

PROMÉTHÉE

un outil d'aide à l'acquisition de relations sémantiques entre termes

Emmanuel MORIN

Institut de Recherche en Informatique de Nantes (IRIN)
2 rue de la houssinière, BP 92208
44322 Nantes Cedex 3, FRANCE
Tél. : 33 02 51 12 58 42
Emmanuel.Morin@irin.univ-nantes.fr

Résumé

Les outils d'aide à la construction de terminologie à partir de corpus ont connu un essor important ces dernières années. D'un autre côté, les outils d'aide à l'acquisition de relations sémantiques entre termes sont peu nombreux. Face à ce problème, nous avons développé le système PROMÉTHÉE qui acquiert incrémentalement un ensemble de patrons lexico-syntaxiques caractéristiques d'une relation sémantique. Ainsi, pour la relation d'hyponymie, PROMÉTHÉE a extrait un ensemble de patrons qui servent à améliorer la couverture d'un thesaurus ou d'une base de connaissances.

1. Introduction

Les outils d'aide à la construction de terminologie à partir de corpus ont connu un essor important ces dernières années, comme en témoignent les systèmes XTRACT (Smadja, 1993), ANA (Enguehard, 1993), ACABIT (Daille, 1994), LEXTER (Bourigault, 1994), FASTER (Jacquemin, 1997). D'un autre côté, les outils d'aide à l'acquisition de relations sémantiques entre termes sont peu nombreux. Face à ce problème, nous avons développé le système PROMÉTHÉE qui acquiert incrémentalement un ensemble de patrons lexico-syntaxiques caractéristiques d'une relation sémantique. Ainsi, pour la relation d'hyponymie, PROMÉTHÉE a extrait un ensemble de patrons qui servent à améliorer la couverture d'un thesaurus ou d'une base de connaissances.

Dans la suite de cet article, nous revenons sur les différentes approches d'acquisition de relations sémantiques entre termes en section 2. La section 3 présente l'architecture du système PROMÉTHÉE. La section 4 décrit la méthode employée pour l'acquisition des patrons lexico-syntaxiques. La section 5 évalue PROMÉTHÉE pour une relation sémantique spécifique : l'hyponymie. Enfin, la section 6 dresse le bilan de ce travail.

2. Acquisition de relations sémantiques à partir de corpus de textes

Dans cette partie, nous nous intéressons aux approches qui supposent qu'un certain niveau d'analyse syntaxique est nécessaire pour acquérir des informations sémantiques.

Une technique classique consiste à étudier le contexte syntaxique où les termes apparaissent. Ainsi, Ruge (1991) regroupe des mots similaires en utilisant des relations tête-modifieur des syntagmes. De même, Hindle (1990) s'intéresse aux noms qui sont à la fois sujets et objets du même verbe en combinant une analyse syntaxique avec des mesures statistiques. Par cette méthode, il détermine que les mots les plus similaires à *boat* (*bateau*) sont *ship* (*navire*), *plane* (*avion*), *bus* (*autobus*), *jet* (*avion à réaction*), *vessel* (*vaisseau*), *truck* (*camion*), *car* (*voiture*), *helicopter* (*hélicoptère*), *ferry* (*bac*) et *man* (*homme*). Comme Hindle, Grefenstette (1994) exploite les contextes syntaxiques des mots pour extraire des mots sémantiquement proches. Dans un premier temps, son système SEXTANT extrait différentes structures syntaxiques : adjectifs-noms, noms-noms et verbes-noms (ces informations fournissent un contexte pour chaque terme du corpus). Ensuite, SEXTANT compare le contexte de chaque terme et détermine, pour un mot donné, quels sont les mots les plus proches. Cette technique n'utilise pas de connaissances spécifiées à la main et est plus performante que les techniques à base de cooccurrences (e.g. Church et Hanks (1990)). Cependant, un mot et son antonyme peuvent se retrouver dans la même classe sémantique. D'une façon générale, ces différentes techniques ont l'inconvénient de ne pas identifier la relation qui existe entre un terme et son cooccurrent.

Une autre alternative consiste à représenter sous la forme d'un patron lexico-syntaxique la relation sémantique qui existe entre un ou plusieurs termes. Robison (1970) a été le premier, à ma connaissance, à utiliser de tels patrons (environ 40 000 pour l'anglais) pour extraire automatiquement des informations sémantiques (e.g. *transformation of S from S into S*, où *S* est un substantif). De la même manière, Hearst (1992) utilise des patrons lexico-syntaxiques pour extraire des relations lexicales entre mots à partir de corpus de textes. Par exemple, le patron $NP \{ , \} \textit{especially} \{ NP , \}^* \{ or | and \} NP$ (où *NP* est un syntagme nominal) appliquée à la phrase «(...) *most European countries, especially France, England and Spain*» extrait trois relations lexicales : hyponymie («France», «European country»), hyponymie («England», «European country») et hyponymie («Spain», «European country»). Par la suite, les couples de termes extraits peuvent être inclus dans un thesaurus. Des travaux similaires ont été réalisés sur le français par (Rousselot *et al.*, 1992) et (Desclés & Jouis, 1993). Dans ce type d'approche, l'acquisition des patrons est un problème majeur.

Dans des domaines plus spécifiques et avec des techniques plus fines, les systèmes d'Extraction d'Informations (Information Extraction, IE) utilisent des structures syntaxiques locales pour extraire des relations caractéristiques d'un domaine. Les systèmes NYU (Grishman, 1995) et FASTUS (Hobbs *et al.*, 1987) repèrent des structures syntaxiques pour les syntagmes nominaux et verbaux. Les systèmes d'IE obtiennent de bons résultats dans un domaine spécifique, mais sont difficilement adaptables à un nouveau domaine.

En modélisation des connaissances, dans la continuité des recherches concernant l'utilisation de LEXTER (Bourigault, 1994), Assadi et Bourigault (1996) proposent au cognicien un ensemble d'outils d'aide à l'exploitation du réseau de candidats termes. Dans ce réseau, chaque terme est relié, d'une part, à sa tête, et d'autre part, à tous les candidats termes dont il est lui-même tête ou expansion. Un filtrage statistique est effectué sur le réseau terminologique et le typage des liens sémantiques est réalisé en combinaison avec les compétences du cognicien.

Dans la suite de cet article, nous présentons PROMÉTHÉE, un système d'acquisition de rela-

tions sémantiques entre termes qui repose sur l'utilisation de patrons lexico-syntaxiques.

3. Architecture de PROMÉTHÉE

PROMÉTHÉE est un système d'Extraction d'Informations (IE) à partir de corpus de textes techniques¹. Son fonctionnement repose principalement sur deux modules : un préprocesseur lexical et un analyseur lexico-syntaxique (cf. figure 1).

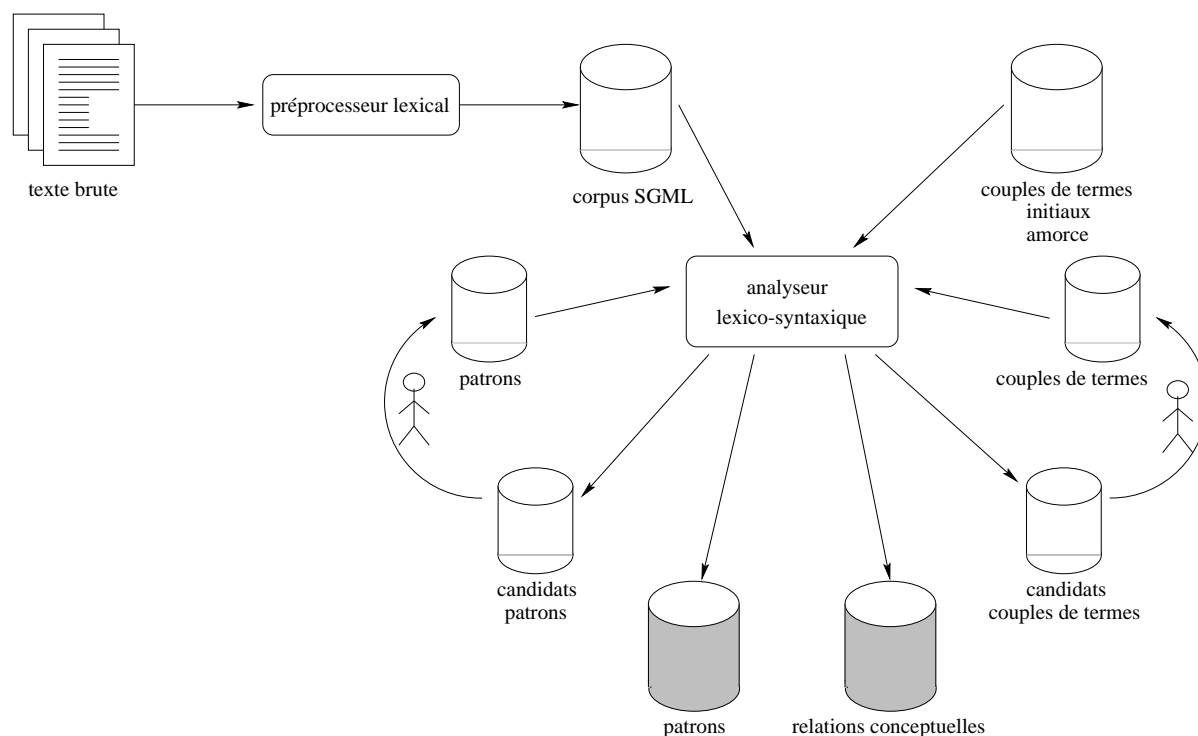


FIG. 1 – Architecture de PROMÉTHÉE

3.1. Préprocesseur lexical

À ce niveau, un ensemble de traitements est réalisé pour mettre en forme le corpus. Ainsi, les différentes phrases du corpus sont repérées et étiquetées. Les syntagmes nominaux, les acronymes et les successions de syntagmes nominaux sont détectées par des expressions régulières (*via* PERL). La sortie de ce module est un corpus au format SGML (Standard Generalized Markup Language).

3.2. Analyseur lexico-syntaxique

L'analyseur lexico-syntaxique a deux fonctionnalités. D'une part l'acquisition de patrons lexico-syntaxiques propres à une relation sémantique, et d'autre part, l'extraction de couples de termes conceptuellement liés.

1. Toutes les expériences évoquées dans cet article ont été réalisées sur le corpus [AGRO] fourni par l'INIST-CNRS. Ce corpus français, de 1,3 millions de mots, est constitué de résumés d'articles scientifiques en agronomie.

2. Nous remercions Évelyne Tzoukermann des Laboratoires Bell de la Société Lucent Technologies pour avoir mis à notre disposition une version étiquetée du corpus [AGRO].

4. Acquisition de patrons lexico-syntaxiques

L'acquisition de patrons lexico-syntaxiques caractéristiques d'une relation sémantique se fait incrémentalement. Ce processus est constitué de 7 étapes :

1. Choisir la relation sémantique pour laquelle on souhaite acquérir des patrons lexico-syntaxiques (e.g. l'hyponymie).
2. Fournir une amorce constituée de couples de termes qui respectent la relation précédemment spécifiée. Cette liste peut être obtenue à partir d'un thesaurus, d'une base de connaissances, ou bien manuellement spécifiée. Par exemple, un thesaurus d'agronomie fournira que le *calcium* EST-UN *cation*.
3. Extraire du corpus l'ensemble des phrases qui contiennent les précédents couples. Ainsi le couple («*cation*»,«*calcium*») sélectionne dans le corpus [AGRO] la phrase «*Des cations tels que le sodium, le potassium, le calcium et le magnésium peuvent être dosés par une méthode de routine*».
4. Trouver un environnement commun qui généralise les phrases extraites à l'étape 3. Cet environnement, écrit sous la forme d'une expression lexico-syntaxique, révèle un candidat patron lexico-syntaxique.
5. Un terminologue retient les candidats patrons les plus pertinents.
6. Utiliser les nouveaux patrons pour extraire de nouveaux candidats couples de termes.
7. Un terminologue retient les candidats couples de termes les plus pertinents. Ces nouveaux couples sont ajoutés à ceux de la liste initiale, puis le processus est réitéré à partir de l'étape 3.

Nous allons maintenant détailler les étapes 3 et 4 qui constituent le cœur de notre méthode.

4.1. Expression lexico-syntaxique

À la troisième étape, nous avons extrait un ensemble de phrases. Celles-ci sont lemmatisées et les syntagmes nominaux sont repérés. Ainsi, nous représentons une phrase sous la forme d'une expression lexico-syntaxique. Par exemple, la phrase «*Des cations tels que le sodium, le potassium, le calcium et le magnésium peuvent être dosés par une méthode de routine*» est représentée par l'expression lexico-syntaxique SN tel que LIST pouvoir être doser par SN³. Par ce processus de simplification, nous avons une représentation plus générale des phrases, et leur comparaison est facilitée.

4.2. Regroupement des expressions lexico-syntaxiques

L'objectif de l'étape 4 est de regrouper dans une même classe les expressions lexico-syntaxiques, extraites à l'étape 3, qui sont *proches* (i.e. qui partagent le même environnement), puis, pour chaque classe, de généraliser un patron lexico-syntaxique qui les subsume.

3. SN est un syntagme nominal, et LIST une succession de syntagmes nominaux, comme dans l'expression régulière SN {, SN*} {et|ou} SN.

4.2.1. Représentation mathématique d’une expression lexico-syntaxique

Soit A une expression lexico-syntaxique, telle que :

$$= A_1 A_2 \cdots A_j \cdots A_k \cdots A_n \tag{1}$$

A_i désigne le i ème élément de l’expression A (*i.e.* un lemme, ou une étiquette syntaxique de type SN ou LIST), et n la longueur en nombre d’éléments de A . De plus, l’expression A possède une relation sémantique entre deux syntagmes nominaux A_j et A_k , que nous notons :

$$RELATION(A_j, A_k) \tag{2}$$

(avec $k - j > 1$, car il y a toujours au moins un élément entre A_j et A_k). Pour l’expression lexico-syntaxique SN tel que LIST pouvoir être doser par SN, nous avons donc : « $A_1 = A_j = \underline{SN}$ », « $A_2 = \text{tel}$ », « $A_3 = \text{que}$ », « $A_4 = A_k = \underline{LIST}$ », ..., « $A_8 = \text{par}$ » et « $A_9 = A_n = \underline{SN}$ ».

4.2.2. Comparaison de deux expressions lexico-syntaxiques

Soient A et B deux expressions lexico-syntaxiques, telles que :

$$A = A_1 A_2 \cdots A_j \cdots A_k \cdots A_n \text{ avec } RELATION(A_j, A_k) \tag{3}$$

et

$$B = B_1 B_2 \cdots B_{j'} \cdots B_{k'} \cdots B_{n'} \text{ avec } RELATION(B_{j'}, B_{k'}) \tag{4}$$

Nous notons $Sim(A, B)$ la fonction mesurant la similarité des expressions A et B . Cette fonction est réflexive et symétrique mais non nécessairement transitive (e.g. si une expression lexico-syntaxique X est similaire à une expression Y , et Y similaire à Z , alors il n’y a pas forcément similarité entre X et Z). Nous avons donc :

$$Sim(A, B) = Sim(A_1 A_2 \cdots A_n, B_1 B_2 \cdots B_{n'}) \tag{5}$$

D’après les définitions 1 et 2, chaque expression lexico-syntaxique possède une relation lexicale qui dénote une relation conceptuelle entre A_j et A_k . Par conséquent, les syntagmes nominaux A_j et $B_{j'}$, ainsi que A_k et $B_{k'}$, occupent potentiellement la même fonction syntaxique (cf. figure 2).

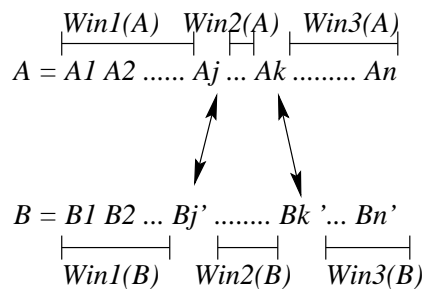


FIG. 2 – Comparaison de deux expressions lexico-syntaxiques

Les composantes A_j , A_k , $B_{j'}$ et $B_{k'}$ repr esentent des fronti eres de fenˆetres. Nous pouvons donc red efinir la fonction (1) par :

$$\begin{aligned} Sim(A, B) = & Sim(A_1 A_2 \cdots A_{j-1}, B_1 B_2 \cdots B_{j'-1}) + \\ & Sim(A_{j+1} \cdots A_{k-1}, B_{j'+1} \cdots B_{k'-1}) + \\ & Sim(A_{k+1} \cdots A_n, B_{k'+1} \cdots B_{n'}) \end{aligned} \quad (6)$$

Soit encore :

$$Sim(A, B) = \sum_{i=1}^3 Sim(Win_i(A), Win_i(B)) \quad (7)$$

avec :

$$\begin{cases} Win_1(A) = A_1 A_2 \cdots A_{j-1} \\ Win_2(A) = A_{j+1} \cdots A_{k-1} \\ Win_3(A) = A_{k+1} \cdots A_n \end{cases} \quad \text{et} \quad \begin{cases} Win_1(B) = B_1 B_2 \cdots B_{j'-1} \\ Win_2(B) = B_{j'+1} \cdots B_{k'-1} \\ Win_3(B) = B_{k'+1} \cdots B_{n'} \end{cases}$$

De plus,   chaque  l ement $Sim(Win_i(A), Win_i(B))$, nous associons un poids diff erent W_i . Ce facteur permet de donner une plus grande importance   l' l ement central qu'aux  l ements ext erieurs :

$$Sim(A, B) = \sum_{i=1}^3 W_i \times Sim(Win_i(A), Win_i(B)) \quad (8)$$

Par exemple, la relation *HYPERONYMIE*(«*cation*», «*calcium*») donne l'expression lexico-syntaxique SN tel que LIST pouvoir ˆtre doser par SN. De la mˆme mani ere, la relation *HYPERONYMIE*(«*mat eriau granulaire*», «*bl e*») extrait la phrase «*L' tude exp erimentale sur des mat eriaux granulaires, tels que le bl e, le ma is et le tournesol a  t   tudi  (...)*» donne l'expression SN sur SN , tel que LIST ˆtre  tudi  (...). Ici, les mots «*cation*» et «*mat eriaux granulaires*», ainsi que les mots «*calcium*» et «*bl e*» ont la mˆme fonction syntaxique (cf. figure 3). La fonction de similarit e est donc donn ee par :

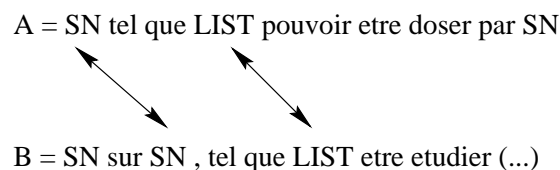


FIG. 3 – Exemple de comparaison de deux expressions lexico-syntaxiques

$$\begin{aligned} Sim(A, B) = & W_1 \times Sim(\text{«} \text{, «SN sur»}) + \\ & W_2 \times Sim(\text{«tel que»} \text{, «tel que»}) + \\ & W_3 \times Sim(\text{«pouvoir ˆtre doser par SN»} \text{, «ˆtre  tudi  (...)»}) \end{aligned}$$

4.2.3. D efinition de la Fonction de Similarit e

Maintenant, nous devons d efinir la fonction $Sim(A, B)$. Pour cela, nous avons  valu e quatre fonctions de similarit e : Salton, Cosinus, Jaccard (Salton & McGill, 1983, chap. 6) et (Wagner & Fischer, 1974). Les mesures de Salton, du Cosinus et de Jaccard sont couramment utilis ees en recherche documentaire pour mesurer la similarit e de deux vecteurs. En revanche, la mesure

de Wagner & Fischer détermine la distance entre deux chaînes comme étant le coût minimum de la composition des opérations nécessaires pour changer une chaîne en une autre chaîne (les opérateurs élémentaires sont : changer un symbole en un autre, effacer un symbole d'une chaîne, ou insérer un symbole dans une chaîne).

À partir d'un jeu de 35 phrases préalablement regroupées en différentes classes, nous avons évalué ces quatre mesures. Pour évaluer la qualité du regroupement, nous calculons trois scores : *Précision*, *Rappel* et *F-Mesure* (MUC-6, 1995). Nous notons par N_{total} le nombre total de réponses correctes, par $N_{correcte}$ le nombre de réponses correctes produit par le système, et par $N_{incorrecte}$ le nombre de réponse incorrectes produit par le système. Nous définissons :

$$Précision = \frac{N_{correcte}}{N_{correcte} + N_{incorrecte}} \quad Rappel = \frac{N_{correcte}}{N_{total}}$$

F-Mesure est un score combinant Précision et Rappel :

$$F = \frac{2 \times Précision \times Rappel}{Précision + Rappel}$$

Le meilleur résultat (avant l'introduction du concept de *Contiguïté*) est donnée pour la mesure de Jaccard (cf. table 1).

	Seuil	Précision	Rappel	F-Mesure
Salton	0,48	87,3	68,6	76,8
Cosinus	0,44	76,8	75,7	76,3
Jaccard	0,44	96,2	71,4	82,0
Wagner & Fischer	0,44	74,4	72,2	73,3
Contiguïté	0,46	93,0	85,7	89,5

TAB. 1 – Évaluation des mesures de similarité

Cependant les mesures utilisées ne prennent pas en compte la nature syntaxique des expressions lexico-syntaxiques. Pour remédier à cet inconvénient nous définissons une mesure spécifique qui introduit la notion de *Contiguïté*. La *Contiguïté* entre deux expressions lexico-syntaxiques A et B peut être vue comme la plus grande chaîne commune :

$$Sim(Win_i(A), Win_i(B)) = \frac{Contiguïté(Win_i(A), Win_i(B))}{Max(\|Win_i(A)\|, \|Win_i(B)\|)} \quad (9)$$

Ainsi, sur le jeu de 35 phrases cette mesure donne le meilleur résultat (cf. table 1).

4.2.4. Extraction des patrons lexico-syntaxiques

À partir de la mesure de *Contiguïté*, toutes les expressions lexico-syntaxiques extraites à l'étape 3 sont comparées deux à deux. Nous obtenons une matrice carrée dont chaque composante connexe reflète une classe au sein de laquelle les expressions lexico-syntaxiques sont proches.

Par exemple, les trois expressions suivantes sont regroupées dans la même classe.

- deux SN de chacun de 3 SN (LIST3) être cultiver avec SN (...)
Deux cultivars de chacune des 3 céréales de printemps (blé, orge et avoine) ont été cultivés avec de l'Avena fatua (...)

2. (...) SN de quatre SN (LIST4) dans le ensemble SN étudié .
(...) *analyse foliaire de quatre espèces ligneuses (chêne, frêne, lierre et cornouiller) dans l'ensemble des sites étudiés.*
3. après cinq SN sur le principal SN (LIST3) (...)
Après cinq années de résultats sur les principales cultures vivrières (sorgho, maïs, mil)
(...)

Pour cette classe, PROMÉTHÉE génère le candidat patron SN (LIST).

5. Évaluation de PROMÉTHÉE

Le système PROMÉTHÉE a été évalué sur la relation d'hyperonymie à partir d'une liste initiale de 41 couples de termes conceptuellement liés par la relation cible, nous avons extrait 20 candidats patrons, dont 11 sont retenus :

1. {deux|trois|...|2|3|4...} SN (LIST)
(...) *analyse foliaire de quatre espèces ligneuses (chêne, frêne, lierre et cornouiller) dans l'ensemble des sites étudiés.*
2. {certain|quelque|de autre...} SN (LIST)
3. {deux|trois|...|2|3|4...} SN : LIST
4. {certain|quelque|de autre...} SN : LIST
(...) *marqueurs moléculaires liés à des gènes de résistance aux principales maladies du pois : Fusariose, Oïdium, Anthraxnose et Mosaïque commune du pois.*
5. {de autre}? SN comme LIST
Le courrier électronique n'est qu'une des fonctions offertes par les réseaux de recherche comme Internet et Bitnet.
6. {de autre}? SN tel que LIST
7. SN tel LIST
8. SN, particulièrement SN,
(...) *on a étudié la densité des espèces de phlébotomes anthropophiles, particulièrement Lutzomyia trapidoi, dans le milieu domestique et les caféières adjacentes (...)*
9. SN {et|ou} de autre SN
10. SN et notamment SN
La fermentation alcoolique de sous-produits de sucrerie, se déroule en conditions non stériles, ce qui entraîne le développement de bactéries et notamment de germes lactiques.
11. chez SN, SN,
Chez les Phalaenopsis, Orchidées monopodiales, l'excision de méristèmes en vue d'une micropropagation détruit ou lèse gravement la plante souche.

Pour chaque patron, nous indiquons le nombre de phrases extraites, le nombre de couples extraits⁴ (après validation), la *Précision*, le *Rappel* et la *F-Mesure* des couples de termes (cf. table 2). En moyenne, la *Précision* sur le corpus [AGRO] est de 79% et le *Rappel* de 40%. Le faible taux de *Rappel* est principalement dû à un mauvais étiquetage du corpus, ce qui génère du bruit lors du repérage des syntagmes nominaux. En privilégiant la *Précision* sur le *Rappel*, on limite le coût du filtrage humain.

Patrons n^0	Phrases extraites	Couples extraits	Précision %	Rappel %	F-mesure %
1	80	186	88	36	51
2	45	134	89	30	45
3	85	167	75	32	45
4	30	82	84	30	44
5	38	59	64	41	50
6	96	147	85	47	66
7	9	23	100	45	73
8	4	4	100	36	53
9	51	17	33	47	40
10	8	6	75	43	55
11	18	11	62	52	57
Total	464	836	79	40	53

TAB. 2 – Évaluation des patrons de l'hyponymie

6. Conclusion et Perspectives

Nous avons présenté une nouvelle méthode pour l'acquisition de relations sémantiques entre termes dans des corpus de textes techniques. Cette approche, qui demande peu de connaissances initiales, combine filtres linguistiques et statistiques. Elle repose sur un processus incrémental qui acquiert des patrons lexico-syntaxiques propres à une relation sémantique. Ces patrons sont ensuite utilisés pour extraire des termes conceptuellement liés.

Nous comptons poursuivre ce travail en utilisant PROMÉTHÉE sur d'autres relations sémantiques, et étudier l'apport des couples extraits dans le cadre de la Recherche d'Informations (e.g. l'extension de requêtes).

Remerciements

Ce travail a bénéficié du soutien de l'AFIRST (Association Franco-Israélienne pour la Recherche Scientifique et Technologique) dans le cadre du projet SKETCHI sur l'acquisition de connaissances terminologiques à partir de corpus. Ce projet est une coopération entre l'Université de BAR ILAN représentée par Ido Dagan et l'IRIN représentée par Béatrice Daille.

4. Une phrase peut produire plusieurs couples de termes, s'il y a une succession de syntagmes nominaux.

Références

- ASSADI H. & BOURIGAULT D. (1996). Acquisition et modélisation de connaissances à partir de textes : outils informatiques et éléments méthodologiques. In *Actes, 9th Congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'96)*, p. 505–514, Rennes, France: A.F.C.E.T.
- BOURIGAULT D. (1994). *LEXTER un Logiciel d'Extraction de Terminologie. Application à l'extraction des connaissances à partir de textes*. Thèse en Mathématiques, Informatique Appliquée aux Sciences de l'Homme, École des Hautes Études en Sciences Sociales, Paris, France.
- CHURCH K. W. & HANKS P. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, **16**(1), 22–29.
- DAILLE B. (1994). *Approche mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistiques*. Thèse en Informatique Fondamentale, Université de Paris 7, Paris, France.
- DESLÉS J.-P. & JOUIS C. (1993). L'exploration contextuelle : une méthode linguistique et informatique pour l'analyse automatique de textes. In *Actes, Informatique & Langue Naturelle (ILN'93)*, p. 339–351, Nantes, France.
- ENGUEHARD C. (1993). Acquisition de terminologie à partir de gros corpus. In *Actes, Informatique & Langue Naturelle (ILN'93)*, p. 373–384, Nantes, France.
- GREFENSTETTE G. (1994). *Explorations in Automatic Thesaurus Discovery*. Boston, MA: Kluwer Academic Publisher.
- GRISHMAN R. (1995). The NYU System for MUC-6 or Where's the Syntax. In *Sixth Message Understanding Conference (MUC-6)*, p. 167–175, Columbia, MD.
- HEARST M. A. (1992). Automatic Acquisition of Hyponyms from Large Text Corpora. In *Actes, 14th International Conference on Computational Linguistics (COLING'92)*, p. 539–545, Nantes, France.
- HINDLE D. (1990). Noun classification from predicate argument structures. In *Actes, 28th Annual Meeting of the Association for Computational Linguistics (ACL'90)*, p. 268–275, Berkeley, CA, USA.
- HOBBS J. R., APPELT D., BEAR J., ISRAEL D., KAMEYAMA M., STICKEL M. & TYSON M. (1987). FASTUS : A Cascaded Finite-State Transducer for Extracting Information from Natural-Language Text. In E. ROCHE & Y. SCHABES, Eds., *Finite-State Language Processing*, p. 383–406. Cambridge, MA: The MIT Press.
- JACQUEMIN C. (1997). *Variations terminologiques : Reconnaissances et acquisition automatique de termes et de leurs variantes en corpus*. Habilitation à Diriger des Recherches, Université de Nantes, Faculté des Sciences, IRIN, Nantes, France.
- MUC-6 (1995). *Sixth Message Understanding Conference*. Columbia, Maryland: Morgan Kaufmann.
- ROBISON H. R. (1970). Computer-detectable semantic structures. *Information Storage and Retrieval*, **6**, 273–288.
- ROUSSELOT F., MIGAULT B. & IGOT P. (1992). Élaboration de techniques d'analyses adaptées à la construction d'une base de connaissances. In *Actes, 14th International Conference on Computational Linguistics (COLING'92)*, p. 483–489, Nantes, France.
- RUGE G. (1991). Experiments on Linguistically Based Term Associations. In *Actes, Intelligent Multimedia Information Retrieval Systems and Management (RIA0'91)*, p. 528–545, Barcelone, Espagne.
- SALTON G. & MCGILL M. J. (1983). *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.
- SMADJA F. (1993). XTRACT: An Overview. *Computer and the Humanities*, **26**, 399–413.
- WAGNER R. A. & FISCHER M. J. (1974). The String-to-String Correction Problem. *Association for Computing Machinery*, **21**(1), 168–173.