

Acquisition de concepts bilingues à partir du Web

Olivier Collin¹, Émmanuelle Pétrier²

¹ France Télécom R&D – Lannion
olivier.collin@francetelecom.com

² Teamlog – Lannion
epe@teamlog.com

Résumé

Nous montrons une utilisation du Web, corpus multilingue de grande taille, pour effectuer une acquisition supervisée de concepts bilingue français/anglais. Cette acquisition utilise comme point initial un verbe français. Nous apparions ensuite des phrases provenant des deux langues à partir de couples de noms propres possédant la même forme dans les deux langues. Cet appariement automatique mais sommaire ne garantit pas l'alignement des phrases. Nous montrons qu'il nous permet cependant d'extraire des termes français et anglais équivalents dans leur contexte d'utilisation. Ces termes constituent des ressources multilingues particulièrement adaptées au Web, notamment pour les applications question réponse « crosslingue ».

Mots-clés : acquisition, concepts bilingue, alignement superficiel, Web.

Abstract

This article describes a way of using the Web as a huge multilingual corpus to perform supervised acquisition of bilingual French/English concepts. Such acquisition is initialized with a French verb. Sentences expressed in both languages are then matched by using couples of proper nouns which are unchanged in these two languages. Such matching is automatic, highly succinct, but does not prevent sentences from being wrongly aligned. However, we show that it makes it possible to extract French and English terms that are equivalent in their context of use. These terms constitute multilingual resources that are particularly adapted to the Web, especially for cross-lingual question answering.

Keywords: acquisition, bilingual concepts, shallow alignment, Web.

1. Introduction

Nous présentons l'utilisation du Web pour acquérir des classes d'équivalence de termes français et anglais. Chaque classe est initialisée à partir d'un verbe français, nous réalisons ensuite une acquisition de termes français et anglais substituables dans des contextes d'emplois contraints. Pour une classe donnée, l'ensemble de ces termes représente un « concept » bilingue. Les termes ainsi visés sont des verbes ou des déverbaux prenant en arguments des couples de noms propres. Nous n'utilisons pas de corpus alignés, les données utilisées sont des phrases provenant des deux langues, globalement mises en correspondance par l'intermédiaire de noms propres possédant la même forme dans les deux langues. Ceci constitue un type d'appariement « faible » réalisé automatiquement par des requêtes effectuées dans les deux langues sur le Web, à partir de mots-clés identiques, les couples de noms propres. Les phrases ainsi obtenues en retour du moteur ne sont pas des traductions mais contiennent suffisamment d'informations exprimant le même concept pour en extraire des termes équivalents. Ce résultat est rendu possible par la taille et le caractère multilingue

du Web : nous ne connaissons aucun autre corpus contenant une quantité aussi importante de noms propres identiques dans plusieurs langues. Ce travail reprend en partie des travaux de Duclaye (2003) pour en réaliser une extension à des classes de termes bilingues. Le principe d'acquisition des données à partir du Web a été conservé mais nous avons entièrement supervisé les phases de classification successives.

2. Une boîte à outils

En pratique, les avantages majeurs du Web sont la quantité de données offerte et le temps d'accès, relativement court, à ces données. Cette quantité offre des perspectives encore peu exploitées. Cependant, l'accès à ces données ne peut être réalisé que par un mode d'indexation quel qu'il soit et ce passage obligé entraîne un biais lors de cet échantillonnage des données. Dans la mesure où nous ne pouvons pas réaliser nous-mêmes cette indexation, nous sommes contraints d'utiliser les moteurs de recherches standards. Nous pouvons toutefois réutiliser le résultat de cet échantillonnage de manière à constituer une image particulière du Web, dans une perspective de modélisation linguistique. Nous accumulons actuellement une image du Web centrée sur la terminologie, le but étant de récupérer un grand nombre de cas d'usage de chaque terme. Les noms propres font partie de cette terminologie : la quantité de contextes d'usage lié à un nom propre ainsi récupérés est, en moyenne, beaucoup plus importante que dans un corpus « standard » (*Le Monde* sur 10 ans). De manière générale, la fréquence de chaque N-grammes observable est beaucoup plus importante. Cette base étant constituée, nous pouvons l'indexer localement, dans un but de modélisation linguistique, en y associant par exemple des données provenant de connaissances « expertes ». Dans cette perspective, nous développons une boîte à outils, articulée autour d'une base SQL, permettant de produire, pour chaque terme, un ensemble d'attributs linguistiques et d'attributs contextuels qualitatifs et quantitatifs. Ces derniers sont le résultat d'un processus d'opérations d'acquisition de phrases sur le Web, de filtrage syntaxique, de comptage et de classification. Des opérations d'analyse contextuelle couplées à nos outils d'analyse linguistique (étiquetage, analyse syntaxique, etc.) sont utilisées. D'autres opérations récurrentes telles que le comptage d'occurrences et la représentation des formes par des identifiants sont aussi intégrées en amont d'algorithmes de classification. Ce travail et le processus d'acquisition associé sont un exemple d'utilisation de notre boîte à outils permettant des traitements couplés sur deux langues.

3. Principe de l'acquisition

Le processus d'acquisition permettant d'acquérir des équivalences à partir d'un verbe est décrit ci-dessous.

3.1. Initialisation à partir des formes françaises du verbe choisi

Nous choisissons un verbe par relation, par exemple le verbe *acheter* pour la relation d'achat. Une requête sur l'Internet français est effectuée pour chacune de ces formes fléchies (*achète, acheta, achètera*, etc.) récupérées dans notre lexique du français. Nous récupérons ainsi en moyenne 1500 phrases contenant chacune de ces formes.

3.2. Extraction et sélection de couples de noms propres syntaxiquement liés au verbe initial

Des patrons syntaxiques permettent d'extraire N couples de noms propres à partir des phrases renvoyées. Ces couples sont ordonnés par leur fréquence absolue. Nous choisissons alors les n meilleurs couples parmi les plus fréquents (validés manuellement), avec n compris entre 1 et 12. Ces couples constituent à leur tour les mots-clés de requêtes envoyées vers le Web français et anglais.

3.3. Extraction et sélection de termes équivalents en français et en anglais

Ces couples sélectionnés permettent de récupérer chacun M phrases françaises et K phrases anglaises. Des verbes et des noms sont extraits dans les deux langues au moyen de patrons syntaxiques contraints par les noms propres. Ces termes sont regroupés par lemmes et ordonnés par leur fréquence absolue. Les plus fréquents sont validés manuellement et, selon les cas, de 1 à 6 termes sont relancés à leur tour sur le Web français et anglais.

3.4. Extraction et sélection de couples de noms propres syntaxiquement liés aux termes

Les patrons syntaxiques permettent d'extraire de nouveaux couples de noms propres à partir des phrases renvoyées. Ces couples sont ordonnés par leur fréquence absolue. Nous sélectionnons 5 nouveaux couples parmi les plus fréquents dans les deux langues : nous sélectionnons de préférence les meilleurs couples occurants simultanément pour chacune des deux langues considérées. Cette stratégie permet, *a priori*, de récupérer plus de termes anglais que lors de la phase précédente analogue.

3.5. Extraction et sélection de termes équivalents en français et en anglais

Les couples sélectionnés précédemment permettent à leur tour d'extraire de nouveaux termes équivalents dans chacune des langues.

Pour réaliser les extractions de termes nous utilisons des patrons syntaxiques : règles basées sur des expressions régulières construites à partir des résultats de notre étiqueteur. L'étiqueteur dépend de chaque langue mais les expressions sont quasiment toutes communes aux deux langues. Ces patrons permettent d'extraire les verbes ou noms à partir des couples de noms propres et inversement, des couples de noms propres à partir des verbes ou noms. Le tableau ci-dessous résume les différents cas de figure, en fonction de la catégorie des termes à extraire et de leur position relative par rapport aux noms propres (avant, entre, après).

Verbales	Nominales
AOL achète acheter Netscape	Le rachat de rachat AOL par Netscape
Après avoir acheté acheter Netscape, AOL est devenu déficitaire	AOL propriétaire de propriétaire Netscape
AOL et Netscape négocient négociier	AOL / Netscape, le rachat rachat

Tableau 1. 6 catégories d'extractions syntaxiques utilisées

Par exemple, pour le cas « AOL achète|acheter Netscape », si les noms propres sont connus, l'expression permettant d'extraire un verbe placé entre les deux noms propres est du type

« AOL * V? * Netscape ». Quelques contraintes sur les * permettent d'éviter de traiter des insertions de relatives. Si le verbe est connu (acheter), le patron est du type « NP1? * acheter * NP2? ». Une étude rapide nous a montré qu'en ce qui concerne nos corpus de phrases provenant du Web, la majorité des relations entre deux NP est réalisée par un verbe. De tels patrons permettent de récupérer la plupart du temps une réelle relation de dépendance syntaxique V(NP1, NP2). Nous n'avons cependant pas extrait la fonction syntaxique (suj, obj) des noms propres. Ce type d'analyse superficielle entraîne inévitablement du bruit et ce n'est bien sûr qu'un pis-aller en l'absence d'outils syntaxiques multilingues plus précis, robustes et rapides. Nos expérimentations sont donc conditionnées par la qualité de cette analyse.

4. Résultats

Nous présentons des résultats qualitatifs, obtenus à partir du processus précédent, pour les relations initialisées par les verbes **acheter**, **assassiner**, **chanter**, **diriger**, **épouser**, **féliciter**, **naître**, **présider**, **rompre**. Chaque tableau résume les résultats obtenus pour chaque étape du processus décrit au paragraphe précédent : formes initiales (FR), couples sélectionnés, termes sélectionnés (FR et EN), nouveaux couple sélectionnés, nouveaux termes validés (FR et EN). Les dernières lignes présentent en plus un certains nombre de termes, pour chacune des langues, que nous n'avons pas considérés comme des équivalences pour la relation visée, mais qui ont une « proximité » avec cette relation : antonymes, concepts « proches », champ lexical.

acheter	Achète, achetait, acheter, acheté, etc.
Couples	(airespace,cisco),(symantec,veritas),(ebay,skype),(lycoris,mandriva), (apple,tivo), (google,kaltix), (logica, unilog)
Synonymes FR	Racheter, acheter, acquérir, achat
Synonymes EN	Acquire, buy, acquisition
Couples	(symantec,veritas software), (airespace,cisco systems), (telecom italia,tiscali france), (ebay,skype technologies), (novell,ximian)
Synonymes FR	Racheter, acheter, acquérir, achat , rachat, acquisition
Synonymes EN	Acquire, buy, acquisition , purchase, buying, takeover, get, take, buyer
Autres FR	Offrir, reprendre, payer, fusion, mariage, avoir des vues, cession, rapprochement, convoiter, marier, implanter, propriétaire, actionnaire majoritaire, mettre la main, titrer, concurrent, porter acquéreur, offre, prendre la tête de, opa, absorber, négocier, avaler
Autres EN	Merger, merge, give, deal, stock, provide, bid, advance, convert, provider

assassiner	Assassine, assassinait, assassiner, assassiné, etc.
Couples	(kennedy,lee harvey oswald), (john wilkes booth,lincoln), (henri iv,ravaillac), (henri iii,jacques clément), (abel,caïn), (john lennon,mark david chapman), (charlotte corday,marat)
Synonymes FR	Assassiner, tuer, assassinat, assassin
Synonymes EN	Assassinate, kill, murder, assassin, assassination, killer
Couples	(john lennon,mark david chapman), (abraham lincoln,john wilkes booth), (john f. kennedy,lee harvey oswald), (charlotte corday,jean - paul marat), (abraham lincoln,john wilkes booth)
Synonymes FR	Assassiner, tuer, assassinat, assassin , meurtre, poignarder, meurtrier, meurtrière
Synonymes EN	Assassinate, kill, murder, assassin, assassination, killer , shoot, murderer, murderess

Autres FR	Assassinat roi, péché originel, crime, torturer, mourir, mort, gardien de prison,
-----------	---

	victime
Autres EN	Accused assassin, execution, death, gun, assassination interview, slayer, stab

chanter	Chante, chantait, chanter, chanté, etc.
Couples	(fredo,renaud), (fela,zombie), (amélie,olympia), (mireille mathieu,paris), (brassens,forestier), (boris vian,négropolitains), (brassens,renaud), (montand,prÉvert), (billy jean,michael jackson), (get up,james brown), (easy,lionel ritchie), (mccartney,your money)
Synonymes FR	Chanter, chanson, interpréter
Synonymes EN	Song, singing
Couples	(nino ferrer,sud), (georges moustaki,métèque), (james brown,sex machine), (forestier,georges brassens), (god save the queen,queen)
Synonymes FR	Chanter, chanson, interpréter , interprétation, interprète, chanteur
Synonymes EN	Song, singing
Autres FR	Lyric, paroles, pochette, faire chanter, rythme, artiste, concert, partition, répertoire, tube, morceau, fan, texte, entendre
Autres EN	Duet

diriger	Dirige, dirigeait, diriger, dirigé, etc.
Couples	(laurent fournier,psg), (george hedare,hedare beauty), (thierry breton,thomson multimedia), (christian duhain,eads international)
Synonymes FR	Diriger, présider, président directeur général, patron, président
Synonymes EN	Chairman, coach, president, lead
Couples	(thierry breton,thomson), (thierry breton,thomson multimédia), (christian duhain,eads), (amérique,israël), (los angeles,new york)
Synonymes FR	Diriger, présider, président directeur général, patron, président, pdg
Synonymes EN	Chairman, coach, president, lead,Executive, ceo
Autres FR	Entraîneur, prendre la tête de, manager
Autres EN	Head

épouser	Epouse, épousait, épouser, épousé, etc.
Couples	(camilla,charles), (camilla parker bowles,charles), (calogero,hortense), (amphion,niobé), (camilla parker - bowles,charles), (charles,diana), (kate moss,pete doherty), (aragorn,arwen), (cruise,katie holmes), (jocaste,laïos), (bérengère de navarre,richard cour de lion)
Synonymes FR	Épouser, mariage, marier, époux, épouse
Synonymes EN	Marry, wed, marriage, wedding
Couples	(camilla parker bowles,prince charles), (katie holmes,tom cruise), (camilla,prince charles), (britney spears,kevin federline), (jennifer lopez,marc anthony)
Synonymes FR	Épouser, mariage, marier, époux, épouse , convoler en justes noces, remariage, marié
Synonymes EN	Marry, wed, marriage, wedding , marriage ceremony, get married
Autres FR	Mari, amour, compagne, couple, liaison, duo, fiancé, mariage prince, coupler, fiancer, réunir, demander la main de, demander le divorce, unir, épouse mère, union, divorce, conquête, aimer, mariage civil, demande en mariage, mairie, divorcé, convoler, famille époux, petit ami, meurtre époux, cérémonie, compagne prince, rupture, rencontre, noces, partenaire, époux prince, petite amie
Autres EN	Husband, couple, relationship, love, mrs, divorce, engage, romance, ceremony, time companion, engagement, date, newlywed, separation, love affair, civil wedding, bride, dating, lovebird, nuptials, term companion, blessing, pregnancy, separate, mariage plan, wedding cover
féliciter	Félicite, félicitait, féliciter, félicité, etc.
Couples	(etats - unis,faure gnassingbé), (france,turquie), (ludovic giuly,ronaldinho), (angela

	merkel,chirac), (arabie saoudite,jalal talabani), (banque mondiale,france), (cpc,vancouver), (ariel sharon,mahmoud abbas), (alexandre vinokourov,lance armstrong), (bernadette chirac,jean - pierre raffarin)
Synonymes FR	Féliciter, félicité
Synonymes EN	Agree
Couples	(mahmoud abbas,prime minister ariel sharon), (president mahmoud abbas,prime minister ariel sharon), (mahmoud abbas,minister ariel sharon), (angela merkel,jacques chirac), (ludovic giuly,year ronaldinho)
Synonymes FR	Féliciter, félicité
Synonymes EN	Agree
Autres FR	Rencontrer, rencontre, poignée de main, homologue, amitié, adhésion, partenaire, faire valoir
Autres EN	Agree, hail

naître	Naître, naissait, naître, né, etc.
Couples	(bethléem,messie), (France,molière), (israël,juda), (ferte - milon,racine), (georges simenon,liège), (bartholdi,colmar), (jacques martin,strasbourg), (comtesse de ségur,saint - pétersbourg), (jérusalem,jésus)
Verbes FR	Naître, naissance
Verbes EN	Bear
Couples	(bethléem,jésus), (colmar,frédéric auguste bartholdi), (jérusalem,salomon), (bethléhem,jésus), (bethléhem,messie)
Synonymes FR	Naître, naissance
Synonymes EN	Bear (be born)
Autres FR	Entrée, enfant, arriver, étable, jésus, arrivée, crèche, vie, faire naître, nativité
Autres EN	Birthplace

présider	Préside, présidait, présider, présidé, etc.
Couples	(convention,valéry giscard), (emir kusturica,festival de cannes), (bertie ahern,union), (csst,gottfried schatz), (cfdt -,françois chérèque -), (comité,solofo rasoarahona), (commission,josé manuel durao barroso), (congrès,dennis hastert), (alex türk,europol), (allergan france,mir nezam), (axa,claudie bébéar), (michel giraud,synafap)
Synonymes FR	Présider, président, patron, présidence, préside
Synonymes EN	President, chairman
Couples	(european commission,josé manuel durao barroso), (Président,république), (Convention,m. valéry giscard), (european union,prime minister bertie ahern), (commission,portugais José Manuel Durao Barroso)
Synonymes FR	Présider, président, patron, présidence, préside, diriger, dirigeant
Synonymes EN	President, chairman
Autres FR	Nommer, élire, fondateur, désigner, élection, premier ministre, chef de l'état, pouvoir, investiture, représentant, mandat, député, responsabilité, président groupe, fondateur groupe, numéro un
Autres EN	board chairman, ceo, prime minister

rompre	Rompre, rompait, rompre, rompu, etc.
Couples	(jude law,sienna miller), (frédéric vi,napoléon), (m. milosevic,radovan karadzic), (henri de navarre,henri iii), (etats - uni s irak), (chris klein,katie holmes), (marc antoine,octave), (jennifer aniston,vince vaughn), (arnold geulincx,descartes)
Synonymes FR	Rompre, séparer, quitter, rupture, larguer
Synonymes EN	Split, break, end engagement
Couples	(brad pitt,jennifer aniston), (ben affleck,jennifer lopez), (kate bosworth,orlando

	bloom), (jessica simpson,nick lachey), (guy chris klein,katie holmes)
Synonymes FR	Rompre, séparer, quitter, rupture, larguer , finir, divorce, divorcer, séparation
Synonymes EN	Split, break, divorce, end engagement , dump, breakup, end romance,
Autres FR	Couple, mari, coupler, mariage, fiancer, reconquérir, retrouver, épouse, marier, réconcilier, épouser, époux, rompre le cou, remplacer, terminer, repartir, retrouvailles, contredire, fiancé, tendon rupture, petite amie, tourner la page, idylle
Autres EN	Couple, marriage, relationship, date, engage, divorce story, romance, marry, separate, engagement, cheat, boyfriend, newlywed, ex, link, wedding, nuptials

Nous remarquons que nous obtenons presque systématiquement, dès la première itération, le déverbal et la traduction anglaise directe. Ainsi, nous récupérons *achat* et *buy* pour *acheter*, *assassinat* et *assassinate* pour *assassiner*, *rupture* et *break* pour *rompre*, etc. Nous retrouvons, par ailleurs, des champs lexicaux identiques pour les relations antonymes, comme *rompre* et *épouser* ou les relations synonymes comme *diriger* et *présider*. Le nombre et la qualité des termes acquis à partir d'un verbe initial varie considérablement d'un verbe à l'autre. Cette productivité semble liée à la généralité du verbe. Ainsi *naître*, assez fréquent dans le vocabulaire courant, donne lieu à peu d'expression. Ceci corrobore les travaux de Duclaye (2003) concernant des équivalences monolingues : 80 % des équivalences extraites lors des 2 premières itérations au cours d'un processus d'une dizaine d'itérations.

Pour un même verbe, nous n'avons pas cherché à contrôler la classe des noms propres en relation. Nous constatons donc l'émergence de plusieurs types de sous-relations liées à une distribution argumentale de types différents. Exemple pour le verbe *chanter* :

- <chanteur>,<chanteur|compositeur> (ex : Montand, Prévert) ;
- <chanteur>,<chanson> (ex : le Sud, Nino Ferrer) ;
- <chanteur>,<lieu> (ex : Amélie, Olympia).

5. Discussion

Les résultats obtenus sont qualitativement intéressants et montrent le potentiel de ce processus qui est relativement simple. La simplicité apparente cependant cache les développements contenus dans notre boîte à outils spécifiés à partir de nos précédentes expériences afin de rendre l'enchaînement des tâches suffisamment simple et rapide pour une utilisation non experte. Ce type d'acquisition est bien entendu contraint par l'utilisation des noms propres et des relations associées et tous les « concepts » ne se prêtent pas à ce type d'acquisition. D'autres relations sont en cours d'étude, initialisée par des verbes (*apprécier, coucher, créer, détester, divorcer, éditer, estimer, fabriquer, interpréter, inventer, jouer, produire, publier, remplacer, succéder, vendre*, etc.) ou par des noms (*match, auteur*).

Les résultats présentés montrent des inégalités dans la qualité et la quantité des termes obtenus pour chacune des relations. Chaque acquisition a été traitée suivant le même protocole. Nous pensons pouvoir améliorer notre processus pour accroître la quantité de données et donc d'hypothèses utilisables. Nous pouvons utiliser d'autres types d'entités nommées (dates, mesures) mieux adaptées à certaines catégories de relations. Pour la relation « être_né_en » nous pourrions utiliser des couples (nom de personne, année/date) comme (Mozart,1756). Nous pensons aussi améliorer nos extracteurs syntaxiques. Nous disposons notamment d'une grammaire de dépendance qui pourra être adaptée à cette tâche. Cette grammaire nous permettra aussi de typer la fonction syntaxique des entités nommées.

Ce travail, entièrement supervisé, permet d'estimer une limite à cette technique, compte tenu des hypothèses générées. Nous pensons cependant intégrer un nouveau modèle de classification peu supervisée (Clérot, 2004) qui pourrait dégrader ces résultats (à nombre d'hypothèses égales) mais qui permettrait de limiter les phases de supervision. L'essentiel du traitement devient automatique et réalisé en tâche de fond.

Les équivalences récupérées pour ces relations sont donc surtout très adaptées aux données provenant du Web qui devient la principale source d'accès et d'extraction d'information. Ces données sont immédiatement réutilisables dans un système des question/réponse translingue et seront intégrées dans nos modèles de patrons d'extraction « adaptés » au Web. Elles doivent aussi permettre de compléter nos ressources linguistiques génériques.

6. Conclusion et perspectives

Nous pensons que la modélisation des ressources linguistiques passe par le traitement de grosses masses de données hétérogènes provenant du Web, de techniques d'apprentissage peu supervisées et par une gestion incrémentale du couplage expertise/apprentissage. Pour aller vers un tel modèle et une meilleure gestion de nos ressources nous développons un outil logiciel adapté à ce point de vue. Cet outil intègre l'acquisition de connaissances à partir de corpus et l'accès à des données expertes dans un seul contexte logiciel qui utilise une structure de stockage homogène : une base SQL. Cette structure nous permet d'atteindre toutes les informations (attributs) liés à un terme en naviguant simplement dans cette base. Les premières utilisations de cet outil nous ont permis de mettre rapidement en œuvre des idées utilisant intensivement le Web comme source de connaissance. Nous avons aussi montré la possibilité de coupler cette connaissance à nos ressources, principalement une analyse syntaxique superficielle, dans un processus global d'acquisition supervisée ou non. Les résultats présentés montrent la faisabilité de l'acquisition de certaines classes de concepts bilingues par l'utilisation de caractéristiques propres au Web permettant de réaliser une forme d'appariement automatique par l'utilisation des noms propres de formes identiques en français et en anglais. Cette acquisition, à notre connaissance unique en son genre, nous permet d'obtenir des données bilingues adaptées à l'extraction d'information sur le Web. Nos efforts à venir portent sur l'intégration d'algorithmes d'apprentissage faiblement supervisé, la réutilisation et la fusion de classifieurs indépendants ainsi que l'intégration des données acquises dans notre modèle sémantique « expert ».

Références

- AKIRA T., TAKENOBU T. (2002). « Automatic disabbreviation by using context information ». In *Proceedings of the NLPRS Workshop on Automatic Paraphrasing : Theories and Applications*.
- BARZILAY R., MCKEOWN K.R. (2001). « Extracting paraphrases from a parallel corpus ». In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*. Toulouse : 50-57.
- BRIN S. (1998). « Extracting patterns and relations from the world wide web ». In *Proceedings of WebDBWorkshop at EDBT*.
- CLÉROT F., COLLIN O., CAPPÉ O., MOULINES E. (2004). « *Le modèle monomaniaque* : un modèle simple pour l'analyse exploratoire d'un corpus de textes ». In *Actes de CFTI 04*.
- COLLIN O., PÉTRIER E. (2005). « Boîte à Outils de Gestion et d'Acquisition de Ressources Linguistiques ». In *Actes des JLC 05*.

- COLLIN O., DUCLAYE F. (2001). *Procédé et système d'enrichissement automatique de ressources sémantiques à partir d'un système électronique temps réel de question-réponse*. Brevet N. 01 09684.
- DUCLAYE F., YVON F., COLLIN O. (2003). « Unsupervised incremental acquisition of a thematic corpus from the Web ». In *Proceedings of NLPKE 03*.
- DUCLAYE F., YVON F., COLLIN O. (2003). « Learning Paraphrases to Improve a Question-Answering System ». In *Proceedings of EACL 03*.
- HOFMANN T., PUZICHA J. (1998). « Statistical models for co-occurrence data ». In *Technical Report AI. 1625. MIT, AI Lab*.
- LÉON S., MILLION C. (2005). « Acquisition semi-automatique de relations lexicales bilingues (français-anglais) à partir du web ». In *Actes de RECITAL 2005*. Dourdan.