

Utilisation de représentations de mots pour l'étiquetage de rôles sémantiques suivant FrameNet

William Léchelle, Philippe Langlais
DIRO, Université de Montréal
{lechellw, felipe}@iro.umontreal.ca

Résumé. D'après la sémantique des cadres de Fillmore, les mots prennent leur sens par rapport au contexte événementiel ou situationnel dans lequel ils s'inscrivent. FrameNet, une ressource lexicale pour l'anglais, définit environ 1000 cadres conceptuels couvrant l'essentiel des contextes possibles.

Dans un cadre conceptuel, un prédicat appelle des arguments pour remplir les différents rôles sémantiques associés au cadre. Nous cherchons à annoter automatiquement ces rôles sémantiques, étant donné le cadre sémantique et le prédicat, à l'aide de modèles à maximum d'entropie.

Nous montrons que l'utilisation de représentations distribuées de mots pour situer sémantiquement les arguments apporte une information complémentaire au modèle, et améliore notamment l'étiquetage de cadres avec peu d'exemples d'entraînement.

Abstract. According to Frame Semantics (Fillmore 1976), words' meaning are best understood considering the semantic frame they play a role in, for the frame is what gives them context. FrameNet defines about 1000 such semantic frames, along with the roles arguments can fill in this frame. Our task is to automatically label arguments' roles, given their span, the frame, and the predicate, using maximum entropy models.

We make use of distributed word representations to improve generalisation over the few training exemples available for each frame.

Mots-clés : rôles sémantiques ; représentations distribuées ; maximum d'entropie.

Keywords: semantic role labelling ; distributed word representations.

Nous avons utilisé les deux corpus de données de FrameNet, le texte complètement annoté, et une partie des phrases exemples³. Au total, l'ensemble d'entraînement comporte 106 926 cadres, avec 1,45 argument par cadre.

Un cadre sémantique qui apparaît dans le jeu de test apparaît en moyenne 300 fois dans l'ensemble d'entraînement (appelé par différents prédicats, avec différents arguments). En pratique, certains cadres sont beaucoup plus représentés que d'autres, ce sur quoi on reviendra en section 5.4.

L'ensemble de test est le même que celui de (Das & Smith, 2011), soit 23 documents complètement annotés. Il comporte 4456 instances de cadres, soit 7209 arguments à classifier (1,6 argument par cadre en moyenne).

Pour permettre une certaine comparaison à l'état de l'art, la performance rapportée pour chacune des méthodes évaluées est la micro-précision sur cet ensemble d'arguments, c'est-à-dire la proportion d'arguments dont le rôle est correctement prédit par le modèle, tous cadres confondus. Une autre mesure de performance pertinente est la macro-précision, au sens de la performance moyenne de chacun des cadres (voir section 5.4).

Formellement, si on note $n(f)$ le nombre d'arguments de l'ensemble de test à annoter pour le cadre f , et $c(f)$ le nombre d'arguments annotés correctement par le modèle du cadre f , $p(f) = \frac{c(f)}{n(f)}$ est la proportion d'arguments annotés correctement pour le cadre f . Nos métriques s'écrivent alors :

$$\text{micro-précision} = \frac{\sum_{f \in \text{cadres}} c(f)}{\sum_{f \in \text{cadres}} n(f)}$$

$$\text{macro-précision} = \text{moyenne}_{f \in \text{cadres}}(p(f)) = \frac{\sum_{f \in \text{cadres}} \frac{c(f)}{n(f)}}{\text{nb cadres}}$$

3 État de l'art

Plusieurs auteurs se sont attachés à identifier les arguments et leurs rôles pour les cadres sémantiques de FrameNet. Les premiers, Gildea & Jurafsky (2000) utilisent une cascade de modèles (*backoff*) s'appuyant sur les comptes de caractéristiques des noeuds de l'arbre syntaxique correspondant aux arguments. Leur modèle final obtient 76.9% de précision sur leur ensemble de test, pour la tâche d'étiquetage des rôles. Leurs données étaient les phrases exemples d'une version préliminaire de FrameNet comportant 67 cadres, soit 49 013 phrases annotées avec un cadre par phrase.

De leur travail ressort un équilibre entre la couverture de chaque modèle et sa précision : le modèle n'utilisant que la cible couvre 100% des cas mais est seulement précis à 41%. À l'inverse, les mots de tête constituent une caractéristique des plus fiables pour déterminer le rôle d'un argument : le modèle utilisant uniquement la cible et le mot de tête pour classifier un rôle obtient 86,7% de précision, mais ne couvre que 56% des données. Pour augmenter la couverture de ce modèle, les auteurs font une expérience pour grouper les noms du lexique (avec la technique de *clustering* décrite dans Lin 1998), ce qui leur permet d'obtenir un modèle couvrant 98% des mots de tête nominaux, et précis à 79.7%. Finalement, cela ajoute 0,8% de précision à leur modèle global, sur l'ensemble de développement.

Notre étude vise à être la continuation de cette expérience, en utilisant les représentations distribuées de mots pour généraliser les mots du lexique.

Cette idée a été employée récemment dans des expériences sur le FrameNet suédois naissant : Johansson *et al.* (2012) utilisent le clustering lexical de Brown (Brown *et al.* 1992) pour augmenter la couverture de caractéristiques lexicales (pour la classification de rôles), et rapportent une légère amélioration de performance.

En 2003, Fleischman *et al.* ont été les premiers à utiliser des modèles à maximum d'entropie pour l'identification du rôle des arguments, en obtenant des résultats proches de ceux de Gildea et Jurafsky (76% de précision avec des données obtenues automatiquement), sur des données comparables (40 000 phrases exemples de FrameNet).

SEMAFOR (Das *et al.* 2010), développé depuis 2010 à l'université Carnegie-Mellon, est un système complet d'annotation sémantique automatique. Pour résoudre notre tâche, ce système détermine, pour chaque rôle sémantique, l'emplacement

3. Nous n'avons considéré que les arguments constitués d'un seul mot.

de l'argument qui le comble (ou aucun argument si le rôle n'est pas exprimé). Ses prédictions sont précises (88%), mais avec un plus faible rappel (75%), c'est-à-dire que le modèle estime que certains rôles ne sont pas exprimés, alors qu'ils le sont. Notre système suit la méthode plus courante (de Johansson & Nugues 2007) qui consiste à détecter d'abord les arguments, pour ensuite les classer en leurs rôles respectifs.

Étant donné la cible, le cadre, et les positions exactes des arguments, SEMAFOR obtient un score F_1 de 81%. Leurs expériences utilisent la version 1.3 de FrameNet, semblable à nos propres données.

4 Représentations de mots

D'après Turian *et al.* 2010, l'utilisation de représentations distribuées de mots – apprises de manière non supervisée – est une méthode simple et générale pour améliorer la précision de systèmes d'apprentissage supervisé pour le traitement des langues.

Nous avons utilisé les représentations de mots fournies par Ronan Collobert, apprises par SENNA⁴ (Collobert *et al.* 2011) à partir d'un grand corpus de texte non étiqueté (provenant essentiellement de Wikipedia). Cette ressource fournit la représentation par un vecteur de valeurs réelles, dans un espace de dimension 50, de 130 000 mots, les plus fréquents dans le corpus. On utilise la distance euclidienne pour mesurer la proximité de mots dans cet espace.

Pour donner une idée, la table 2 montre les plus proches voisins de quelques mots pris au hasard.

TABLE 2 – Mots les plus proches d'exemples choisis au hasard, d'après leur représentation vectorielle.

| Exemple | plus proches voisins |
|-----------------|---|
| characteristics | traits indicators measurements phenotypes ... |
| chalkboard | sofa washroom hallway bathroom darkroom skit ... |
| charlene | cynthia cathy benji angie ronnie julie caitlin cheryl ... |
| delighted | amazed thrilled dismayed ridiculed astonished ... |
| deregulate | liberalise reallocate unsettle penalise unnerve ... |
| falsifiability | teleology rationality holism causality ... |
| memorizing | deciphering interpreting embodying unlocking ... |
| parrots | wasps cormorants beetles lizards newts ... |
| planet | earth universe portal basestar mothership galaxy ... |
| retirement | tenure graduation incarceration signing ... |
| visible | confusing common hidden standing peculiar ... |

5 Expériences

Dans l'idée, si des mots sont proches dans l'espace des représentations, c'est qu'ils sont sémantiquement proches – synonymes au sens large, typiquement. L'apprentissage devrait pouvoir profiter de cette connaissance pour annoter des exemples de test inconnus, proches sémantiquement d'exemples d'entraînement connus, avec le même rôle que ces derniers. Autrement dit, généraliser la connaissance des exemples d'entraînement aux arguments qui en sont sémantiquement proches.

5.1 Modèle de référence

Comme système de référence, nous avons entraîné des modèles à maximum d'entropie, un par cadre (les rôles que peuvent remplir les arguments sont différents pour chaque cadre).

Notre implémentation utilise Python et NLTK⁵. L'entraînement du modèle utilise l'algorithme du gradient conjugué ('CG').

4. <http://ronan.collobert.com/senna/>

5. <http://nltk.org/>

Le modèle s'appuie sur une quinzaine de caractéristiques de surface de l'argument. Avec l'exemple de l'Agent dans le cadre Cause_de_bruit (figure 1), les caractéristiques employées sont regroupées en catégories et présentées dans la table 3.

TABLE 3 – Caractéristiques du modèle de référence, ainsi que leur valeur dans l'exemple présenté en figure 1.

| Caractéristiques par catégorie | Valeur dans l'exemple |
|--|-----------------------|
| Caractéristiques de base | |
| le texte de l'argument | enough ringers |
| le texte de la cible | ring |
| la position (en caractères) de l'argument dans la phrase | 23 |
| Position relative | |
| est-ce que l'argument est avant ou après la cible | avant |
| si l'argument est à la même place que la cible | non |
| la distance (en mots) entre la cible et l'argument | 1 |
| Nombre de mots | |
| le nombre de mots de l'argument | 2 |
| le nombre de mots pleins ⁶ de l'argument | 2 |
| Contenu | |
| le premier mot de l'argument | enough |
| le premier mot plein de l'argument | enough |
| Parties du discours | |
| partie du discours du premier mot de l'argument | JJ (adjectif) |
| partie du discours du dernier mot de l'argument | NNS (nom pluriel) |
| la partie du discours majoritaire dans l'argument | JJ |
| Arguments précédents de cadre | |
| nombre d'arguments déjà étiquetés dans le cadre | 0 |

Ce système de référence arrive à 80.0% de précision, à comparer aux 76% de précision de Fleischman & Hovy (2003), ou aux 80.97% de F_1 -mesure de Das *et al.* (2010), sur d'autres versions des données de FrameNet.

Dans notre travail, l'utilisation de représentations de mots vient améliorer ce système de référence en apportant de l'information supplémentaire (obtenue par apprentissage non supervisé).

5.2 Plus proches voisins

On cherche à utiliser la proximité sémantique d'arguments pour prédire leur rôle. Pour mesurer la proximité et calculer des distances dans l'espace des représentations, il faut situer les arguments dans cet espace. Nous sommes partis du principe qu'un mot appartenant à chaque argument devait le représenter sémantiquement. 50% des arguments du jeu de test sont constitués de plusieurs mots.

Prenons un exemple, dans le cadre de l'Engagement (Commitment : un Orateur prend un engagement auprès d'un Destinataire). L'argument de test *l'ambassadeur iraquien* est inconnu à l'entraînement, mais peut être représenté par *ambassadeur*, proche des exemples connus *personne* ou *porte-parole*. Ces mots remplissent généralement le même rôle dans ce cadre : Orateur. On peut alors conclure que *l'ambassadeur iraquien* remplit aussi le rôle d'Orateur.

Comme représentants sémantiques des arguments, nous avons choisi d'utiliser leurs têtes syntaxiques, déterminées grâce au Stanford Parser (de Marneffe *et al.* 2006). Par exemple, l'argument *their ignorance which was based on prominent views* est représenté par *ignorance*, car toute la proposition subordonnée dépend (indirectement) de *based*, qui dépend de *ignorance*, et *their* dépend aussi d'*ignorance*. Dans le cadre d'un Jugement (fait par une personne, pouvant être positif ou négatif), cet argument est ensuite classifié comme *Celui_ou_ce_qui_est_jugé*.

D'autres approches seraient possibles pour représenter les arguments : Surdeanu *et al.* (2003) en particulier propose un choix de représentant sémantique plus élaboré et probablement plus adapté. En particulier, dans le cas de syntagmes prépositionnels, la préposition n'est pas nécessairement sémantiquement significative.

6. de plus de 5 caractères

Chaque argument est situé à l'emplacement de son mot représentatif dans l'espace des représentations⁷. On peut alors prédire le rôle d'un argument avec le modèle des k plus proches voisins.

Le modèle du 1-plus-proche-voisin, qui prédit pour un argument le rôle du mot annoté le plus proche dans l'ensemble d'entraînement (pour le cadre considéré) obtient 70% de précision, expérimentalement. C'est assez remarquable pour un modèle aussi simple : pour comparaison, le modèle qui prédit pour un argument le rôle le plus fréquemment annoté à l'entraînement arrive seulement à 50% de précision.

Prendre en compte plusieurs voisins dans la prédiction du rôle d'un argument réduit le bruit dans les données et améliore nettement les performances. La figure 4 montre les résultats du modèle des plus proches voisins en fonction de k . Dans nos expériences, utiliser cette prédiction comme caractéristique unique d'un classifieur à maximum d'entropie est un peu meilleur⁸ que de l'utiliser directement, et c'est ce que nous avons fait ici.

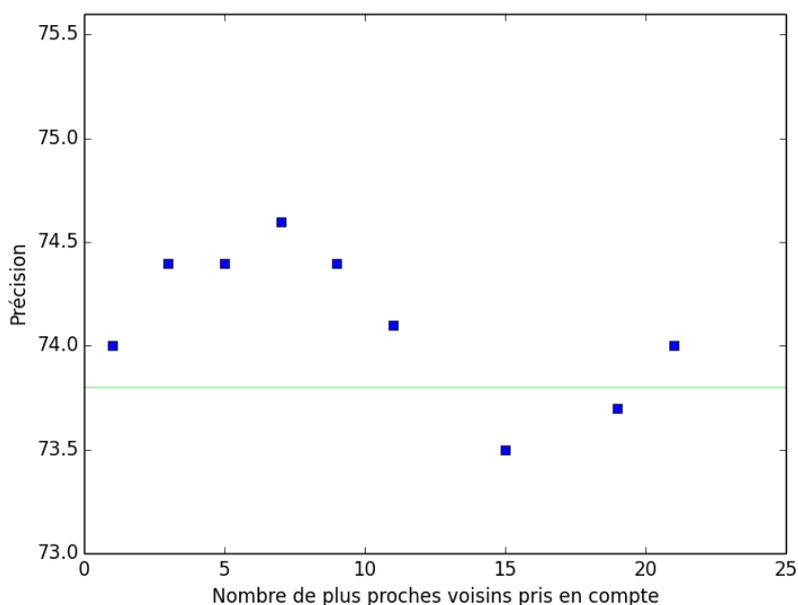


FIGURE 4 – Performance du modèle utilisant le rôle majoritaire parmi les k plus proches voisins de chaque argument comme seule caractéristique. La ligne représente la performance du modèle qui utiliserait uniquement la meilleure des caractéristiques du modèle de référence (la position relative de l'argument par rapport à la cible).

Pour intégrer la prédiction du modèle des plus proches voisins au modèle de référence, on peut simplement l'ajouter comme caractéristique d'un argument. La figure 5 montre les performances obtenues : avec plusieurs voisins, cette information améliore le modèle de référence (en vert), et permet d'arriver au niveau du système SEMAFOR, à l'état de l'art (en rouge pointillé), sur des données d'entraînement semblables.

5.3 Centres des exemples d'un rôle

Une autre façon de prédire le rôle d'un argument consiste à trouver quelle classe il représente le mieux : dans l'espace des représentations, on situe la position moyenne des mots représentant un rôle, et on assigne alors à un argument le rôle dont il est le plus proche.

Cela revient à partitionner l'espace des représentations en zones, une par rôle. Une zone est l'ensemble des points les plus proches du "représentant moyen" d'un rôle sémantique (comme un diagramme de Voronoï). On classe alors les arguments en fonction de la zone dans laquelle ils se situent.

7. les mots de tête de 2% des arguments sont hors du vocabulaire des représentations, et alors seul le modèle de référence est utilisé

8. la différence est de l'ordre de 3-4%

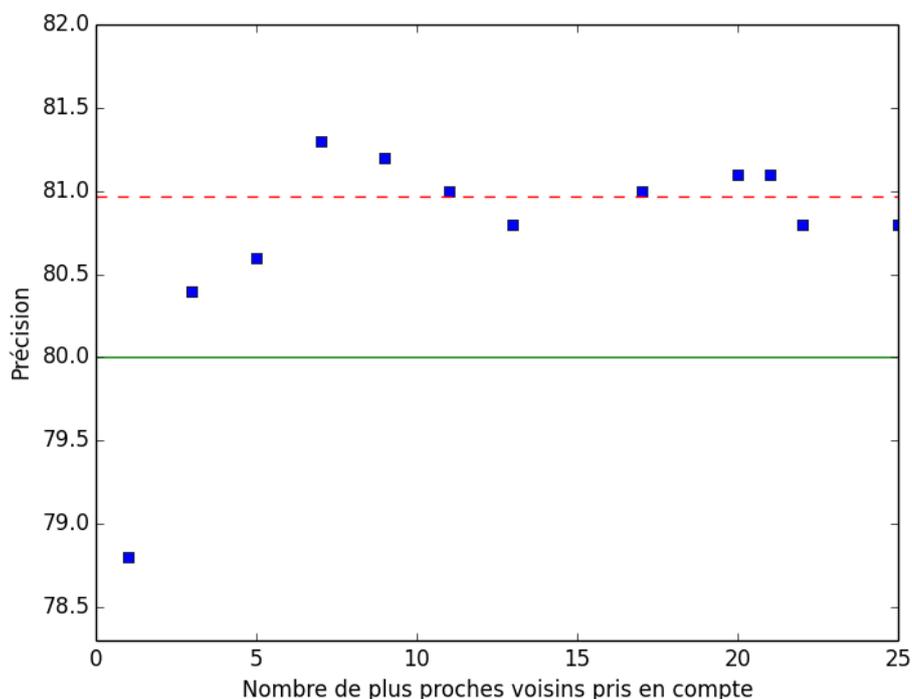


FIGURE 5 – Précision du modèle de référence informé de la prédiction du modèle des k plus proches voisins. La ligne verte représente le système de référence, et la ligne rouge pointillée SEMAFOR.

Ajoutée au modèle de référence, cette caractéristique permet d’atteindre 81.1% de précision (soit une amélioration de 1.1 point).

On peut combiner ce modèle avec celui des plus proches voisins, et ajouter les deux prédictions au modèle de référence. La figure 6 montre les résultats d’une telle combinaison. La meilleure performance est de 81.5%, avec 20 voisins.

5.4 Discussion

La table 7 récapitule les différents résultats, en termes de micro-précision (cf section 2). Rappelons que Fleischman & Hovy utilisent une version de FrameNet datant de 2002, c’est-à-dire sensiblement moins de données. Nous mesurons nos performances sur le même ensemble de test que SEMAFOR, en utilisant des données similaires à l’entraînement.

TABLE 7 – Récapitulatif des performances (micro-précision).

| Modèle | Performance |
|----------------------------------|-------------|
| k plus proches voisins | 74.6% |
| Fleischman et Hovy (2003) | 76% |
| Référence | 80.0% |
| SEMAFOR (2010) | 81.0% |
| Référence + centres | 81.1% |
| Référence + plus proches voisins | 81.3% |
| Référence + centres + PPV | 81.5% |

En regardant les résultats plus en détail, on observe que les gains en performance proviennent en large part des cadres avec le moins d’exemples. Il est alors intéressant de mesurer la macro-précision qu’obtiennent les différents modèles (cf section 2). Par rapport aux performances rapportées précédemment, c’est comme si la performance du modèle de chaque cadre n’était plus pondérée par le nombre d’exemples de ce cadre dans le jeu de test.

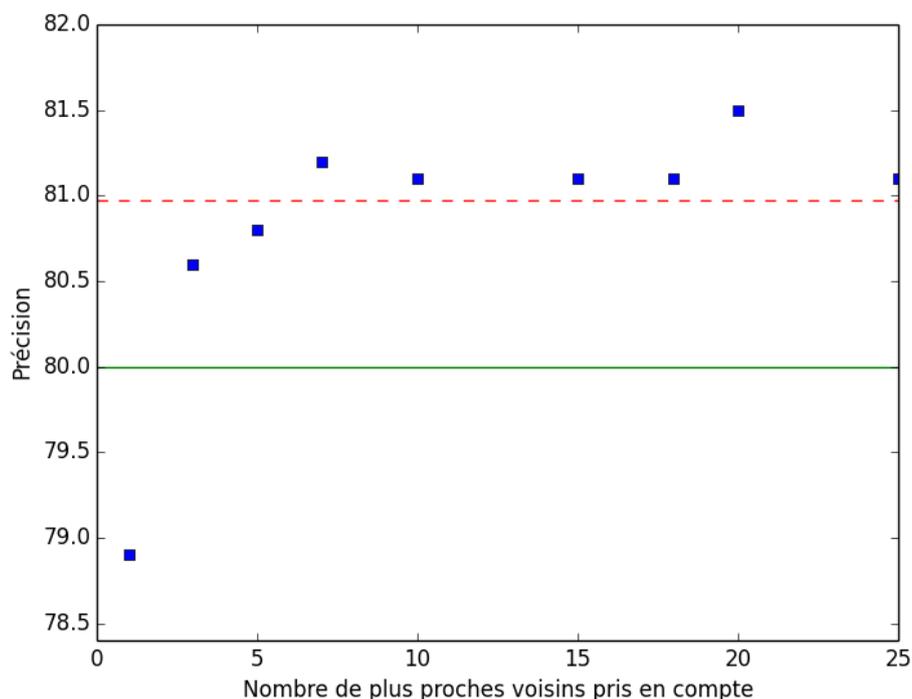


FIGURE 6 – Précision des modèles combinés de référence, des plus proches voisins et des centres des rôles.

TABLE 8 – Macro-précision pour chacune des méthodes. Les cadres avec moins d'exemples prennent davantage d'importance, par rapport au calcul de la micro-précision.

| Modèle | Précision moyenne des cadres |
|----------------------------------|------------------------------|
| Centre des rôles le plus proche | 69.6% |
| k plus proches voisins | 70.0% |
| Référence | 73.2% |
| Référence + centres | 75.8% |
| Référence + plus proches voisins | 75.2% |
| Référence + centres + PPV | 76.1% |

La table 8 montre la macro-précision des modèles que nous avons testés. Comme on peut voir, l'apport des représentations de mots est plus net avec cette mesure. En particulier, les 20% de cadres sémantiques avec le moins d'exemples d'entraînement améliorent leur précision moyenne de 5 points lorsque l'on rajoute la prédiction utilisant les centres des rôles au modèle de référence.

Cette observation est cohérente avec la vision évoquée plus haut, à savoir que l'abstraction sur les mots du lexique permet de mieux généraliser les données disponibles, surtout lorsqu'elles sont peu importantes (peu d'exemples par classe). Cette dernière mesure est d'autant plus pertinente que les cadres peu représentés dans l'ensemble de test sont aussi les cadres les plus difficiles à entraîner, du fait de leur plus faible nombre d'exemples d'entraînement.

6 Travaux futurs

L'implémentation des règles proposées dans Surdeanu *et al.* (2003) pour déterminer le mot le plus représentatif du contenu d'un argument serait une direction naturelle pour poursuivre ce travail. En particulier, les règles qui déterminent le mot de tête de syntagmes prépositionnels (la préposition) sont inadaptées à l'usage qu'on souhaite en faire ici, ce que Surdeanu propose d'améliorer.

Du point de vue de l'apprentissage, on pourrait améliorer l'entraînement des modèles. Actuellement, les performances globales sont peu sensibles à la variation du nombre de voisins pris en compte, passé un certain stade (voir figure 5). Un ensemble de développement permettrait de mieux adapter les paramètres à chaque méthode, et en particulier de faire varier la valeur du nombre de voisins considérés en fonction du cadre sémantique et des données disponibles.

Dans l'espace des représentations, pondérer l'algorithme des plus proches voisins (par exemple par l'inverse de la distance, ou par la fréquence du rôle du voisin) peut permettre de capturer davantage d'information. On pourrait surpondérer les rôles sémantiques sous-représentés, ou difficiles à détecter, notamment.

Il serait également intéressant d'utiliser d'autres représentations distribuées de mots (entraînées par d'autres systèmes que SENNA), pour comparer les résultats. Des expériences préliminaires avec les représentations distribuées par Turian *et al.* (2010)⁹ montrent des résultats semblables et encourageants. Les clusters lexicaux de Brown pourraient être employés avec la même idée.

Enfin, notre modèle prend actuellement toutes ses décisions de manière indépendante. Les arguments d'un même cadre (dans la même phrase) gagneraient à être étiquetés conjointement. Das *et al.* (2010) explorent cette idée, et gagne en précision, contre une petite perte de rappel à cause des contraintes supplémentaires. Les gains sont limités, notamment à cause du faible nombre d'arguments à annoter par cadre (moins de 2 en moyenne). Nous avons mené quelques expériences dans cette direction, mais les résultats ne sont pas concluants.

7 Conclusion

FrameNet définit un ensemble de cadres sémantiques appelés par des prédicats, ainsi que les rôles pouvant être remplis par les arguments du dit prédicat. Nous avons employé des représentations distribuées de mots, entraînées par SENNA, pour améliorer la tâche de classification des arguments en rôles, en supposant connus le prédicat, le cadre sémantique, et la position des arguments.

Les représentations de mots situent les mots du lexique, *via* leurs coordonnées, dans un espace, l'espace des représentations. La représentation des arguments à classifier par leur mot le plus représentatif – dans nos expériences, leur tête syntaxique – permet de les situer eux-mêmes dans l'espace des représentations. Dès lors, on peut utiliser l'algorithme des plus proches voisins, ou bien partitionner l'espace suivant le rôle *en moyenne* le plus proche, pour classifier les arguments, et faire des prédictions raisonnables, un peu inférieures à l'état de l'art.

En utilisant ces prédictions dans le cadre d'un modèle à maximum d'entropie utilisant des caractéristiques descriptives de l'argument dans la phrase (notre système de référence), on obtient un modèle performant, légèrement supérieur à SEMAFOR, système à l'état de l'art, sur la tâche évaluée. En particulier, on remarque que les cadres sémantiques avec le moins d'exemples d'entraînement profitent davantage de la généralisation apportée par les représentations de mots, par rapport au modèle de référence.

Références

- BROWN P. F., DESOUZA P. V., MERCER R. L., PIETRA V. J. D. & LAI J. C. (1992). Class-based n-gram models of natural language. *Comput. Linguist.*, **18**(4), 467–479.
- COLLOBERT R., WESTON J., BOTTOU L., KARLEN M., KAVUKCUOGLU K. & KUKSA P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, **12**, 2493–2537.
- DAS D., SCHNEIDER N., CHEN D. & SMITH N. A. (2010). Probabilistic frame-semantic parsing. In *Human Language Technologies : The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, p. 948–956, Stroudsburg, PA, USA : Association for Computational Linguistics.
- DAS D. & SMITH N. A. (2011). Semi-supervised frame-semantic parsing for unknown predicates. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies - Volume 1*, HLT '11, p. 1435–1444, Stroudsburg, PA, USA : Association for Computational Linguistics.
- DE MARNEFFE M.-C., MACCARTNEY B. & MANNING C. D. (2006). Generating typed dependency parses from phrase structure parses. In *IN PROC. INT'L CONF. ON LANGUAGE RESOURCES AND EVALUATION (LREC)*, p. 449–454.

9. <http://metaoptimize.com/projects/wordreprs/>

- FILLMORE C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, **280**(1), 20–32.
- FLEISCHMAN M. & HOVY E. (2003). A maximum entropy approach to framenet tagging. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology : companion volume of the Proceedings of HLT-NAACL 2003–short papers - Volume 2*, NAACL-Short '03, p. 22–24, Stroudsburg, PA, USA : Association for Computational Linguistics.
- FLEISCHMAN M., KWON N. & HOVY E. (2003). Maximum entropy models for framenet classification. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, EMNLP '03, p. 49–56, Stroudsburg, PA, USA : Association for Computational Linguistics.
- GILDEA D. & JURAFSKY D. (2000). Automatic labeling of semantic roles. In *ACL : ACL*.
- JOHANSSON R., HEPPIN K. F. & KOKKINAKIS D. (2012). Semantic role labeling with the swedish framenet. In N. C. C. CHAIR), K. CHOUKRI, T. DECLERCK, M. U. DOĞAN, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS, Eds., *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey : European Language Resources Association (ELRA).
- JOHANSSON R. & NUGUES P. (2007). Lth : semantic structure extraction using nonprojective dependency trees. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, SemEval '07, p. 227–230, Stroudsburg, PA, USA : Association for Computational Linguistics.
- LIN D. (1998). Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 2*, ACL '98, p. 768–774, Stroudsburg, PA, USA : Association for Computational Linguistics.
- PUNYAKANOK V., ROTH D. & YIH W. (2008). The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, **34**(2).
- SURDEANU M., HARABAGIU S., WILLIAMS J. & AARSETH P. (2003). Using predicate-argument structures for information extraction. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, p. 8–15, Stroudsburg, PA, USA : Association for Computational Linguistics.
- TURIAN J., RATINOV L. & BENGIO Y. (2010). Word representations : a simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, p. 384–394, Stroudsburg, PA, USA : Association for Computational Linguistics.