

Annotation précise du français en sémantique de rôles par projection cross-linguistique

Sebastian Padó (1), Guillaume Pitel (2)

(1) Computerlinguistik – Université de la Sarre
pado@coli.uni-sb.de

(2) Equipe TALARIS, LORIA – INRIA
Guillaume.Pitel@loria.fr

Résumé Dans le paradigme FrameNet, cet article aborde le problème de l'annotation précise et automatique de rôles sémantiques dans une langue sans lexique FrameNet existant. Nous évaluons la méthode proposée par Padó et Lapata (2005, 2006), fondée sur la projection de rôles et appliquée initialement à la paire anglais-allemand. Nous testons sa généralisabilité du point de vue (a) des langues, en l'appliquant à la paire (anglais-français) et (b) de la qualité de la source, en utilisant une annotation automatique du côté anglais. Les expériences montrent des résultats à la hauteur de ceux obtenus pour l'allemand, nous permettant de conclure que cette approche présente un grand potentiel pour réduire la quantité de travail nécessaire à la création de telles ressources dans de nombreuses langues.

Abstract This paper considers the task of the automatic induction of role-semantic annotations for new languages with high precision. To this end we test the generalisability of the language-independent, projection-based annotation framework introduced by Padó and Lapata (2005, 2006) by (a) applying it to a new, more distant, language pair (English-French), and (b), using automatic, and thus noisy, input annotation. We show that even under these conditions, high-quality role annotations for French can be obtained that rival existing results for German. We conclude that the framework has considerable potential in reducing the manual effort involved in creating role-semantic resources for a wider range of languages.

Mots-clés : Multilingue, FrameNet, Annotation sémantique automatique, Sémantique lexicale, Projection d'annotation de rôles, rôles sémantiques

Keywords: Multilingual, FrameNet, Automatic semantic annotation, Lexical semantics, Annotation projection, Semantic roles

1 Introduction

L'analyse sémantique de surface (en anglais, *shallow semantic parsing*) consiste à reconnaître les rôles sémantiques attribuables aux différents constituants d'un énoncé, sans décrire avec précision la sémantique interne de ces constituants. Les rôles sémantiques correspondent aux arguments des prédicats évoqués par certains mots, notamment les verbes. Ce type d'analyse présente un intérêt tout particulier dans les applications utilisant les informations sémantiques à grande échelle, telles que l'extraction d'information (Bouillon et al., 2000 ; Surdeanu et al., 2003), la traduction automatique (Boas, 2002) et les systèmes de question/réponse (Narayanan et Harabagiu, 2004).

Le projet FrameNet (Fillmore et *al.*, 2003) peut jouer un rôle central dans les entreprises de ce type, en mettant à disposition une ressource lexicale de grande couverture, précisément articulée autour de la notion de rôle sémantique. Dans la sémantique des frames de Fillmore (1982) sur laquelle s'appuie FrameNet, le sens prédicatif est représenté à partir de *frames*¹ qui peuvent être considérées comme des représentations schématiques de situations. Dans ce cadre, les rôles sémantiques sont appelés *frame elements* (FEs), ils sont attachés de manière unique à une frame (on compte en moyenne 7 à 8 FEs par frame). Le projet FrameNet en lui-même consiste à construire une base de données mettant en relation les frames, les lemmes qui les évoquent (*Unités Lexicales ou ULs*) et les informations détaillées sur la réalisation de surface des Fes, sous la forme de cadres syntaxiques et grammaticaux et d'exemples d'annotations sur le British National Corpus. Ces informations ont créé un fort intérêt du côté de l'analyse sémantique de surface et ont rendu possible le développement d'analyseurs automatiques pour le texte libre (initiés par Gildea et Jurafsky, 2002) qui peuvent être utilisés pour les applications citées ci-dessus.

Frame: ARRIVING (Un THEME se rapproche d'un GOAL)			
Frame Elements	THEME	The officer approached the house Amy arrived home from school After she arrived home, ...	L'officier s'approcha de la maison De l'école, Amy rentra à la maison Après son arrivée à la maison, ...
	GOAL	The officer approached the house Amy arrived home from school He had arrived there from London	L'officier approcha de la maison De l'école, Amy rentra à la maison Clarke est arrivé là de Londres
Unités Lexicales	approach.n, approach.v, arrival.n, arrive.v, come.v, crest.v, descend_(on).v, enter.v, entrance.n, entry.n, get.v, make it.v, make.v, reach.v, return.n, return.v, visit.n, visit.v	aborder.v, aboutir.v, accéder_à.v, approche.n, approcher.v, approcher_de.v, arriver.v, arrivée.n, atteindre.v, descendre_à.v, descendre_sur.v, entrer.v, entrée.n, gagner.v, parvenir.v, passer.v, rapprochement.n, regagner.v, rejoindre.v, rentrer.v, rentrée.n, retour.n, retourner.v, revenir.v, venir.v	

Tableau 1: Exemple simplifié de *frame* dans le paradigme FrameNet (français et anglais)

Le tableau 1 illustre, pour la frame ARRIVING, certaines des informations contenues dans la base FrameNet (les ULs pour le français sont tirées d'une méthode semi-automatique de construction de lexique sémantique (Pitel, 2006)). Cette frame, qui modélise une situation où un objet en mouvement se rapproche d'un endroit, possède deux FEs principaux : le thème (THEME) et le but (GOAL). D'autres FEs secondaires lui sont rattachés, parmi lesquels : MODE_OF_TRANSPORTATION, CIRCUMSTANCES et GOAL_CONDITION². Comme le montrent les exemples d'annotation, les mêmes FEs peuvent être évoqués par des constituants de différentes natures syntaxiques et grammaticales et ce sont ces informations qui seront particulièrement exploitées par les systèmes d'annotation sémantique automatique. Dans sa version 1.1 utilisée dans l'expérimentation que nous présentons, FrameNet décrit 513 frames liées à 7125 unités lexicales, avec en moyenne 7.5 FEs par frame.

Malheureusement, l'anglais est actuellement la seule langue dans laquelle une telle ressource existe à grande échelle. Un petit nombre de projets existent dans d'autres langues (allemand, espagnol et japonais principalement), mais ceux-ci n'ont pas encore atteint la maturité nécessaire à une exploitation automatique. La principale cause de ce déficit en ressources est le haut niveau d'exigence en temps et en attention requis pour l'annotation sémantique manuelle. Le fait que l'analyse sémantique de surface soit limitée à un petit nombre de langues est un obstacle important à son usage comme stratégie d'analyse générale pour le TAL. Il est donc

¹ Afin d'éviter la confusion avec les autres usages de « cadre », nous conserverons les expressions usuelles anglaises lorsqu'elles comprennent le terme *frame*, et traduisons dans les autres cas.

² De nombreux exemples de frames ainsi que des documents sur FrameNet sont accessibles sur <http://framenet.icsi.berkeley.edu>

impératif de concevoir des méthodes pour réduire l'effort nécessaire à l'amorçage de telles ressources. Dans cet article nous nous intéressons à l'induction des informations sémantiques sur les FEs, problème pour lequel (Padó et Lapata, 2005 ; 2006) – ci-après, P&L – ont suggéré l'utilisation de la *projection d'annotations*. Ce paradigme s'articule autour de l'exploitation de ressources parallèles entre des langues L1 et L2, où L1 est dotée en ressources de type FrameNet et L2 ne l'est pas. En supposant que l'on puisse obtenir une analyse sémantique pour le côté L1, la projection d'annotations consiste à *recopier* simplement les annotations de L1 sur le côté L2. et permet ainsi de réutiliser le travail manuel effectué sur L1.

P&L ont montré l'efficacité de cette méthode en utilisant FrameNet pour induire l'annotation de FEs en allemand, mais leur étude a été limitée sur deux aspects:

1. Seule une langue cible est considérée : l'allemand. Johansson et Nugues (2006) rapportent, avec une stratégie identique, un succès similaire pour le suédois. Cependant, le suédois et l'allemand étant deux langues germaniques, typologiquement proches de l'anglais, l'approche dans le cas général n'est pas validée.
2. P&L n'ont pris en compte que le cas où le côté L1 est annoté manuellement, ce qui ne permet pas de généraliser les résultats pour un passage à grande échelle. Bien que Johansson et Nugues (2006) l'aient fait pour une source automatique, les différences avec la méthode expérimentale de P&L ne permettent pas d'évaluer l'impact réel d'une annotation automatique.

Dans cet article, nous montrons que le paradigme de projection d'annotations des FEs se généralise au-delà du cas étudié dans P&L. Nous reproduisons les expériences réalisées dans P&L sur une langue romane, le français. Nous étendons de plus le champ d'investigation en réalisant une comparaison avec une projection à partir de FEs annotés automatiquement. Nous montrons que même cette situation permet d'obtenir des résultats de haute qualité pour le français. La structure de cet article est la suivante : nous présentons tout d'abord quelques travaux ayant exploré la projection cross-linguistique et décrivons la méthode de projection utilisée. Après avoir vérifié l'hypothèse fondamentale de parallélisme cross-linguistique des frames et Fes posée par cette méthode, nous présentons les résultats obtenus pour la projection de FEs de l'anglais vers le français. Nous concluons en évoquant les différentes pistes ouvertes par la projection cross-linguistique de *frame elements*.

2 Approches existantes

Le paradigme de la projection d'annotation a été introduit par Yarowsky et al. (2001), en utilisant un corpus parallèle pour adapter des outils monolingues (POS taggers, chunkers et analyseurs morphologiques) à des nouvelles langues. Le transfert effectif entre les langues a été rendu possible en utilisant les alignements de mots individuels entre les phrases, alignements qui peuvent aujourd'hui être obtenus automatiquement grâce à des outils comme GIZA++ (Och et Ney, 2003). Cette approche a été ensuite adaptée à d'autres niveaux de description linguistique, principalement pour la grammaire et la syntaxe. Par exemple, Hwa et al. (2002) ont projeté les informations de dépendance syntaxique de l'anglais au chinois.

La première tentative de transfert cross-linguistique d'informations sémantique a été présentée par Fung et Chen (2004) dans un projet de construction de FrameNet pour le chinois. Ceux-ci exploitent les informations du FrameNet anglais en les mettant en relation avec des concepts tirés d'une ontologie en chinois, HowNet, sans d'ailleurs exploiter de corpus alignés. Leur stratégie requiert donc l'existence d'une grande ontologie pour la langue cible, ce qui n'est pas toujours disponible. L'approche de P&L se donne la même tâche, mais sans nécessiter une telle ressource, en se focalisant sur la projection d'annotations, moins gourmande en connaissances.

Cependant, la réussite de la projection d'annotations dépend en grande partie du *parallélisme cross-linguistique* de ces annotations. Comme la projection d'annotations basique consiste à copier l'information de l'annotation source, il y a erreur si l'annotation « idéale » de la cible n'est pas identique à l'annotation source. Le degré de parallélisme cross-linguistique est connu pour être très dépendant du niveau de description en question. Alors que Yarowsky et al. (2001) rapportent un parallélisme allant jusqu'à 85 % pour les étiquettes de partie du discours (anglais-français), Hwa et al. (2002) n'ont pas mesuré plus de 40 % de liens de dépendance syntaxique pouvant être projeté directement de l'anglais au chinois. P&L ont trouvé une bonne correspondance entre les frames et les FEs entre l'anglais et l'allemand, (voir section 4), ce qui donne une idée de la pertinence de l'approche. La question de la généralisation à d'autres langues, donc de l'évaluation du parallélisme sémantique, est cependant posée, car la proximité anglais-allemand est particulièrement importante.

3 Méthode de projection

Cette section décrit rapidement la méthode de projection proposée par P&L, qui définit un modèle général voulu indépendant de la langue, pour projeter les FEs de la langue source vers la langue cible. Contrairement aux études précédentes, P&L ont trouvé que les alignements mot à mot donnent de moins bons résultats pour la projection de FEs. Ces études concernaient la projection d'informations résidant au niveau des mots (Hwa et al., 2002) ou de syntagmes courts (Yarowsky et al., 2001), alors que les FEs peuvent s'étendre sur des syntagmes longs. Du fait des erreurs et des omissions dans les alignements automatiques de mots, la projection de longs constituants est délicate. Nous considérons donc comme P&L que les modèles de projections par alignement mot à mot (M) doivent être pris comme modèle de référence uniquement en l'absence d'informations plus riches.

D'autre part, P&L ont montré que la projection peut bénéficier énormément d'informations sur les constituants. Les transferts par constituants permettent à la fois d'espérer projeter les FEs sur des étendues pertinentes et d'avoir une meilleure robustesse que M, puisqu'un certain nombre d'erreurs d'alignement peut être compensé par les bons alignements. Par ailleurs, le fait que l'alignement soit recalculé au niveau des constituants permet des stratégies alternatives à l'alignement un à un. P&L proposent ainsi d'évaluer trois classes d'alignement de constituants : les alignements Total (T), Couvrant (C) et Exact (E) qui diffèrent dans la force des contraintes qu'ils imposent sur l'alignement.

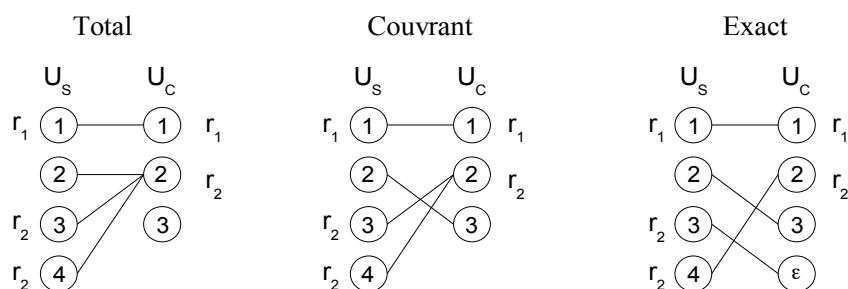


Figure 1: Modèles d'alignement de constituants (U_s et U_c sont les ensembles de constituants sources et cibles, r₁ et r₂ sont deux FEs)

Ces classes sont représentées figure 1 : T impose que chaque constituant source soit projeté au moins une fois, C impose en plus que chaque constituant cible soit lié à au moins une source et E impose que les constituants soient projetés un à un, introduisant éventuellement des constituants vides (ε). Un compromis doit être trouvé entre les alignements plus stricts comme E, qui peut corriger plus d'erreurs et un alignement plus souple comme T qui peut mieux modéliser les glissements dûs à la traduction. Ce compromis pouvant être dépendant de la langue, nous comparons toutes ces différentes classes pour le français.

Parallèlement aux différents modèles d'alignements, P&L proposent différents filtres pré- ou post-projection afin de réduire l'impact des erreurs d'alignements sur le résultat.³

1. Le filtre correctif de *convexité de couverture*, appliqué après projection, consiste à étendre artificiellement l'étendue d'un FE sur tous les éléments non annotés situés entre le premier et le dernier élément qui lui est attribuable. Johansson et Nugues (2006) étendent une telle heuristique en y ajoutant des caractéristiques spécifiques au suédois, ce qui leur permet d'obtenir d'excellents scores de projection.
2. Le filtrage de mots permet de corriger les erreurs dues à l'alignement : les mots grammaticaux souvent mal alignés et les nombreux mots laissés non alignés. La présence de ces erreurs dans un constituant peut le pénaliser inutilement dans le transfert. Pour palier ce problème, P&L proposent un filtre sur les mots grammaticaux (*MG*) et un autre sur les mots non alignés (*NA*).
3. Le *filtrage des non-arguments* (*Arg*) permet, à partir d'informations syntaxiques profondes, de ne pas prendre en compte les constituants qui ont peu de chances d'être des arguments du prédicat. L'efficacité de cette stratégie repose fortement sur la qualité de l'analyse syntaxique, afin que les constituants légitimes ne soient pas filtrés.

4 Parallélisme sémantique cross-linguistique

Nous avons argumenté dans la section 2 sur le fait que le parallélisme sémantique entre deux langues représente une borne maximum à la performance d'un système de projection de l'annotation. Pour cette raison, le parallélisme sémantique doit être évalué auparavant afin de déterminer la possibilité d'appliquer la projection à une langue particulière. Afin d'estimer ce parallélisme entre l'anglais et le français, nous avons produit un corpus annoté en sémantique des frames. Pour faire la comparaison avec les résultats de P&L, nous avons produit un sous-corpus en annotant les phrases correspondants au sous-corpus extrait pour leurs travaux. Ceci a été possible du fait que ce sous-corpus est tiré de Europarl (Koehn, 2005), corpus des minutes du parlement européen dans 11 langues. Le sous-corpus original a été extrait afin d'évaluer la projection entre l'allemand et l'anglais et est soumis à des contraintes spéciales qui introduisaient des biais que nous discuterons plus loin.

Etant donné la nature de l'expérimentation, consistant à mettre en relation une unique frame entre deux traductions, nous avons construit un guide pour l'annotation décrivant pour chaque phrase à annoter en français, le mot qui était le plus probablement le prédicat de la frame (trouvé par alignement à partir de la version anglaise annotée), ainsi que les frames potentiellement évoquées. En utilisant ce guide, deux annotateurs⁴ ont produit une annotation du sous-corpus français de 1076 phrases, en utilisant 60 phrases pour la mise au point. Le sous-corpus était initialement analysé syntaxiquement par Syntex (Bourigault et al., 2005) et a été annoté en utilisant l'outil SALTO (Burchardt et al., 2006). L'évaluation de l'accord inter-annotateurs (avant adjudication) est résumée dans la colonne gauche du tableau 2 ; la colonne de droite présente les résultats de P&L. En général l'accord pour le français est élevé et correspond aux résultats pour l'allemand. Une des deux raisons auxquelles nous attribuons le score plus faible pour l'étendue des FEs est la nature plus fragmentaire de l'arbre syntaxique français (seuls 82 % des FEs français ont pu être assignés à des constituants uniques).

³ Nous ne testons pas les combinaisons possibles de filtres, suivant en cela P&L.

⁴ Nous remercions à ce propos Christiane Jadelot de l'ATILF pour son implication.

	Français	Allemand (P&L)
Acc. frames	0.87	0.87
Acc. FEs	0.89	0.95
Acc. étendue	0.72	0.83

Tableau 2: Accords inter-annotateurs pour les sous-corpus français (sur 500 phrases), comparés aux résultats de P&L pour l'allemand.

Le tableau 3 montre l'accord pour le français par rapport à l'anglais. Les résultats de P&L pour l'anglais-allemand sont donnés à titre de comparaison. La première ligne donne l'accord sur la frame choisie, la seconde sur les FEs. Nous avons trouvé que la correspondance cross-linguistique est quasiment identique pour les deux paires de langues. C'est le premier résultat important de notre étude, qui montre que le paradigme de projection est applicable aussi pour la paire anglais-français, malgré une plus grande distance typologique.

	Français/anglais	Allemand/anglais (P&L)
Corresp. frames	0.69	0.71
Corresp. FEs	0.88	0.91

Tableau 3: Correspondance interlingue des sous-corpus annotés

Il est intéressant de constater des différences dans la distribution des frames entre l'annotation française d'une part et les annotations anglaise et allemande d'autre part. Le tableau 4 donne le nombre de frames ayant un nombre d'annotation compris dans la fourchette décrite dans la colonne de gauche et entre parenthèses le total des annotations pour ces frames. Il montre des différences de répartition des frames dans l'annotation, qui s'expliquent par le fait que le sous-corpus a été initialement sélectionné pour maximiser les chances d'avoir des phrases avec des correspondances de frames entre l'allemand et l'anglais.

Nb annotations par frame	Français	Allemand	Anglais
25-160	8 (418)	13 (578)	9 (447)
10-24	20 (315)	14 (228)	25 (389)
1-9	93 (233)	45 (133)	49 (151)
Total	121 (966)	73 (987)	83 (987)

Tableau 4: Distribution des frames en fonction du nombre d'annotations dans les trois sous-corpus: français, allemand et anglais.

La distribution des frames est gonflée vers le bas, ce qui est une seconde raison pour le plus faible accord sur les étendues, les frames plus rares étant a priori plus difficiles à annoter.

5 Evaluation expérimentale

5.1 Conditions de l'expérimentation

Dans nos expériences, nous appliquons la méthode de projection au sous-corpus anglais-français que nous avons décrit dans la section 3. L'information sur les FEs est projetée de l'anglais sur le français. L'annotation manuelle en français sert de référence pour évaluer les annotations projetées. Nous comparons deux sources de projection différentes :

- *Condition 1 : annotation manuelle.* Cette annotation correspond à la configuration de P&L, dont le matériel est disponible. Comme nous l'avons évoqué, la pertinence de

cette disposition est incertaine pour une application pratique, puisqu'en général, aucune annotation manuelle n'est disponible pour des corpus parallèles.

- *Condition 2 : annotation automatique.* Dans cette configuration nous avons utilisé un analyseur sémantique de surface de dernière génération (Giuglea et Moschitti 2006) entraîné sur les données de FrameNet 1.1, pour annoter les FEs sur le côté anglais du corpus⁵. Giuglea et Moschitti rapportent une précision de 85.2 % sur un jeu de données tiré de FrameNet ; dans notre évaluation avec l'annotation anglaise de référence de P&L, nous obtenons une f-mesure⁶ de 65.1 (préc.: 78.1 %, rap.: 55.8 %). Les sources de différence sont les suivantes : (a) notre jeu de données « standard » n'inclut pas les traits de PropBank utilisés par Giuglea et Moschitti, (b) l'application à un corpus d'un autre domaine et (c) la couverture restreinte aux verbes, qui ne concernent que 87 % de notre corpus d'évaluation. Avec une évaluation limitée aux verbes, le système obtient un rappel de 62.4 %.

Afin de pouvoir rendre nos résultats comparables avec P&L, nous suivons leur démarche: le corpus parallèle est divisé en un corpus de développement et un corpus de test (50 % chacun). Dans les deux conditions, nous utilisons l'ensemble de développement pour comparer le modèle M de référence avec les modèles sur constituants, chacun combinant une classe d'alignement avec une procédure de filtrage. Les résultats des meilleurs modèles pour chaque alignement sont ensuite vérifiés sur l'ensemble de test. Toutes les évaluations utilisent l'alignement automatique intersectif produit par GIZA++.

5.2 Résultats

Modèle \ Filtre	∅	NA	MG	Arg
Mots	<u>50.7</u> (53.3/48.3)	<u>50.7</u> (53.3/48.3)	30.4 (32.5/28.6)	-
Total	53.5 (57.2/50.2)	57.8 (68.1/50.2)	45.6 (60.5/36.6)	<u>64.1</u> (71.4/58.1)
Couvrant	55.9 (60.0/52.3)	62.6 (65.9/59.7)	61.9 (66.8/57.6)	64.2 (71.5/58.3)
Exact	54.7 (60.9/49.6)	63.4 (68.9/58.7)	62.3 (69.4/56.6)	60.6 (84.2/47.3)

Tableau 5: Evaluations dans l'ensemble de développement, source manuelle (condition 1)

Nous présentons les résultats sur l'ensemble de développement dans les tableaux 5 (condition 1) et 6 (condition 2). Les résultats sur l'ensemble de test sont dans le tableau 7. Le format des mesures reproduites est : « F-mesure (%Précision/%Rappel) ». Les meilleures configurations globales d'après l'ensemble de développement sont en grisé, les meilleures f-mesures par filtre sont en gras, les meilleures f-mesures par modèle sont soulignées.

Modèle \ Filtre	∅	NA	MG	Arg
Mots	45.4 (54.6/38.9)	45.4 (54.6/38.6)	28.1 (34.0/24.0)	-
Total	47.9 (58.4/40.6)	51.3 (69.7/40.6)	42.8 (66.3/31.6)	59.5 (74.4/49.6)
Couvrant	51.7 (62.9/43.9)	57.6 (68.8/49.6)	56.7 (69.5/47.9)	59.5 (74.4/49.6)
Exact	51.5 (65.0/42.6)	58.3 (71.7/49.1)	57.1 (72.0/47.3)	55.9 (84.6/41.7)

Tableau 6: Evaluations dans l'ensemble de développement, source automatique (condition 2)

⁵ Nous remercions Ana-Maria Giuglea et Alessandro Moschitti pour l'accès à leur logiciel.

⁶ F-mesure = $(2 \times \text{Rappel} \times \text{Précision}) / (\text{Rappel} + \text{Précision})$

Modèle	Condition 1	Condition 2
Mots	∅ : 49.3 (50.6/48.1)	∅ : 45.4 (54.6/38.9)
Total	Arg : 62.7 (68.3/57.9)	Arg : 56.1 (72.0/47.9)
Couvrant	Arg : 63.0 (68.8/58.3)	Arg : 55.9 (71.6/45.9)
Exact	NA : 63.1 (66.2/60.3)	NA : 57.2 (70.2/48.3)

Tableau 7: Evaluation des projections dans l'ensemble de test, comparaison des modèles dans chaque condition (avec les meilleurs filtres selon l'ensemble de développement).

5.3 Comparaisons et discussion

Nous considérons tout d'abord les résultats sur l'ensemble de développement pour la condition 1, tableau 5. On constate que les modèles à base de constituants dépassent systématiquement la référence du modèle M, ce qui signifie que la segmentation est utile aussi pour le français. Les résultats de la condition 1 sont largement similaires aux résultats de P&L dans plusieurs aspects. La qualité globale est proche : le meilleur modèle sur le français ($f=64.2\%$) est seulement 3 % en dessous du score pour l'allemand obtenu par P&L ($f=67.3\%$). Les résultats sur le français sont encore plus favorables quand ils sont comparés à leur borne supérieure, qui est l'accord inter-annotateur (à défaut d'avoir une évaluation manuelle des projections) : ils sont à 6 % sous le plafond de 72 %, là où les résultats allemands sont 16 % sous l'accord à 83 %. Ensuite, on observe le même impact des procédures de filtrage. Les résultats montrent clairement un impact positif des filtres NA et Arg. En revanche, le filtre NA qui limite le bruit dû à l'alignement montre de meilleurs résultats quand le modèle est plus restrictif (Couvrant et Exact par rapport à Total), alors que le filtre Arg améliore nettement la précision mais fait baisser le rappel, et favorise les modèles relativement moins restrictifs (Total et Couvrant par rapport à Exact). Ces observations montrent que les modèles d'alignement et les filtres de P&L sont pertinents au-delà de la paire de langues pour laquelle ils ont été créés. La généralité de ce paradigme doit être prise en compte pour comparer nos résultats avec ceux de Johansson et Nugues (2006), qui annoncent une f -mesure de 82.0 (préc. : 84.0, rap. : 81.0) pour leur projection. Ils utilisent des heuristiques spécifiques au suédois⁷, qui devront être recrées pour chaque nouvelle langue cible et sont probablement plus difficile à identifier pour des langues plus distantes que l'anglais et le suédois, qui sont des langues très proches (Koehn 2005).

Ensuite, nous nous intéressons aux résultats utilisant une source automatique (condition 2), qui sont exposés dans le tableau 6 pour l'ensemble de développement. On note que le passage de la version manuelle à celle annotée automatiquement conduit à une perte de performance allant de seulement 3 à 7 points sur la f -mesure. Cette différence correspond assez bien à la différence observée entre les annotations manuelles ($f=72\%$, d'après l'accord inter-annotateurs), et l'analyseur sémantique de surface ($f=65\%$). Une comparaison des tableaux 5 et 6 montre que les caractères de l'annotation automatique (haute précision, rappel bas) sont très clairement reproduits dans les propriétés de la projection : alors que la source automatique amène une nette chute du rappel, la précision reste constante voire s'améliore par rapport à la source manuelle. Cet apparent paradoxe vient du fait que la projection à partir de la source automatique ne tente pas de projeter certains FE difficiles, par exemple ceux qui s'étendent sur plusieurs constituants, simplement parce que les FE n'ont pas été annotés par le système automatique sur l'anglais. En somme, les résultats de la condition 2 montrent la possibilité d'utiliser des analyseurs sémantiques de surface comme entrée de la projection, ce qui est indispensable pour appliquer ce paradigme à grande échelle. En plus, les annotations obtenues sont de grande précision, caractéristique essentielle pour la création de ressources.

⁷ L'accord inter-annotateur de leur corpus de référence n'étant pas connu, et comme celui-ci ne comporte que 150 phrases, il n'est pas possible de faire une comparaison des deux méthodes.

Les résultats sur l'ensemble de test sont exposés dans le tableau 7. Moins riches d'enseignement que les résultats de l'ensemble de développement et par manque de place, nous n'en ferons qu'une synthèse courte. Une baisse de 2 à 3 % est observée en moyenne sur la f-mesure, par rapport à l'ensemble de développement, mais cette différence peut s'expliquer par une variation aléatoire naturelle sur le partitionnement du corpus (P&L obtiennent des résultats légèrement meilleurs avec le même découpage). Les différences entre les méthodes basées sur constituants ne sont pas significatives, mais le meilleur modèle (E+NA) pour le français est le même que le meilleur modèle pour l'allemand trouvé par P&L.

6 Conclusion et perspectives

Dans cette étude nous avons montré la possibilité de produire des annotations sémantiques de grande précision pour le français en appliquant la démarche proposée par Padó et Lapata (2005, 2006). Cette démonstration procède en trois étapes : (1) la vérification que le parallélisme français-anglais du point de vue de la sémantique des frames est suffisant pour permettre la projection de l'annotation, (2) la démonstration que les résultats obtenus par P&L pour la paire anglais-allemand se reproduisent bien pour la paire anglais-français aussi bien en terme de performance absolue que du point de vue des effets des différents filtrages – ce qui montre la généralité dans la démarche suivie et valide une fois encore l'aspect multilingue de FrameNet – et (3) la démonstration que même l'utilisation d'une source bruitée provenant d'une annotation automatique résulte en une annotation qui, en particulier, montre une haute précision. D'ores et déjà, une analyse plus approfondie de nos données a montré que 124 phrases du corpus ont une annotation parfaitement identique à l'annotation manuelle de référence. Ces résultats ont de fortes chances de s'améliorer avec des avancées dans les technologies d'alignement, d'analyse syntaxique et sémantique de surface.

Avec un tel résultat de référence, relativement facile à obtenir (alignement automatique, analyseur syntaxique et corpus aligné avec l'anglais), on peut très raisonnablement envisager d'entraîner dès maintenant des annotateurs automatiques de FEs en français à moindre coût. Les résultats obtenus par Johansson et Nugues (2006) nous indiquent que des techniques spécifiques à la langue cible, qui éliminent les FEs non plausibles projetés durant l'induction de l'analyseur sémantique, permettent de produire un annotateur automatique ayant une performance à peine dégradée par rapport aux versions natives. Ces bons résultats autorisent à envisager une méthode de construction de ressources fondée sur un travail itératif, utilisant à la fois un outil automatique pour produire des annotations et une phase de révision manuelle pour obtenir un corpus de bonne qualité lexicographique. En effet, l'effort requis pour démarrer une telle ressource est rendu abordable par cette approche, même si une évaluation plus précise du coût de correction par rapport au coût de production reste à faire afin de valider totalement cette démarche.

Ces travaux ont été financés par le fonds France-Berkeley (projet FR.FrameNet sous la responsabilité de C. J. Fillmore et L. Romary) et le DFG (Padó; bourse PI-154/9-2).

Références

- BOAS H.C. (2002). Bilingual FrameNet dictionaries for machine translation. Actes de *LREC 2002*, pp. 1364–1371, Las Palmas, Iles Canaries.
- BOUILLON P., FABRE C., SÉBILLOT P., JACQMIN L. (2000). Apprentissage de ressources lexicales pour l'extension de requêtes, *Traitement Automatique des Langues*, 41:2, pp. 367–393.
- BOURIGAUULT D., FABRE C., FRÉROT C., JACQUES M.-P., OZDOWSKA S. (2005). Syntex, analyseur syntaxique de corpus, Actes de *TALN 2005*, Dourdan, France.

- BURCHARDT A., ERK K., FRANK A., KOWALSKI A., PADO S. (2006). SALTO – A Versatile Multi-Level Annotation Tool. Actes de *LREC 2006*, Gênes, Italie.
- FILLMORE C.J. (1982). Frame Semantics. *Linguistics in the Morning Calm*, pp. 111—38. Hanshin, Seoul.
- FILLMORE C.J., JOHNSON C.R., PETRUCK M.R. (2003). Background to FrameNet. *International Journal of Lexicography*, 16:235–250.
- FUNG P., CHEN B. (2004). BiFrameNet: Bilingual frame semantics resources construction by cross-lingual induction. Actes de *COLING 2004*, pp. 931—935, Genève, Suisse.
- GILDEA D., JURAFSKY D. (2002). Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- GIUGLEA A.-M., MOSCHITTI A. (2006). Semantic role labeling via FrameNet, VerbNet and PropBank. Actes de *COLING/ACL 2006*, pp. 929—936, Sydney, Australie.
- HWA R., RESNIK P., WEINBERG A., KOLAK O. (2002). Evaluation translational correspondance using annotation projection. Actes de *ACL 2002*, pp. 392—399, Philadelphia.
- JOHANSSON R., NUGUES P. (2006). A FrameNet-based Semantic Role Labeler for Swedish. Actes de *COLING/ACL 2006 Main Conf. Poster Sessions*, pp. 436—443, Sydney, Australie.
- KOEHN P. (2005). Europarl: A parallel corpus for statistical machine translation. Actes de *MT Summit X*. Phuket, Thaïlande.
- NARAYANAN S., HARABAGIU S. (2004). Question answering based on semantic structures. Actes de *COLING 2004*, pp. 693–701, Genève, Suisse.
- OCH F. J., NEY H. (2003). A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics*, 29(1):19-51.
- PADÓ S., LAPATA M. (2005). Cross-lingual projection of role-semantic information. Actes de *HLT/EMNLP 2005*, Vancouver, Canada.
- PADÓ S., LAPATA M. (2006). Optimal Constituent Alignment with Edge Covers for Semantic Projection. Actes de *COLING/ACL 2006*, pp. 1161—1168, Sydney, Australie.
- PITEL G. (2006). Using bilingual LSA for FN annotation of French text from generic resources. Workshop on Multilingual Annotation : Theory and Applications, Saarbrücken.
- SURDEANU M., HARABAGIU S., WILLIAMS J., AARSETH P. (2003). Using predicate-argument structures for information extraction. Actes de *ACL 2003*, pp. 8—15, Sapporo, Japon.
- YAROWSKY D., NGAI G., WICENTOWSKI R. (2001). Inducing multilingual text analysis tools via robust projection across aligned corpora. Actes de *HLT 2001*, pp. 161—168. San Francisco.