# Les adjectifs relationnels dans les lexiques informatisés : formalisation et exploitation dans un contexte multilingue

Bruno Cartoni

Université de Genève
cartonib@gmail.com

**Résumé** Dans cet article, nous nous intéressons aux adjectifs dits relationnels et à leur statut en traitement automatique des langues naturelles (TALN). Nous montrons qu'ils constituent une « sous-classe » d'adjectifs rarement explicitée et donc rarement représentée dans les lexiques sur lesquels reposent les applications du TALN, alors qu'ils jouent un rôle important dans de nombreuses applications. Leur formation morphologique est source d'importantes divergences entre différentes langues, et c'est pourquoi ces adjectifs sont un véritable défi pour les applications informatiques multilingues. Dans une partie plus pratique, nous proposons une formalisation de ces adjectifs permettant de rendre compte de leurs liens avec leur base nominale. Nous tentons d'extraire ces informations dans les lexiques informatisés existants, puis nous les exploitons pour traduire les adjectifs relationnels préfixés de l'italien en français.

**Abstract** This article focuses on a particular type of adjectives, relational adjectives, and especially on the way they are processed in natural language processing systems. We show that this class of adjectives is barely recorded in an explicit manner in computer lexicons. There is an important discrepancy in the way those adjectives are morphologically constructed in different languages, and therefore, they are a real challenge for multilingual computing applications. On a more practical side, we propose a formalisation for the adjectives that shows their semantic link with their nominal base. We make an attempt to extract this kind of information in existing machine lexica, and we exploit their semantic links in the translation of prefixed relational adjectives from Italian into French.

**Mots-clés :** Adjectifs relationnels, ressources lexicales, morphologie constructionnelle
**Keywords:** Relational adjectives, lexical resources, constructional morphology

# 1 Introduction

Cet article s'interroge sur l'attention que le TALN a porté aux adjectifs d'un genre particulier : les adjectifs relationnels. Nous montrons que ces adjectifs, de par la relation (morphologique et sémantique) qu'ils entretiennent avec leur base nominale, représentent une information importante pour de nombreuses applications du TALN, et que les lexiques informatisés négligent, à tort, ce type d'information.

Dans cette optique, nous avons procédé à une expérience d'extension semi-automatique de deux lexiques informatisés, pour pouvoir ensuite exploiter les informations relatives aux adjectifs relationnels dans une application précise : un système de traduction automatique (TA) des mots morphologiquement construits.

Dans un premier temps, nous resituons la question des adjectifs relationnels, en précisant les différentes caractéristiques qui les définissent, puis nous montrons (section 2) comment les informations morphosémantiques qu'ils contiennent (c'est-à-dire le lien qu'ils entretiennent avec la base nominale sur laquelle ils sont construits) constituent une information que l'on peut acquérir facilement de manière automatique.

Dans la seconde partie, nous présentons notre prototype de traducteur automatique des mots construits, pour lequel une extension des lexiques a été réalisée. Nous montrons que l'information « adjectif relationnel » est particulièrement importante dans un contexte multilingue, car ces adjectifs sont souvent la source de nombreuses divergences.

# 2 Les adjectifs relationnels

Les adjectifs relationnels sont souvent décrits dans les études linguistiques comme un type d'adjectif à part, qui possède un certain nombre de propriétés bien connues (même si elles ont toutes été nuancées à plusieurs reprises), à savoir (Mélis-Puchulu, 1991) :

- Ils sont morphologiquement construits sur une base nominale par suffixation (*président* → *présidentiel*), parfois avec la forme supplétive de cette base (*cœur* → *cardiaque*).

- Ils possèdent un certain nombre de caractéristiques syntaxiques (impossibilité d'être utilisés comme attribut **elle est présidentielle*, non-gradabilité **une élection très présidentielle*, …).

- Quand ils sont employés avec un nom, ils instaurent une relation entre le sens du nom recteur du groupe nominal et le nom base de l'adjectif. Ainsi, la séquence *élection présidentielle* peut être paraphrasée par *élection du président*.

- Comme l'a récemment montré (Fradin (2008), les adjectifs relationnels possèdent la même représentation sémantique que leur base nominale.

A ces caractéristiques bien connues, s'ajoute un « comportement » très particulier, qu'il nous faut ici souligner : les adjectifs relationnels sont très souvent employés comme base dans des règles de préfixation qui permettent la formation de lexèmes dont le sens construit est un peu particulier. Dans ce type de préfixation, « la signification du lexème construit résulte de la

combinaison du lexème-base et du préfixe, et non de la combinaison du lexème-suffixé et du préfixe » (Fradin, 2003). Ainsi, *anticonstitutionnel* qualifie le fait d'être *contre la constitution* ; l'instruction sémantique de la règle de préfixation porte donc sur la base *constitution* de l'adjectif relationnel (qui lui sert de base formelle). Ce phénomène est étroitement lié au fait que les adjectifs relationnels désignent une relation entre l'entité dénotée par leur nom base et l'entité modifiée. Dans la séquence *loi anticonstitutionnelle,* l'adjectif relationnel permet de désigner la relation entre *loi* et le fait d'être contre la *constitution*.

De plus, c'est dans les règles de préfixation que l'adjectif relationnel se révèle être le plus proche de sa base nominale. Ainsi, il existe des cas où la base nominale peut être utilisée comme input formel de la règle de préfixation (*un traitement anticancer*). Ces cas surviennent généralement lorsque l'adjectif relationnel n'existe pas (*un plan antidrogue*, *drogue* n'ayant pas d'adjectif relationnel possible en français), mais parfois les deux possibilités de formation semblent interchangeables (*un traitement anticancéreux*).

# 3 Les adjectifs relationnels en TALN

En traitement automatique des langues, les adjectifs relationnels ont plusieurs fois été étudiés, notamment dans le cadre de l'extraction terminologique, par exemple dans (Daille 1999). En effet, les termes complexes sont souvent composés d'un nom recteur et d'un adjectif relationnel, et leur traitement permet de repérer des liens entre différentes variantes d'un même terme, comme entre *production de céréales* et *production céréalière* (l'exemple est également tiré de (Daille, 1999)). Tout le travail est alors ici, pour l'analyseur de l'extracteur de terme, de *déconstruire* l'adjectif afin de retrouver sa base nominale. Si cette tâche est à effectuer à chaque extraction, cela veut dire que les lexiques sur lesquels reposent ces analyseurs ne possèdent pas cette information, alors que, comme nous le montrons dans la suite, cette information peut facilement être acquise, même avec un minimum de ressources.

## 3.1 Extension automatique d'un lexique informatisé

Pour les besoins de notre expérience (la construction d'un prototype de traducteur automatique de mots construits), dont nous décrivons les tenants et aboutissants dans la deuxième partie de ce travail, nous avons besoin de l'information lexicale décrite ci-dessus, à savoir le lien entre les adjectifs relationnels et leur base nominale, et ce pour les deux langues de notre système, l'italien et le français. Notre système de TA repose sur un lexique bilingue et deux lexiques monolingues utilisés par l'analyseur morphosyntaxique *Mmorph* (Bouillon, Lehmann et al. 1998). D'un point de vue quantitatif, le lexique français contient 279 007 formes fléchies, et le lexique italien 739 000 formes fléchies. Nous décrivons dans cette section l'extension effectuée pour le lexique de l'italien, celle du lexique français ayant suivi les mêmes principes.

Pour étendre le lexique sur lequel repose l'analyse, nous avons cherché la base nominale potentielle de tous les adjectifs relationnels. Ainsi, nous avons implémenté une routine basée sur les suffixes typiques des adjectifs relationnels de l'italien : *-ale, -are, -ario, -ano, -ico, -ile, -ino, -ivo, -orio, -esco, -asco, -iero, -izio, -aceo* (Wanddruszka, 2004). Cette routine recherche, pour chaque adjectif se terminant par l'un de ces suffixes, une possible base nominale en opérant une sorte de « fuzzy matching » (en *désuffixant* l'adjectif puis en effectuant une série de calcul sur la chaîne de caractère restante pour retrouver une base

potentielle enregistrée dans le lexique). L'extracteur, proposé par (Daille, 1999), vérifiait ensuite la présence de la base nominale trouvée dans le corpus analysé, ce qui empêchait le repérage de « monstre ». Dans notre contexte d'extension du lexique, nous avons dû vérifier toutes les paires « adj_rel–nom » manuellement.

Dans un deuxième temps, nous avons évalué le nombre de cas dans lesquels l'appariement « adj_rel–nom » était correct, afin d'avoir une idée du fonctionnement d'une routine aussi simple. Si certains suffixes donnent d'excellents résultats avec une précision supérieure à 90 %, d'autres sont beaucoup plus ambigus et provoquent trop de bruit — c'est notamment le cas de *-ile, -ano, -iano, -iario*. Pour ces derniers, la correction manuelle est donc très importante.

Cette implémentation fort simple nous a permis d'ajouter au lexique 8 466 liens entre des adjectifs relationnels et des noms dans le lexique italien et 1 822 dans le lexique français. Nos deux lexiques sont ainsi enrichis d'une information supplémentaire, qui sera ensuite exploitée dans notre système de TA (cf. section 4).

Malheureusement, ce type d'implémentation basée uniquement sur une approche d'extension automatique (comme décrit dans (Gdaniec et Manandise, 2000)), ne permet pas de retrouver des paires « adj_rel–nom » où l'adjectif est construit sur une forme supplétive du nom (*cœur* → *cardiaque*). Pour ce type de formation, il faut alors envisager la création d'une ressource *ad hoc* contenant tous les supplétifs nécessaires, comme l'a d'ailleurs fait (Namer, 2005) dans son analyseur *Dérif*.

# 4 Exploitation des adjectifs relationnels dans un contexte multilingue

Comme nous l'avons déjà mentionné, l'information « adjectifs relationnels » peut jouer un rôle important dans de nombreuses applications du TALN. Dans cette section, nous présentons l'application pour laquelle l'extension du lexique présentée ci-dessus a été réalisée : un système de traduction automatique des néologismes construits. Ce système, et les différentes études qui en sont découlées, ont déjà été présenté, notamment dans (Cartoni, 2005) et (Cartoni, à paraître).

## 4.1 Un système de TA des néologismes construits

Le système dans lequel s'inscrit cette étude est un système de traduction automatique des néologismes construits, qui est encore à l'état de prototype. Il repose sur le présupposé que les procédés de construction des néologismes peuvent être transposés d'une langue à l'autre et ainsi être « traduit », palliant ainsi l'absence de ces mots dans les lexiques. Ainsi, pour tout mot néologique dans une langue (p. ex. *ricostruire* en italien), le système (i) procède d'abord à une analyse morphologique, permettant de retrouver la règle qui a produit le mot inconnu (*ri+costruire*), puis (ii) par l'intermédiaire de règles de transfert morphologique (ici la *règle*

*de réitération*), produit un équivalent de traduction, soit en reconstruisant un équivalent de traduction (*reconstruire*) ou en proposant une glose (*construire à nouveau*).[1]

Ce système repose donc sur un ensemble de Règles de Construction des Lexèmes (RCL) bilingues, qui permettent d'analyser les néologismes construits dans une langue et de transférer les informations morphosémantiques dans l'autre langue pour générer une traduction possible. En outre, le système repose, comme nous l'avons mentionné, sur un lexique bilingue construit semi-automatiquement pour les besoins de l'expérience, et sur deux lexiques monolingues de l'italien et du français.

Le schéma ci-dessous représente le fonctionnement global de ce système et les deux types de ressources impliquées (les RCL et les lexiques).

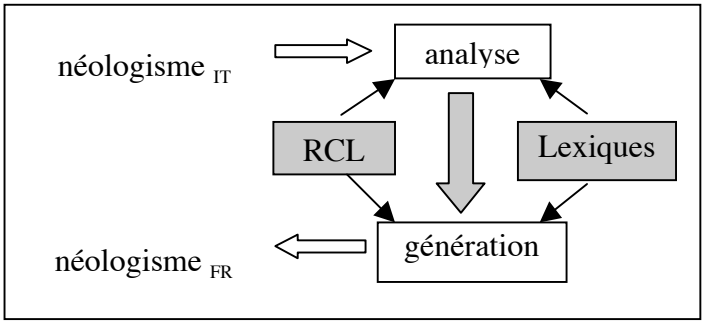



Figure 1: Le système de traduction automatique des néologismes

Pour ce système, nous avons implémenté plus d'une centaine de RCL bilingues, permettant d'analyser la plupart des néologismes construits par préfixation. Avant leur implémentation dans les programmes d'analyse de transfert, les RCL bilingues ont été formalisées en suivant le modèle proposé par (Fradin, 2003). Nous reproduisons dans la figure 2 ci-dessous la RCL bilingue de réitérativité.

| | *italien* | *français* |
|---|---|---|
| | input | input |
| (G) | X | Y |
| (F) | /X/ | /Y/ |
| (SX) | cat :v | cat :v |
| (S) | X'(...) | Y'(...) |
| | ↕ | ↔ ↕ |
| | output | output |
| (G) | riX | reY |
| (F) | /ri/⊕/X/ | /Rə/⊕/Y/ |
| (SX) | cat :v | cat :v |
| (S) | réitérativité (X'(...)) | réitérativité (Y'(...)) |

où X' = Y', équivalent de traduction

Figure 2: RCL bilingue de réitérativité

Dans ce formalisme, un lexème est représenté par 4 rubriques (2 formelles : sa forme graphique (G) et sa forme phonologique (F), et 2 profondes : la structure syntaxique (SX) et la

[1] Le système a été conçu pour deux langues proches, le français et l'italien, dans le but de pouvoir individualiser les problèmes méthodologiques d'une telle approche. L'extension à d'autres langues est également envisagée.

représentation sémantique (S))[2]. La RCL applique sur une ou plusieurs rubriques du lexème-base (input) une série d'instructions simultanées et indépendantes pour construire un lexème dérivé (output). La règle ci-dessus propose donc de mettre en « équivalence » les RCL monolingues de l'italien et du français (les deux colonnes) pour former une RCL bilingue.

## 4.2 Les règles de formalisation bilingues des adjectifs relationnels préfixés

Comme nous l'avons déjà mentionné, les adjectifs relationnels sont très souvent impliqués dans des règles de préfixation. En outre, le procédé de formation des adjectifs relationnels n'est pas toujours identique dans les deux langues de notre travail. Le modèle présenté ci-dessus doit donc être adapté pour représenter ce type de préfixation et les différentes possibilités de sélection de la base. Mais avant de présenter les RCL bilingues pour les préfixations d'adjectifs relationnels (section 4.2.2), nous présentons une petite étude sur les divergences entre les langues dans la formation de ces adjectifs.

### *4.2.1 Divergences entre les langues*

Dans les deux langues de notre expérience, l'italien et le français, les adjectifs relationnels sont fortement présents dans les procédés de préfixation. De même, il existe, dans les deux langues, la possibilité d'utiliser alternativement les adjectifs relationnels ou leur « équivalent » base nominale (par exemple, en italien : *trattamento antiparassiti* / *trattamento antiparassitario*).

En revanche, la possibilité de former des adjectifs relationnels n'est pas présente de manière égale dans les deux langues. Une évaluation effectuée sur le dictionnaire bilingue italien-français Garzanti (2006) nous a permis d'estimer à plus de 1 000 les adjectifs relationnels de l'italien qui ne disposaient pas d'équivalent en français. Pour tous ces adjectifs, un équivalent sous forme de locution prépositionnelle est à chaque fois proposée, comme le montrent les exemples suivants :

- *aziendale* → *de l'entreprise*
- *creditizio* → *de crédit*
- *gattesco* → *de chat*
- *partitico* → *de parti*
- *congressuale* → *du congrès*

Ainsi, d'un point de vue formel tout comme d'un point de vue traductionnel, la RCL bilingue doit rendre compte de cette divergence, et garantir l'accès à la base nominale et permettre la génération d'une traduction dans la langue qui n'aurait pas d'adjectif relationnel.

---

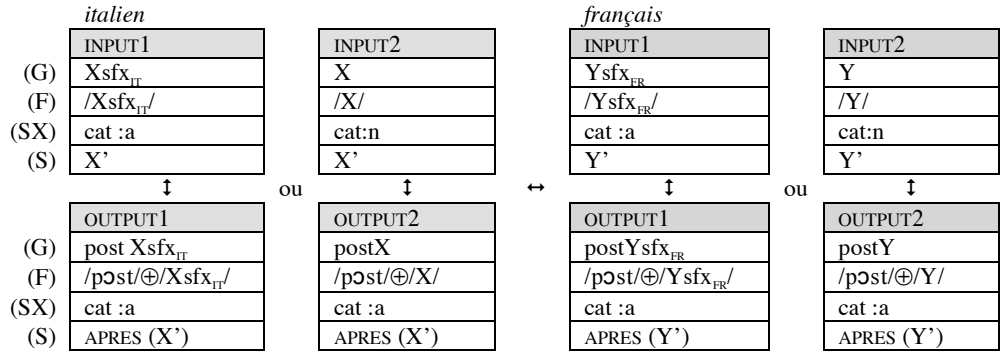
[2] Dans cet exemple, nous décrivons uniquement les informations utilisées dans notre système. Il va de soi que les différentes rubriques du modèle de B Fradin peuvent contenir des informations plus complexes.

### *4.2.2 Des RCL bilingues pour la préfixation des adjectifs relationnels*

Pour formaliser les RCL impliquant un adjectif relationnel, la RCL bilingue doit pouvoir rendre compte de la double possibilité des bases dans les deux langues (bases nominales et bases adjectivales, les deux partageant la même sémantique). Ce phénomène est présent dans un certain nombre de RCL, à savoir :

- Les règles de préfixation quantitative (par exemple en italien : *pluri, poli, tri, uni, mono, multi, bi, di*)
- Les règles de préfixation locative (par exemple en italien : *neo, oltre, para, ex, extra, inter, intra, meta, post, pre, pro, sopra, sovra, sotto, sub, super, trans*)
- Quelques règles de préfixation négative (par exemple, en italien : *a, anti*).

La proposition de règle présentée dans la figure 3 pour la *préfixation temporelle APRÈS*, permet, par exemple, de rendre compte de la double possibilité de base (input). Elle contient, pour chaque langue, deux *input* - l'un adjectival, construit sur une base nominale (par exemple Xsfx pour l'italien), et l'autre nominal (X) qui partagent la même représentation sémantique (X'). La représentation sémantique de l'output est également identique, quelque soit sa représentation formelle (APRES (X) pour les formes post$Xsfx_{IT}$ et postX).




où X'=Y', équivalents de traduction

où $sfx_{IT}$ et $sfx_{IT}$ représentent l'ensemble des suffixes permettant de former des adjectifs relationnels sur des bases nominales.

Figure 3: RCL bilingue de position temporelle APRÈS

Une telle règle permet donc de passer d'une forme italienne de type pfx$Xsfx_{IT}$ (comme *post-adolescienzale*) à une forme française pfxX (*post-adolescence*), les deux formes étant construites sémantiquement sur la base nominale X. Une fois formalisées, ces règles sont implémentées dans notre prototype de système de TA, mais le mécanisme de traduction doit encore être adapté en conséquence.

## 4.3 Mécanisme de traduction

Pour tout néologisme préfixé impliquant un adjectif relationnel, les règles de traduction tirent profit du lexique « étendu », notamment quand l'adjectif relationnel n'existe pas dans l'une des deux langues. La figure 4 ci-dessous décrit le traitement des adjectifs relationnels préfixés dans notre système de TA. Quand un néologisme construit arrive dans le système (ici *anticostituzionale*), il est analysé par la partie « analyse » de la règle, et la base formelle (c.-à-d. l'adjectif) est recherchée dans le lexique bilingue (Biling_lex - étape 1). Si cette base est présente dans ce lexique, le néologisme est directement généré en français. Dans le cas

contraire, le système recherche la base nominale de l'adjectif dans le lexique italien étendu (*costituzione* - étape 2). Si celle-ci est retrouvée, la base nominale peut être recherchée dans le lexique bilingue (étape 3). Ensuite, deux options sont possibles : soit la traduction est générée en reconstruisant un adjectif sur une base nominale (étape 4 : *anticonstitution*), soit l'adjectif relationnel français est retrouvé à partir de la base nominale dans le lexique français (étape 5 : *constitution* → *constitutionnel*), et un néologisme est généré en français (étape 6 : *anticonstitutionnel*).

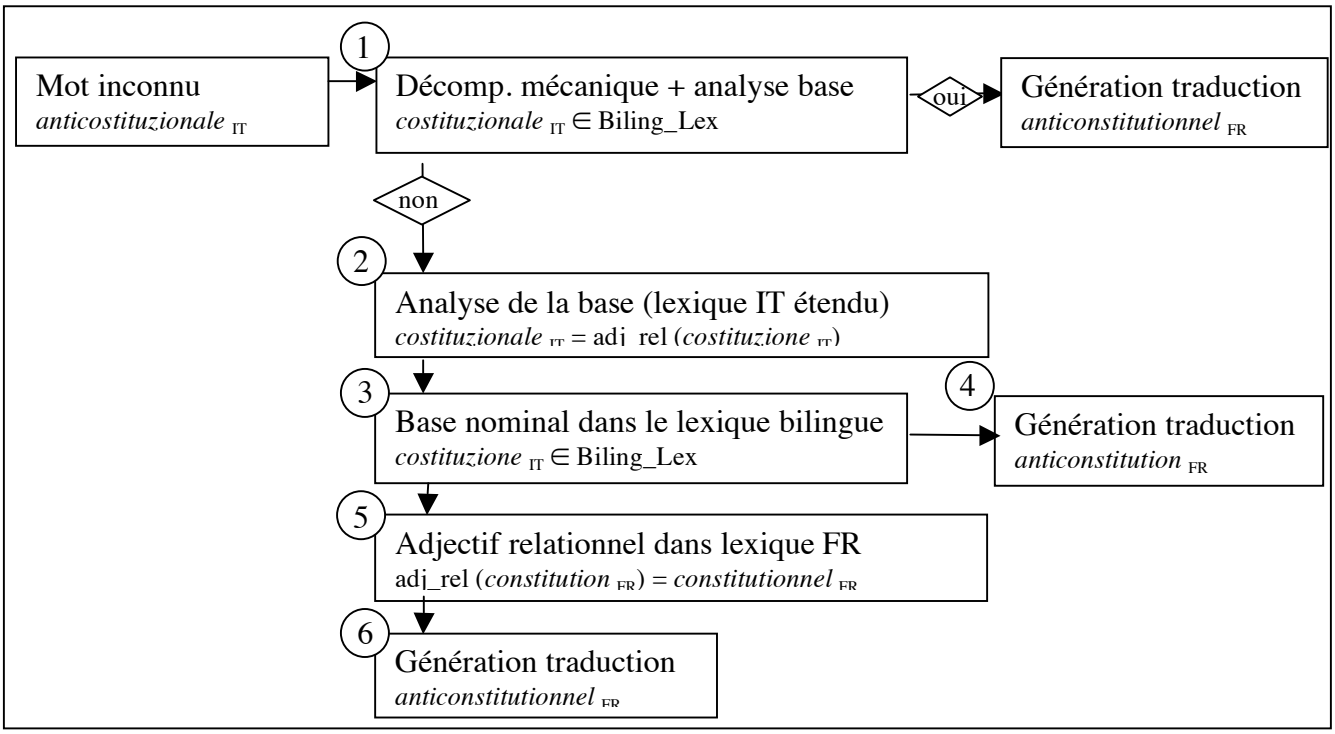



Figure 4: Mécanisme de traduction des adjectifs relationnels préfixés

Dans certain cas, le lexique étendu permet de proposer une traduction sous la forme d'un lexème préfixé sur une base nominale quand l'adjectif relationnel est possible en français, mais, pour d'autres raisons, est absent du dictionnaire bilingue. Ainsi, le néologisme italien *antileucemico* est construit sur l'adjectif relationnel *leucemico* qui dérive du nom *leucemia*. Le dictionnaire bilingue ne contenait pas d'entrée pour *leucemico,* mais seulement pour le nom (*leucemia=leucémie*) et grâce au lexique étendu et aux nouvelles informations qu'il fournit, le système a pu au moins générer l'équivalent *antileucémie* construit sur la base nominale.

# 5 Évaluation

Pour évaluer plus globalement ce système, nous avons extrait du corpus italien *La Reppublica* (Baroni, Bernardini, et al. 2009) un ensemble de 24 247 mots inconnus qui étaient potentiellement des néologismes prefixés. Le système de traduction « avant extension des lexiques » traduisait 17 034 néologismes (68,7%). Parmi eux, 5 025 étaient construits avec un préfixe qui peut potentiellement entrer dans une préfixation d'une base adjectif relationnel (cf. liste section 4.2.2). Parmi ces derniers, le lexique étendu a été capable d'identifier 1 783

adjectifs relationnels, ce qui constitue déjà en soi une amélioration significative en termes de qualité de l'analyse.

Par exemple, grâce au lexique étendu, l'analyse propose désormais une décomposition mécanique telle que :

- *multidisciplinare* → *multi#disciplinare*$_A$/*disciplina*$_N$
- *sottoministeriali* → *sotto#ministeriali*$_A$/*ministero*$_N$
- *antidemocratico* → *anti#democratico*$_A$/*democrazia*$_N$

Du point de vue de la génération, tous les néologismes ont été traduits. Une majorité d'entre eux (1 570) par un adjectif relationnel préfixé et le reste (213) par une construction sur une base nominale en français, soit parce que l'adjectif n'était pas dans le dictionnaire bilingue, soit parce qu'il n'est tout simplement pas possible en français.

Parmi ce dernier groupe, il est intéressant de noter les exemples suivants, qui soulignent bien les divergences de construction entre les deux langues :

- *precongressuale* → *précongrès*
- *post-transfuzionale* → *post-transfusion*
- *predibatimentale* → *prédébat*

Ainsi, l'extension de nos lexiques avec l'information sur les adjectifs relationnels présente deux avantages. Premièrement, les adjectifs relationnels sont mieux analysés, et deuxièmement, quand la base n'est pas dans le lexique bilingue, cette analyse plus « profonde » permet de produire une traduction tout à fait satisfaisante.

# 6 Discussion et perspectives

L'application présentée ici n'est sans doute pas la plus prototypique des utilisations du TALN, mais cette expérience a le mérite de montrer que les adjectifs relationnels sont à l'origine de nombreuses divergences entre les langues, même proches historiquement. De plus, nous avons vu qu'un traitement formel ces procédés de construction permet d'améliorer non seulement la qualité des lexiques, mais leur robustesse dans un contexte multilingue. A ce titre, l'information acquise s'avère essentielle pour traiter un nombre important de néologismes construits en TA.

Par ailleurs, l'exploitation de cette information pour la traduction des adjectifs relationnels préfixés a permis de confirmer l'ampleur de la divergence, et a apporté une certaine validation à la proposition de B. Fradin de définir les adjectifs relationnels comme une simple forme supplétive de leur noyau nominal. Cette expérience a également montré que le traitement des adjectifs relationnels par une formalisation du lien qu'ils entretiennent avec leur base nominale est une solution très intéressante, qui mériterait d'être implémentée de manière plus fine et plus systématique dans tous les lexiques informatisés. Ce lien fournit des informations morphosémantiques régulières non négligeables pour un traitement « intelligent » des langues naturelles.

Mes remerciements vont à ma collègue Selja Seppällä pour sa relecture attentive et ses conseils, ainsi qu'aux relecteurs anonymes pour leurs précieux commentaires. Evidemment, les erreurs qui devraient rester dans la présente version sont miennes.

## Références

(2006) *Garzanti francese : francese-italiano, italiano-francese*. I grandi dizionari Garzanti. Milano, Garzanti Linguistica.

BARONI, M., S. BERNARDINI, F. COMASTRI, L. PICCIONI, A. VOLPI, G. ASTON et M. MAZZOLENI (2004) Introducing the « La Repubblica » corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. Actes de *LREC 2004*, Lisbon : 1771-1774.

BOUILLON P., S. LEHMANN, S. MANZI et D. PETITPIERRE (1998) Développement de lexiques à grande échelle. Actes de *Colloque des journées LTT de TUNIS*, Tunis : 71-80.

CARTONI, B. (2005) Traduction de règles de construction des mots pour résoudre les problèmes d'incomplétude lexicale en traduction automatique Étude de cas. Actes de *RECITAL 2005*, Dourdan, Atala : 565-574.

CARTONI, B. (à paraître) Lexical Morphology in Machine Translation: a Feasibility Study. Actes de *EACL 2009*, Athènes

DAILLE, B. (1999) Identification des adjectifs relationnels en corpus. Actes de *TALN 1999*, Cargèse :105-114.

FRADIN, B. (2003) *Nouvelles approches en morphologie*. Paris, Puf.

FRADIN, B. (2008) On the semantics of Denominal Adjectives. Actes de *6th Mediterranean Morphology Meeting*, Ithaca

GDANIEC, C. et E. MANANDISE (2000) Using word formation rules to extend MT lexicons. Actes de *5th conference of the Association for Machine Translation in the Americas, AMTA 2000*, Tiburon, CA, USA, Springer Verlag : 64-73.

MELIS-PUCHULU A. (1991) Les adjectifs dénominaux: des adjectifs de « relation ». Lexique 10 : 33-60.

NAMER F. (2005) Morphosémantique pour l'appariement de termes dans le vocabulaire médical: approche multilingue. Actes de *TALN 2005* Dourdan : 63-72.

WANDRUSZKA U. (2004) Derivazione aggettivale. *La Formazione delle Parole in Italiano*. M. Grossman et F. Rainer(éds). Tübingen, Niemeyer.