Apport des dépendances syntaxiques et des patrons séquentiels à l'extraction de relations

Kata Gábor[1], Nadège Lechevrel[2], Isabelle Tellier[3], Thierry Charnois[1], Haïfa Zargayouna[1], Davide Buscaldi[1]
(1) LIPN, CNRS (UMR 7030), Université Paris 13
(2) Université Paris-Ouest Nanterre La Défense
(3)LaTTiCe, CNRS (UMR 8094), ENS Paris, Université Sorbonne Nouvelle - Paris 3
PSL Research University, Université Sorbonne Paris Cité
utrucmuche@lab.fr, umachinchose@adresse-academique.fr

Résumé

Dans cet article, nous étudions la contribution de propriétés syntaxiques à la tâche de clustering d'instances de relations sémantiques. Les instances, constituées de couples de concepts apparaissant dans des textes scientifiques, sont représentées dans une matrice où on les croise avec une représentation de leur contexte de co-occurrence. Différentes variantes de représentations sont envisagées pour ce contexte, en faisant appel à la fouille de données séquentielles et à l'analyse syntaxique en dépendances. Nos comparaisons suggèrent que les attributs issus d'analyses syntaxiques permettent d'améliorer la qualité du clustering final.

Abstract

Integrating Dependency Parses with Sequential Patterns to Improve Relation Extraction

In this paper, we investigate the contribution of syntactic features to the task of unsupervised clustering of semantic relation instances. Instances, i.e. couples of concepts appearing in scientific texts, are represented in a couple-pattern matrix over co-occurrence contexts. Various possible contextual representation features are compared, using sequential pattern mining and syntactic path extraction. We compare the purely lexical feature space with a combined representation, and conclude that adding syntactic features has the potential to improve clustering performance.

Mots-clés : Extraction d'Information, relations sémantiques, apprentissage non supervisé.

Keywords: Information Extraction, semantic relations, clustering.

1 Introduction

La tâche d'extraction de relations vise à reconnaître automatiquement la nature des relations sémantiques qui relient des tuples d'entités ou de concepts présents dans un corpus. Cette tâche est une composante essentielle de l'extraction d'information et un préalable indispensable à l'alimentation automatique de bases de connaissances à partir de textes.

L'extraction de relations est le plus souvent abordée par apprentissage automatique supervisé,

en se servant en entraînement d'instances de tuples dont la relation sémantique est déjà connue et annotée (Hobbs & Riloff, 2010; Zhou *et al.*, 2005; Weeds *et al.*, 2014; Turney & Mohammad, 2014; Turney, 2012). Dans ce cas, le nombre et la nature des relations sémantiques possibles sont fixés à l'avance. L'extraction d'information en domain ouvert (OpenIE) (Banko *et al.*, 2007; Del Corro & Gemulla, 2013; Ferret, 2015) est beaucoup moins contrainte. Son but est d'inférer des relations sémantiques entre couples d'entités ou de concepts de manière non supervisée à partir des données textuelles quelconques. Certaines approches se focalisent sur l'extraction des instances de relations en contexte (Angeli *et al.*, 2015). Dans d'autres approches, les couples sont automatiquement regroupés dans des clusters, et la liste de leurs relations sémantiques possibles n'est pas fixée à l'avance. Cette méthode passe aussi par la représentation des instances dans un espace d'attributs. Mais, quand on opère de façon non supervisée, il est plus risqué d'utiliser un espace de représentation hétérogène pour les couples de concepts. C'est pourquoi, alors que les approches supervisées tentent de combiner toutes sortes d'attributs, les non supervisées exploitent en général essentiellement les segments de textes qui les relient (Hearst, 1992; Yangarber *et al.*, 2002; Béchet *et al.*, 2012; Turney, 2005, 2006). Les patrons ne se limitent pour autant pas nécessairement à de simples séquences de mots ; ils peuvent intégrer des combinaisons d'informations lexicales et syntaxiques (Fader *et al.*, 2011). D'autre part, il est également possible de construire l'espace d'attributs en s'appuyant non pas sur les patrons qui caractérisent les co-occurrences des deux entités, mais sur la représentation vectorielle des entités individuelles. De telles approches se sont montrées exploitables pour le calcul non supervisé des analogies relationnelles (Mikolov *et al.*, 2013), des relations lexicales comme l'hypéronymie (Santus *et al.*, 2014) et pour des relations sémantiques génériques (Gábor *et al.*, 2017), mais leur performance reste limitée sur les domaines de spécialité (Gábor *et al.*, 2016b).

C'est aussi dans un cadre non supervisé que nous nous plaçons. Nous ne cherchons toutefois pas à extraire des connaissances générales de textes quelconques, comme en OpenIE. Nous avons choisi de nous focaliser sur un corpus constitué d'articles du domaine du TAL, pour lesquels nous pouvons servir d'experts en validation. En ce sens, nos travaux se rapprochent de ceux menés dans le domaine des données médicales et de la bioinformatique, où l'analyse automatique d'articles scientifiques est déjà largement explorée. Or, dans ce domaine, il est courant d'intégrer dans les attributs le résultat d'analyseurs syntaxiques en dépendances (Fundel *et al.*, 2007; Porumb *et al.*, 2015). Le "plus court chemin syntaxique" en dépendances reliant deux entités ("shortest path hypothesis") est ainsi souvent utilisé (éventuellement en combinaison avec des méthodes à base de noyaux) pour représenter ce couple d'entités (Bunescu & Mooney, 2005; Mooney & Bunescu, 2005). Néanmoins, si cette approche a fait ses preuves en apprentissage supervisé et semi-supervisé (Nakamura-Delloye & Clergerie, 2010; Nakamura-Delloye & Stern, 2011), elle a été à notre connaissance encore peu explorée en clustering.

Dans une étude portant sur des articles de TAL, il a été récemment montré qu'une représentation à base de patrons séquentiels fréquents permettait un meilleur clustering des couples de concepts que la prise en compte du texte complet qui les sépare (Gábor *et al.*, 2016b). Nous proposons ici d'évaluer l'apport d'informations syntaxiques issues d'analyses en dépendances à cette tâche de clustering de couples de concepts scientifiques en domaine de spécialité.

Dans ce qui suit, nous décrivons tout d'abord (section 2) nos données et la nature de la tâche traitée. Les attributs syntaxiques utilisés sont présentés en section 3, tandis que la section 4 est consacrée aux conditions expérimentales de nos expériences et à leurs résultats.

La section 5 conclut en indiquant la direction de futures investigations.

2 Définition de la tâche

Soit a_1, a_2, b_1, b_2 des concepts extraits d'un corpus et pertinents pour le domaine considéré. a_1 et a_2 apparaissent dans une même phrase, de même que b_1 et b_2. Nous voulons réunir les couples $a = (a_1, a_2)$ et $b = (b_1, b_2)$ dans un même groupe (ou cluster) si les relations sémantiques qui lient a_1 à a_2 d'une part, et b_1 à b_2 d'autre part, sont similaires. Pour cela, nous avons besoin de représenter (a_1, a_2) et (b_1, b_2) dans un même espace vectoriel permettant de calculer la similarité $sim(a, b)$ sur laquelle se basera un algorithme de clustering.

Les conditions expérimentales et les données utilisées pour nos expériences sont les mêmes que celles décrites dans (Gábor *et al.*, 2016b), cela permettra les comparaisons. Le corpus initial est ACL-RelAcS (Gábor *et al.*, 2016a), issu du ACL Anthology Corpus (Radev *et al.*, 2009). Les concepts scientifiques pertinents pour faire partie des couples d'instances sont supposés connus. Les couples pris en compte sont ceux dont les concepts appartiennent à une même phrase. La Table 1 donne des exemples de différentes portions de textes reliant deux mêmes concepts dans le corpus.

argument 1	context/relation content	argument 2
domain task		*(language) data*
parser	*trained to minimize cost over*	sentences
learning algorithm	*is trained to simultaneously chunk*	sentences
parser	*is learned given a set of*	sentences
algorithm	*is presented for learning a*	phrase-structure

TABLE 1 – Exemple : relation "task_applied" sur un même couple de concepts dans différents contextes

Dans (Gábor *et al.*, 2016b), les attributs ayant permis les meilleurs clusterings étaient ceux extraits par fouille de données séquentielle (Srikant & Agrawal, 1996; Béchet *et al.*, 2012). Dans ce domaine, une séquence est une liste ordonnée de "littéraux" (ou items), dont un patron séquentiel est une sous-liste. Dans notre application, le rôle des littéraux est joué par les mots du corpus, auxquels aucun traitement linguistique n'a été appliqué. Seuls les patrons séquentiels "fermés", c'est-à-dire qui ne sont pas des sous-listes de patrons de même support, ont été considérés. Dans ce qui suit, nous cherchons à enrichir cet espace de nouveaux attributs en tenant compte de la syntaxe.

3 Represéntation syntaxique

Pour réaliser l'analyse syntaxique des phrases de notre corpus contenant des concepts, nous avons utilisé un parser en dépendances. Cette approche est la plus courante en extraction de relations (cf. l'état de l'art complet disponible dans (Valsamou, 2017)). Les analyses en dépendances via les parsers de type "shift-reduce" (Nivre, 2003) ne nécessitent pas la définition d'une grammaire formelle et sont plus adaptées aux langues à ordre souple. Dans

ce paradigme, la structure d'une phrase est décrite en termes de relations binaires typées entre mots (ou unités lexicales). Chaque mot est associé à un unique autre mot qui est son "gouverneur", dont il est soit un argument soit un modifieur. La Figure 1 montre une telle analyse pour une phrase du corpus.

Ayant choisi la représentation basique des dépendances universelles de Stanford, tous les arbres de dépendances obtenus sont des graphes orientés acycliques ou DAG en anglais (*Directed Acyclic Graphs*), même dans les cas de coordination. Les arcs sont étiquetés par la nature de la relation syntaxique qui relie un "gouverneur" à son "gouverné". Nous avons choisi d'utiliser l'analyseur syntaxique de Stanford version 3.8.0 car il offre une représentation en dépendances orientée vers la sémantique avec un jeu d'étiquettes proche du projet des Universal Dependencies qui anime une partie de la communauté TAL [1]. La version *collapsed* des dépendances n'a pas été utilisée dans cette expérience. Cette version produit des graphes orientés qui ne sont pas forcément des arbres. Cela permet de prendre en compte la variété des contextes, mais entraîne la possiblité d'avoir plus d'un plus court chemin entre deux entités. Or dans nos arbres, contrairement aux graphes généraux qui peuvent contenir un ou plusieurs chemins les plus courts, il n'y a qu'un seul plus court chemin possible.

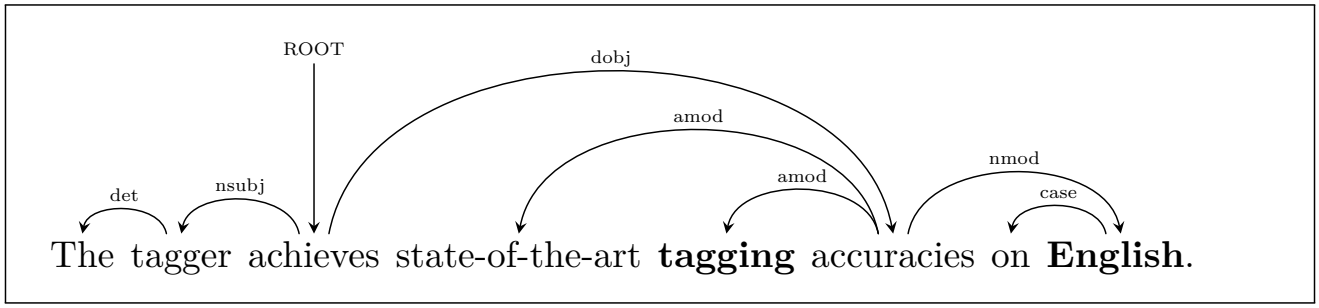

FIGURE 1 – Analyse syntaxique en dépendances d'une phrase du corpus

Bunescu et Mooney (Bunescu & Mooney, 2005) ont montré que le plus court chemin entre deux concepts ou entités dans une structure en dépendances est une information utile pour identifier la nature de la relation sémantique qui relie ces deux concepts. Dans l'analyse de la Figure 1, les deux mots "tagging" et "English" sont en gras parce que ce sont des concepts susceptibles de former un couple exprimant la relation sémantique *task_applied* (une tâche est appliquée sur des données) dont des exemples sont fournis en Table 1. Le plus court chemin qui les relie dans cette analyse passe par le mot "accuracies", dont ils sont tous les deux des modifieurs. Les informations syntaxiques que l'on peut extraire de l'analyse en dépendances pour caractériser ce couple de concepts sont détaillées en Figure 2.

1. Il s'agit d'un projet qui développe des corpus arborés pour de nombreuses langues dans le but de faciliter le développement d'analyseurs syntaxiques automatiques multilingues, l'apprentissage multilingue en ligne et les travaux de recherche en analyse syntaxique automatique dans une perspective de typologie linguistique. Le schéma d'annotation est hérité des dépendances de Stanford développées par de Marneffe et al. (en 2006, 2008, et 2014), du jeu d'étiquettes universelles de catégories *Part-of-speech* de Google et de l'*Interset interlingua for morphosyntactic tagsets*. Ici, nous nous référons uniquement aux *USD*, c'est-à-dire aux dépendances universelles de l'analyseur syntaxique de Stanford, et non au projet des *Universal Dependencies* (v2) dont on trouve un historique en ligne.

Analyse du plus court chemin entre les entités (tagging-5, English-8)
Path nodes: tagging-5, accuracies-6, english-8
Shortest path: **tagging** <– amod <– accuracies –> nmod –> **English**
Distance: 2

FIGURE 2 – Informations syntaxiques extraites de l'analyse pour un couple de concepts

4 Expériences

Nos expériences visent à mesurer l'impact de la prise en compte d'attributs syntaxiques sur les résultats du clustering de couples de concepts. Dans (Gábor *et al.*, 2016b,c), une classification manuelle des couples en 20 classes sémantiques avait été réalisée afin d'évaluer la qualité des clusterings. Nous reprenons ces données pour nos évaluations [2].

Pour les analyses syntaxiques, nous avons utilisé la version 3.8.0 du parser de Stanford, entraîné sur le Penn Treebank (Marneffe *et al.*, 2006). Les analyses en dépendances qu'il produit sont orientées vers la sémantique (Marneffe & Manning, 2008b). La version de base des types de dépendances (voir (Marneffe & Manning, 2008a) pour une description des étiquettes) a été choisie comme paramètre par défaut.

La sortie de l'analyseur syntaxique appliqué sur une phrase est une liste de relations binaires typées entre les mots de cette phrase. Pour chaque couple de concepts apparaissant dans une même phrase, nous avons extrait le plus court chemin en dépendances issu de ces relations binaires. La séquence des étiquettes de dépendances figurant sur ce plus court chemin (indépendamment de leur orientation) est devenu un attribut de l'espace de représentation des couples. Ainsi, l'exemple de la Figure 2 produit comme attribut la séquence "amod-nmod". Pour construire l'espace global de représentation des couples, le même filtre a été appliqué sur les séquences de mots et sur les séquences de dépendances : seuls les attributs apparaissant au moins 5 fois dans les données dont au moins 2 fois dans un couple en relation sémantique non vide ont été conservés. L'espace de représentation utilisé dans nos expériences était constitué soit des patrons séquentiels de mots seuls (1930 patrons séquentiels distincts), soit des seules séquences de dépendances (1191 séquences distinctes), soit d'un mélange des deux.

Le clustering a été réalisé par une approche hiérarchique agglomérative avec initialisation bissective (Zhao & Karypis, 2002) implementée dans cluto (Zhao *et al.*, 2005). L'initialisation par bissections successives produit un certain nombre de centroïdes qui viennent s'ajouter aux dimensions de l'espace de représentation initial. La valeur des couples sur ces nouvelles dimensions est leur distance au centroïde.

Les clusters obtenus ont été évalués relativement aux données de référence en termes de précision, rappel et F1-mesure. Les matrices auxquelles ont été soumis l'algorithme de clustering avec divers paramètres (3 valeurs distinctes du nombre de classes à trouver ont été testées) contenaient soit les simples comptes en nombre d'occurrences des attributs représentés dans chaque couple (Table 2), soit une variante avec les pondérations PPMI_α, fondées sur un calcul d'informations mutuelles (Levy *et al.*, 2015) (Table 3).

2. Les concepts exprimés par des expressions multimots ont été retirés de cette liste pour ces premières expériences, car ils peuvent poser des problèmes lors de l'analyse syntaxique. Nos données annotées comprennent

Input	#clusters	weight	Prec	Recall	F-measure
word seq	100	none	0.2590	0.1702	0.2054
word seq	50	none	0.2364	0.1873	0.2090
word seq	25	none	0.2293	0.2127	0.2207
parse	100	none	0.2760	0.1672	0.2082
parse	50	none	0.2443	0.1866	0.2116
parse	25	none	0.2346	0.2078	0.2204
combined	100	none	**0.2946**	0.1973	0.2364
combined	50	none	0.2755	0.2249	0.2476
combined	25	none	0.2640	**0.2390**	**0.2509**

TABLE 2 – Clustering : comparaison des différentes représentations sans pondération

Sans pondération, on observe des résultats très similaires en utilisant l'espace de représentation à base de patrons séquentiels de mots et celui à base de séquences de dépendances. Mais si on combine les deux représentations dans un seul espace, les résultats s'améliorent systématiquement d'environ 0,03 de F1-mesure. Cela suggère que les deux représentations capturent des informations différentes et complémentaires. Leur combinaison ne pénalise pas le clustering, malgré des différences de nature et d'échelles (les attributs syntaxiques sont plus génériques que les autres, et tendent donc à recevoir des valeurs plus grandes).

Features	#clusters	weight	Prec	Recall	F-measure
word seq	100	PPMI_α	**0.5337**	0.1906	0.2809
word seq	50	PPMI_α	0.4161	0.2759	0.3318
word seq	25	PPMI_α	0.3756	**0.2993**	**0.3332**
parse	100	PPMI_α	0.2300	0.1121	0.1507
parse	50	PPMI_α	0.1996	0.1461	0.1687
parse	25	PPMI_α	0.1952	0.1661	0.1795
combined	100	PPMI_α	0.4252	0.1738	0.2467
combined	50	PPMI_α	0.3552	0.2501	0.2935
combined	25	PPMI_α	0.3370	0.2889	0.3111

TABLE 3 – Clustering : comparaison des différentes représentations avec pondération PPMI_α

Un des résultats rapportés dans (Gábor *et al.*, 2016b) était que la pondération PPMI_α était particulièrement adaptée aux représentations à base de patrons séquentiels de mots. Nos expériences confirment ce constat, mais montrent aussi qu'il ne se généralise pas aux représentations à base de séquences de dépendances, pour lesquelles on note au contraire une dégradation. Nous espérions que la pondération permettrait d'équilibrer les échelles dans la variante combinée des réprésentations, et aboutirait aux meilleurs résultats, mais ce n'est pas confirmé par nos expériences. En observant de plus près les clusters obtenus, nous pensons que cette pondération est bénéfique à la prise en compte des patrons séquentiels peu fréquents, mais ne suffit pas à compenser les écarts d'échelles. Une explication possible est que l'information mutuelle redonne de l'importance aux événements rares (les patrons

donc un total de 567 couples, au lieu de 614 dans (Gábor *et al.*, 2016b).

séquentiels peu fréquents) mais que, pour ce qui concerne les attributs syntaxiques, ce sont les plus fréquents qui sont les plus porteurs d'information.

5 Conclusion et travaux futurs

Nous avons présenté dans cet article les premiers résultats d'expériences destinées à évaluer l'apport d'attributs syntaxiques pour le clustering de relations sémantiques dans un corpus de spécialité. Les résultats sont encourageants : une même relation sémantique peut être exprimée de façons très différentes dans un corpus, et il est donc important de prendre en compte des attributs divers susceptibles de la caractériser. Combiner des attributs de natures diverses est toujours risqué en clustering, nos premières expériences montrent que la combinaison envisagée est profitable. Il est aussi satisfaisant de constater que les seuls attributs syntaxiques permettent d'aboutir à d'aussi bons clusters que les seuls patrons séquentiels de mots.

La pondération PPMI_α, quant à elle, ne semble adaptée qu'à un seul type d'attributs parmi les deux familles testées. Cela ne remet pas en cause l'intérêt de la combinason (dans la Table 3, la combinaison des attributs reste meilleure que quand on utilise les attributs syntaxiques seuls) mais montre simplement que PPMI_α n'est pas adaptée aux dépendances syntaxiques. La suite de notre travail consistera donc notamment à chercher d'autres pondérations plus performantes pour cette représentation, et pour la combinaison des attributs.

Références

Angeli G., Premkumar M. J. J. & Manning C. D. (2015). Leveraging linguistic structure for open domain information extraction. In *ACL 2015*, p. 344–354.

Banko M., Cafarella J., Soderland S., Broadhead M. & Etzioni O. (2007). Open information extraction from the web. In *IJCAI*.

Béchet N., Cellier P., Charnois T. & Crémilleux B. (2012). Discovering linguistic patterns using sequence mining. In *CICLing '12*.

Bunescu R. C. & Mooney R. J. (2005). A shortest path dependency kernel for relation extraction. In *HLT-EMNLP'05*.

Del Corro L. & Gemulla R. (2013). Clausie : Clause-based open information extraction. In *International Conference on World Wide Web*, WWW '13.

Fader A., Soderland S. & Etzioni O. (2011). Identifying relations for open information extraction. In *EMNLP '11*.

Ferret O. (2015). *Language Production, Cognition, and the Lexicon*, chapter Typing Relations in Distributional Thesauri, p. 113–134. Springer International Publishing.

Fundel K., Küffner R. & Zimmer R. (2007). Relex—relation extraction using dependency parse trees. *Bioinformatics*, **23**(3), 365.

Gábor K., Zargayouna H., Buscaldi D., Tellier I. & Charnois T. (2016a). Semantic annotation of the acl anthology corpus for the automatic analysis of scientific literature. In *LREC '16*.

Gábor K., Zargayouna H., Buscaldi D., Tellier I. & Charnois T. (2016b). Unsupervised relation extraction in specialized corpora using sequence mining. In *Advances in Intelligent Data Analysis XV (IDA 2016), LNCS 9897*.

Gábor K., Zargayouna H., Tellier I., Buscaldi D. & Charnois T. (2016c). A typology of semantic relations dedicated to scientific literature analysis. In *SAVE-SD Workshop at the 25th World Wide Web Conference, LNCS 9792*.

Gábor K., Zargayouna H., Tellier I., Buscaldi D. & Charnois T. (2017). Exploring vector spaces for semantic relations. In *EMNLP 2017*, p. 1814–1823.

Hearst M. (1992). Automatic acquisition of hyponyms from large text corpora. In *COLING '92*, p. 539–545.

Hobbs J. R. & Riloff E. (2010). Information extraction. In N. Indurkhya & F. J. Damerau, Eds., *Handbook of Natural Language Processing, Second Edition*. Boca Raton, FL : CRC Press, Taylor and Francis Group.

Levy O., Goldberg Y. & Dagan I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the ACL*, **3**.

Marneffe M.-C. d., MacCartney B. & Manning C. D. (2006). Generating typed dependency parses from phrase structure parses. In *LREC '06*.

Marneffe M.-C. d. & Manning C. D. (2008a). Stanford typed dependencies manual. The Stanford NLP Group. revised for the Stanford Parser v. 3.7.0 in September 2016.

Marneffe M.-C. d. & Manning C. D. (2008b). The stanford typed dependencies representation. In *COLING Workshop on Cross-framework and Cross-domain Parser Evaluation*.

Mikolov T., Yih W. & Zweig G. (2013). Linguistic regularities in continuous space word representations. In *NAACL*.

Mooney R. J. & Bunescu R. (2005). Mining knowledge from text using information extraction. *SIGKDD Explor. Newsl.*, **7**(1), 3–10.

Nakamura-Delloye Y. & Clergerie E. d. l. (2010). Exploitation de résultats d'analyse syntaxique pour extraction semi-supervisée des chemins de relations. In *17e Conférence sur le Traitement Automatique des Langues Naturelles - TALN 2010*.

Nakamura-Delloye Y. & Stern R. (2011). Extraction de relations et de patrons de relations entre entités nommées en vue de l'enrichissement d'une ontologie. In *TOTh 2011 : Terminologie & Ontologie : Théories et Applications*, p.$\tilde{5}$0.

Nivre J. (2003). An efficient algorithm for projective dependency parsing. p. 149–160.

Porumb M., Barbantan I., Lemnaru C. & Potolea R. (2015). Remed : Automatic relation extraction from medical documents. In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, iiWAS '15 : ACM.

Radev D., Muthukrishnan P. & Qazvinian V. (2009). The ACL Anthology Network Corpus. In *ACL Workshop on Text and Citation Analysis for Scholarly Digital Libraries*.

Santus E., Lenci A., Lu Q. & Schulte Im Walde S. (2014). Chasing hypernyms in vector spaces with entropy. In *EACL'14*.

Srikant R. & Agrawal R. (1996). Mining sequential patterns : Generalizations and performance improvements. In *EDBT*, p. 3–17.

Turney P. D. (2005). Measuring semantic similarity by latent relational analysis. In *IJCAI-05*.

Turney P. D. (2006). Similarity of semantic relations. *CoRR*, **abs/cs/0608100**.

Turney P. D. (2012). Domain and function : A dual-space model of semantic relations and compositions. *Journal of Artificial Intelligence Research*, **44**.

Turney P. D. & Mohammad S. M. (2014). Experiments with three approaches to recognizing lexical entailment. *Natural Language Engineering*.

Valsamou D. (2017). *Extraction d'Information pour les réseaux de régulation de la graine chez Arabidopsis Thaliana*. Thèse de doctorat, Université Paris-Saclay, Ecole doctorale 580 Sciences et technologies de l'information et de la communication (STIC).

Weeds J., Clarke D., Reffin J., Weir D. & Keller B. (2014). Learning to distinguish hypernyms and co-hyponyms. In *COLING '14*.

Yangarber R., Lin W. & Grishman R. (2002). Unsupervised learning of generalized names. In *COLING '02*.

Zhao Y. & Karypis G. (2002). Evaluation of hierarchical clustering algorithms for document datasets. In *CIKM*.

Zhao Y., Karypis G. & Fayyad U. (2005). Hierarchical clustering algorithms for document datasets. *Data Mining for Knowledge Discovery*, **10**.

Zhou G., Su J., Zhang J. & Zhang M. (2005). Exploring various knowledge in relation extraction. In *ACL '05*.