Corela

HS-2 (2005)
Le traitement lexicographique des noms propres

Denis Maurel et Mickaël Tran

Une ontologie multilingue des noms propres

Référence électronique
Denis Maurel et Mickaël Tran, « Une ontologie multilingue des noms propres », *Corela* [En ligne], HS-2 | 2005, mis en ligne le 02 décembre 2005, consulté le 16 décembre 2015. URL : http://corela.revues.org/1203

Éditeur : Université de Poitiers
http://corela.revues.org
http://www.revues.org

Document accessible en ligne sur :
http://corela.revues.org/1203
Document généré automatiquement le 16 décembre 2015.

Denis Maurel et Mickaël Tran

Une ontologie multilingue des noms propres

Introduction

1 Les ressources dictionnairiques sont en général indispensables au traitement automatique des langues, même si certains composants logiciels n'utilisent que des corpus d'apprentissage et des traitements statistiques. Ces dictionnaires contiennent des noms communs (comme le système DELA de dictionnaires électroniques (Courtois, Silberztein, 1990)) ou des termes spécifiques à un domaine, même si, dans ce cas, le dictionnaire sert souvent d'amorce à un système de découverte de nouveau termes. Mais qu'en est-il des noms propres ? Bien qu'ils constituent, à eux seuls, plus de 10 % des textes journalistiques (Coates-Stephens, 1993), ceux-ci sont souvent absents de ces dictionnaires, même dans les applications multilingues où prévaut parfois l'idée fausse que les noms propres ne se traduisent pas.

2 Faut-il associer les noms propres et les mots inconnus capitalisés (du moins pour le français), comme (Ren, Perrault, 1992) ? Cette association n'est vraie qu'une fois sur deux, à cause de la présence d'homographes et de mots composés (Maurel, 2004). Or la recherche d'information, l'extraction d'information ou l'aide à la traduction nécessitent de délimiter précisément les noms propres, de les catégoriser et même, parfois, de les relier entre eux. Faut-il alors n'utiliser pratiquement que des règles avec une liste minimum de noms propres, comme le défend (Mikheev et al., 1999) ? Bien sûr, une liste exhaustive de noms propres est impossible, mais un juste équilibre entre listes et règles est certainement souhaitable. Nous avons choisi de développer les deux approches, dans le cadre d'une plate-forme technologique consacrée au traitement automatique des noms propres. Cette présentation est consacrée à notre approche par liste ou, plus précisément, à la définition de notre dictionnaire relationnel multilingue de noms propres. Notre système de règle est présenté dans (Friburger, Maurel, 2004).
Précisons tout d'abord que nos entrées correspondent à la définition de (Jonasson, 1994 :21) pour qui :

> Toute expression associée dans la mémoire à long terme à un particulier en vertu d'un lien dénominatif conventionnel stable est un nom propre.

3 Cette définition inclut les noms propres descriptifs qui résultent souvent de la composition d'un nom propre avec une expansion, comme *Tour Eiffel* ou *Musée Rodin*, ou semblent être des descriptions définies figées ou en cours de figement, comme *Pont Neuf* ou *Médecins sans frontière*. Elle est proche de celle des entités nommées (Chinchor, 1997), largement utilisée dans le monde du *traitement automatique des langues* depuis les conférences *MUC* (aux dates et unités chiffrées près).

4 Pour permettre une création et une gestion cohérente de ce dictionnaire, il est nécessaire d'identifier les concepts et les relations du domaine des noms propres, tout en distinguant ce qui dépend de la langue de ce qui en est indépendant. Ceci nous a conduit à adopter une démarche ontologique (Gruber, 1995).

5 La section décrit notre ontologie et la section , les relations entre noms propres. Dans la section , nous envisageons des liens vers d'autres structures lexicales.

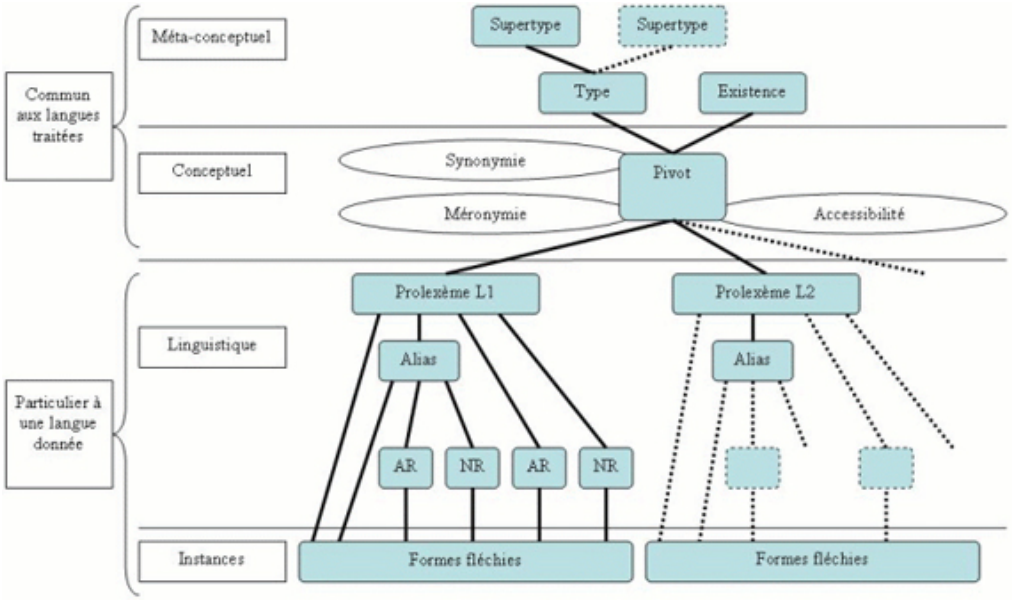
Une ontologie en deux parties

6 Notre ontologie est divisée en deux parties (Figure1), chacune de ces parties étant elle-même divisée en deux niveaux[1] :

- Une partie supérieure commune aux langues traitées : les niveaux conceptuel (le référent suivant différents points de vue) et méta-conceptuel (la typologie et l'essence), hiérarchisés par la relation d'hyperonymie. Elle contient trois relations entre noms propres (Prédication, Méronymie, Synonymie) .

- Une partie inférieure particulière à une langue donnée : le niveau linguistique (morphologie dérivationnelle) et celui des instances (morphologie flexionnelle). L'arborescence de cette partie, plus ou moins complexe selon la langue, a pour tête la forme canonique d'un nom propre, que nous appelons *prolexème*.

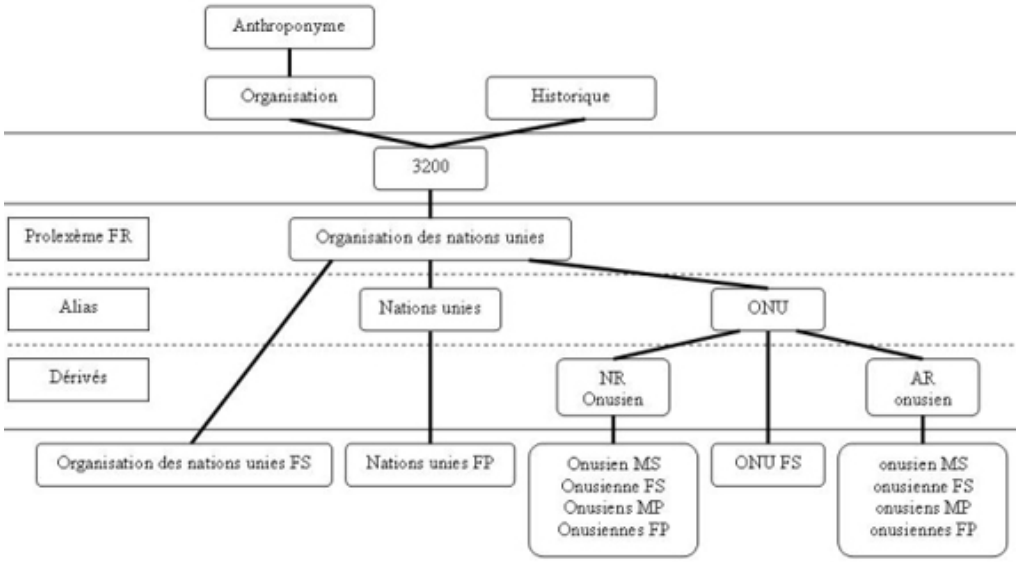
Figure 1 : L'architecture en deux parties de Prolexbase



La partie propre à une langue donnée

7 La partie propre à une langue donnée est d'abord constituée du niveau des instances qui correspond à l'ensemble des formes fléchies, c'est-à-dire l'ensemble des mots réellement rencontrés dans les textes, que (Polguère, 2003) désigne par le terme de *lexie*. Par exemple, la lexie *onusien* (Figure 3) regroupe les quatre mots formes *onusien*, *onusienne*, *onusiens* et *onusiennes*. En français, le nom propre ne possède, en général, qu'une forme fléchie, mais ce n'est pas le cas des langues casuelles (déjà l'allemand, mais surtout les langues slaves[2]).

Figure 2 : Exemple en français avec le prolexème Organisation des nations unies



8 Les instances sont donc déduites de leur lemme qui est placé au niveau linguistique. Les différents lemmes correspondant à un même nom propre sont regroupés autour du prolexème qui est non seulement le lemme correspondant aux instances d'un nom propre, mais aussi la forme canonique d'un certain nombre d'alias et de formes dérivées.
Les alias représentent différentes écritures possibles à partir du prolexème :

- Des variantes sur les caractères : hauteur de casse (*Suisse* ou *SUISSE*), diacritiques (*Étretat* ou *Etretat*), ligatures (*la Basilique du Sacré-Cœur* ou *la Basilique du Sacré-Coeur*), esperluette (*Sciences et Vie* ou *Sciences & Vie*)…
- Des abréviations (*Nations Unies*, voir Figure 2), des acronymes (*Onu*), des sigles (*SNCF*) …
- Des transcriptions différentes[3] (le nom russe Чехов aura pour prolexème français *Tchekhov* et pour alias *Tchékhov*).

9 Les formes dérivées proviennent quant à elles du prolexème ou de ses alias (par exemple *onusien* sur la Figure 2). En français, il en existe deux sortes, les noms relationnels (éventuellement diastratiques) et les adjectifs relationnels[4], mais d'autres langues en possède plus (en serbe, par exemple, un adjectif possessif dérive de chaque nom propre et, même, de chaque alias).

Le prolexème est accompagné de quelques indications supplémentaires (en fonction de la langue) :

- La détermination, à supprimer ou à rétablir, suivant la langue.
- La phonétique, qui, lorsqu'un nom propre n'a pas de traduction dans une langue cible utilisant un système d'écriture différent de la langue source, permet souvent de proposer une transcription.
- L'ordre de tri des noms propres polylexicaux qui se classent en permutant une partie des mots qui le composent (*Mer d'Aral* est classé à la lettre *A* et non la lettre *M*).
- Un indicateur de notoriété, qui varie en fonction de la langue et de critères extralinguistiques (le pays, la culture, la période considérée…), et un indicateur de fréquence d'apparition dans un certains corpus… Les noms propres indispensables pour l'étude d'un corpus journalistique d'une année donnée peuvent se révéler inutiles quelques années plus tard.
- Les règles d'aliasisation ou de dérivation qui permettent la création d'alias et de dérivés réguliers.
- Les expansions notoires, associées à des grammaires locales (Gross, 1989), qui sont une sorte d'hyperonymie vers le lexique général. La traduction des expansions n'est pas triviale (par exemple, en allemand, *Rechtsanwalt Paul Bischof* devient en français *Maître Paul Bischof* et non *Avocat Paul Bischof*). Lorsqu'elles sont omises, ces expansions sont parfois à rétablir dans la traduction (*la Seine* et *la rivière Kwaï* deviennent en anglais *the river Seine* et *the river Kwai*).

10 Ces deux dernières indications se rapprochent des *structures interne et externe* de (MacDonald, 1996). Ajoutons pour terminer trois informations éponymiques éventuelles :

- L'antonomase du nom propre est sa lexicalisation comme synonyme d'un nom commun. Une antonomase peut exister dans certaines langues et ne pas être présente dans d'autres. Le nom propre *Pampers* existe en français, mais celui-ci n'a pas donné lieu à une antonomase, contrairement au polonais (*pampersy* pour une *couche jetable*).
- Les tournures idiomatiques à partir de noms propres. Ces formes devront être reconnues pour elle-même. Leur traduction n'implique pas nécessairement la présence d'un nom propre. Notons de plus que ce figement n'a pas toujours le même sens d'une langue à l'autre : *zwischen Scylla und Charybdis sein* signifie *être entre deux dangers* alors que *tomber de Charybde en Scylla* signifie *quitter un mal pour un autre pire encore*.
- Les formes terminologiques, comme *maladie de Parkinson* ou *théorème de Pythagore*, pour lesquelles le nom propre n'est plus une clé d'indexation. Il se peut aussi que la traduction soit différente dans une langue cible ; ainsi, en allemand, on utilisera plutôt un dérivé que le prolexème (*der pythagoreische Lehrsatz*, *die parkinsonsche Krankheit*). De plus, des variations sont possibles sur l'ordre d'apparition des noms propres cités : *maladie de Legg-Perthes-Calvé* sera traduit en anglais par *Legg-Perthes-Calvé disease*, mais, en allemand, par *Perthes-Legg-Calvé-Krankheit* (Bodenreider, Zweigenbaum, 2000a :733). Ou même sur la liste des noms cités : *maladie de Weber-Christian* reste

en anglais *Weber-Christian disease*, mais devient en allemand *Pfeifer-Weber-Christian-Krankheit* (Bodenreider, Zweigenbaum, 2000b :445).

La partie commune aux langues traitées

11 Pour relier les différentes langues, nous avons adopté, comme le recommande l'Afnor[5], une approche par pivot. Déjà présente dans le projet *Eurotra* pour la traduction automatique (Danlos, 1989), elle est utilisée dans de nombreux projets : *EuroWordnet* (Vossen, 1998) et *Balkanet* (Tufiş et al., 2004) (Krstev et al., 2004), *Crossmarc* (Hachey et al., 2003), *Papillon* (Mangeot-Lerebours et al., 2003), etc.

12 Pour nous, ce pivot représente un *nom propre conceptuel*, qui ne correspond pas au référent linguistique, mais à un certain point de vue sur ce référent. Nous aurons donc deux pivots distincts pour *Saint-Pétersbourg* et *Leningrad* (point de vue diachronique), pour *Marseille* et *Cité phocéenne* (point de vue diaphasique), etc. Chaque prolexème (pour une langue donnée) correspond à un unique nom propre conceptuel. Deux prolexèmes (de deux langues différentes) reliés au même pivot sont donc en relation de traduction.

13 Aux noms propres conceptuels, sont associés, par des relations d'hyperonymie, les métaconcepts de type et d'essence.

14 La classification des noms propres a donné lieu a de nombreuses études, en particulier (Bauer, 1985), (Chinchor, 1997) et (Paik et al., 1996). Nous souhaitions adopter un nombre restreint de types, afin que ceux-ci restent seulement une indication sur le nom propre conceptuel et non une note encyclopédique. La Figure 3 énumère notre liste de vingt-huit types, eux-mêmes en relation d'hyperonymie avec quatre supertypes : les anthroponymes[6] (trait *humain*), les toponymes (trait *locatif*), les ergonymes (trait *inanimé*) et les pragmonymes (trait *événement*). On peut considérer ces quatre supertypes comme hyponymes d'un supertype commun, les noms propres.

15 La plupart des noms propres appartiennent au domaine historique, mais d'autres cependant relèvent du domaine de la croyance religieuse ou de celui de la fiction. Nous avons considéré cette distinction comme une relation d'hyperonymie sur les noms propres conceptuels.

Figure 3 : Hyperonymie entre les types et les supertypes

Types	Supertypes hyperonymes
Célébrité Dynastie Ethnonyme Patronyme Prénom Pseudo anthroponyme	Anthroponyme
Association Ensemble[7] Entreprise Institution Organisation Ville	Anthroponyme Ergonyme Toponyme
Pays Région Supra[8]	Anthroponyme Toponyme
Œuvre Produit	Ergonyme
Fête Histoire Manifestation	Ergonyme Pragmonyme
Edifice Vaisseau Voie	Ergonyme Toponyme
Catastrophe Météorologie	Pragmonyme
Astronyme Géonyme	Toponyme

Hydronyme	

Les relations

16 Les relations entre noms propres sont fréquemment à l'origine d'anaphores et sont de ce fait indispensables au TAL. Nos définitions se sont inspirés du *Dictionnaire explicatif et combinatoire du français contemporain* (Mel'cuk, 1984, 1988, 1992) et du système *Wordnet* (Miller et al., 1990). Nous avons implanté ces relations autour des noms propres conceptuels, de manière indépendante des langues[9].

17 Nous allons présenter dans cette partie les quatre relations sémantiques que nous avons introduites dans notre ontologie.

La synonymie

18 (Polguère, 2003) propose deux définitions de la synonymie, les synonymes exacts (rarissimes) et les synonymes approximatifs[10], considérées comme ayant une valeur sémantique suffisamment proche pour utiliser l'un à la place de l'autre en exprimant sensiblement la même chose.

19 Les alias pourraient être considérés comme des synonymes exacts de leur prolexème, comme *Organisation des nations unies*, *Nations unies* et *Onu* (voir Figure 2) [Gross, 1997]). Mais l'existence même de ces alias est particulière à une langue donnée, c'est pourquoi la formation des alias a été placée dans l'arborescence du prolexème.

20 Pour nous, la relation de synonymie doit donc associer des noms propres conceptuels, indépendamment de la langue. Elle s'applique à des noms propres ayant approximativement le même référent linguistique, mais selon différents points de vue, comme nous l'avons déjà présenté précédemment, section 1.2. Pour distinguer ces différents points de vue, nous avons adopté un marquage diasystématique emprunté à la métalexicographie [Blanco, 2001 :152]. Nous retenons les quatre catégories de [Coseriu, 1998 :9] :

> On constate, dans chaque état de langue (c'est-à-dire, même en faisant abstraction du développement de cette langue dans le temps) trois types fondamentaux de variétés (à savoir : la variété dans l'espace, la variété relative à la stratification socio-culturelle de la communauté parlante et la variété concernant les occasions, circonstances et finalités de l'emploi de la langue dans le discours.

Donnons quelques exemples :

- Diachronique (variété dans le temps) : *Zaïre* et *République démocratique du Congo*. Par contre, nous n'avons pas considéré *Molière* et *Jean-Baptiste Poquelin* comme des synonymes, mais comme des alias d'un seul prolexème, *Jean-Baptiste Poquelin dit Molière* (comme dans les dictionnaires édités).
- Diatopique (variété dans l'espace) : *Nantes* et *Naoned*. Cette variété se retrouve aussi dans les alias (Pour un Tourangeau, *Saint-Cyr* désignera *Saint-Cyr-sur-Loire* et non une des cinq villes française qui portent effectivement ce nom…).
- Diatrastique (variété relative à la stratification socio-culturelle) : Nous n'avons pour l'instant pas trouvé d'exemple de prolexème correspondant à cette variété. Les dérivés diastratiques dépendent de la langue et seront intégrés au prolexème (par exemple *Parisien* et *Parigot*).
- Diaphasique (variété concernant les finalités de l'emploi) : *France* et *République française*.

La méronymie

21 La relation de méronymie est une relation partie-tout qui concerne tous les types de noms propres et peut donner lieu à des reprises anaphoriques :

- Les toponymes (Tours ⊂ Indre-et-Loire ⊂ Région Centre ⊂ France).
- Les entreprises, filiales, marques et produits (LU ⊂ Danone, Mégane ⊂ Renault).
- Les pays et les organisations internationales (France ⊂ Onu).
- Les célébrités et les anthroponymes collectifs (John Lenon ⊂ les Beatles, Zinédine Zidane ⊂ le Real Madrid).

- Les évènements historiques (la Prise de la Bastille ⊂ la Révolution française).
- etc.

22 On parlera par exemple (surtout dans un contexte sportif) des *Bretons* pour désigner les *Nantais*… Ce qui montre encore une fois l'intérêt de placer les relations au niveau du prolexème.

La prédication

23 Une autre source importante d'anaphores est la relation qu'on peut établir entre deux noms propres sur la base d'un prédicat notoire, comme :

- Les toponymes et leurs capitales (*Tours* est le *chef-lieu* de l'*Indre-et-Loire*, *Paris* est la *capitale* de la *France*).
- Les politiques et les institutions (*Jacques Chirac* est le *président* de la *République française*).
- Les dirigeants et les entreprises (*Ray Norda* est le *patron* de *Novell*).
- Les œuvres et les auteurs (*Mozart* est le *compositeur* de *La flûte enchantée*).
- Certains bâtiments officiels et les dirigeants (*Jacques Chirac* est le *locataire* de *l'Elysée*).
- Les entreprises, associations ou organisations et le toponyme correspondant au siège social (*Peugeot* est la *firme sochalienne*).
- Les personnes et les membres de leur famille (*Aaron* est le frère de *Moïse*).
- etc.

24 Cette relation s'inspire au départ de la fonction lexicale appelée Cap, que l'on trouve dans le *Dictionnaire Explicatif et Combinatoire du français contemporain*. Le contexte d'apparition de cette relation est décrit par des grammaires locales.

Liens vers d'autres dictionnaires

25 Il nous a paru nécessaire que notre dictionnaire des noms propres puisse s'intégrer dans d'autres structures lexicales, par une sorte de pointeur que nous avons appelé *export*.

26 Nous envisageons actuellement deux liens, un vers une encyclopédie (Wikipédia[11]) et un vers le lexique général (EuroWordnet). Ce dernier lien est présent aux niveaux conceptuel et linguistique.

Au niveau conceptuel :

27 Si le nom propre est présent dans la base Wordnet, on lui associera son propre numéro ILI (*Inter Lingual Index*). C'est le cas de Paris, qui porte le numéro 0558236n :

entity
location
region
area, country
center, middle, heart
seat
capital
national capital
Paris, City of Light, French capital, capital of France

28 Dans le cas contraire, on lui associera le numéro ILI correspondant à son insertion dans la hiérarchie, à partir d'une des expansions notoires (voir section 1.1.). Par exemple, Jules Verne sera inséré sous le concept *writer*, qui porte le numéro 06438760n :

entity
object, physical object
living thing, animate thing
organism, being
person, individual, someone
communicator
writer

Au niveau linguistique :

29 Dans le cas où, dans une langue donnée, il existe une antonomase, on lui associera le numéro ILI correspondant au nom commun concerné. Prenons l'exemple du mot *bic* qui sera associé au numéro ILI des stylobilles, 02108168n :

entity:1
object:1, physical object:1
whole:2, whole thing:1, unit:6
artifact:1, artefact:1
instrumentality:3, instrumentation:1
implement:1
writing implement:1
pen:1
ballpoint:1, ballpoint pen:1, ballpen:1, Biro:1

Conclusion

30 Le projet Prolex a pour but la réalisation d'une plate-forme technologique pour le traitement automatique des noms propres. L'ontologie que nous venons de présenter en est le fondement.

31 A partir de l'ontologie, nous avons implanté une base de données relationnelle multilingue de noms propres, qui est accessible par Internet (http://tln.li.univ-tours.fr/tln_prolex/prolex.php). Pour le francais, notre base contient à ce jour (juin 2005) plus de 51 000 prolexèmes, c'est-à-dire plus de 119000 instances. Pour la partie multilingue, nous avons la traduction de 234 pays (allemand : 219, anglais : 234, espagnol : 231, italien : 227, hollandais : 210, portugais : 229) et de 552 villes (allemand : 524, anglais : 552, espagnol : 515, italien : 529, hollandais : 481, portugais : 298). Nous travaillons aussi sur la mise en place d'un format XML d'échange de données (Bouchou et al., 2005).

32 L'organisation de notre ontologie en deux parties, conceptuelle et morphologique, et la présence de relations entre noms propres permettra le développement d'outils d'aide à l'utilisateur (pour la rédaction ou la traduction) ou de traitement automatique des langues (étiquetage, traitement des coréférences, recherche d'information, traduction automatique, alignement de textes multilingues…).

Bibliographie

BAUER G. (1985), *Namenkunde des Deutschen*, Bern, Germanistische Lehrbuchsammlung Band 21.

BODENREIDER O., ZWEIGENBAUM P. (2000a), Stratégies d'identification de noms propres à partir de nomenclatures médicales parallèles, *Traitement automatique des langues*, 41-3 :727-757.

BODENREIDER O., ZWEIGENBAUM P. (2000b), Identifying proper names in parallel medical terminologies, *Medical Infobahn for Europe* (*MIE2000*), 443-447.

BOUCHOU B., TRAN M., MAUREL D. (2005), Towards an XML Representation of Proper Names and Their Relationships, *Tenth International Conference on Applications of Natural Language to Information Systems* (*NLDB'2003*), Alicante, Spain, 15-17 juin.

CHINCHOR N. (1997), *Muc-7 Named Entity Task Definition*, http://www.itl.nist.gov.

COATES-STEPHENS S. (1993) *The Analysis and Acquisition of Proper Names for the Understanding of Free Text*, Kluwer Academic Publishers, Hingham, MA.

COURTOIS B., SILBERZTEIN M. (1990), Dictionnaires électroniques du français, *Langues française*, n° 87, 11-22.

DANLOS L. (1989), La traduction automatique, *Annales des télécommunications*, 44, n° 1-2, 101-110.

FRIBURGER N., MAUREL D. (2004), Finite-state transducer cascades to extract named entities in texts, *Theoretical Computer Science*, vol. 313, 94-104.

GROSS M. (1989), The Use of Finite Automata in the Lexical Representation of Natural Language. Electronic Dictionaries and Automata in Computational Linguistics, LNCS 377 :34-50.

GROSS M. (1997), Synonymie, morphologie dérivationnelle et transformations, *Langage*, n° 128, 72-90.

GRUBER T. R. (1995), Toward Principles for the Design of Onthologies Used for Knowledge Sharing, Int. Journal of Human-Computer Strudies, vol. 43, 907-928.

HACHEY B., GROVER C., KARKALETSIS V., VALARAKOS A., PAZIENZA M. T., VINDIGNI M., CARTIER E., COCH J. (2003), Use of Ontologies for Cross-lingual Information Management in the Web, *Ontologies and Information Extraction International Workshop*, *EUROLAN 2003*, Bucarest, Roumanie.

JONASSON K. (1994), *Le nom propre. Constructions et interprétations*, Duculot, Paris.

KRSTEV C., PAVLOVIĆ-LAŽETIĆ G., VITAS D., OBRADOVIĆ I. (2004), Using Textual and Lexical Resources in Developing the Serbian Wordnet, *Romanian journal of Information science and technology*, 147-162.

MACDONALD D. (1996), Internal and external evidence in the identification and semantic categorisation of Proper Names, *Corpus Processing for Lexical Acquisition*, 21-39, Massachussetts Institute of Technology.

MANGEOT-LEREBOURS M., SÉRASSET G., LAFOURCADE M. (2003), Construction collaborative d'une base lexicale multilingue, le projet Papillon, TAL, 44/2, 151-176.

MAUREL D. (2004), Les mots inconnus sont-ils des noms propres ?, *Septièmes Journées internationales d'Analyse statistique des Données Textuelles* (*JADT 2004*), Louvain-la-Neuve, Belgique.

MEL'CUK I. (1984-I, 1988-II, 1992-III), *Dictionnaire explicatif et combinatoire du français contemporain*, Les presses de l'Université de Montréal.

MIKHEEV A., MOENS M., GROVER C. (1999), Named entity Recognition without Gazetteers, EACL'99:1-8.

MILLER G., BECKWITH R., FELLBAUM C., GROSS D., MILLER K. (1990), Introduction to WordNet: an on-line lexical database, *International Journal of Lexicography*, n° 3, p. 235-244.

PAIK W., LIDDY E. D., YU E., MCKENNA M. (1996), Categorizing and Standardizing Proper Nouns for Efficient Information Retrieval, *Corpus Processing for Lexical Acquisition*, 61-73, Massachussetts Institute of Technology.

POLGUERE A. (2003), *Lexicologie et sémantique lexicale. Notions fondamentales*, Presses de l'Université de Montréal.

REN X., PERRAULT F. (1992), The typology of Unknown Words: An Experimental Study of Two Corpora, *COLING 92*, Nantes.

TUFIŞ D., CRISTEA D., STAMOU S. (2004), BalkaNet: Aims, Methods, Results and Perspectives. A General Overview, *Romanian journal of Information science and technology*, volume 7:1-2, 9-44.

VOSSEN P. (1998), EuroWordNet: A Multilingual Database with Lexical Semantic Networks, Kluwer Academic Publishers, Dordrecht.

Notes

1 Cette division en quatre niveaux s'inspire de (Hachey et al., 2003) qui en définit trois ; la particularité de la morphologie dérivationnelle des noms propres nous a amené à créer un niveau intermédiaire, le niveau linguistique.

2 Par exemple, le nom propre *Bouvard* apparaît sous seize formes différentes dans la traduction en serbe du roman *Bouvard et Pécuchet* de *Gustave Flaubert*.

3 Remarquons au passage que la traduction du français *Tchekhov* en allemand est *Tschechow* et en anglais *Chekhov*… Chaque pays ayant développé ses propres règles de transcription…

4 Ils sont le plus souvent homographes (à la majuscule près) ; une exception bien connue est le nom relationnel *Suisse* qui a pour lexie féminine *Suissesse*, alors que l'adjectif relationnel *suisse* a pour lexie féminine *suisse*. Mais, dans d'autres langues, comme le serbe, les adjectifs et les noms relationnels sont systématiquement différents.

5 CN RNIL N 7 : 2003-11-25.

6 Un niveau intermédiaire existe : les anthroponymes individuels et les anthroponymes collectifs, hyponymes des anthroponymes.

7 Ensemble artistique, club sportif…

8 Région supranationale.

9 Les relations du *Dictionnaire explicatif et combinatoire* concernent les *lexies* et celles de *Wordnet*, les *synsets*.

10 Wordnet parle de *near synonyms*, mais place les deux notions dans les *synsets*.

11 Wikipédia est un projet d'encyclopédie gratuite, écrite coopérativement et dont le contenu est réutilisable selon les conditions de la Licence de documentation libre GNU (http://www.wikipedia.org/).

Pour citer cet article

Référence électronique

Denis Maurel et Mickaël Tran, « Une ontologie multilingue des noms propres », *Corela* [En ligne], HS-2 | 2005, mis en ligne le 02 décembre 2005, consulté le 16 décembre 2015. URL : http://corela.revues.org/1203

À propos des auteurs

Denis Maurel
Université François-Rabelais de Tours, Laboratoire d'informatique
Mickaël Tran
Université François-Rabelais de Tours, Laboratoire d'informatique

Résumés

Cet article décrit une ontologie multilingue de noms propres divisée en deux parties, une partie supérieure partagée par toutes les langues traitées et une partie inférieure particulière à chacune d'elles. Elle comprend, d'une part, trois relations sémantiques (Synonymie, Méronymie et Prédication) et, d'autre part, des informations morphosyntaxiques.

This paper describes a multilingual ontology of proper names divided into two parts, a first part shared by all the treated languages and a second part specific to each language. It includes, on the one hand, three semantic relations (Synonymy, Meronymy and Predication) and, on the other hand, some morphosyntactical information.

Entrées d'index

Mots-clés : noms propres, dictionnaires électroniques, ontologies, dictionnaires multilingues, synonymie, méronymie, typologie, alias, dérivation