

Résolution de Métonymie des Entités Nommées : proposition d'une méthode hybride

Caroline Brun¹ Maud Ehrmann¹ Guillaume Jacquet¹

(1) Xerox Research Center Europe - XRCE

6, Chemin de Maupertuis, 38240 Meylan

{Caroline.Brun, Maud.Ehrmann, Guillaume.Jacquet}@xrce.xerox.com

Résumé. Dans cet article, nous décrivons la méthode que nous avons développée pour la résolution de métonymie des entités nommées dans le cadre de la compétition SemEval 2007. Afin de résoudre les métonymies sur les noms de lieux et noms d'organisation, tel que requis pour cette tâche, nous avons mis au point un système hybride basé sur l'utilisation d'un analyseur syntaxique robuste combiné avec une méthode d'analyse distributionnelle. Nous décrivons cette méthode ainsi que les résultats obtenus par le système dans le cadre de la compétition SemEval 2007.

Abstract. In this paper, we describe the method we develop in order to solve Named entity metonymy in the framework of the SemEval 2007 competition. In order to perform Named Entity metonymy resolution on location names and company names, as required for this task, we developed a hybrid system based on the use of a robust parser that extracts deep syntactic relations combined with a non supervised distributional approach, also relying on the relations extracted by the parser. We describe this methodology as well as the results obtained at SemEval 2007

Mots-clés : Entités Nommées, métonymie, méthode hybride, analyse syntaxique robuste, approche distributionnelle.

Keywords: Named Entities, metonymy, hybrid method, robust parsing, distributional approach.

1 Introduction

Le traitement des entités nommées fait aujourd'hui figure d'incontournable en Traitement Automatique des Langues (TAL). Apparue au milieu des années 1990 à la faveur des dernières conférences MUC (*Message Understanding Conferences*, (MUC, 1995) et (MUC, 1998)), la tâche de reconnaissance et de catégorisation des noms de personnes, de lieux, d'organisations, etc. apparaît en effet comme fondamentale pour diverses applications participant de l'analyse de contenu et nombreux sont les travaux se consacrant à sa mise en oeuvre, obtenant des résultats plus qu'honorables (F-mesure¹ dépassant généralement 0.9), et ce pour diverses langues. Fort de ce succès, le traitement des entités nommées s'oriente désormais vers de nouvelles perspectives avec, entre autres, la catégorisation fine (Ehrmann & Jacquet, 2006), la normalisation et la désambiguïsation. En effet, à l'instar des autres unités lexicales, plusieurs phénomènes de glissement ou de superposition de sens peuvent avoir lieu au regard des entités nommées et il importe de pouvoir les résoudre afin de traiter au mieux ces unités. Nous nous intéressons plus particulièrement à la métonymie des entités nommées et présentons dans cet article une méthode hybride pour la résolution de ce type particulier de polysémie.

La première section reviendra tout d'abord sur la définition et la caractérisation de la métonymie des entités nommées. La suivante présentera rapidement la campagne d'évaluation SemEval (*Semantic Evaluation*), dans le cadre de laquelle les travaux ici présentés ont été réalisés et évalués, puis la troisième section s'attachera à décrire le système mis au point. Enfin, la dernière section rendra compte de l'évaluation.

2 La métonymie des entités nommées

La métonymie correspond au fait d'employer un mot (par exemple, le mot *récolte*) attaché à une certaine entité (l'action de récolter) pour en désigner une autre (les produits cueillis), la seconde étant liée à la première par une relation de type partie-tout ou une relation fonctionnelle (ici relation de cause à effet). Si la métonymie permet un nombre indéfini de glissements de sens, il existe néanmoins des changements de sens réguliers ou systématiques, au regard notamment des entités nommées. Examinons les exemples suivants :

La politique américaine est plombée par l'Irak.

La France a gagné en demi-finale.

Dans la première phrase, il n'est bien sûr pas question du pays proprement dit mais de l'événement qui s'y déroule, tout comme dans la seconde où il s'agit non pas de la France en tant que telle mais d'une équipe sportive française. Cet usage des unités *Irak* et *France* en tant qu'événement et équipe sportive respectivement est possible pour d'autres noms de pays dans des situations similaires. Il existe bien d'autres exemples possibles, sur lesquels nous reviendrons par la suite, mais il est d'ores et déjà possible de postuler une certaine régularité et productivité des phénomènes de métonymie pour les entités nommées.

Outre ces caractéristiques, des études conduites par K. Markert, U. Hahn et M. Nissim (Markert & Hahn, 2002), (Markert & Nissim, 2006) ont fait état de la fréquence de ce phénomène, montrant que 17% de l'ensemble des occurrences dans un corpus de 27 magazines allemands étaient

¹Moyenne harmonique de la précision et du rappel : $(2 \times \text{précision} \times \text{rappel}) / (\text{précision} + \text{rappel})$.

métonymiques, tout comme 20% des occurrences des noms de pays et 30% de celles des noms d'organisation sur des extraits significatifs du *British National Corpus*. Régulière, productive et fréquente, la métonymie des entités nommées constitue ainsi un réel intérêt pour le traitement automatique des langues.

3 Résolution de métonymie pour la campagne SemEval

La campagne SemEval2007 proposait 18 tâches d'évaluations autour de problèmes sémantiques tels que la désambiguïsation de prépositions, l'annotation d'expressions et de relations temporelles (*TempEval*) ou encore la désambiguïsation des noms de personne sur le web (*Web People Search*). La tâche proposée par K. Markert et M. Nissim est une tâche lexicale sur l'anglais, portant plus précisément sur la résolution de métonymie pour deux classes sémantiques, la classe LOCATION avec des noms de lieux et la classe ORGANISATION avec des noms d'entreprises. L'objectif pour les participants est de classer automatiquement des occurrences pré-sélectionnées et en contexte de noms de lieux et d'entreprises, et ce en fonction de leur interprétation littérale ou non-littérale. Cette première alternative correspond à l'annotation "gros grain" ou *coarse-grained annotation*. Deux autres niveaux d'annotation sont possibles : un niveau "moyen" (*medium*) pour lequel il faut faire la distinction entre des interprétations littérales, métonymiques et mixtes et, enfin, un niveau "fin" (*fine*), pour lequel il importe de préciser, en cas d'interprétation métonymique, le patron métonymique dont il est question. À titre d'illustration, nous pouvons reprendre les exemples donnés par les organisatrices dans le document décrivant la tâche (Markert & Nissim, 2007) :

At the time of Vietnam, increased spending led to inflation.

BMW slipped 4p to 31p.

The BMW slowed down.

Pour la première phrase, le nom *Vietnam* ne renvoie pas au pays mais à la guerre qui s'y est déroulée, il convient donc d'annoter cette entité comme *non-literal* (niveau 1), comme *metonymic* (niveau 2) ou comme *place-for-event* (niveau 3). De même, les occurrences de *BMW* dans les exemples suivants ne renvoient pas à l'entreprise mais à l'action de l'entreprise pour la première (*org-for-index*) et au produit de cette entreprise pour la seconde (*org-for-product*). Les trois niveaux d'annotation pour les noms d'entreprise et pour les noms de pays sont représentés ci-après dans le tableau 1².

4 Un système de résolution de métonymie pour les EN

Notre participation à la tâche de résolution de métonymie pour les noms de pays et d'entreprises (Brun *et al.*, 2007) a consisté en l'élaboration d'un système automatique hybride reposant sur la combinaison d'un composant symbolique et d'un composant distributionnel. Nous présenterons successivement ces deux composants.

²Davantage de précisions sur les catégories d'annotation (ou patrons métonymiques) sont disponibles dans (Markert & Nissim, 2007).

Catégorie : ORGANISATION	<i>coarse</i>	<i>medium</i>	<i>fine</i>	
	literal	literal	literal	
	non-literal	metonymic	mixed	mixed
			othermet	
			object-for-name	
			object-for-representation	
			organisation-for-members	
			organisation-for-event	
			organisation-for-product	
			organisation-for-facility	
organisation-for-index				
Catégorie : LOCATION	<i>coarse</i>	<i>medium</i>	<i>fine</i>	
	literal	literal	literal	
	non-literal	metonymic	mixed	mixed
			othermet	
			object-for-name	
			object-for-representation	
			place-for-people	
			place-for-event	
			place-for-product	

TAB. 1 – Niveaux de granularité et catégories d’annotation pour les classes ORGANISATION et LOCATION.

4.1 Composant symbolique

Analyse syntaxique robuste avec XIP L’élément fondamental sur lequel repose notre approche est l’analyseur syntaxique robuste XIP (Aït-Mokhtar & Chanod, 1997), (Aït-Mokhtar *et al.*, 2002). *Xerox Incremental Parser* prend en entrée du texte tout venant et produit en sortie de façon robuste une analyse syntaxique profonde. À partir d’un ensemble de règles, l’analyseur est capable de faire de la désambiguïisation de catégories, de construire des syntagmes noyaux et d’extraire des relations de dépendances syntaxiques entre unités lexicales simples et/ou complexes. En plus de l’analyse des relations syntaxiques de surface, XIP effectue également une analyse syntaxique dite “profonde” ou “normalisée” (prise en compte des sujets et objets de verbes non finis, normalisation de la forme passive en forme active, etc (Hagège & Roux, 2003)), l’exploitation d’information de sémantique lexicale, avec les classes verbales de Levin (Levin, 1993) et quelques éléments de Framenet (Ruppenhofer *et al.*, 2006). Ci-dessous un exemple de sortie XIP (avec l’arbre des chunks et les relations de dépendance) pour la phrase suivante :

Iran wanted the two centers to generate part of its electricity needs. TOP { SC{ NP{Iran} FV{wanted}} NP{the two centers} IV{to generate} NP{part} PP{of NP{its electricity needs}} . }

MOD_PRE (needs, electricity) SUBJ-N (generate, centers)
 LOCATION (Iran) EXPERIENCER-PRE(wanted, Iran)
 CONTENT(wanted, centers) OBJ-N(generate, needs)

L'analyseur décrit ci-dessus contient déjà un module de reconnaissance d'entités nommées (Brun & Hagège, 2004), mais ce dernier ne permet pas le traitement de la métonymie ; il a fallu procéder à son adaptation.

Adaptation à la tâche de résolution de métonymie Pour ce faire, nous avons étudié les corpus d'entraînement à l'aide de XIP, en tenant compte des directives d'annotation. Pour les noms de pays et d'entreprises, notre attention s'est focalisée sur les types de relations grammaticales dans lesquelles étaient impliquées les unités à étudier et sur les informations lexicales ou sémantiques attachées aux arguments de ces relations. Pour chaque patron métonymique, nous avons donc analysé l'ensemble de ses occurrences (attachées à une unité lexicale) dans le corpus d'entraînement pour dégager des configurations grammaticales et lexicales jouant le rôle d'“ amorces ” d'interprétations métonymiques.

Au terme de cette étude de corpus, il fut possible de prendre en compte l'ensemble des indices récoltés et d'élaborer un module de résolution de métonymie dans XIP. Intervenant à la fin de la chaîne de traitements de l'analyseur, cette adaptation consiste en l'écriture de règles traduisant les configurations discriminantes relevées pour tel ou tel patron d'une part, et à l'ajout de lexique d'autre part. À titre d'exemple, prenons l'hypothèse suivante : “ *Si un nom de pays est sujet d'un verbe renvoyant à une action économique, alors le patron loc-for-people doit s'appliquer* ” ; cette hypothèse se retrouve dans XIP sous la forme de règle :

```
if (LOCATION(#1) & SUBJ-N(#2[v_econ],#1))  
LOC-FOR-PEOPLE (#1)
```

Règle qui se lit ainsi : si le parseur a détecté un nom de pays (#1) qui est le sujet d'un verbe (#2) portant le trait “ v_econ ”, alors il crée une relation unaire LOC-FOR-PEOPLE pour le nom de pays.

En sus des indices récoltés lors de l'étude de corpus, nous avons exploité des informations lexicales déjà codées dans XIP, comme par exemple les traits attachés aux verbes de communication (*say, deny, comment*) et les catégories relevant du cadre “experiencer” de Framenet, à savoir des verbes tels que *feel, sense, see*, les noms *despair, compassion, adoration* ou les adjectifs *sympathetic, sad*, etc. En effet, eu égard au fait que ce cadre “experiencer” renvoie à des personnes ou à des groupes de personnes, dès lors qu'un nom de pays ou d'entreprise a ce rôle, il peut être annoté en tant que *loc-for-people* ou *organisation-for-members*. Ci-dessous un exemple de sortie de l'analyseur avec le nouveau développement :

It was the largest Fiat everyone had ever seen.

```
ORG_FOR_PRODUCT (Fiat)           MOD_PRE (seen, ever)  
SUBJ-N_PRE (was, It)             ATTRIB(It, Fiat)  
EXPERIENCER_PRE(seen, everyone)  QUALIF(Fiat, largest)
```

Dans cet exemple, la présence de la relation QUALIF entre le nom d'entreprise *Fiat* et l'adjectif *largest*, renforcée par la présence d'un article défini, conduit à l'annotation en tant que ORG_FOR_PRODUCT. Ce composant symbolique constitue le premier volet de notre méthode de résolution de métonymie, il est complété par un composant distributionnel, qu'il convient à présent de détailler.

4.2 Composant distributionnel

Intervient en “deuxième passe” de notre système, un composant distributionnel. L’idée principale de cette combinaison est d’exploiter deux méthodes relevant d’approches différentes mais complémentaires, autrement dit de pallier les manquements de l’analyse symbolique, focalisée sur des données précises nécessitant une étude minutieuse, par une analyse distributionnelle, apte à récolter des informations à grande échelle sur d’importantes données textuelles. Le principe est donc, lorsqu’une unité n’a pu être traitée par le composant symbolique, d’essayer de trouver son annotation en exploitant les informations présentes à propos de cette unité ou d’une unité de même type dans un grand corpus. Nous présentons rapidement l’analyse distributionnelle avant de détailler sa mise en œuvre pour la résolution de métonymie.

L’analyse distributionnelle La notion d’analyse distributionnelle a été introduite par Z. S. Harris. En linguistique de corpus, cette méthode est aujourd’hui largement exploitée, notamment dans les travaux de terminologie, de structuration de terminologie et de construction d’ontologies (Faure & Nedellec, 1999) (Assadi, 1998) (Bourigault, 2002). L’hypothèse est la suivante : il serait possible, à partir de régularités syntaxiques observées pour un ensemble de mots, de déduire des propriétés sémantiques pour ces mots. Concrètement, il s’agit d’étudier la distribution des mots, c’est à dire les contextes lexico-syntaxiques dans lesquels ils apparaissent, pour ensuite tenter de dégager des parentés sémantiques. S’inscrivant dans ce cadre, G. Jacquet et F. Venant ont élaboré une méthode automatique de désambiguïsation du sens d’un mot en contexte, reposant sur la prise en compte de l’influence des éléments syntaxiques et lexicaux présents dans l’énoncé (Jacquet & Venant, 2003). Ainsi, il ne cherchent plus seulement à créer des classes de mots relevant du même champ sémantique “*mais des classes de mots dont le comportement sémantique influence de la même façon un contexte donné*”. Le composant distributionnel élaboré pour la résolution de métonymie s’inscrit dans la perspective de ces travaux.

Méthode distributionnelle pour la résolution de métonymie L’objectif ici est de rapprocher des contextes et d’exploiter les résultats du composant symbolique. Cette méthode comporte deux processus. Il s’agit d’une part de construire un espace distributionnel pour être en mesure de rapprocher des contextes en fonction d’une entité donnée et, d’autre part, de capitaliser l’information du composant symbolique sous la forme d’une sorte de “base de donnée” de contextes avec annotation. Nous détaillons ces deux processus successivement.

Construction de l’espace distributionnel et rapprochement des contextes Ce premier processus comporte 5 étapes. Le point de départ est le corpus BNC dans son entier (100 millions de mots), duquel on a extrait les corpus d’entraînement et de test de la tâche de résolution de métonymie.

La première étape correspond à l’analyse syntaxique de ce corpus (à l’aide de l’analyseur XIP) afin d’en extraire des dépendances syntaxiques. C’est à partir de ces dépendances que sont construits les contextes et les unités lexicales. Prenons un exemple, avec la proposition *provide Albania with food aid*. Les dépendances extraites par XIP sont les suivantes :

```
IND-OBJ-N3(VERB :provide,NOUN :Albania)
PREP_WITH(VERB :provide,NOUN :aid)
```

³la relation *ind-obj-n* correspond à la relation syntaxique *objet indirect*.

Résolution de Métonymie des Entités Nommées : proposition d'une méthode hybride

PREP_WITH(VERB :provide,NOUN :food aid)

Où l'on peut voir que les arguments des dépendances sont de simples unités lexicales (*aid*) ou des syntagmes (*food aid*).

La deuxième étape de ce processus est la construction d'un espace distributionnel à partir de ces dépendances. Cet espace distributionnel est l'inverse de celui habituellement construit en analyse distributionnelle : puisque l'objectif est de rapprocher des contextes, chaque point de l'espace est une contexte syntaxique et chaque dimension est une unité lexicale. Cet inversion constitue une première différence avec l'approche de G. Jacquet et F. Venant (Jacquet & Venant, 2003). Chaque dépendance obtenue lors de l'étape précédente permet de construire plusieurs contextes simples et/ou composés. Un contexte syntaxique comporte une relation et une unité lexicale. La méthode prévoit également de construire des contextes composés (autre différence par rapport à (Jacquet & Venant, 2003)), soit des contextes combinant plusieurs contextes, dont le premier comporte nécessairement un syntagme verbal et le second a pour recteur ce même syntagme. Pour la phrase analysée ci-dessus, les unités lexicales, les contextes simples (1. pour recteur et 2. pour régi) et les contextes composés sont les suivants :

Unités lexicales	Contextes simples	Contextes composés
VERB :provide	1.VERB :provide.IND-OBJ-N	1.VERB :provide.IND-OBJ-N+2.NOUN :aid.PREP_WITH
NOUN :Albania	1.VERB :provide.PREP_WITH	1.VERB :provide.IND-OBJ-N+2.NOUN :aid.PREP_WITH
NOUN :aid	2.NOUN :Albania.IND-OBJ-N	1.VERB :provide.IND-OBJ-N+2.NOUN :food aid.PREP_WITH
NOUN :food aid	2.NOUN :aid.PREP_WITH	1.VERB :provide.PREP_WITH+2.NOUN :Albania.IND-OBJ-N
	2.NOUN :food aid.PREP_WITH	

Une heuristique permet de filtrer ces données en fonction de leur productivité : chaque unité lexicale doit être présente au moins 100 fois dans le corpus, tout comme les contextes (y compris les contextes composés). Avec le corpus BNC de 100 millions de mots, on obtient au final 60 849 unités lexicales et 140 634 contextes. Il est donc possible de construire un espace distributionnel comportant 140 634 points (les contextes) et 60 849 dimensions (les unités lexicales). Cet espace est le matériau de base à partir duquel les autres traitements viennent s'effectuer.

À partir de la troisième étape intervient la prise en compte d'une unité lexicale précise, pour laquelle le composant symbolique n'a pu trouver d'annotation. En fonction de cette unité et de son contexte d'apparition (soit telle qu'elle apparaît dans un extrait de SemEval), il s'agit tout d'abord de construire un sous-espace, de la manière suivante :

Pour un couple donné formé d'un contexte i et d'une unité lexicale j :

- Le sous-espace des contextes équivaut à la liste des contextes occurring avec l'unité lexicale j . S'il y a plus de k contextes, alors on ne garde que ces k contextes les plus fréquents ;
- Le sous-espace des dimensions équivaut à la liste des unités lexicales occurring avec au moins un des contextes du sous-espace des contextes. S'il y a plus de n unités, alors on ne garde que ces n unités les plus fréquentes.

La quatrième étape consiste à réduire les dimensions de ce sous-espace, à l'aide d'une analyse factorielle des correspondances (AFC ; équivaut à une ACP avec la métrique du Chi2).

Enfin, la cinquième étape consiste à rapprocher les contextes restants, c'est à dire ayant passés le filtre des deux étapes précédentes. Ce rapprochement est calculé à l'aide de la distance euclidienne. On obtient ainsi, au final, une liste des contextes proches de celui de l'unité considérée.

Pour l'unité *Albania* dans le contexte *provide*, les contextes les plus proches sont présentés dans la première colonne du tableau 2.

Pour pouvoir utiliser cette liste de contextes, encore faut-il savoir s'ils ont reçu ou non une annotation par le composant symbolique. Intervient alors le second processus.

Capitalisation de l'information du composant symbolique Ce second processus a pour objectif de construire une sorte de base de données de contextes avec annotations. Deux étapes sont nécessaires. Le corpus est analysé BNC avec XIP, augmenté cette fois-ci du module de résolution de métonymie. Cette analyse permet d'obtenir des dépendances syntaxiques mettant en jeu des noms de pays et des noms d'entreprise avec leurs annotation métonymique (application du module de résolution de métonymie sur les entités nommées reconnues par l'analyseur, indépendamment des données de SemEval). Le résultat de cette première étape est donc une série de contextes impliquant des unités lexicales avec leur annotation. La deuxième étape correspond à la sélection de contextes discriminants au regard des annotations. Par exemple, si le contexte VERB :allow.IND-OBJ-N se retrouve pour 10% des cas avec une unité comportant l'annotation *literal* et pour 90% des autres avec une unité comportant l'annotation *loc-for-people*, alors le contexte est considéré comme discriminant vis-à-vis de cette dernière annotation. En revanche, si un contexte se trouve dans 50% des cas avec telle annotation et 50% des cas avec telle autre, alors il n'est pas conservé. À l'issue de cette sélection de contextes discriminants, on dispose alors d'un stock de contextes avec leurs annotations, pouvant être exploité en parallèle avec le processus précédent.

Annotation d'une unité Le premier processus permet de déterminer une liste de contextes plus ou moins proches du contexte dans lequel apparaît une unité lexicale non annotée. Le second permet de collecter un certain nombre de contextes avec leur annotation. Ces deux types de données peuvent dès lors être croisées pour permettre l'annotation d'une unité.

Ci-dessous (tableau de gauche) la liste de contextes proches du contexte VERB :provide.IND-OBJ-N pour la proposition *provide Albania with food aid*, avec l'indication de leur distance et de leur annotation correspondante dans la base de données de contextes :

Contexte	Distance	Annotation
VERB :provide.IND-OBJ-N	0.00	
VERB :allow.OBJ-N	0.76	LOC-FOR-PEOPLE
VERB :include.OBJ-N	0.96	
ADJ :new.MOD_PRE	1.02	
VERB :be.SUBJ-N	1.43	
VERB :supply.SUBJ-N_PRE	1.47	LITERAL
VERB :become.SUBJ-N_PRE	1.64	
VERB :come.SUBJ-N_PRE	1.69	
VERB :support.SUBJ-N_PRE	1.70	LOC-FOR-PEOPLE
...

Annotation	Score
LOC-FOR-PEOPLE	3.11
LITERAL	1.23
LOC-FOR-EVENT	0.00
...	...

TAB. 2 – Listes des contextes les plus proches de VERB :provide.IND-OBJ-N et Scores des annotations.

Pour pouvoir déterminer une annotation pour le contexte *provide.IND-OBJ-N*, il faut regarder les annotations attribuées à ses contextes les plus proches et examiner si elles sont pertinentes ou non. Pour ce faire, il importe de calculer un score pour chaque annotation, valorisant sa

présence dans les contextes de “tête de liste” : le score d’une annotation donnée est égal à l’inverse de la somme des distances des contextes portant cette annotation. Pour notre cas, les scores sont présentés dans le tableau ci-dessus à gauche ; il est alors possible, en fonction de ces scores, d’attribuer l’annotation LOC-FOR-PEOPLE à l’unité *Albania* dans la proposition *provide Albania with food aid*.

5 Evaluation et analyse des résultats

Ci-dessous les taux de précisions pour l’ensemble des participants pour les noms de lieux (tableau 3) et les noms d’entreprises (tableau 4).

Tâche	Systèmes					
	baseline	up13	FUH	UTD-HLT-CG	XRCE-M	GYDER
LOCATION-coarse	0.794	0.754	0.778	0.841	0.851	0.852
LOCATION-medium	0.794	0.750	0.772	0.840	0.848	0.848
LOCATION-fine	0.794	0.741	0.759	0.822	0.841	0.844

TAB. 3 – Taux de Précision pour tous les systèmes pour la classe LOCATION.

Tâche	Systèmes			
	baseline	XRCE-M	UTD-HLT-CG	GYDER
ORGANISATION-coarse	0.618	0.732	0.739	0.767
ORGANISATION-medium	0.618	0.711	0.711	0.733
ORGANISATION-fine	0.618	0.700	0.711	0.728

TAB. 4 – Taux de Précision pour tous les systèmes pour la classe ORGANISATION.

Malgré certaines imperfections, XRCE-M est relativement bien placé au sein des autres systèmes ayant participé à l’évaluation (cf tableaux 3 et 4), avec notamment une seule erreur de plus pour la classe LOCATION par rapport au système le plus performant GYDER. Cinq systèmes ont participé à la tâche sur les LOCATION et trois à celle sur les ORGANISATION, pour tous les niveaux de granularité à chaque fois. Hormis XRCE-M, les systèmes reposaient sur des méthodes statistiques par apprentissage, mais surtout, notre système est le seul à ne pas avoir utilisé les annotations syntaxiques manuelles fournies par les organisatrices. Deux systèmes ne sont pas parvenus à atteindre la baseline pour les noms de lieux.

6 Conclusion

Dans cet article, nous avons décrit un système hybride pour la résolution de métonymies. Notre approche, quoique perfectible, a d’ores et déjà montré des résultats encourageants lors de la compétition SemEval. Les perspectives de développement sont nombreuses, pour notre système tout comme pour la tâche de résolution de métonymie en général. À notre échelle, nous envisageons de poursuivre le travail entrepris afin d’améliorer les performances du système pour les glissements de sens déjà pris en compte et d’étendre ce traitement à d’autres types d’entités nommées.

Références

- (1995). *MUC-6. Proceedings of the Sixth Message Understanding Conference*. Informations disponibles à l'adresse suivante : <http://www.cs.nyu.edu/cs/faculty/grishman/muc6.html>.
- (1998). *MUC-7. Proceedings of the Seventh Message Understanding Conference*. Informations disponibles à l'adresse suivante : http://www-nlpir.nist.gov/related_projects/muc/.
- ASSADI H. (1998). *Construction d'ontologies à partir de textes techniques. Application aux systèmes documentaires*. PhD thesis, Université Paris VI.
- AÏT-MOKHTAR S. & CHANOD J. P. (1997). Incremental finite-state parsing. In *Proceedings of Applied Natural Language Processing*, Washington, DC.
- AÏT-MOKHTAR S., CHANOD J. P. & ROUX C. (2002). Robustness beyond shallowness : incremental dependency parsing. *NLE Journal*.
- BOURIGAULT D. (2002). Upery : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus. In *Actes de TALN 2002*, Nancy.
- BRUN C., EHRMANN M. & JACQUET G. (2007). XRCE-M : A hybrid system for named entity metonymy resolution. In *4th International Workshop on Semantic Evaluations*, Prague : ACL-SemEval 2007.
- BRUN C. & HAGÈGE C. (2004). Intertwining deep syntactic processing and named entity detection. In *ESTAL*, Alicante, Espagne.
- EHRMANN M. & JACQUET G. (2006). Vers une double annotation des entités nommées. *Traitement Automatique des Langues*, **47**(3). Disponible sur : <http://www.atala.org>.
- FAURE D. & NEDELLEC C. (1999). Knowledge acquisition of predicate argument structures from technical texts using machine learning : the system asium. In *Proceedings of the 11th European Workshop, Knowledge Acquisition, Modelling and Management (EKAW'99)*, Juan-les-Pins, France.
- HAGÈGE C. & ROUX C. (2003). Entre syntaxe et sémantique : Normalisation de la sortie de l'analyse syntaxique en vue de l'amélioration de l'extraction d'information à partir de textes. In *TALN 2003*, Batz-sur-Mer.
- JACQUET G. & VENANT F. (2003). Construction automatique de classes de sélection distributionnelle. In *Actes de la 10ème conférence annuelle sur le Traitement Automatique des Langues (TALN 2003)*, Dourdan, France.
- LEVIN B. (1993). *English Verb Classes and Alternations - A Preliminary Investigation*. The University of Chicago Press,.
- MARKERT K. & HAHN U. (2002). Understanding metonymies in discourse. *Artificial Intelligence*, **135**(1/2), 145–198.
- MARKERT K. & NISSIM M. (2006). Metonymic proper names : a corpus based account. In A. STEFANOWITSCH & S. GRIES, Eds., *Corpus-based approaches to Metaphor and Metonymy*, p. 152–174. Mouton de Gruyter.
- MARKERT K. & NISSIM M. (2007). *Metonymy Resolution at SemEval I : Guidelines for Participants*. Rapport interne, SemEval.
- RUPPENHOFER J., ELLSWORTH M., PETRUCK M., JOHNSON C. & SCHEFFCZYK J. (2006). *Framenet II : Extended Theory and Practice*. Rapport interne, University of Berkeley. Disponible sur : <http://framenet.icsi.berkeley.edu/>.