

# Les noms propres de personne en français et en grec : reconnaissance, extraction et enrichissement de dictionnaire

Tita Kyriacopoulou, Claude Martineau, Thanassis Mavropoulos

## ► To cite this version:

Tita Kyriacopoulou, Claude Martineau, Thanassis Mavropoulos. Les noms propres de personne en français et en grec : reconnaissance, extraction et enrichissement de dictionnaire. 30th International Conference on Lexis and Grammar, Oct 2011, Nicosia, Chypre. pp.467-479. hal-00790515

HAL Id: hal-00790515

<https://hal-upec-upem.archives-ouvertes.fr/hal-00790515>

Submitted on 20 Feb 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# LES NOMS PROPRES DE PERSONNE EN FRANÇAIS ET GREC : RECONNAISSANCE, EXTRACTION ET ENRICHISSEMENT DE DICTIONNAIRES

TITA KYRIACOPOULOU(1), CLAUDE MARTINEAU(1), THANASSIS MAVROPOULOS(2)

(1) LIGM, Université Paris-Est, LTTL,  
{tita, claudemartineau}@univ-mlv.fr

(2) Université Aristote de Thessalonique  
rip@frrl.auth.gr

**Résumé.** Notre recherche s'intéresse à la reconnaissance, l'extraction des noms propres de personne dans des corpus français et grecs mais aussi à l'enrichissement des dictionnaires existants. Pour cela nous sommes partis des grammaires développées pour le français et nous les avons adaptées et enrichies pour le grec. Notre méthode s'appuie en particulier sur les graphes dictionnaires qui rendent les grammaires produites aisément réutilisables.

**Mots-clés:** noms propres, dictionnaire électronique, transducteur, graphe dictionnaire

## Introduction

Le travail présenté ici s'inscrit dans le cadre d'une collaboration de recherche entre le LTTL<sup>1</sup> et le LIGM<sup>2</sup> et s'intéresse à la reconnaissance, l'extraction des noms propres de personne dans des corpus français et grecs mais aussi à l'enrichissement des dictionnaires existants.

Le besoin de créer des outils qui automatisent l'exploitation et/ou l'extraction d'informations pertinentes, notamment dans les textes, est aujourd'hui un fait. Les systèmes d'extraction d'information (Hobbs *et al.* 1996; Fourour 2002), de recherche d'information (Sekine et Isahara 1998) ou de fouille de textes (Jacquemin et Bush 2000) sont de plus en plus nombreux. Dans cet article, nous proposons une méthode symbolique permettant de traiter les noms propres de personne du français et du grec.

Après une brève présentation de l'objectif et de l'intérêt de ce travail nous présenterons les particularités du grec ainsi que les dictionnaires des noms propres développés par le LTTL. Ensuite nous parlerons des limites des grammaires existantes pour le grec et enfin de la méthodologie que nous avons adoptée. Notre objectif étant de proposer une solution commune pour ces deux langues et adaptable aux autres langues du système UNITEX (Paumier, 2003, 2011).

## 1. Objectif de la recherche

La reconnaissance de nom de personne a donné lieu à de nombreux travaux, notamment la thèse de Nathalie Friburger sur les noms propres (Friburger 2002), ou ceux de Denis Maurel (Maurel *et al.* 2007) qui se situent dans le cadre plus large de la reconnaissance d'entités nommées. Des ressources comprenant des dictionnaires de prénoms et des grammaires existent pour plusieurs langues dont le français.

Pour la reconnaissance des patronymes des langues romanes ou de l'anglais qui ne possèdent pas de déclinaisons, le prénom ou une suite de prénom (dans le cas de prénom composé) sert de déclencheur et le mot voire la séquence de mots commençant par une majuscule suivant ce prénom (composé ou non) est considéré comme le patronyme (composé ou non). Mais ceci n'est pas applicable pour les langues à cas comme le grec par exemple.

---

<sup>1</sup>Laboratoire de Traduction et de Traitement Automatique du Langage de l'Université Aristote de Thessalonique.

<sup>2</sup>Laboratoire d'Informatique Gaspard-Monge de L'Université Paris-Est Marne-la-Vallée.

En ce qui concerne le grec, qui possède quatre cas (nominatif, génitif, accusatif, vocatif), le prénom comme le nom patronymique se déclinent et ils possèdent des formes différentes selon le cas, le genre et le nombre. Un dictionnaire de patronymes a donc été constitué qui comme pour les prénoms comportent toutes les formes fléchies (*Βαλετόπουλος, Βαλετόπουλο, Βαλετόπουλου, Βαλετόπουλε*).

Néanmoins ce dictionnaire de patronymes ne pouvant être exhaustif, des règles de reconnaissances fondées sur la morphologie respective des formes nominales des quatre cas du grec ont été développées dans l'environnement Unitex<sup>3</sup>.

Notre objectif étant de valoriser l'existant, nous sommes partis des grammaires développées notamment dans le cadre du projet Infom@gic (Martineau et al. 2007) et nous les avons adaptées au grec; d'autres ont été mises au point spécialement pour cette langue. En outre, constatant la présence de mots en alphabet latin dans les textes grecs il nous a paru intéressant de construire des ressources couvrant le français et le grec. Nos grammaires permettent aujourd'hui non seulement de reconnaître des noms propres dans les textes mais aussi d'enrichir les dictionnaires existants en attribuant le plus correctement possible, le cas, le genre et le nombre à une forme reconnue comme nominale qui serait absente du dictionnaire électronique grec ou français des prénoms et des patronymes.

Dans la présentation qui suit nous mettons principalement l'accent sur les spécificités du grec qui doivent être prises en compte et nous ne ferons des références au français que lorsque c'est nécessaire pour la clarté du texte.

## 2. Brève présentation des dictionnaires des noms propres grecs

Rappelons qu'en grec, pour les noms propres de personne, nous disposons aujourd'hui de 50 000 patronymes et 500 prénoms. Les noms propres se déclinent comme les autres mots grecs et *a priori* n'ont pas de pluriel. Cependant, dans certains cas, les noms propres de personnes (prénoms, noms de famille) peuvent avoir un pluriel; nous l'avons donc généré :

- (1) *Οι Γιώργηδες ήρθαν* (singulier: *ο Γιώργος*, Pluriel : *οι Γιώργηδες*)  
\*Les Georges sont arrivés
- (2) *Οι Αφοι Παπαδόπουλοι δουλεύουν μαζί*  
Les frères Papadopoulos travaillent ensemble

Il faut rappeler ici que pour engendrer toutes les formes fléchies des mots grecs nous faisons appel à un programme de flexion qui se sert de codes flexionnels (Kyriacopoulou 2002, 2003).

Les pluriels des noms propres étant souvent spécifiques (*ο άνθρωπος*/l'homme-*οι άνθρωποι*/les hommes, mais *ο Γιώργος*/Georges-*οι Γιώργηδες*/\*les Georges), nous avons créé des codes flexionnels spécifiques. Voici un extrait de nos dictionnaires :

### Prénoms

Άγγελος,.N+Prenom:Nms	Αθανάσιος,.N+Prenom:Nms	Ακάκιος,.N+Prenom:Nms
Άγγελου,Άγγελος.N+Prenom:Gms	Αθανάσιου,Αθανάσιος.N+Prenom:Gms	Ακάκιου,Ακάκιος.N+Prenom:Gms
Άγγέλου,Άγγελος.N++Prenom:Gms	Αθανασίου,Αθανάσιος.N+Prenom:Gms	Ακακίου,Ακάκιος.N+Prenom:Gms
Άγγελο,Άγγελος.N+Prenom:Ams	Αθανάσιο,Αθανάσιος.N+Prenom:Ams	Ακάκιο,Ακάκιος.N+Prenom:Ams
Άγγελε,Άγγελος.N+Prenom:Vms	Αθανάσιε,Αθανάσιος.N+Prenom:Vms	Ακάκιε,Ακάκιος.N+Prenom:Vms
Άγγελοι,Άγγελος.N+Prenom:Nmp:Vmp	Αθανάσιοι,Αθανάσιος.N+Prenom:Nmp:Vmp	Ακάκιοι,Ακάκιος.N+Prenom:Nmp:Vmp
Άγγελων,Άγγελος.N+Prenom:Gmp	Αθανάσιων,Αθανάσιος.N+Prenom:Gmp	Ακάκιων,Ακάκιος.N+Prenom:Gmp
Άγγέλων,Άγγελος.N+Prenom:Gmp	Αθανασίων,Αθανάσιος.N+Prenom:Gmp	Ακακίων,Ακάκιος.N+Prenom:Gmp
Άγγελους,Άγγελος.N+Prenom:Amp	Αθανάσιους,Αθανάσιος.N+Prenom:Amp	Ακάκιους,Ακάκιος.N+Prenom:Amp
Άγγέλους,Άγγελος.N+Prenom:Amp	Αθανασίους,Αθανάσιος.N+Prenom:Amp	Ακακίους,Ακάκιος.N+Prenom:Amp

<sup>3</sup> La version utilisée est Unitex3.0beta, 2011, <http://igm.univ-mlv.fr/~unitex/index.php?page=3&html=beta.html>.

## Patronymes

Εβλίδης, Εβλίδης. N+Hum+Nprop+[Ln]:Nms  
Εβλίδη, Εβλίδης. N+Hum+Nprop+[Ln]:Gms:Ams:Vms:Nfs:Gfs:Afs:Vfs  
Εβλίδηδες, Εβλίδης. N+Hum+Nprop+[Ln]:Nmp:Amp:Vmp  
Εβλίδηδων, Εβλίδης. N+Hum+Nprop+[Ln]:Gmp  
Εβλίδου, Εβλίδης. N+Hum+Nprop+[Ln]:Nfs:Gfs:Afs:Vfs  
Εβραίνου, Εβραίνου. N+Hum+Nprop+[Ln]:Nms:Gms:Ams:Vms:Nmp:Gmp:Amp:Vmp:Nfs:Gfs:Afs:Vfs:Nfp:Gfp:Afp:Vfp  
Εβρεμιτίδη, Εβρεμιτίδης. N+Hum+Nprop+[Ln]:Gms:Ams:Vms:Nfs:Gfs:Afs:Vfs  
Εβρεμιτίδηδες, Εβρεμιτίδης. N+Hum+Nprop+[Ln]:Nmp:Amp:Vmp  
Εβρεμιτίδηδων, Εβρεμιτίδης. N+Hum+Nprop+[Ln]:Gmp  
Εβρεμιτίδου, Εβρεμιτίδης. N+Hum+Nprop+[Ln]:Nfs:Gfs:Afs:Vfs  
Εβρέν, Εβρέν. N+Hum+Nprop+[Ln]:Nms:Gms:Ams:Vms:Nmp:Gmp:Amp:Vmp:Nfs:Gfs:Afs:Vfs:Nfp:Gfp:Afp:Vfp  
Εβρενέζου, Εβρενέζου. N+Hum+Nprop+[Ln]:Nms:Gms:Ams:Vms:Nmp:Gmp:Amp:Vmp:Nfs:Gfs:Afs:Vfs:Nfp:Gfp:Afp:Vfp

Les noms propres de personne peuvent apparaître dans les textes sous les formes suivantes :

- (3) Prénom Nom : *Σάκης Ρουβάς* (Sakis Rouvas)
- (4) Prénom abrégé Nom : *Σ. Ρουβάς* (S. Rouvas)
- (5) Nom : *Ρουβάς* (Rouvas)
- (6) Prénom : *Σάκης* (Sakis)

et même précédés d'un article : *Ο Σάκης Ρουβάς, Ο Σάκης, Ο Ρουβάς*

Ces variations peuvent être exprimées à l'aide d'une grammaire spécifique qui vérifie que les cas du prénom ou du déterminant et du patronyme sont compatibles. En revanche, les considérer comme des multimots devant être fléchis serait une approche trop coûteuse.

### 3. Traitement des cas particuliers

Les surnoms ou les diminutifs sont très courants en grec. Actuellement, ils n'ont pas été pris en compte dans nos dictionnaires mais ils peuvent aider à la désambiguïsation. Par exemple, pour la famille *Papandreou* (entre le grand-père (ancien premier Ministre) et le petit-fils (Premier Ministre actuel) nous avons :

- (7) Premier Ministre: *Γιωργάκης + Γιώργος Παπανδρέου*  
(*Giorgakis + Georges Papandreou*)
- (8) Ancien Premier ministre : *Γεώργιος Παπανδρέου* (Georges Papandreou)  
\**Γιωργάκης Παπανδρέου* (*Giorgakis Papandreou*)

Pour les noms de personne complexes (plus d'un prénom ou plus d'un nom de famille) nous avons opté pour une représentation par graphe. Ces noms peuvent comporter également des traits d'union, des lettres initiales de noms ou prénoms comme dans :

- (9) *Αγγελοπούλου-Δασκαλάκη, Αγγελοπούλου-Δ., Α.-Δασκαλάκη, Αγγελοπούλου, Δασκαλάκη*  
*Άννα-Μαρία, Άννα-Μ., Α.Μ., Α.-Μ.*

A noter aussi qu'en grec les patronymes sont précédés de formes abrégées, utilisées comme déclencheurs pour la reconnaissance automatique des noms propres. Ils peuvent indiquer notamment, la civilité de la personne *κ. Κύριος* (M. Monsieur) mais peuvent aussi avoir des mentions particulières comme dans :

Στρατηγός ε.α. Ιωάννης Καλέργης (Stratigos e.a. Ioannis KALERGIS)

Στρατηγός εν αποστρατεία Ιωάννης Καλέργης (Stratigos en apostrateia Ioannis KALERGIS)

Général en retraite Ioannis KALERGIS

Enfin il faut signaler qu'en grec, entre le prénom et le patronyme, il apparaît souvent une lettre majuscule suivie d'un point terminal qui renvoie au prénom du père :

(10) Γιώργος Χ. Γεωργιάδης (Yiorgos X. Gueoryadis)

Mais cette lettre est ambiguë. Par exemple la lettre « X. » de l'exemple précédent peut correspondre à deux prénoms différents : *Χαράλαμπος* (*Xaralampous*) ou *Χαρίλαου* (*Xarilaou*). Pour la reconnaissance de ces séquences nous avons créé le graphe de la figure 1.



Figure 1

En outre, dans certains textes, notamment journalistiques ou issus du web, on trouve souvent pour les noms de personnes étrangers, à côté de leur transcription grecque, leur graphie latine d'origine. Par exemple, pour l'ex-premier ministre français, Dominique de Villepin on trouve sur le *wikipedia* grec la phrase suivante :

*Ο Ντομινίκ ντε Βιλπέν (γαλλ. Dominique de Villepin) είναι Γάλλος διπλωμάτης και πολιτικός, Πρωθυπουργός της Γαλλίας από τις 31 Μαΐου 2005 μέχρι το 2007.*

(Dominique<sub>grec</sub> de<sub>grec</sub> Villepin<sub>grec</sub> (français. *Dominique de Villepin*) est un diplomate français et homme politique, premier ministre français du 13 Mai 2005 à 2007.)

dans laquelle la forme grecque *Ντομινίκ ντε Βιλπέν* est suivie de sa forme latine entre parenthèses. Pour cette raison, les ressources que nous avons développées permettent de reconnaître des noms de personne écrits en alphabet grec ou en alphabet latin. Remarquons qu'en grec, les patronymes étrangers, même écrits en alphabet grec, sont invariables et possèdent la même forme à tous les cas.

Comme pour le français, les noms propres de personnes peuvent créer des ambiguïtés avec les mots simples :

(11) *Αγάπη γιατί με προσπερνάς ;*  
Agapi pourquoi tu m'évites ?  
Amour pourquoi tu m'évites ?

Mais ces cas n'ont pas été traités pour l'instant.

#### 4. Approche et ressources développées

Pour l'enrichissement et la maintenance des dictionnaires de prénoms et de patronymes existants nous avons opté pour l'utilisation de graphe dictionnaire qui permet de capturer les formes (fléchies) absentes du dictionnaire mais présentes dans le texte.

Rappelons qu'une entrée de dictionnaire Dela « statique » se présente généralement sous la forme :

*forme fléchie, forme canonique.catég gram+traits syntaxico-sémantiques: infos flexionnelles*

Les entrées produites par le graphe dictionnaire comportent quelques différences par rapport à celles d'un dictionnaire Dela.

En particulier, la forme canonique d'un nom de personne grec n'est pas toujours la stricte séquence des formes canoniques de ses constituants. Considérons le nom féminin au génitif suivant :

*Μαρίας Γιακοπούλου* (*Maria<sub>Gen</sub> Yiacopoulou<sub>Gen</sub>*). La séquence des formes canoniques serait *Μαρία Γιακόπουλος* (*Maria Yiacopoulos*) or la forme canonique d'un nom féminin se forme avec le nominatif féminin du nom patronymique. La forme correcte est donc : *Μαρία Γιακοπούλου* (*Maria Yiacopoulou*).

Les noms propres de personne présents dans un texte sont souvent utilisés pour son indexation. Pour que celle-ci soit correcte, il faut que les différentes formes (cas/déclinaisons, utilisation de majuscules/minuscules) prises par ces noms soient ramenées à une forme unique pour que ces formes différentes ne soient pas prises pour des noms de personnes différents. La forme canonique peut constituer cette forme unique ou normalisée. Ainsi les quatre formes :

*Φρειδερίκος Βαλετόπουλος, Φρειδερίκο Βαλετόπουλο, Φρειδερίκων Βαλετόπουλων, Φρειδερίκε Βαλετόπουλε*

désignent la même personne (Freiderikos Valetopoulos) respectivement au nominatif, à l'accusatif, au génitif et au vocatif.

La normalisation doit également prendre compte l'unicité graphique (majuscule/minuscule) : les noms et prénoms commencent par une majuscule et se poursuivent par des minuscules, les particules présentes dans les noms sont normalisées en minuscules et certains préfixes (dans des noms d'origine arabe) commencent eux par une majuscule. Ainsi des noms présents dans des titres de journaux, écrits entièrement en majuscules ont une forme normalisée où la graphie est définie sans ambiguïté. Par exemple des noms présents dans les titres comme :

(12) ΦΡΑΝΣΟΥΑ ΝΤΕ ΛΑ ΡΟΣΦΟΥΚΩ (FRANCOIS DE LA ROCHEFOUCAULD)

(13) ΝΑΝΤΙΑ ΕΛ-ΜΑΜΠΡΟΥΚ (NADIA EL-MABROUK)

ont les formes canonique/normalisées respectives : *Φρανσουά ντε λα Ροσφουκώ* (François de la Rochefoucauld) et *Νάντια Ελ-Μαμπρουκ* (Nadia El-Mabrouk).

L'approche adoptée ici et réalisable dans l'environnement Unitex consiste à utiliser la notion de graphe dictionnaire qui permet de créer dynamiquement des entrées de dictionnaires similaires à celles présentes dans les dictionnaires Dela (Courtois 1990) réalisés manuellement. La figure 2 ci-après présente une version ultra simplifiée d'un graphe dictionnaire qui reconnaît un nom de personne et qui produit dans le dictionnaire du texte des entrées comme : Jean Dupond,.NPersonne.



Figure 2

Les grammaires existantes sous Unitex développées notamment dans le cadre du projet Infom@gic (Martineau et al. 2007) dont le but était la reconnaissance d'entités nommées comportait des graphes ou plutôt des sous-graphes pour chaque type d'entité. Mais ces graphes ne permettaient pas facilement de reconnaître, suivant les besoins, des noms à particule, d'exiger ou non la présence du prénom, ou d'extraire les noms de personnes comportant des prénoms composés ce qui nous a conduit à en développer de nouveaux.

La reconnaissance de noms propres de personnes peut avoir de nombreuses utilisations: ils peuvent être utilisés en tant que constituants de groupes nominaux (*Οι γονείς της Μαρίας Γιακοπούλου* (*Les parents de Maria Yiacopoulou*) et prendre leur part dans le cadre de systèmes d'analyse syntaxique, être incorporés dans des outils de reconnaissances d'entités nommées, enfin être utilisés dans des systèmes fondés sur des requêtes comme par exemple : « quels sont les noms à particules écrits en grec et comportant un prénom composé présents dans le document traité ? ».

Pour répondre à ces différentes utilisations, les entrées produites par les ressources que nous avons développées comportent des traits qui gardent une trace du chemin suivi dans le graphe dictionnaire qui les a créées et qui donnent ainsi des informations sur les mots reconnus qui sont les constituants de ces entrées complexes. Le tableau 1 donne un large extrait des traits éventuellement présents dans les entrées dynamiquement produites.

<i>Trait</i>	<i>Indique la présence de</i>
+prenom	<i>un prénom</i>
+prenom_composé	<i>un prénom composé</i>
+nom	<i>un nom</i>
+nom composé	<i>un nom composé</i>
+particule	<i>une particule dans le nom</i>
+préfixe	<i>un préfixe dans le nom</i>
+saint	<i>Saint (latin/grec/abrev) avant le prénom</i>
+prenom_grec	<i>un prénom du dico grec</i>
+prenom_etr	<i>un prénom étranger du dico grec</i>
+nom_grec	<i>un nom du dico grec</i>
+prenom_latin	<i>un prénom écrit en alphabet latin</i>
+nom_latin	<i>un nom écrit en alphabet latin</i>

Tableau 1

Les entrées produites possèdent comme celles présentes dans les dictionnaires Dela statiques des informations flexionnelles. Celles-ci sont calculées par des grammaires grâce à des ressources existantes. Le dictionnaire électronique développé par le LTTL<sup>1</sup> et le LIGM<sup>2</sup> ne comporte pas de prénoms composés. La constitution de telles listes ne saurait être exhaustive et donc seuls les prénoms simples y figurent. La reconnaissance, par exemple, d'une forme fléchie d'un prénom composé utilise les ressources disponibles sur ses composants simples.

#### 4.1 Traitement des noms complets<sup>4</sup> de personne

A titre d'exemple, la reconnaissance et l'attribution d'infos flexionnelles pour la forme fléchie composée *Κώστα-Λάζαρου Γιακογιάννη* (*Kosta-Lazarou Yacoyanni*) commence par le traitement du prénom composé qui s'effectue grâce aux données suivantes<sup>5</sup> : *Κώστα* :Gms :Ams :Vms et *Λάζαρου* :Gms. D'où *Κώστα-Λάζαρου* : Gms. C'est le prénom qui a la forme la moins ambiguë qui détermine les infos flexionnelles du composé. Ensuite, elles se combinent avec celles du patronyme<sup>6</sup> *Γιακογιάννη* :Gms :Ams :Vms :Gfs :Afs :Vfs qui suit pour aboutir à celles du nom complet. Ainsi la reconnaissance de *Κώστα-Λάζαρου Γιακογιάννη* donne lieu à la création de l'entrée de dictionnaire munie de traits suivante :

*Κώστα-Λάζαρου Γιακογιάννη, Κώστας-Λάζαρος Γιακογιάννης*. Person +prenom+prenom\_composé +prenom\_grec +nom+nom\_grec+Prenom=Κώστας-Λάζαρος +Nom=Γιακογιάννης:Gms.

	<i>Singulier</i>	<i>Pluriel</i>		<i>Singulier</i>	<i>Pluriel</i>		<i>Masc. Sing.</i>	<i>Fém..Sing.</i>	<i>Pluriel</i>
<i>Nom.</i>	Κώστας	Κώστηδες	<i>Nom.</i>	Λάζαρος	Λάζαροι	<i>Nom.</i>	Γιακογιάννης	Γιακογιάννη	Γιακογιάννηδες
<i>Gén.</i>	Κώστα/Κώστος	Κώστηδων	<i>Gén.</i>	Λάζαρου/Λαζάρου	Λάζαρων/Λαζάρων	<i>Gén.</i>	Γιακογιάννη	Γιακογιάννη	Γιακογιάννηδες
<i>Acc.</i>	Κώστα	Κώστηδες	<i>Acc.</i>	Λάζαρο	Λάζαρους/Λαζάρους	<i>Acc.</i>	Γιακογιάννη	Γιακογιάννη	Γιακογιάννηδων
<i>Voc.</i>	Κώστα	Κώστηδες	<i>Voc.</i>	Λάζαρε	Λάζαροι	<i>Voc.</i>	Γιακογιάννη	Γιακογιάννη	Γιακογιάννηδες

Tableau 2

Tableau 3

Tableau 4

<sup>4</sup> Nous entendons par nom complet un couple prénom nom, chacun d'eux éventuellement composé.

<sup>5</sup> Voir les tableaux 2 et 3.

<sup>6</sup> Voir tableau 4.

En ce qui concerne le traitement des noms écrits en caractères latin, bien que le dictionnaire français des prénoms utilisé pour les extraire ne contienne que les marque *:ms* et *:fs*, la grammaire que nous avons développée les remplace par les infos flexionnelles équivalentes adéquates selon leur genre. Les patronymes qui les suivent qui ne figurent pas dans des dictionnaires (il n'en existe pas pour le français qui ne possède pas de cas) se voient donc attribuées toutes infos flexionnelles du masculin et du féminins. Voici à titre d'exemples les formes « hellénisées » attribuées à quelques prénoms français :

Pierre,.N+prenom:Nms:Ams:Gms:Vms  
 Marie,.N+prenom:Nfs:Afs:Gfs:Vfs  
 Dominique,.N+prenom:Nms:Ams:Gms:Vms:Nfs:Afs:Gfs:Vfs

Une phrase comme : « *Η Μαρία λατρεύει τον Sébastien* » (*Marie adore Sébastien*) peut donc être analysée par une grammaire « exclusivement » écrite pour le grec sans problème. Le graphe dictionnaire développé permet de produire dans le dictionnaire du texte des entrées pour un nom complet écrit en caractères latins. On trouvera dans le dictionnaire du texte des entrées comme :

Eric Rouleau, Eric Rouleau. Person+prenom+prenom\_latin+nom+nom\_latin+latin+  
 Prenom=Eric+Nom=Rouleau:Nms:Ams:Gms:Vms

Ces entrées sont ensuite utilisables par des grammaires au même titre que les dictionnaires statiques. Par exemple, si l'on recherche des noms de personne comportant un prénom et un nom en évitant les édifices religieux ou les fêtes nominales, il est possible d'écrire une expression régulière ou un graphe comportant : <Person+prenom+nom~saint> ou si l'on recherche des noms à particules : <Person+particule>.

## Conclusion et perspectives

Cette présentation très sommaire illustre déjà l'intérêt de ce travail et en particulier de la méthode des graphes dictionnaires. Il sera très intéressant de comparer cette méthode avec celle nommée *CasEN* fondée sur le système *Cassys* dernièrement inclus dans Unitex et développé par Nathalie Friburger et Denis Maurel (Friburger & Maurel 2004, Maurel et al. 2011). Notre travail doit être complété pour le grec et le français mais aussi s'appliquer aux autres langues présentes dans Unitex. Nous sommes persuadés que même si le système de reconnaissance des noms de personne est commun les règles de désambiguïsations seront propres à chaque langue voire à chaque pays.

## Bibliographie

- COURTOIS, Blandine, « Un système de dictionnaires électroniques pour les mots simples du français », in Courtois B. et Silberztein, M. (éds), *Dictionnaires électroniques du français, Langue Française*, n° 87, Larousse, Paris, 1990, p. 11-22.
- FOUOUR, Nordine, « Nemesis, un système de reconnaissance incrémentielle des entités nommées pour le français », in *Actes de la 9<sup>ème</sup> Conférence Nationale sur le Traitement Automatique des Langues Naturelles (TALN 2001)*, Nancy, 2002, vol. 1, p. 265-274.
- FRIBURGER, Nathalie, *Reconnaissance automatique des noms propres : Application à la classification automatique des textes journalistiques*, Thèse de doctorat, Université de Tours, Paris, 2002.
- FRIBURGER, Nathalie, MAUREL Denis, « Finite-state transducer cascade to extract named entities in texts », *Theoretical Computer Science*, 2004, vol. 313, p. 94-104.
- HOBBS, Jerry, APPELT, Douglas, BEAR, John, ISRAEL, David, KAMEYAMA, Megumi, STICKEL Mark et TYSON, Mabry, « Fastus : a cascaded finite-state transducer for extracting information from natural-language text », in ROCHE, Emmanuel, et SCHABES, Yves. (éds), *Finite State Devices for Natural Language Processing*, MIT Press, Cambridge, USA, 1996, p. 383-406.



JACQUEMIN, Christian, et BUSH, Caroline, « Fouille du Web pour la collecte d'entités nommées », in *Actes de la 8<sup>ème</sup> Conférence Nationale sur le Traitement Automatique des Langues Naturelles (TALN 2000)*, Lausanne, 2000, p. 187-196.

KYRIACOPOULOU, Tita, MRABTI, Safia, et YANNAKOPOULOU, Anastasia, « Le dictionnaire électronique des noms composés en grec moderne », *Linguisticae Investigationes* 25:1, Amsterdam/Philadelphia : John Benjamins, 2002, p. 7-28.

KYRIACOPOULOU, Tita, *Analyse automatique des textes écrits : le cas du grec moderne*, Mémoire d'Habilitation à diriger des recherches, Université de Marne-la-Vallée, 2003.

MARTINEAU, Claude, TOLONE, Elsa, et VOYATZI, Stavroula, « Le statut référentiel des entités nommées », in : *Actes du XXVI<sup>e</sup> Colloque International Lexique et Grammaire, Bonifacio, Corse du Sud, 2-6 octobre 2007*, éd. Catherine Camugli, Matthieu Constant et Anne Dister, 2007, p. 105-112.

MAUREL, Denis, VITAS, Duško, KRSTEV, Cvetana, et KOEVA, Svetla, « PROLEX: A Lexical Model For Translation Of Proper Names Application To French, Serbian And Bulgarian », 2007

MAUREL, Denis, FRIBURGER, Nathalie, ANTOINE, Jean-Yves, ESHKOL-TARAVELLA, Iris, NOUVEL, Damien, « Cascades autour de la reconnaissance des entités nommées », *TAL* p. 52-1, 2011.

SEKINE, Satoshi, et NOBATA, Chikashi, « An Information Extraction System and a Customization Tool », in *Proceedings of the New Challenges in Natural Language Processing and its Application*, p. 25-26 May 1998, Tokyo, Japan, 1998.

PAUMIER, Sébastien, « De la reconnaissance de formes linguistiques à l'analyse syntaxique », Paris : Université de Marne-la-Vallée, [Thèse de doctorat], 2003.

PAUMIER, Sébastien, <[www-igm.univ-mlv.fr/~unitex/manuel.htm](http://www-igm.univ-mlv.fr/~unitex/manuel.htm)>, 2011.

Tita KYRIACOPOULOU  
Claude MARTINEAU  
Université Paris-Est  
77454 Marne-la-Vallée Cedex 2

Thanassis MAVROPOULOS  
Université Aristote de Thessalonique  
54006 Thessalonique