du discours : les synsets verbaux ont deux fois plus de candidats que les noms et les adjectifs (Table 1). Cela montre l'importance du dictionnaire pour limiter le nombre initial de littéraux parmi lesquels les algorithmes doivent choisir.

Le sélecteur par synonymie est le seul sélecteur syntaxique appliqué aux verbes. Il utilise les relations syntaxiques de second ordre pour trois types de dépendances syntaxiques verbales : si deux verbes partagent les mêmes objets, ils sont susceptibles d'être synonymes ou quasi-synonymes. C'est le cas des verbes *dévorer* et *manger* qui acceptent tous deux l'objet *pain*. Les autres sélecteurs syntaxiques n'ont pas été retenus pour les verbes en raison de leurs faibles résultats. En effet, alors que la détection de l'hyponymie en utilisant seulement l'inclusion de contextes a été efficace sur les noms, elle a les performances d'un classifieur aléatoire pour les verbes. Cela met en évidence la complexité de la polysémie des verbes.

© ATALA

Pour les adjectifs, comme pour les verbes, seul le sélecteur de synonymie a été appliqué. Pour les ressources à haut F-score et haute couverture, ce sont les mêmes sélecteurs (initiaux et syntaxiques) qui sont appliqués, ce qui explique que les résultats sont les mêmes. Alors que l'accord inter-annotateurs était plus bas sur les adjectifs que sur les verbes, les résultats eux sont bien meilleurs pour les adjectifs. Cela s'explique principalement par le nombre de candidats parmi lesquels sélectionner : il y en a deux fois moins pour les adjectifs. Cela met en avant l'importance des dictionnaires.

5.4 Évaluation par rapport à WOLF

	WOLF 0.1.4			WOLF 1.0b		
	pP	pR	Ajouts	pP	pR	Ajouts
Noms	50.7	40.0	9 646	73.6	46.4	6 842
Verbes	33.0	23.9	1 064	41.7	17.5	1 084
Adjectifs	41.7	46.1	3 009	64.4	53.8	3 172
Adverbes	56.2	44.4	3 061	76.5	41.9	2 835

TABLE 5 – Évaluation de la ressource à haute précision en considérant WOLF 0.1.4 et 1.0b comme des références.

Il n'est pas possible de comparer WOLF et WoNeF en utilisant notre annotation de référence : tout mot correct de WOLF non présent dans les dictionnaires pénalisera WOLF injustement. Nous avons décidé d'évaluer WoNeF en considérant WOLF 0.1.4 et WOLF 1.0b comme des références (Table 5). Les mesures ne sont pas de véritables précision et rappel puisque WOLF lui-même n'est pas entièrement validé. Le dernier article pF donnant des chiffres globaux (Sagot et Fišer, 2012) : iindique un nombre de paires autour de 77 000 pour une précision de 86% [6]. Nous appelons donc pseudo-précision (pP) le pourcentage des éléments présents dans WoNeF qui sont également présents dans WOLF, et pseudo-rappel le pourcentage d'éléments de WOLF qui sont présents dans WoNeF. Ces chiffres montrent que même si WoNeF est encore plus petit que WOLF, il s'agit d'une ressource complémentaire, surtout quand on se souvient que le WoNeF utilisé pour cette comparaison est celui présentant une précision élevée, avec une précision globale de 93,3%. Il convient également de noter que la comparaison de la différence entre WOLF 0.1.4 et WOLF 1.0b est instructive puisque elle montre l'étendue des améliorations apportées à WOLF.

La colonne « Ajouts » donne le nombre de traductions qui sont présentes dans WoNeF mais pas dans WOLF. Pour les noms, les verbes et les adjectifs, cela signifie que nous pouvons contribuer 11 098 nouvelles paires (littéral, synset) de haute précision en cas de fusion de WOLF et WoNeF, soit 94% des paires du WoNeF haute précision ce qui montre la complémentarité des approches : ce sont des littéraux différents qui sont ici choisis. Cela produira un wordnet français 13% plus grand que WOLF avec une précision améliorée. Une fusion avec la ressource de F-score élevée aurait une précision légèrement inférieure, mais fournirait 57 032 nouvelles paires (littéral, synset) par rapport à WOLF 1.0b, résultant en une fusion contenant 73 712 synsets non vides et 159 705 paires (littéral, synset), augmentant la couverture de WOLF de 56% et celle de WoNeF de 83%.

6. Les résultats détaillés pour WOLF 1.0b ne sont pas actuellement disponibles.

© ATALA

Conclusion

Dans ce travail, nous avons montré que l'utilisation d'un modèle de langue syntaxique pour identifier des relations lexicales entre des lexèmes est possible dans un environnement contraint et conduit à des résultats ayant une précision au niveau de l'état de l'art pour la tâche de traduction de WordNet. Nous offrons trois ressources différentes, chacune d'elles ayant un objectif différent. Enfin, nous fournissons une annotation de référence validée de haute qualité qui nous a permis de montrer à la fois la validité de l'approche de traduction de WordNet par extension et la validité de notre approche spécifique. Cette annotation de référence peut également être utilisée pour évaluer et développer d'autres traductions françaises de WordNet. WoNeF est disponible librement au format XML DEBVisDic [7] sur `http://wonef.fr/` sous la licence CC-BY-SA.

Les travaux futurs sur WoNeF mettront l'accent sur les verbes, les adjectifs et les adverbes, pour lesquels de nouveaux sélecteurs efficaces peuvent être envisagés pour améliorer la couverture. Par exemple, le sélecteur de similarité peut être étendu à la relation de quasi-synonymité que partagent certains adjectifs dans WordNet. En effet, la synonymie entre les adjectifs est limitée par rapport à la quasi-synonymie : alors que *fast* est le seul mot dans son synset, c'est le quasi-synonyme de 20 synsets. Puisque les techniques de sémantique distributionnelle ont plutôt tendance à identifier des quasi-synonymes plutôt que des synonymes, utiliser cette relation de WordNet pour identifier de nouveaux adjectifs fait partie de nos objectifs.

Une autre source importante d'amélioration sera l'enrichissement de notre modèle de langue syntaxique qui pourra prendre en compte les verbes pronominaux et les expressions multi-mots. Nous aimerions aussi nous orienter vers un modèle de langue continu (Le *et al.*, 2012) plus performant. Cela sera couplé avec la collecte d'un corpus issu du Web plus récent et plus grand analysé avec une version récente de notre analyseur linguistique LIMA. Cela nous permettra de mesurer l'impact de la qualité du modèle de langue sur la traduction de WordNet.

Le wordnet français WOLF a été construit en utilisant plusieurs techniques. La fusion de WOLF et de WoNeF permettra de bientôt améliorer à nouveau le statut de la traduction française de WordNet : nous travaillons avec les auteurs de WOLF afin de fusionner WOLF et WoNeF.

Références

Apidianaki, M. et Sagot, B. (2012). Applying cross-lingual WSD to wordnet development. *In LREC 2012*.

Besançon, R., de Chalendar, G., Ferret, O., Gara, F., Laib, M., Mesnard, O. et Semmar, N. (2010). LIMA : A multilingual framework for linguistic analysis and linguistic resources development and evaluation. *In LREC 2010*.

Boyd-Graber, J., Fellbaum, C., Osherson, D. et Schapire, R. (2006). Adding dense, weighted connections to wordnet. *In GWC 2006*.

Copestake, A. et Herbelot, A. (2012). Lexicalised compositionality. Unpublished draft.

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Commun. ACM*, 7(3):171–176.

7. `http://nlp.fi.muni.cz/trac/deb2/wiki/WordNetFormat`

© ATALA

de Melo, G. et Weikum, G. (2008). On the Utility of Automatically Generated Wordnets. *In GWC 2008*.

de Melo, G. et Weikum, G. (2009). Towards a universal wordnet by learning from combined evidence. *In CIKM 2009*, pages 513–522. ACM.

Dyvik, H. (2004). Translations as semantic mirrors : from parallel corpus to wordnet. *Language and computers*, 49(1):311–326.

Fellbaum, C., éditeur (1998). *WordNet : an Electronic Lexical Database*. The MIT Press.

Fellbaum, C. et Vossen, P. (2007). Connecting the universal to the specific : Towards the global grid. *Intercultural Collaboration*, pages 1–16.

Finkenstaedt, T., Wolff, D., Neuhaus, H. et Herget, W. (1973). *Ordered profusion : Studies in dictionaries and the English lexicon*, volume 13. C. Winter.

Grefenstette, G. (2007). Conquering language : Using NLP on a massive scale to build high dimensional language models from the web. *In CICLing 2007*, pages 35–49.

Gwet, K. (2001). *Handbook of inter-rater reliability*. Advanced Analytics, LLC.

Hanoka, V. et Sagot, B. (2012). Wordnet extension made simple : A multilingual lexicon-based approach using wiki resources. *In LREC 2012*.

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. *In Proceedings of the 14th conference on Computational linguistics - Volume 2*, pages 539–545. ACL.

Jacquin, C., Desmontils, E. et Monceaux, L. (2007). French EuroWordNet Lexical Database Improvements. *In CICLing 2007*, volume 4394 de *LNCS*, pages 12–22.

Le, H.-S., Allauzen, A. et Yvon, F. (2012). Continuous Space Translation Models with Neural Networks. *In NAACL-HLT 2012*, pages 39–48. ACL.

Lenci, A. et Benotto, G. (2012). Identifying hypernyms in distributional semantic spaces. *In *SEM 2012*, pages 75–79. ACL.

Mouton, C. (2011). *Ressources et méthodes semi-supervisées pour l'analyse sémantique de texte en français*. Thèse de doctorat.

Mouton, C. et de Chalendar, G. (2010). JAWS : Just Another WordNet Subset. *In TALN 2010*.

Navigli, R. et Ponzetto, S. (2010). BabelNet : Building a very large multilingual semantic network. *In ACL 2010*, pages 216–225.

Pianta, E., Bentivogli, L. et Girardi, C. (2002). MultiWordNet : developing an aligned multilingual database.

Pierrel, J. (2003). Un ensemble de ressources de référence pour l'étude du français : TLFi, Frantext et le logiciel Stella. *Revue québécoise de linguistique*, 32(1):155–176.

Powers, D. (2012). The Problem with Kappa. *In EACL 2012*, page 345.

Sagot, B. et Fišer, D. (2012). Cleaning noisy wordnets. *In LREC 2012*.

Sagot, B. et Fišer, D. (2012). Automatic Extension of WOLF. *In GWC 2012*.

Sagot, B. et Fišer, D. (2008). Building a free French wordnet from multilingual resources. *In Ontolex 2008*.

Tufis, D., Cristea, D. et Stamou, S. (2004). BalkaNet : Aims, methods, results and perspectives. a general overview. *Romanian Journal of Information Science and Technology*, 7(1-2):9–43.

Vossen, P. (1998). *EuroWordNet : a multilingual database with lexical semantic networks*. Kluwer Academic.

© ATALA