

Reconnaissance automatique de formes dérivées dans les textes grecs

Tita Kyriacopoulou¹, Claude Martineau²,
Anastasia Yannacopoulou²

¹ Université Aristote de Thessaloniki, Faculté des Lettres
Laboratoire de traduction et de traitement des langues
tita@frl.auth.gr

² Université de Marne-la-Vallée
Laboratoire d'Informatique linguistique de l'Institut Gaspard-Monge
{claude.martineau ; anastasia.annacopoulou}@univ-mlv.fr

Résumé

Notre objectif est la reconnaissance automatique de certaines formes dérivées, *i.e.* des diminutifs et des augmentatifs des noms et des adjectifs simples, ainsi que des comparatifs et des superlatifs des adjectifs simples du grec moderne. Il s'agit de formes qui sont généralement produites par l'adjonction d'un suffixe à la forme standard correspondante. Nous justifions notre choix de les ajouter dans le dictionnaire électronique. Leur traitement a nécessité une nouvelle représentation du dictionnaire qui utilise désormais un système de règles permettant de générer aisément les formes fléchies dérivées, de les étiqueter en tant que telles, et de les mettre en relation avec leur forme de base. Il en résulte une meilleure structuration des ressources lexicales et une production de dictionnaires flexible.

Mots-clés : dictionnaire électronique, flexion, formes dérivées.

Abstract

Our study concerns the automatic recognition of certain derived forms, *i.e.* the diminutives and the augmentatives of simple nouns and adjectives, as well as the comparatives and the superlatives of simple adjectives of Modern Greek. These forms are generally produced by the adjunction of a suffix to the corresponding standard form. We justify our choice to add them to the electronic dictionary of Modern Greek. Their processing implies an evolution of the representation of the dictionary which now uses a system of rules in order to easily generate the derived inflected forms, to label them as such, and to clearly put them in relation to their base form. We achieve better structuring of the lexical resources and flexible modular dictionary production.

Keywords : electronic dictionary, inflection, derived forms.

1. Introduction

Dans cet article, nous étudions les moyens permettant de reconnaître automatiquement les diminutifs et les augmentatifs des noms et des adjectifs simples, ainsi que les comparatifs et les superlatifs des adjectifs simples du grec moderne. Désormais, nous appellerons ces formes dérivées par opposition aux formes standards, *i.e.* les formes non marquées par des suffixes spécifiques. Nous considérons, dans un premier temps, la reconnaissance de suffixes caractéristiques adjoints aux mots d'un texte à partir de règles morphologiques. Nous montrons les limites de cette approche et justifions ainsi la génération (semi-)automatique de ces formes dérivées en les reliant à leur forme standard (cf. § 2). Ceci implique leur ajout au dictionnaire électronique du grec moderne (cf. Kyriacopoulou, 2003 ; Sklavounou, 1999). Ce dernier utilise le formalisme DELA mis en œuvre à l'origine pour le français (Courtois *et al.*,

1990) et adopté depuis pour de nombreuses langues. Nous expliquons le format utilisé par ce système et la manière dont un dictionnaire de formes fléchies (DelafGR)¹ est généré automatiquement à partir des formes canoniques correspondantes (DelasGr) (cf. § 3.1). Ce dernier sert uniquement à la maintenance des données linguistiques et il n'y a donc aucune redondance entre le DelasGR et le DelafGr (cf. Laporte, 2005). Nous proposons une extension du format DELA afin de traiter efficacement la génération des formes dérivées qui nous intéressent (cf. § 3.2) et de nous permettre une grande flexibilité quant au type du DelafGr produit (cf. § 3.3).

2. Reconnaissance des formes dérivées par règles morphologiques

Les formes dérivées sont nombreuses dans les corpus et leur reconnaissance automatique par application de règles morphologiques portant sur les suffixes semblerait possible. Dans les paragraphes qui suivent, nous présentons les suffixes en question et les limites de cette approche.

2.1. Les suffixes des augmentatifs et des diminutifs des noms et adjectifs

Pour marquer le diminutif des noms, le grec moderne dispose de suffixes particuliers (cf. Holton *et al.*, 1997 ; Triantafyllidis, 2000 ; Clairis *et al.*, 2005). Il s'agit des suffixes : *-άκης*, *-άκος*, *-κας*, *-ούλης*, *-τζίκος*, *-ντζίκος*, *-αράκος*, *-ίσκος* qui forment des noms masculins, *-ίτσα*, *-οπούλα*, *-ούλα*, *-ούδα* des féminins, et *-άκι*, *-δάκι*, *-ουδάκι*, *-αράκι*, *-αλάκι*, *-όπουλο*, *ούδι*, *-ούλι*, *-έλι*, *-ίδιο* des neutres. Ces suffixes s'associent à des noms simples masculins, féminins ou neutres (formes standards), par effacement soit de la dernière lettre de la forme standard soit des deux dernières lettres². Par exemple :

| | |
|-------------------------------------|--|
| <i>αυτοκίνητο</i> ,.N:Nns (voiture) | <i>αυτοκινητάκι</i> ,.N+Dim:Nns (petite voiture) |
| <i>άγγελος</i> ,.N:Nms (ange) | <i>αγγελούδι</i> ,.N+Dim:Nns (petit ange) |

où *Nns* signifie nominatif neutre singulier et *Nms* nominatif masculin singulier.

Le diminutif des adjectifs se construit à l'aide de trois groupes de suffixes : *-ούλης*, *-ούλικος* et *-ούτσικος*. Par exemple, l'adjectif *όμορφος* (beau) peut former trois diminutifs : *ομορφούλης*, *ομορφούλικος* et *ομορφούτσικος*, l'adjectif *ωραίος* (joli) accepte uniquement la forme *ωραιούτσικος*.

En ce qui concerne les augmentatifs des noms simples, ils possèdent, eux aussi, des suffixes particuliers : *-άκλα*, *-άρα*, *-αρόνα*, *-ούκλα*, *-ούρα* qui forment des féminins, *-αράς*, *-αρος* et *-ούκλος* des masculins. Par exemple :

| | |
|-------------------------------------|--|
| <i>αυτοκίνητο</i> ,.N:Nns (voiture) | <i>αυτοκινητάρα</i> ,.N+Aug:Nfs (grande voiture) |
|-------------------------------------|--|

où *Nfs* indique la forme du nominatif féminin au singulier.

2.2. Les suffixes des comparatifs et des superlatifs des adjectifs

Les formes canoniques des comparatifs des adjectifs simples (nominatif masculin singulier) sont générées par l'adjonction des suffixes suivants : *-ότερος*, *-ύτερος*, *-ούστερος*, et *-έστερος*, après effacement des deux dernières lettres de la forme canonique standard. Par exemple, la forme du comparatif de l'adjectif *απλός* (simple) est *απλούστερος* (plus simple). De même,

¹ Le DELAF est comprimé sous la forme d'un automate acyclique (D.A.G. – Direct acyclic graph).

² Pour certains emprunts, l'adjonction du suffixe n'entraîne l'effacement d'aucune lettre. Par exemple : *μπαρ* (bar), *μπαράκι* (petit bar).

les formes du superlatif sont produites par l'adjonction des suffixes : *-ότατος*, *-ύτατος*, *-ούστατος*, et *-έστατος*, par exemple : *απλούστατος* (le plus simple).

2.3. Reconnaissance à l'aide de motifs : évaluation de la méthode

Afin de repérer dans des textes grecs les noms diminutifs et augmentatifs, nous avons effectué des requêtes basées sur les suffixes sus-exposés dans tous les cas grammaticaux³. Nous avons appliqué nos requêtes sur un extrait du corpus journalistique *TA NEA* (Janvier 2003 et rubrique « Avis-Commentaires » de l'année 1997, 2 113 902 tokens), à l'aide de l'analyseur Unitex⁴. Les requêtes ont été effectuées à l'aide de filtres morphologiques appliqués aux unités lexicales recherchées. Par exemple, pour reconnaître les formes augmentatives des noms simples nous avons appliqué, à notre corpus, les filtres de la figure 1⁵ :

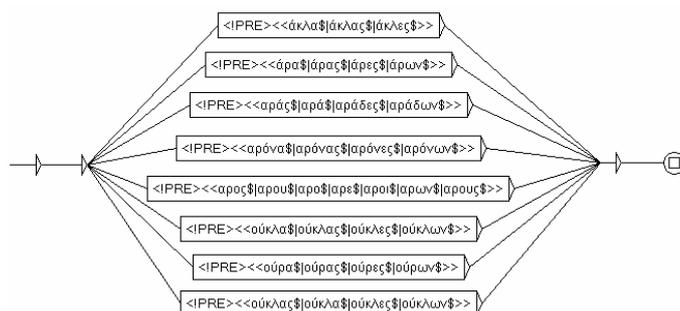


Figure 1. Suffixes des noms augmentatifs

En ce qui concerne les augmentatifs, nous avons repéré 993 formes et plus précisément 179 formes différentes, dont 8 % sont effectivement des augmentatifs. Quant aux diminutifs, nous avons trouvé 2 675 formes, c'est-à-dire 832 formes différentes, mais seulement 37,5 % d'entre elles sont de vrais diminutifs.

Nous constatons alors que les suffixes seuls ne constituent pas un critère pour attribuer l'étiquette de diminutif (+Dim) ou d'augmentatif (+Aug) à un nom. Ainsi, les mots : *σκάκι* (échecs), *χαντάκι* (douve), *τασάκι* (cendrier), *ράκος* (loque), *τραγούδι* (chanson), *λουλούδι* (fleur), etc. ne désignent pas des diminutifs en grec moderne, les mots : *σφαλιάρα* (gifle), *μπάρα* (barre), *κούκλα* (poupée), etc. ne désignent guère des augmentatifs.

De plus, la majorité des formes fléchies que l'on trouve dans les textes posent, également, des « problèmes » de genre (*κορίτσι*/fille est au neutre et *κορίτσαρος*/grande fille au masculin) et de confusion entre forme dérivée et forme standard. Par exemple, le nominatif pluriel *φυλάκια* (postes de garde) se termine par *-άκια*, qui est une des formes du diminutif (*αυτοκινητάκια* (petites voitures), *βαποράκια* (petits navires), *αγγελάκια* (petits anges), etc.) mais sa forme standard n'est pas **φυλάκι* mais *φυλάκιο* (nominatif singulier) qui ne correspond à aucune forme de diminutif ou d'augmentatif. De même, alors que le mot *παλικαράς* (celui qui est très brave) est bien l'augmentatif de *παλικάρι* (le brave), *σφουγγαράς* (pêcheur d'éponges) ne constitue guère l'augmentatif de *σφουγγάρι* (éponge). Nous avons

³ Remarquons que le grec moderne est une langue à cas. Ainsi, pour le mot *αυτοκίνητο* (voiture), les formes fléchies correspondantes de l'augmentatif sont : *αυτοκινητάρα* (Nfs:Afs:Vfs), *αυτοκινητάρας* (Gfs), *αυτοκινητάρες* (Nfp:Afp:Vfp) et *αυτοκινητάρων* (Gfp), où *N, A, V, G*, désignent respectivement le nominatif, l'accusatif, le vocatif et le génitif.

⁴ Il s'agit d'un analyseur automatique de textes multilingue, téléchargeable à l'adresse : <http://www-igm.univ-mlv.fr/~unitex/download.html>.

⁵ <PRE> signifie qu'on exclut les mots en lettres minuscules commençant par une lettre majuscule.

observé que les formes en majuscules posent également des problèmes, puisque l'accent tonique n'est pas marqué explicitement. Ainsi, le motif <<ούλη\$>> reconnaît des formes comme *ΒΟΥΛΗ* (parlement), alors que l'accent se trouve sur la dernière voyelle (*βουλή*).

En outre, la forme diminutive d'un nom peut avoir un sens différent de la forme standard. Par exemple, le nom *βαποράκι* a deux sens : soit il constitue le diminutif du mot *βαπόρι* (navire) et dans ce cas il s'agit d'un « petit navire », soit il désigne un humain, signifiant « le trafiquant de drogue » et il n'a pas de forme standard :

βαποράκι, βαπόρι. N+Dim:Nns (petit navire)
βαποράκι, βαποράκι. N+Hum:Nns (trafiquant de drogues)

Ces arguments nous ont conduits à ne pas opter pour la reconnaissance automatique des formes diminutives et augmentatives à base de règles morphologiques, c'est-à-dire par simple consultation du suffixe des mots repérés, mais pour la construction d'un dictionnaire électronique qui comporterait tous les mots dérivés fléchis et leurs formes standards correspondantes (Dal *et al.*, 1999).

À noter cependant qu'une telle démarche serait utile pour repérer de « nouveaux » mots obtenus par des règles de génération très productives, tels que les mots *zipάκι*⁶ et *zip-άκι* (petit zip), *cdάκι* (petit cd), *FIATάκι* (petite voiture FIAT) qui constituent des diminutifs « hellénisés » des mots *zip*, *cd* et *FIAT*, respectivement. Les requêtes par motifs peuvent dans ce cas s'avérer utiles pour l'enrichissement aisé du dictionnaire.

Concernant le traitement par motifs des formes du comparatif et du superlatif, la situation est différente puisque le taux de reconnaissance est très élevé. Nous avons décrit de la même façon les différents suffixes concernés (cf. figure 2) :



Figure 2. Suffixes des formes du comparatif

Nous constatons que dans le même corpus, 3 157 formes ayant un des suffixes indicatifs du comparatif ont été repérées (637 formes différentes), dont 95 % constituent des formes du comparatif. En ce qui concerne les superlatifs, toutes les formes reconnues, *i.e.* 202 formes (125 formes différentes), constituent des formes du superlatif. Si, dans les règles précédemment exposées, nous incluons les quelques exceptions qui sont en nombre restreint comme *καλός* (bon) – *κάλλιστος* (le meilleur), *κακός* (mauvais) – *κάκιστος* ou *χειρίστος* (le pire), nous garantissons un taux de reconnaissance de 100 %. Mais dans un souci d'homogénéité de la méthode de traitement, nous avons opté pour l'incorporation de ces formes dans le dictionnaire électronique. De plus, une représentation par dictionnaire nous

⁶ Nous remarquons ici le mélange possible entre alphabet latin et grec.

permet de relier aisément les formes dérivées en question (comparatifs, superlatifs et adjectifs de base), information utile, par exemple, dans le cadre d'une analyse sémantique.

3. Le dictionnaire électronique du grec moderne

3.1. Le système DELA

Rappelons que le dictionnaire électronique du grec moderne suit le formalisme DELA mis en œuvre au Laboratoire d'Automatique Documentaire et Linguistique (LADL) depuis plus de 20 ans, pour le français (Courtois *et al.*, 1990 ; Clemenceau, 1993, 1995). Il est construit par l'équipe d'informatique linguistique du Laboratoire de traduction et traitement des langues (LTTL, Grèce) et celle de l'Institut Gaspard-Monge (IGM, France). À l'heure actuelle, il comprend 1 850 000 entrées (simples et composées) appartenant à toutes les catégories grammaticales : noms, adjectifs, verbes, adverbes, noms propres, prépositions, conjonctions, déterminants, pronoms, particules. Le système DELA s'appuie sur une approche de génération morphologique, ce qui nous permet de contrôler pour chaque lemme toutes les formes produites et d'éviter par conséquent la surgénération.

Le dictionnaire électronique se présente sous la forme d'une liste la plus exhaustive possible de formes fléchies du grec moderne, munies d'informations morphologiques, syntaxiques et sémantiques (DelafGr). Il est obtenu à partir d'un dictionnaire de formes canoniques dont la taille est nettement plus faible (DelasGr)⁷. Une entrée du DelasGr se présente sous la forme :

*forme_canonique.code_flexionnel(+info_syntactico-sém.+...)*⁸ (*lcomment*)

Le code flexionnel associé à la forme canonique est composé d'une partie alphabétique, la catégorie du discours du mot, et d'une partie numérique qui fait référence à une table de flexion représentée sous la forme d'une liste de vecteurs flexionnels. Après flexion, on obtient une ligne de type DelafGR de la forme :

forme_fléchie,forme_canonique.cat_gram(+info_syntactico-sém.+...):info_flex:... (*lcomment*)

Les éventuelles informations sémantiques associées aux forme canoniques se présentent sous la forme de traits sémantiques tels que +*Hum* (humain), +*Abst* (abstrait), +*Conc* (concret), +*[Geo]* (géographie), etc. Lors de la flexion d'une entrée, elles sont reproduites à l'identique dans le DelafGR, comme le commentaire (sauf option contraire donnée au programme de flexion).

Considérons le mot *θύελλα* (tempête), la ligne du Delas le décrivant est :

Delas : *θύε"λλα. N235+[Meteo] /Climat*

Le code flexionnel N235 fait apparaître tout d'abord sa catégorie grammaticale *N* (nom), ce qui permet au programme de flexion de retrouver le vecteur d'étiquettes :

$$N.:Nfs,:Gfs,:Afs,:Vfs,:Nfp,:Gfp,:Afp,:Vfp \quad (1)$$

où *Nfs* signifie nominatif (*N*) d'un nom féminin (*f*) au singulier (*s*), *G* est le symbole utilisé pour indiquer le génitif, *A* pour l'accusatif, *V* pour le vocatif et *p* pour le pluriel.

Le vecteur flexionnel de code N235 (cf. Kyriacopoulou *et al.*, 2002) est le suivant :

⁷ La forme canonique des noms est celle du nominatif singulier, des adjectifs le nominatif masculin singulier et des verbes la première personne singulier de l'indicatif présent. À l'heure actuelle, le Delas des noms simples du grec moderne comprend 64 951 formes canoniques et celui des adjectifs simples 37 790 formes canoniques. Ces formes donnent lieu à 268 645 formes fléchies pour les noms et 399 388 formes fléchies pour les adjectifs.

⁸ Les données entre parenthèses sont optionnelles.

$$N235. \alpha, \alpha\varsigma. \textcircled{\cdot} \textcircled{\cdot} \textcircled{\cdot} \eta\varsigma, \alpha, \alpha, \epsilon\varsigma, \textcircled{\cdot} \acute{\omega}\nu, \epsilon\varsigma, \epsilon\varsigma \quad (2)$$

La première composante (ici α) joue un rôle particulier ; elle est toujours considérée comme un opérateur qui indique le suffixe à enlever du mot traité pour obtenir sa racine. C'est à cette dernière, que toutes les composantes du vecteur flexionnel (y compris la première) seront successivement ajoutées pour obtenir les formes fléchies voulues. En ce qui concerne les symboles $\textcircled{\cdot}$ et $\textcircled{\cdot}$, ils représentent respectivement l'effacement de l'accent tonique et son déplacement sur la voyelle indiquée dans la forme canonique par le symbole $\textcircled{\cdot}$. Quant au point « . » présent dans la seconde composante ($\alpha\varsigma. \textcircled{\cdot} \textcircled{\cdot} \eta\varsigma$), il sert à séparer deux variantes flexionnelles⁹ pour le génitif féminin ($\theta\acute{\upsilon}\epsilon\lambda\lambda\alpha\varsigma$ et $\theta\upsilon\acute{\epsilon}\lambda\lambda\eta\varsigma$).

Les vecteurs (1) et (2) sont ainsi utilisés pour générer automatiquement le Delaf, par exemple :

Delaf : $\theta\acute{\upsilon}\epsilon\lambda\lambda\alpha, \theta\acute{\upsilon}\epsilon\lambda\lambda\alpha.N:Nfs:Afs:Vfs+[Meteo] /Climat$
 $\theta\acute{\upsilon}\epsilon\lambda\lambda\alpha\varsigma, \theta\acute{\upsilon}\epsilon\lambda\lambda\alpha.N:Gfs+[Meteo] /Climat$
 $\theta\upsilon\acute{\epsilon}\lambda\lambda\eta\varsigma, \theta\acute{\upsilon}\epsilon\lambda\lambda\alpha.N:Gfs+[Meteo] /Climat$
 $\theta\acute{\upsilon}\epsilon\lambda\lambda\epsilon\varsigma, \theta\acute{\upsilon}\epsilon\lambda\lambda\alpha.N:Nfp:Afp:Vfp+[Meteo] /Climat$
 $\theta\upsilon\epsilon\lambda\lambda\acute{\omega}\nu, \theta\acute{\upsilon}\epsilon\lambda\lambda\alpha.N:Gfp+[Meteo] /Climat$

3.2. Extension du format DELA

Comme il a été mentionné précédemment, dans le format actuel du Delas chaque forme canonique est pourvue d'un **unique** code flexionnel, suivi par des informations sémantiques et éventuellement d'un commentaire.

Format actuel :

forme_canonique.code_flexionnel(+info_syntactico-sém.+...)(/comment)

Le traitement des formes fléchies qui nous intéressent nous obligerait à ajouter dans le vecteur flexionnel associé à la forme de base traitée toutes les terminaisons nécessaires à leur génération. Etant donné que d'une part, un même mot peut accepter, par exemple, un ou plusieurs diminutifs ou augmentatifs, et que d'autre part, deux mots différents acceptent des combinaisons différentes de formes dérivées, il découlerait du format actuel, un accroissement considérable du nombre de codes flexionnels et de la taille des vecteurs flexionnels associés. Pour cette raison, nous considérons une extension du format DELA dans lequel le code flexionnel associé à la forme canonique est suivi d'une liste de règles séparées par des « ; ». Le nouveau format a donc la forme générale suivante :

Format étendu :

forme_canonique.code_flexionnel;R₁;R₂;...;R_i(+info_syntactico-sém.+...)(/comment)

Les règles R_i peuvent se présenter sous deux formes :

Un couple de codes $D_m\#F_n$ où D_m est une règle de dérivation qui permet d'obtenir une forme canonique dérivée à partir de la forme canonique standard, et F_n est un code flexionnel qui s'applique à la forme canonique dérivée obtenue.

Un code, C_p ou S_p , fait référence à une règle de substitution permettant notamment de produire une forme composée à partir des formes simples précédemment générées.

⁹ Les variantes flexionnelles sont très nombreuses en grec moderne. Certaines d'entre elles correspondent à des formes savantes provenant du grec ancien.

Les règles de type $D_m\#F_n$ ou C_p et S_p se présentent sous la forme *opérateurs/infos* où *opérateurs* désigne une suite de transformations à effectuer sur la forme canonique standard et *infos* est une suite d'informations à ajouter à celles préexistantes dans la ligne du Delas. Notons que le nom des règles de dérivation est constitué par une ou deux lettres suivies d'un chiffre. Les lettres utilisées sont indicatives du type de règle : D pour les diminutifs, M pour les augmentatifs, C pour les comparatifs, S pour les superlatifs. Par exemple, la règle $D15.-2^\circ\acute{o}\lambda\eta\varsigma/+Dim$ signifie que l'on enlève deux lettres (-2) à la forme canonique, puis son accent ($^\circ$), et que l'on ajoute le suffixe $\acute{o}\lambda\eta\varsigma$. Enfin, on insère le trait sémantique $+Dim$ dans les informations déjà associées à la forme canonique standard.

Ainsi, si l'on considère l'adjectif $\mu\kappa\rho\acute{o}\varsigma$ (petit), la ligne du Delas : $\mu\kappa\rho\acute{o}\varsigma.A20;D15\#A80;Cp1;S1^{10}$ comporte :

- 1) le code flexionnel A20 qui permet la génération des formes fléchies de l'adjectif $\mu\kappa\rho\acute{o}\varsigma$. Nous obtenons donc les formes fléchies suivantes :

| | |
|--|---|
| $\mu\kappa\rho\acute{o}\varsigma, \mu\kappa\rho\acute{o}\varsigma.A:Nms$ | $\mu\kappa\rho\acute{o}\varsigma, \mu\kappa\rho\acute{o}\varsigma.A:Amp$ |
| $\mu\kappa\rho\acute{o}\upsilon, \mu\kappa\rho\acute{o}\varsigma.A:Gms:Gns$ | $\mu\kappa\rho\acute{o}\eta, \mu\kappa\rho\acute{o}\varsigma.A:Nfs:Afs:Vfs$ |
| $\mu\kappa\rho\acute{o}, \mu\kappa\rho\acute{o}\varsigma.A:Ams:Nns:Ans:Vns$ | $\mu\kappa\rho\acute{o}\eta\varsigma, \mu\kappa\rho\acute{o}\varsigma.A:Gfs$ |
| $\mu\kappa\rho\acute{\epsilon}, \mu\kappa\rho\acute{o}\varsigma.A:Vms$ | $\mu\kappa\rho\acute{\epsilon}\varsigma, \mu\kappa\rho\acute{o}\varsigma.A:Nfp:Afp:Vfp$ |
| $\mu\kappa\rho\acute{o}\iota, \mu\kappa\rho\acute{o}\varsigma.A:Nmp:Vmp$ | $\mu\kappa\rho\acute{o}\alpha, \mu\kappa\rho\acute{o}\varsigma.A:Nnp:Anp:Vnp$ |
| $\mu\kappa\rho\acute{o}\acute{\omega}\nu, \mu\kappa\rho\acute{o}\varsigma.A:Gmp:Gfp:Gnp$ | |

- 2) le couple $D15\#A80$ qui permet d'obtenir le diminutif par la règle $D15$ ($D15.-2^\circ\acute{o}\lambda\eta\varsigma/+Dim$) puis sa flexion par $A80$. Les formes fléchies générées sont :

| | |
|---|---|
| $\mu\kappa\rho\acute{o}\lambda\eta\varsigma, \mu\kappa\rho\acute{o}\varsigma.A+Dim:Nms$ | $\mu\kappa\rho\acute{o}\lambda\alpha\varsigma, \mu\kappa\rho\acute{o}\varsigma.A+Dim:Gfs$ |
| $\mu\kappa\rho\acute{o}\lambda\eta, \mu\kappa\rho\acute{o}\varsigma.A+Dim:Gms:Ams:Vms$ | $\mu\kappa\rho\acute{o}\lambda\epsilon\varsigma, \mu\kappa\rho\acute{o}\varsigma.A+Dim:Nfp:Afp:Vfp$ |
| $\mu\kappa\rho\acute{o}\lambda\eta\delta\epsilon\varsigma, \mu\kappa\rho\acute{o}\varsigma.A+Dim:Nmp:Amp:Vmp$ | $\mu\kappa\rho\acute{o}\lambda\iota, \mu\kappa\rho\acute{o}\varsigma.A+Dim:Nns:Ans:Vns$ |
| $\mu\kappa\rho\acute{o}\lambda\eta\delta\acute{o}\nu, \mu\kappa\rho\acute{o}\varsigma.A+Dim:Gmp$ | $\mu\kappa\rho\acute{o}\lambda\iota\alpha, \mu\kappa\rho\acute{o}\varsigma.A+Dim:Nnp:Anp:Vnp$ |
| $\mu\kappa\rho\acute{o}\lambda\alpha, \mu\kappa\rho\acute{o}\varsigma.A+Dim:Nfs:Afs:Vfs$ | |

- 3) le code $Cp1$ qui construit les formes composées du diminutif précédées par $\pi\iota\omicron$ (plus). Il renvoie à la règle : $Cp1.\pi\iota\omicron \sim/+CompC$ où le symbole \sim représente une forme fléchie, issue de la forme canonique de base, que l'on substitue. Ainsi, à partir de ce code, nous obtenons les formes fléchies suivantes :

| | |
|---|--|
| $\pi\iota\omicron \mu\kappa\rho\acute{o}\varsigma, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Nms$ | $\pi\iota\omicron \mu\kappa\rho\acute{o}\upsilon\varsigma, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Amp$ |
| $\pi\iota\omicron \mu\kappa\rho\acute{o}\upsilon, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Gms:Gns$ | $\pi\iota\omicron \mu\kappa\rho\acute{o}\eta, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Nfs:Afs:Vfs$ |
| $\pi\iota\omicron \mu\kappa\rho\acute{o}, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Ams:Nns:Ans:Vns$ | $\pi\iota\omicron \mu\kappa\rho\acute{o}\eta\varsigma, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Gfs$ |
| $\pi\iota\omicron \mu\kappa\rho\acute{\epsilon}, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Vms$ | $\pi\iota\omicron \mu\kappa\rho\acute{\epsilon}\varsigma, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Nfp:Afp:Vfp$ |
| $\pi\iota\omicron \mu\kappa\rho\acute{o}\iota, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Nmp:Vmp$ | $\pi\iota\omicron \mu\kappa\rho\acute{o}\alpha, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Nnp:Anp:Vn$ |
| $\pi\iota\omicron \mu\kappa\rho\acute{o}\acute{\omega}\nu, \mu\kappa\rho\acute{o}\varsigma.A+CompC:Gmp:Gfp:Gnp$ | |

Une règle de type D_m peut être précédée d'un caractère « > » à la place de « ; », dans ce cas la dérivation ne s'applique pas à la forme canonique standard mais à la dernière forme canonique dérivée. Par exemple, l'entrée $\acute{\alpha}\gamma\gamma\epsilon\text{"}\lambda\omicron\varsigma.N125;D13\#N3331>D1\#N3331$

où $N125$ renvoie à la règle de flexion de la forme standard, $D13$ à la règle de dérivation de la forme diminutive $D13.-2^\circ\acute{o}\acute{\upsilon}\delta\iota/+Dim$, $>D1$ à la règle de dérivation $D1.-1^\circ\acute{\alpha}\kappa\iota/+Dim$ qui s'appliquera à la forme diminutive précédente, générée par le code $D13$, et $N3331$ à la règle de flexion de deux formes diminutives. Nous obtenons ainsi les formes fléchies suivantes :

¹⁰ Le superlatif (S) est traité de façon similaire au comparatif (C).

| | |
|-------------------------------|---|
| ἀγγελος, ἀγγελος. N: Nms | αγγέλους, ἀγγελος. N: Amp |
| αγγέλου, ἀγγελος. N: Gms | αγγελοῦδι, ἀγγελος. N+Dim: Nns: Ans: Vns |
| ἀγγελο, ἀγγελος. N: Ams | αγγελοῦδια, ἀγγελος. N+Dim: Nnp: Anp: Vnp |
| ἀγγελε, ἀγγελος. N: Vms | αγγελουδάκι, ἀγγελος. N+Dim: Nns: Ans: Vns |
| ἀγγελοι, ἀγγελος. N: Nmp: Vmp | αγγελουδάκια, ἀγγελος. N+Dim: Nnp: Anp: Vnp |
| αγγέλων, ἀγγελος. N: Gmp | |

Au total nous comptons 29 règles pour générer les diminutifs des noms et adjectifs, 10 pour les augmentatifs des noms, 5 règles pour les comparatifs réguliers simples et 5 pour les superlatifs réguliers simples.

En ce qui concerne l'augmentatif, l'adjonction du suffixe qui le caractérise implique, dans certains cas, une remontée de l'accent sur la racine de la forme canonique, par exemple : *παιδί* (enfant), *παιδαρος*, alors que la flexion de la forme standard (*παιδί*) n'implique pas de tel déplacement de l'accent. Pour résoudre ce problème, nous avons créé l'opérateur * qui effectue une remontée de l'accent de la voyelle qui le porte à la voyelle précédente, puisque (i) le suffixe de l'augmentatif est dissyllabique et (ii) l'accent tonique peut remonter jusqu'à la troisième syllabe à partir de la fin du mot. La règle permettant de produire la forme canonique de l'augmentatif s'écrit donc : *M6.1. *-I°άρος/+Aug.*

3.3. Génération de dictionnaires « à la demande »

Les règles sont strictement identifiables par leur nom, et peuvent ainsi être **activées** ou **désactivées** lors de la flexion du dictionnaire. Ceci permet de générer les formes désirées selon les applications du Delaf à produire. Il est également possible de choisir quelle forme canonique sera associée à une forme fléchie dérivée : la forme canonique de base ou la forme canonique dérivée. Cette dernière option permet de générer un dictionnaire de type « classique » dans lequel les formes dérivées ne sont mises en relation qu'avec leur propre forme canonique.

De plus, on peut gérer avec précision la production des informations sémantiques associées aux formes générées. Ces informations peuvent provenir de quatre sources distinctes :

Les traits sémantiques qui sont présents dans le Delas : *θύελλα.N235+[Meteo]*. Ces traits apparaissent dans toutes les formes fléchies.

Les traits sémantiques issus du vecteur flexionnel. C'est le cas du trait *+sv* qui signifie que la forme correspondante appartient – morphologiquement – à la langue savante (ou *katharévoussa*). Par exemple : *N235.α,ας. °""ης/+sv,α,α,ες, °ών,ες,ες*

Les traits produits lors de la dérivation : *+Dim* pour les formes du diminutif, *+Aug* pour celles de l'augmentatif, *+Comp* pour celles du comparatif, *+Sup* pour celles du superlatif, etc.

L'adjonction de traits sémantiques spécifiques aux formes dérivées. Par exemple, dans *παιδί.N33403;M6+fam#N121+Hum*, le trait *+fam* indique que la forme *παιδαρος,παιδί. N+Hum+fam:Nms* dénote une nuance familière.

Cette structuration nous permet de choisir quelles sources d'informations seront prises en compte et donc de pouvoir générer des Delaf adaptés à chaque application.

4. Conclusion

Nous avons démontré qu'une grande partie des formes dérivées ne peuvent pas être reconnues automatiquement par des règles morphologiques à cause des problèmes d'ambiguïtés qu'elles engendrent. Pour cette raison, mais aussi dans un souci d'homogénéisation de traitement, nous avons incorporé, dans le dictionnaire électronique, les diminutifs et les augmentatifs des

noms et des adjectifs simples, ainsi que les comparatifs et les superlatifs des adjectifs simples du grec moderne. Ceci nous permet non seulement de traiter aisément les formes dérivées mais aussi de leur associer des informations sémantiques spécifiques qui sont nécessaires à l'analyse automatique. De plus, cette solution permet la liaison des formes dérivées entre elles, puisqu'elles proviennent de la même forme canonique standard (sans que ceci nous empêche de générer un dictionnaire de format classique où chaque mot est associé à sa forme canonique dérivée correspondante). Le format du dictionnaire proposé offre une réduction du nombre des classes flexionnelles mais surtout une meilleure structuration des données et par conséquent une production des formes et des informations modulable en fonction des applications. En effet, un dictionnaire constitue une ressource linguistique nécessaire à diverses applications comme par exemple l'enseignement, la correction orthographique, la génération et bien d'autres. Il est évident qu'on ne souhaite pas proposer une forme savante (+sv) dans le cas d'un correcteur orthographique ou d'un logiciel de didactique des langues.

Références

- CLAIRIS C., BABINIOTIS G. (2005). *Γραμματική της Νέας Ελληνικής. Δομολειτουργική-Επικοινωνιακή*. Ellinika Grammata, Athènes.
- CLEMENCEAU D. (1993). *Structuration du lexique et reconnaissance de mots dérivés*. Thèse de doctorat, Université Paris 7, LADL, CERIL.
- CLEMENCEAU D. (1995). « Étude sur la dérivation en français contemporain ». In *Linguisticae Investigationes* 17 : 115-133.
- COURTOIS B., SILBERZTEIN M. (éds) (1990). « Les dictionnaires électroniques du français ». In *Langue Française* 87 : 11-22.
- DAL G., NATHOUT N., NAMER F. (1999). « Construire un lexique dérivationnel : théorie et réalisation ». In *TALN'99*. Cargèse : 115-124.
- GROSS M. (1989). « La construction de dictionnaires électroniques ». *Annales des Télécommunications* 44 (1-2) : 4-19.
- HOLTON D., MACKRIDGE P., PHILIPPAKI-WARBURTON I. (1997). *Greek grammar. A Comprehensive Grammar of the Modern Language*. Routledge, Londres (éditée en grec : *Γραμματική της Ελληνικής Γλώσσας*, 2000. Editions Patakis, Athènes).
- KYRIACOPOULOU T. (2003). *Analyse automatique des textes écrits : le cas du grec moderne*. Mémoire d'Habilitation à diriger des recherches, Université de Marne-la-Vallée.
- KYRIACOPOULOU T., MRABTI S., YANNAKOPOULOU A. (2002). « Le dictionnaire électronique des noms composés en grec moderne ». In *Linguisticae Investigationes* 25 (1) : 7-28.
- LAPORTE E. (2005). « Lexicon management and standard formats ». In Zygmunt Vetulani (éd.). *Proceedings of the 2nd Language & Technology Conference, April 21-23, 2005, Poznań, Poland*, Wydawnictwo Poznańskie Sp. z o.o. : 318-322.
- SILBERZTEIN M. (1993). *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX*. Masson, Paris.
- SKLAVOUNOU E. (1999). « Les classes de déclinaison des noms et des adjectifs du grec moderne ». In *Actes du 16^e Colloque Européen sur les Lexiques et Grammaires Comparés* (24-27 septembre 1997, Louvain-la-Neuve) : 191-200.
- TRIANAFYLLIDIS M. (2000). *Νεοελληνική Γραμματική της δημοτικής*. Institut d'Etudes néo-helléniques, Fondation Manolis Triantafyllidis, Thessaloniki.