

Étude lexicographique des noms de professions

Sandrine Fuentes

Universidad Autònoma de Barcelona

FLexSem

0. Introduction

Le présent article rend compte de recherches menées dans le cadre d'une thèse doctorale, au sein du groupe de recherche en Linguistique Appliquée aux Langues Romanes de l'Université Autonome de Barcelone, et plus exactement de la section *Lexicologie contrastive et Sémantique formalisée (FLexSem¹)*. Les études réalisées par cette section se centrent principalement sur la lexicologie contrastive à travers l'élaboration de grammaires locales et de dictionnaires électroniques monolingues coordonnés. Ces outils sont à la base d'applications informatiques en rapport aussi bien avec la génération qu'avec la reconnaissance automatique de textes. Ils constituent également un support important pour des recherches en lexicologie théorique.

Dans un premier temps, nous évoquerons le cadre théorique sur lequel repose notre travail, ensuite nous passerons à l'étude lexicographique des noms de professions et plus précisément à la présentation de notre dictionnaire: *DicPro* («Dictionnaire électronique des noms de professions et métiers»)². Pour terminer, nous offrirons une ébauche de l'étude lexicologique qui prend appui sur ce dictionnaire. Nous étudierons tout d'abord l'environnement phrastique dans lequel se trouvent impliqués les noms de professions (<Nprof>), et nous terminerons en montrant de quelle manière il est possible de créer des sous-classes de <Nprof> à partir de critères syntactico-sémantiques.

¹ <http://www.oaid.uab.es/lexsem>

² Précisons que cette étude a reçu le support financier de l'AUF, et plus précisément de son réseau *Lexicologie, Terminologie et Traduction* («Dictionnaire électronique français-espagnol-catalan-arabe des noms des professions et métiers»), ainsi que de l'AECE (projet de coopération hispano-tunisienne: «Estudio lexicológico de los nombres de profesión»)

1. Cadre théorique

La perspective théorique que nous avons adoptée se base en grande partie sur la méthodologie du *Laboratoire de Linguistique Informatique* (CNRS, Paris 13) pour qui le concept de *classes d’objets* (cf. Gross, 1999) est fondamental. Nous reviendrons en détail sur ce concept dans cet article, et nous nous contenterons pour le moment d’en donner une définition: «il s’agit de classes sémantiques [homogènes] élaborées à partir de critères syntaxiques [communs]» (Le Pesant et Mathieu-Colas, 1998). Nous nous inspirons également de la théorie du lexique-grammaire telle qu’elle a été conçue par M. Gross. Ce modèle théorique part du principe que l’unité minimale d’étude doit être la phrase simple, autrement dit que le mot doit être décrit non pas de manière isolée mais dans son environnement phrastique. D’autre part, la phrase simple doit être composée au minimum d’un prédicat, autour duquel gravitent d’éventuels arguments. Ainsi, la phrase *Luc aime Léa* est constituée d’un prédicat, le verbe *aimer*, qui sélectionne deux arguments élémentaires *Luc* et *Léa*.

Un prédicat peut prendre la forme d’un verbe, comme c’est le cas ici, mais aussi d’un nom ou d’un adjectif. Parfois, le même prédicat peut apparaître sous les trois formes; c’est le cas du prédicat *aimer*. En effet, dans *Luc est amoureux de Léa* et *Luc éprouve de l’amour pour Léa*, les prédicats sont l’adjectif *amoureux* et le nom *amour* respectivement. Cependant, les trois phrases sont paraphrastiques.

À la différence des prédicats verbaux, les prédicats adjectivaux et nominaux ont besoin d’un verbe sémantiquement vide, appelé *verbe support*, pour être actualisés. Dans notre exemple, ce sont l’auxiliaire *être* (support approprié aux prédicats d’états) et le verbe *éprouver* (approprié aux noms de sentiments) qui jouent ce rôle. Les verbes supports servent à transmettre des informations sur le temps, la personne etc. mais peuvent également véhiculer des valeurs aspectuelles comme par exemple l’inchoatif dans *tomber amoureux*.

Certains prédicats imposent des contraintes de sélection plus strictes que d’autres à leurs arguments. Ainsi, le prédicat *adorer* n’impose quasiment aucune restriction à l’argument qui se trouve en position de complément: *adorer (la lune + la télévision + le sport + le chocolat + les enfants + Paris)*. Quant à lui, le prédicat verbal *converser* impose des contraintes fortes à l’argument sujet, qui ne peut être qu’un être humain. Le trait ‘humain’, tout comme les traits ‘animal’, ‘végétal’, ‘concret’, ‘abstrait’, ‘locatif’ etc., sont des traits syntactico-sémantiques auxquels on a recours pour décrire l’ensemble du lexique. Pour revenir à notre exemple *converser*, nous pouvons observer que, à moins d’être intégrées dans un type de discours particulier (fantastique, onirique etc.), les phrases suivantes semblent peu naturelles: *³(*la table + le mur + le géranium + le lézard + le ciel...*) *converse avec Luc*.

³ El símbolo *indica que se trata de una frase gramatical y semánticamente inaceptable, el signo? significa que la frase es dudosa y los dos ?? que la frase es muy dudosa.

Le trait syntactico-sémantique ‘humain’ a fait l’objet d’une étude de la part de Gaston Gross (1995c), sur laquelle nous nous sommes basée pour l’espagnol (Fuentes, 2004). C’est à la suite de cette étude que nous nous sommes intéressée à la classe des noms de professions. La classe d’objets <Nprof> est en effet une des 54 classes que nous avons dégagées pour ce trait syntactico-sémantique.

2. Étude lexicographique des noms de professions

2.1. L’unité lexicale

La base de notre travail repose sur un dictionnaire électronique dans lequel sont recensés tous les noms appartenant à la classe <Nprof>. Nous allons présenter dans un premier temps la macrostructure de ce dictionnaire, autrement dit sa nomenclature.

Décrire la macrostructure d’un dictionnaire revient à décrire la méthode ainsi que les critères qui justifient la présence ou non d’un terme en tant que lemme. En ce qui nous concerne, chaque entrée du dictionnaire doit correspondre à une seule et unique *unité lexicale*. Rappelons qu’une unité lexicale se définit par une *forme*, un ‘sens’ et une *combinatoire* (Σ) (Mel’čuk, 1995). Ce principe nous oblige à ce que toutes les formes polysémiques soient automatiquement dédoublées. Ainsi, les noms *archivista* et *archivero* ont une entrée indépendante dans la mesure où, même si le sens qu’ils véhiculent et la combinatoire qu’ils impliquent sont identiques, leur forme est différente. De la même manière, la forme *portero* apparaîtra plusieurs fois dans notre dictionnaire dans la mesure où nous sommes en présence de plusieurs sens et de combinaisons distinctes. Pour ce qui est du sens, voici deux des acceptions que l’on peut trouver dans le DRAE⁴:

- (1) Persona que, en las casas de vecinos, tiene a su cargo el guardar, cerrar y abrir el portal y vigilar la entrada y salida de personas, limpiar la entrada, escalera, etc.
- (2) Jugador que en algunos deportes defiende la portería de su bando.

Il est important de préciser tout d’abord que *portero* (4) entre dans la définition de ce que nous considérons comme <Nprof> dans la mesure où il peut se combiner avec des opérateurs appropriés aux noms de profession comme nous le voyons dans les exemples qui suivent. Afin d’éviter les ambiguïtés causées par la polysémie espagnole, nous utiliserons son équivalent de traduction français *gardien de but* :

En 1977, il arrive à Moissej avec toute sa famille, l’Association Sportive de Moissej l’ayant embauché comme gardien de but dans son équipe de football

⁴ Diccionario de la Real Academia Española en ligne (www.rae.es).

*Père dévoué et philanthrope, Kolzig connaît aussi une **carrière prestigieuse comme gardien de but***

Professionnel à 19 ans comme gardien de but à Montpellier

*Nom : Tim Howard; **Profession: gardien de but** à Manchester United*

*[...] tout comme Jean-Loup Leplat, **gardien de but de profession**, qui s'est lui aussi reconverti dans la formation des jeunes*

En ce qui concerne la combinatoire de ces deux unités lexicales, on constate qu'elle n'est pas non plus la même. En effet, seule la première (*portero* = sportif) peut se combiner avec des adjectifs comme *professional* ou *amateur*:

Il travaille comme gardien de but (professionnel + amateur)

**Il travaille comme concierge (professionnel + amateur)*

Il va sans dire que le phénomène de la polysémie ne se répète pas obligatoirement dans la langue d'arrivée. Comme nous venons de le voir, le français a recours à deux lexèmes différents pour rendre les deux acceptions de *portero*: *concierge* et *gardien de but*. C'est une des raisons pour lesquelles il est important de dédoubler les entrées et de donner à chacune des unités lexicales la description syntactico-sémantique qui lui correspond.

2.2. Les sources documentaires

L'élaboration d'un dictionnaire comme DicPro implique l'établissement d'une liste la plus exhaustive possible de noms de *professions* ou, pour être plus exacte, de noms de *professionnels*. C'est ce qui va constituer le corpus du dictionnaire. Pour ce faire, notre recherche s'est centrée sur les organismes officiels consacrés à la recherche d'un emploi. Nos langues de travail étant l'espagnol et le français, nous nous sommes servis principalement de la classification de l'Agence Nationale Pour l'Emploi française (le ROME) et de celle de son équivalent espagnol, l'Institut National pour l'Emploi (la CNO-94). Cependant, la *Classification Nationale des Professions* proposée par le *Ministère des Ressources Humaines et du Développement des Compétences* du Canada, qui regroupe plus de 20 000 dénominations, nous a également été fort utile, notamment au niveau du recensement de professionnels liés aux domaines sportif, artistique et religieux.

2.3. Le dictionnaire DicPro et sa microstructure

Après avoir recensé tous les noms de professions existant dans les deux langues, il s'agit ensuite de transformer cette liste en un dictionnaire électronique. Pour ce faire, il s'agit d'introduire dans la base de données des champs d'informations lexicographiques.

Ces champs lexicographiques sont de différents types: certains nous donnent des informations de nature morphologique, d'autres de type sémantique et d'autres décrivent les propriétés syntaxiques du lemme. Signalons au passage que notre dictionnaire se compose actuellement d'environ 4000 formes simples telles que *acrobate, boulanger cardiologue, électricien, philologue, webmaster* etc., et de plus de 40 000 formes composées comme *agent de maintenance, aide-ambulancier, danseur classique, conducteur d'autobus, directeur général d'entreprise de transport* etc.

Nous présentons ci-dessous deux lemmes, l'un simple et l'autre composé, tels qu'ils sont décrits dans notre dictionnaire:

Tableau 1: Description de deux noms de professions (simple et composé)

Lemme	médico	agente de policía nacional
Code INEM	2121.001.1	5221.002.1
Catég. Gram.	nm	nm
Trait syntactico-Sém.	Hum	Hum
Classe d'objets	<professions>	<professions>
Domaine	médecine	société
Sous-Domaine	médecine générale	police
Synonymes	doctor, facultativo (antonomase) matasanos, medicastro (péj.)	agente de policía, agente poli (fam.)
Français	médecin	agent de police nationale
Code ANPE	31111	11311
Code CNP	3112	6261
Morphologie	-	ND(NA)

Le champ G – catégorie grammaticale du lemme – correspond toujours à un nom (masculin ou féminin) dans le dictionnaire-noyau, le champ T – trait syntactico-sémantique – correspond toujours à un Hum (humain) et le champ C – classe d'objets – correspond toujours à <professions>. Le fait de disposer de ces données n'est pas du tout trivial si l'on considère que la description réalisée dans le cadre de **DicPro** peut s'intégrer de façon immédiate dans le système de dictionnaires du LLI ou de notre groupe de recherches LexSem. En effet, ces informations deviennent alors différentielles et, donc, nécessaires.

Les champs D (*domaines*) et SD (*sous-domaines*) sont des champs de nature sémantique. On y spécifie, sur la base des 90 domaines et 750 sous-domaines dégagés par Buvet et Mathieu-Colas (1999), le domaine de spécialité auquel se rattache le lemme. Pour en revenir à notre exemple de départ *portero*, dans une de ces acceptions il sera mis en rapport avec le domaine ‘services’ et le sous-domaine ‘entretien et maintenance’, et dans l’autre avec le domaine ‘sport’ et le ‘sous-domaine ‘football’. Ce genre d’informations est important dans le cas des noms de professions car ils englobent un grand nombre de spécialités différentes. D’autre part, les indications de D et SD permettent d’effectuer des recherches précises et d’identifier rapidement les unités lexicales liées à un domaine donné. Il est possible d’établir par exemple une liste de toutes les professions liées au domaine du bois, de l’enseignement universitaire ou de l’hydraulique. De plus, comme nous l’avons vu avec *portero*, elles nous servent souvent à la désambiguïsation de certaines unités lexicales.

Pour ce qui est des codes qui apparaissent dans le tableau, il s’agit des codes officiels utilisés par l’ANPE, l’INEM et la CNP. Cette information sera réservée à l’une des possibles applications de ce dictionnaire: la recherche d’un emploi sur le Web. Grâce à ces codes, il est en effet possible de mettre en correspondance des professions et ce, dans des pays différents.

Le champ M, où est indiqué le type morphologique du lemme, est réservé aux noms composés, c’est-à-dire aux lemmes comportant des caractères séparateurs. Pour les noms de profession, nous avons dégagé environ 20 types de composition, allant des plus ‘simples’ comme *afinador de pianos* (NDN), *guardia municipal* (NA), *panadero- pastelero* (NN), aux plus ‘complexes’⁵: *gestor comercial de servicios financieros* (NA)D(NA), *ingeniero técnico en confección industrial* (NA)P(NA), *abastecedor de la cadena de envasado en el calzado* ((ND(NDN))PN).

Pour finir, il est important de signaler que **DicPro** n’est pas un dictionnaire bilingue espagnol-français mais un système de dictionnaires monolingues coordonnés à travers le champ ‘équivalents de traduction’. Ainsi, la traduction que nous proposons dans le champ FR du dictionnaire espagnol renvoie au dictionnaire monolingue français dans lequel le lemme proposé aura sa propre description linguistique.

Pour que ce dictionnaire électronique puisse être implémenté dans des systèmes de traitement automatique des langues, il est indispensable que tous les lemmes soient dotés des informations que nous venons de mentionner.

⁵ Les termes *simple* et *complexe* figurent entre guillemets puisque, de fait, tous les lemmes de notre dictionnaire sont des noms composés ‘complexes’.

3. Grammaires locales des noms de professions

Comme nous l'avons expliqué auparavant, le point commun entre tous les noms que nous avons recensés dans le dictionnaire est qu'ils font tous partie de la classe d'objets <Nprof>. Cependant, comment justifie-t-on linguistiquement la création d'une classe de ce type?

Tout d'abord, élaborer une classe d'objets revient à construire une grammaire locale autour des éléments qui constituent la classe. C'est-à-dire qu'il ne s'agit pas simplement de recenser les unités lexicales qu'implique le nom de la classe mais aussi d'indiquer leur syntaxe. Cette grammaire locale peut être représentée formellement par des automates à états finis (sous forme de graphes) puis implémentés par la suite dans des systèmes de traitement automatique des langues. Nous allons donc analyser par la suite dans quels environnements phrastiques apparaissent les noms de professions.

En premier lieu, les noms de professions sont susceptibles d'apparaître dans la réponse à la question: «Qu'est-ce que tu fais dans la vie?» et cette réponse peut être formulée de plusieurs manières comme nous le voyons dans les constructions détaillées ci-dessous:

Nhum être Ø Nprof⁶

Luc est (technicien + standardiste + chef comptable)

Être Nprof (de profession + de métier)

Luc est (technicien + standardiste + chef comptable) (de profession + de métier)

Être Nprof dans la vie

Luc est (technicien + standardiste + chef comptable) dans la vie

Nhum (travailler + exercer + gagner sa vie) comme Nprof

Luc (travaille + exerce + gagne sa vie) comme (technicien + standardiste + chef comptable ...)

Ensuite et comme nous le voyons dans les exemples ci-dessous, les noms de professions peuvent figurer comme compléments après l'expression «*embrasser (la profession + le métier + la carrière) de*»:

*De nombreux commissaires de l'air ont ainsi **embrassé la carrière de magistrat, d'auditeur à la cour des comptes, de conseiller de tribunal administratif...***

⁶ Les noms de professions ont la particularité de pouvoir être introduits sans article après le verbe copule *être*, ce qui n'est pas le cas des noms prototypiques.

*Après avoir voyagé à travers toute l'Italie, le jeune homme **embrasse la profession de médecin***

*Il a pour cela **embrassé le métier de charretier** et le fait sans complexe
[source: Google]*

Ensuite, on les trouve dans des phrases relatives à la recherche d'un emploi:

Nhum (chercher + rechercher + postuler à + aspirer à + solliciter + briguer + viser) (un emploi + poste) de Nprof

*Je **recherche** un emploi de vendeur à domicile*

Ou à l'obtention d'un emploi:

Nhum (décrocher + obtenir + gagner) un (poste + emploi) de Nprof

*Grâce à son oncle allemand, il **décroche** un emploi de contrôleur des wagons-lits dans la compagnie ferroviaire Zentropa*

En ce qui concerne l'embauche et le licenciement, des opérateurs appropriés peuvent également être recensés. Notons que dans ces deux cas, le sujet éventuel peut être un humain mais aussi un humain collectif. Il peut s'agir par exemple d'une entreprise ou d'une institution:

(Nhum + Nhum_coll) (embaucher + employer + engager + recruter) Nhum comme Nprof

Cette entreprise a (embauché + employé + engagé + recruté) Luc comme (technicien + standardiste + chef comptable)

Du point de vue contrastif, il est intéressant de remarquer que le verbe *recruter* a un sens beaucoup plus large que son équivalent espagnol *reclutar* qui lui, a gardé son sens premier et ne s'utilise que dans le domaine militaire :

El ejército ha reclutado a 1000 soldados para resolver este conflicto

**El director ha reclutado a (una nueva secretaria + un nuevo responsable de ventas)*

Le directeur a recruté (une nouvelle secrétaire + un nouveau responsable des ventes)

Il est évident que des variantes diastématiques peuvent entrer en jeu. Ainsi, les opérateurs appropriés relevés pour le licenciement appartiennent à des registres différents:

(Nhum + Nhum_coll) (renvoyer + licencier + virer + mettre à la porte + congédier) Nprof

Le directeur a (renvoyé + licencié + viré + mis à la porte + congédié) le (technicien + standardiste + chef comptable)

Pour ce qui est des conditions d'emploi, les noms de professions peuvent se combiner avec des expressions adverbiales du type *à temps partiel, à plein temps, à mi temps...* ou avec des adjectifs comme *intérimaire, contractuel, vacataire, stagiaire* etc. ou bien *salarié, free-lance, autonome*. Les exemples qui suivent le montrent:

Léa est secrétaire médicale à (temps partiel + plein temps + mi-temps)

Luc est professeur (intérimaire + contractuel + vacataire + stagiaire)

Luc est traducteur (salarié + free-lance + autonome)

Comme nous l'avons vu au début de cet article, pour les professions du domaine sportif ou artistique, les adjectifs *professionnel* ou *amateur* sont des compléments possibles:

Luc est (footballeur + guitariste + danseur) (professionnel + amateur)

Le type de rémunérations est également un élément à prendre en compte lorsque l'on décrit l'environnement phrastique des professions. La construction la plus commune serait:

Nprof (toucher + percevoir + gagner) un <salaire> de X

dans laquelle <salaire> représente à son tour une classe d'objets incluant des noms comme: *appointement, émoluments, traitement, honoraires, cachet, vacation, gages, solde, salaire, paye, commissions...*

Si nous observons cet extrait de Marie-Éva de Villers, nous constatons à quel point cette information est importante pour des applications informatiques comme la reconnaissance ou la génération automatique de textes:

Les avocats, les médecins, les consultants **touchent des honoraires**; les représentants, les agents, les courtiers reçoivent des **commissions**. Les comédiens, les musiciens, les artistes touchent des **cachets**, les serveurs, des **pourboires**, les enseignants, les cadres, les fonctionnaires, un **traitement**. Les mères de famille ne vont pas loin avec leurs **allocations familiales** tandis que les domestiques recevaient des **gages** (le terme n'est plus usité). Les administrateurs se contentent de **jetons de présence**, les écrivains, les compositeurs encaissent parfois des **droits d'auteur**. Pour

leur part, les militaires touchent une **solde**. À ce propos, les **congés sans solde** sont réservés aux soldats; tous les autres travailleurs peuvent bénéficier de **congés non payés**.

En effet, si dans les descriptions il est précisé que le type de salaire que les personnes exerçant une profession libérale perçoivent, porte le nom d'*honoraires* et non pas de *cachet*, l'ordinateur sera capable de générer sans aucune difficulté l'énoncé:

Cet avocat a touché 10 000 € d'honoraires pour défendre ce client

D'autre part, l'ordinateur saura reconnaître que *cachet* équivaut à 'argent' et pas à 'médicament', c'est-à-dire sera à même de résoudre l'ambiguïté existante, si dans l'entourage phrastique du lexème il trouve des verbes comme 'percevoir', 'toucher' ou 'gagner', et si ce mot est étiqueté correctement dans le dictionnaire électronique comme membre de la classe <salaire>.

Enfin, le dernier aspect, mais non le moindre en ce qui concerne cette recherche, est le fait que derrière ces noms de rétributions se cache obligatoirement un nom de profession. Ainsi, pour paraphraser les noms de la classe <salaire>, il est possible d'utiliser une glose explicative comme: c'est le *salaire* de <Nprof>, dans laquelle le nom de profession apparaît de manière explicite.

Pour finir, il nous reste à mentionner le lieu de travail qui, même s'il n'est pas toujours explicité dans la phrase, fait également partie de l'environnement phrastique des noms de professions:

Luc travaille comme technicien chez Fagor

Léa travaille comme standardiste à l'hôpital

Luc travaille comme comptable dans une entreprise de construction

Le schéma correspondant à ces énoncés serait:

Nhum travailler comme Nprof (à + chez + dans) Nloc

où Nloc renverrait à l'une des classes de locatifs suivantes: <entreprise>, <administration>, <établissement>, <usine>, <atelier>, <cabinet>, etc. Ces classes, à leur tour, renvoient à des sous-classes. Ainsi, la classe <établissement> renvoie aux sous-classes: <établissement d'enseignement>, <établissement hospitalier>, <établissement de spectacle>, etc.

D'après Denis Le Pesant (2000), « les noms d'établissements sont corrélés d'une part à des noms d'humains *usagers* et d'autre part, à des noms d'humains *professionnels* ». Ainsi, si nous prenons comme exemple le domaine de l'enseignement, nous pouvons arriver à des observations comme celles qui suivent:

	Université	École primaire
usagers	<i>élève</i> <i>*écolier</i> <i>Étudiant</i>	<i>élève</i> <i>écolier</i> <i>*étudiant</i>
professionnels	<i>enseignant</i> <i>*professeur des écoles</i> <i>maître de conférence</i> <i>*directeur</i> <i>président</i> <i>*instituteur</i> <i>professeur de <discipline scolaire, professionnelle ou artistique></i>	<i>enseignant</i> <i>professeur des écoles</i> <i>*maître de conférence</i> <i>directeur</i> <i>*président</i> <i>instituteur</i> <i>?professeur de <discipline scolaire, professionnelle ou artistique></i>

Il ressort par exemple de ces observations que, même si appartenant à des domaines très proches, les noms *professeur des écoles* et *maître de conférences* ne pourront pas être traités de la même manière dans notre dictionnaire. Le lieu d'enseignement sera en effet un critère distinctif à prendre en compte.

Finalement, on constate que très souvent le lieu est utilisé comme une paraphrase approximative du nom de profession. Ainsi, il est possible d'établir les correspondances suivantes:

Je travaille dans un bar (= je suis serveur/se)

Je travaille dans une boutique (= je suis vendeur/se)

Je travaille dans un garage (= je suis mécanicien)

Toutes les descriptions que nous venons d'exposer nous conduisent à une liste d'opérateurs appropriés aux noms de professions; certains sont des opérateurs appropriés généraux, c'est-à-dire s'appliquant à tous les éléments de la classe <Nprof>, et d'autres sont des opérateurs spécifiques. Ce sont ces derniers auxquels nous aurons recours pour procéder à l'établissement de sous-classes d'objets. Ce sous-classement est nécessaire dans la mesure où, comme nous allons le voir, certains noms de professions se distinguent clairement des autres de par leur comportement syntactico-sémantique.

4. Les sous-classes d’objets <Nprof: spécialiste> et <Nprof: fonctionnaire>

Une des sous-classes que nous avons dégagées est <Nprof: spécialiste>, qui contient des noms comme *biologiste*, *archéologue*, *linguiste*, *informaticien* etc. Ces derniers se distinguent des autres noms de professions pour les raisons qui apparaissent dans les paraphrases suivantes:

Luc est (biologiste + archéologue + linguiste + informaticien)

Luc est spécialiste en (biologie + archéologie + linguistique + informatique)

Luc est licencié en (biologie + archéologie + linguistique + informatique)

Luc a fait des études de (biologie + archéologie + linguistique + informatique)

Luc est (boulangier + assureur + coiffeur + plombier)

??Luc est spécialiste en (boulangerie + assurances + coiffure + plomberie)

**Luc est licencié en (boulangerie + assurances + coiffure + plomberie)*

??Luc a fait des études de (boulangerie + assurances + coiffure + plomberie)

Luc a suivi une formation en (boulangerie + assurances + coiffure + plomberie)

La sous-classe <Nprof: fonctionnaire> quant à elle, renferme des noms comme *professeur*, *policier*, *douanier*, *greffier*. Le premier critère linguistique pour lequel ces noms ont été regroupés est qu’ils peuvent tous se combiner aussi bien avec l’adjectif *titulaire* qu’avec l’adjectif *stagiaire*, ce qui n’est pas le cas d’autres professions:

Luc est professeur (stagiaire + titulaire)

*Luc est électricien (stagiaire + *titulaire)*

D’autre part, un complément adverbial du type *en détachement* ou *en disponibilité* peut leur être attaché:

Luc est policier (en détachement + en disponibilité)

**Luc est fleuriste (en détachement + en disponibilité)*

D’autre part, les noms de la classe <Nprof: fonctionnaire> entrent tous dans la construction: <Nprof: fonctionnaire> être (radié + révoqué + limogé), ce que nous voyons à travers ces exemples recueillis sur le Web:

*Un magistrat **radié** pour pédophilie*

*Un magistrat **révoqué** pour actes pédophiles*

*Le magistrat **limogé** avait à plusieurs reprises proféré des menaces et des injures*

En revanche, d'autres noms de professions comme *éleveur*, *ouvrier* ou *vendeur* n'acceptent pas cette construction:

**Un (éleveur + ouvrier + vendeur...) (radié + révoqué + limogé) pour actes pédophiles*

Ces critères sont ceux que nous utiliserons pour justifier la création d'une sous-classe de ce type.

5. Conclusion et perspectives

Notre but était d'une part de montrer la démarche que nous avons suivie pour élaborer le dictionnaire électronique des noms de professions *DicPro*, et d'autre part de mettre en valeur l'importance des descriptions linguistiques pour le traitement automatique des langues. Décrire l'ensemble de la classe <Nprof> revient à construire toutes les grammaires locales qui sont liées aux noms de professions, puis à dégager des sous-classes pertinentes et basées sur des critères linguistiques le plus fiables possibles. Notre but dans l'avenir est de parvenir à implémenter notre dictionnaire et les grammaires locales qui l'accompagnent dans des systèmes informatiques afin qu'ils soient utilisés dans des applications comme la recherche d'un emploi sur le Web ou la reconnaissance des entités nommées dans un texte.

Bibliographie

- BLANCO, X. y LAJMI, D. (2004). “Dictionnaire électronique français-espagnol-catalan-arabe des noms de professions et métiers”. *Actes des Premières Journées Scientifiques des Réseaux de Chercheurs de l’AUF*. Agence Universitaire de la Francophonie.
- BUVET, P-A. et MATHIEU-COLAS, M. (1999). “Les champs *Domaine et Sous-Domaine* dans les dictionnaires électroniques”. *Cahiers de Lexicologie* 75. Paris: INaLF.
- FUENTES, S. (2006). “Les dictionnaires monolingues coordonnés DicPro”. *Les noms de professions. Approches linguistiques, contrastives et appliquées*. Servei de Publicacions : Universitat Autònoma de Barcelona.
- LE PESANT, D. (2000) *Six études de sémantique lexicale sur les noms communs de lieux*. Ouvrage rédigé en vue d’obtenir l’Habilitation à Diriger des Recherches. Villetaneuse: Université Paris 13.
- LE PESANT, D. et MATHIEU-COLAS, M. (1998): “Introduction aux classes d’objets”. *Langages* 131. Larousse: Paris.