

21. Ορολογία και ηλεκτρονικός ορθογραφικός έλεγχος

**Άννα Αναστασιάδη – Συμεωνίδη
Τίτα Κυριακοπούλου
Γεωργία Νικολάου
Άννα Παναγιώτου – Τριανταφυλλοπούλου
Βασιλική Φούφη**

ΠΕΡΙΛΗΨΗ

Η παρούσα έρευνα, που εκπονείται από το Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης και το Πανεπιστήμιο Κύπρου, συνίσταται στη δημιουργία ενός ελληνικού ηλεκτρονικού ορθογράφου ο οποίος θα διορθώνει γραπτά κείμενα με βάση τους αποδεκτούς τύπους της κοινής νεοελληνικής, της γραφόμενης κυπριακής και της αθλητικής και οικονομικής-χρηματιστηριακής ορολογίας. Η λειτουργία του «Ελληνικού Ορθογράφου» πρόκειται να βασιστεί σε ηλεκτρονικά λεξικά, καθώς και σε τοπικές γραμματικές. Στην ανακοίνωση αυτή θα προβούμε σε μία αδρομερή περιγραφή (α) των γλωσσικών μονάδων που έχουν συλλεγεί και (β) των εργαλείων που πρόκειται να χρησιμοποιηθούν για την επεξεργασία τους.

Terminology and automatic spelling correction

**Anna Anastassiadis-Symeonidis
Tita Kyriacopoulou
Georgia Nikolaou
Anna Panayotou-Trantaphyllopoulou
Vasiliki Foufi**

SUMMARY

This paper deals with the development of a Greek spelling corrector which will be based on electronic simple-word and compound-word dictionaries and on finite-state automata. The spelling correction will be compatible to the grammatical rules of the official Modern Greek language (koine), the written variety of the Cypriot dialect and the specific economic and sports terminology. In the present paper we should give a broad description (a) of our corpus of lexical units and (b) of the tools which are going to be used for its processing.

0 Εισαγωγή

Ο «Ελληνικός Ορθογράφος», μια επιστημονική εφαρμογή που αναπτύσσεται στο πλαίσιο του ερευνητικού προγράμματος ηλεκτρονικής λεξικογραφίας μεταξύ του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης και του Πανεπιστημίου Κύπρου, αποτελεί μια προσπάθεια να καλυφθούν οι ανάγκες των φυσικών και των μη φυσικών ομιλητών της ελληνικής, όσον αφορά την παραγωγή και την επεξεργασία κειμένου σε ηλεκτρονικό περιβάλλον γραμματισμού. Ο ορθογράφος αυτός σχεδιάστηκε και υλοποιείται με τη φιλοδοξία να αντιμετωπίσει τις ελλείψεις και τα προβλήματα άλλων παρόμοιων ηλεκτρονικών προϊόντων που σήμερα κυκλοφορούν στην αγορά και που, παρά τη βοήθεια που προσφέρουν στους χρήστες ηλεκτρονικού υπολογιστή, αδυνατούν να αποτελέσουν ένα πλήρως αξιόπιστο εργαλείο γλωσσικής επεξεργασίας. Πιο συγκεκριμένα, η πρωτοπορία του «Ελληνικού Ορθογράφου» έγκειται σε τρία χαρακτηριστικά: (α) στην κατάρτισή του με βάση ηλεκτρονικά λεξικά της νέας ελληνικής, (β) στον εμπλουτισμό του σε μια μεταγενέστερη φάση εργασίας με στοιχεία από τη γραφόμενη κυπριακή, ιδιότητα που απουσιάζει από όλους τους υπάρχοντες ηλεκτρονικούς διορθωτές ελληνικού κειμένου, και (γ) στη δυνατότητά του να αναγνωρίζει και να διορθώνει όρους ειδικών λεξιλογίων.

1 Γενικά χαρακτηριστικά του ορθογράφου

1.1 Το corpus του ορθογράφου

Η λειτουργία του «Ελληνικού Ορθογράφου» βασίζεται στα ηλεκτρονικά λεξικά της κοινής νέας ελληνικής¹ που σχεδιάστηκαν και αναπτύσσονται από τη μονάδα αυτόματης επεξεργασίας φυσικών γλωσσών (<http://linginfo.frl.auth.gr>) του Εργαστηρίου Μετάφρασης και Επεξεργασίας του Λόγου² υπό την εποπτεία της επίκουρης καθηγήτριας του Α.Π.Θ. Τίτας Κυριακοπούλου. Αξίζει να σημειωθεί ότι η επεξεργασία των μονάδων στα λεξικά αυτά γίνεται με το πρόγραμμα αυτόματης κλίσης *GenereFlexion*, το οποίο, με την προσθήκη κωδικών, παράγει όλους τους κλιτούς τύπους των καταχωρισμένων γλωσσικών μονάδων. Πιο συγκεκριμένα, τα ηλεκτρονικά αυτά λεξικά θα χρησιμοποιηθούν ως το κύριο μέσο για την αναγνώριση όσο το δυνατόν περισσότερων μονάδων από τον ηλεκτρονικό διορθωτή, ενώ για τη διαδικασία του ορθογραφικού ελέγχου στην παραγωγή λόγου, όπου ισχύουν οι ορθογραφικοί και συντακτικοί περιορισμοί των επίσημων έργων κωδικοποίησης της νεοελληνικής (σχολική γραμματική, συντακτικό, λεξικά), δημιουργήθηκε ένα λιγότερο ευρύ

1 18 000 ρήματα, 70 000 απλά ουσιαστικά, 28 000 πολυλεκτικές μονάδες, 40 000 επίθετα, 50 000 κύρια ονόματα, 16 000 απλά και σύνθετα επιρρήματα, 3 000 τοπωνύμια, 1 000 γραμματικές λέξεις.

2 Τμήμα Γαλλικής Γλώσσας και Φιλολογίας, Φιλοσοφική Σχολή, Α.Π.Θ.

corpus, με σημεία αναφοράς το *Λεξικό της Κοινής Νεοελληνικής* του Ινστιτούτου Νεοελληνικών Σπουδών (Ίδρυμα Μανόλη Τριανταφυλλίδη) (1998, 50 000 λήμματα) και το *Αντίστροφο Λεξικό της Νέας Ελληνικής* της Άννας Αναστασιάδη-Συμεωνίδη (2002, 180 000 λήμματα)³. Όσον αφορά την αναγνώριση τύπων της γραφόμενης κυπριακής, διαδικασία η οποία πρόκειται να ολοκληρωθεί σε ένα δεύτερο στάδιο επεξεργασίας του ορθογράφου, η ερευνητική ομάδα προτίθεται να αποδελτιώσει κείμενα του κυπριακού τύπου⁴ και πιο συγκεκριμένα των εφημερίδων *Φιλελεύθερος*, *Πολίτης*, *Χαραυγή* και *Σημερινή*. Από τη μέχρι τώρα διερεύνηση του δείγματος που η ομάδα έχει συλλέξει, έχει προκύψει ότι η κυπριακή ποικιλία, ακόμα και στην επίσημη πραγμάτωσή της σε γραπτό λόγο, παρουσιάζει ορισμένες αποκλίσεις σε σχέση με την κοινή νεοελληνική σε αντίστοιχη χρήση. Οι αποκλίσεις αυτές μπορούν να συνοψιστούν στα εξής:

- αγγλισμοί που έχουν ενσωματωθεί στην κυπριακή σε επίπεδο λεξιλογίου, σύνταξης, μορφολογίας και ορθογραφίας, π.χ. *Θαύλάνδη*
- διπλά σύμφωνα στην αρχή κάποιων λέξεων -κυρίως ονομάτων και δανείων από την τουρκική-, π.χ. *πενεκές*
- άηχα σύμφωνα [p], [t], [k] αντί για τα αντίστοιχα ηχηρά της κοινής νεοελληνικής [b], [d], [g], π.χ. *χάμπερκερ* (<hamburger)
- διαφορές ως προς το χαρακτηριστικό [+/- λόγιο], π.χ. ονόματα μηνών του επίσημου γραπτού λόγου: *Γενάρης*, *Φεβράρης*, *Απρίλλης*
- προσθήκη του γράμματος *ι* σε λέξεις που περιέχουν παχύ συριστικό ή προστριβόμενο, π.χ. *Χατζηγιάννης* (για την απόδοση παχέος προστριβόμενου).

Για το χειρισμό των φαινομένων αυτών, ο ορθογράφος θα πραγματοποιεί τρία είδη σημαδέματος: ένα για τους μη αποδεκτούς τύπους, αυτούς δηλαδή που θεωρούνται εντελώς λαθεμένοι είτε για την κοινή νεοελληνική είτε για τη γραφόμενη κυπριακή, ένα για τους τύπους που είναι αποδεκτοί σύμφωνα με τους κανόνες της κυπριακής ποικιλίας και ένα τρίτο σημάδεμα για ονόματα κυπριακά, τα οποία, εκτός από τον τόνο, είναι ίδια με τα αντίστοιχα ελλαδικά, π.χ. *Παπάμιχαηλ* (κυπριακή) - *Παπαμιχαήλ* (κοινή νεοελληνική).

³ Τα κριτήρια με βάση τα οποία επιλέχτηκαν τα λεξικά αυτά είναι τρία: (α) βιβλιογραφική αναγνώριση (β) εύρος της μακροδομής (χρησιμοποιήσαμε λεξικά που υπερβαίνουν τα 30 000 λήμματα) και (γ) πρόσφατη χρονολογία έκδοσης, για μια όσο το δυνατόν πιο αξιόπιστη περιγραφή της σύγχρονης γλωσσικής πραγματικότητας.

⁴ Υπενθυμίζεται ότι σε όλες τις βαθμίδες της επίσημης εκπαίδευσης στην Κύπρο χρησιμοποιείται ως κανονιστικό εργαλείο της γλωσσικής χρήσης η *Νεοελληνική Γραμματική*, αναπροσαρμογή της Μικρής Νεοελληνικής Γραμματικής του Μανόλη Τριανταφυλλίδη.

Με τη μέθοδο αυτή, ελπίζουμε ότι όχι μόνο θα προβλέψουμε έναν μέγιστο αριθμό γλωσσικών τύπων που πραγματώνονται κατά τη χρήση, αλλά επιπλέον ότι για πρώτη φορά θα ανταποκριθούμε στις απαιτήσεις ενός ευρύτερου κοινού χρηστών, χωρίς να υπεραπλουστεύουμε τη γλωσσική πραγματικότητα και χωρίς να προβαίνουμε σε αυθαίρετες ρυθμίσεις της παραγωγής γραπτού λόγου.

1.2 Το πρόγραμμα αυτόματης κλίσης *GenereFlexion*

Το πρόγραμμα αυτόματης κλίσης *GenereFlexion* [3,4] κατασκευάστηκε στο LADL (Laboratoire d'Automatique Documentaire et Linguistique), είναι γραμμένο σε γλώσσα προγραμματισμού C και δέχεται για επεξεργασία αρχεία μόνο σε μορφή απλού κειμένου (*.txt).

Για την αυτόματη παραγωγή όλων των κλιτών τύπων, το πρόγραμμα χρησιμοποιεί τρία αρχεία. Το πρώτο αρχείο περιλαμβάνει τους λημματικούς τύπους απλών και πολυλεκτικών μονάδων, οι οποίοι συνοδεύονται από ένα σύμβολο που δηλώνει σε ποιο μέρος του λόγου ανήκουν (*N* για το ουσιαστικό, *A* για το επίθετο, *DET* για το άρθρο, *ADV* για το επίρρημα, *CONJ* για το σύνδεσμο και *PREP* για την πρόθεση⁵), τον αντίστοιχο μορφολογικό κωδικό και το σύμβολο «*»* σε περίπτωση που υπάρχει μετατόπιση του τόνου κατά την κλίση.

Μετά το σύμβολο που δηλώνει σε ποιο μέρος του λόγου ανήκει ο λημματικός τύπος τοποθετούνται τα φίλτρα: *S*, όταν το συγκεκριμένο λήμμα χρησιμοποιείται μόνο στον ενικό (π.χ. στην πετοσφαίριση *άσσος* N121,N,S), και *P*, όταν το λήμμα χρησιμοποιείται μόνο στον πληθυντικό⁶ (π.χ. στην αθλητική ορολογία *αποδυτήρι*^α. N301,N,P και *κάμπεις*. N261,N,P).

Στην περίπτωση των πολυλεκτικών μονάδων, μετά το κάθε συνθετικό τοποθετούμε τον αντίστοιχο κωδικό κλίσης, ενώ στο τέλος τοποθετούμε το σύμβολο που δηλώνει σε ποιο μέρος του λόγου ανήκει η σύνθετη λέξη ως σύνολο (π.χ. *αλλαγή*.N251 *σερβίς*.N305,N). Τέλος, στις σύνθετες λέξεις μπορεί να έχουμε περιπτώσεις στις οποίες ένα από τα συνθετικά χρησιμοποιείται σε μία μόνο πτώση. Στην περίπτωση αυτή χρησιμοποιούμε και πάλι φίλτρο που δηλώνει την πτώση και τον αριθμό στον οποίο χρησιμοποιείται το συνθετικό (*N* για την ονομαστική, *G* για τη γενική, *A* για την αιτιατική, *V* για την κλητική⁷ και *S* για τον ενικό και *P* για τον πληθυντικό). Για παράδειγμα, στην περίπτωση του σύνθετου ονόματος *επιτροπή διαίτησίας* το συνθετικό *διαίτησίας* χρησιμοποιείται μόνο στη γενική ενικού και για το λόγο αυτό κάνουμε την εξής καταχώριση:

5 Τα σύμβολα του λατινικού αλφαβήτου προκύπτουν από τις αντίστοιχες αγγλικές ή γαλλικές λέξεις.

6 Τα σύμβολα του λατινικού αλφαβήτου «S» για τον ενικό και «P» για τον πληθυντικό προκύπτουν από τις γαλλικές λέξεις *Singulier* και *Pluriel* αντίστοιχα.

7 Τα σύμβολα του λατινικού αλφαβήτου «N» για την Ονομαστική, «G» για τη Γενική, «A» για την Αιτιατική και «V» για την Κλητική προκύπτουν από τις γαλλικές λέξεις *Nominatif*, *Génitif*, *Accusatif* και *Vocatif* αντίστοιχα.

επιτροπή. N251 δαιτησίας. N232, N, -GS2

Το δεύτερο αρχείο περιλαμβάνει τα κλιτικά διανύσματα. Παρακάτω παρατίθενται τα κλιτικά διανύσματα που αντιστοιχούν στους παραπάνω λημματικούς τύπους:

N251.2.ή,ής,ή,ή,ές,ών,ές,ές

N232.2.α,ας,α,α,ες,°ών,ες,ες

Στο τρίτο αρχείο παράγονται αυτόματα όλοι οι κλιτοί τύποι. Για τα παραπάνω παραδείγματα το πρόγραμμα δίνει τα ακόλουθα αποτελέσματα:

επιτροπή δαιτησίας, .N:Nfs:Afs:Vfs

επιτροπής δαιτησίας, επιτροπή δαιτησίας. N:Gfs

επιτροπές δαιτησίας, επιτροπή δαιτησίας. N:Nfp:Afp:Vfp

επιτροπών δαιτησίας, επιτροπή δαιτησίας. N:Gfp

Όπως προαναφέραμε, κάθε κλιτός τύπος συνοδεύεται από γραμματικές και μορφολογικές πληροφορίες, δηλαδή το λημματικό τύπο, τη γραμματική κατηγορία, το γένος, τον αριθμό και την πτώση. Με κεφαλαία λατινικά γράμματα συμβολίζονται η γραμματική κατηγορία και η πτώση και με μικρά λατινικά γράμματα συμβολίζονται το γένος και ο αριθμός (*ms* για το αρσενικό στον ενικό και *mp* για το αρσενικό στον πληθυντικό, *fs* για το θηλυκό στον ενικό και *fp* για το θηλυκό στον πληθυντικό, *ns* για το ουδέτερο στον ενικό και *np* για το ουδέτερο στον πληθυντικό⁸). Όταν ένας κλιτός τύπος είναι ίδιος σε περισσότερες από μία πτώσεις, τότε εμφανίζεται μία μόνο φορά, αλλά συνοδεύεται από περισσότερους συνδυασμούς συμβόλων που αντιστοιχούν στις πτώσεις, τα γένη και τους αριθμούς. Η αναγνώριση των μορφολογικών και ορθογραφικών παραλλαγών επιτυγχάνεται χάρη στο μορφολογικό λεξικό των κλιτών κατηγοριών (ρημάτων, ουσιαστικών και επιθέτων).

2 Ορολογία

2.1 Περιεχόμενο των ειδικών λεξιλογίων

Στο πεδίο της αναγνώρισης και διόρθωσης ειδικών όρων, ο «Ελληνικός Ορθογράφος» πρόκειται να εμπλουτιστεί με μονάδες από αθλητική⁹ και οικονομική-χρηματιστηριακή ορολογία [6], έτσι όπως αυτές καταγράφηκαν, ομαδοποιήθηκαν και ελέγχθηκαν από τη μονάδα αυτόματης επεξεργασίας φυσικών γλωσσών του Εργαστηρίου Μετάφρασης και

8 Τα σύμβολα του λατινικού αλφαβήτου «ms» για το αρσενικό στον ενικό και «mp» για το αρσενικό στον πληθυντικό, «fs» για το θηλυκό στον ενικό και «fp» για το θηλυκό στον πληθυντικό, «ns» για το ουδέτερο στον ενικό και «np» για το ουδέτερο στον πληθυντικό προκύπτουν από τις γαλλικές λέξεις *masculin singulier*, *masculin pluriel*, *feminin singulier*, *feminin pluriel*, *neutre singulier* και *neutre pluriel* αντίστοιχα.

9 Université de Rennes II, Πρόγραμμα Euradic, στο πλαίσιο του προγράμματος *Technolanguages* (μονόγλωσσες και δίγλωσσες γλωσσολογικές πηγές) / Ειδικό Λεξικό: Παραγωγή πολύγλωσσου λεξικού αθλημάτων.

Επεξεργασίας του Λόγου¹⁰. Οι αθλητικοί όροι ανέρχονται σε 37 000 και διακρίνονται σε δύο ευρείες κατηγορίες: αθλήματα θερινών ολυμπιακών αγώνων και αθλήματα χειμερινών ολυμπιακών αγώνων. Οι οικονομικοί-χρηματιστηριακοί όροι που πρόκειται να ενσωματωθούν στον ορθογράφο αποτελούν ένα σύνολο 8 500 περίπου μονάδων.

2.2 Τυπολογία των μονάδων ειδικού λεξιλογίου

2.2.1 Γενικές κατηγορίες

Από την άποψη της μορφής, οι μονάδες ειδικού λεξιλογίου που θα αναγνωρίζονται από τον ορθογράφο χωρίζονται στις ακόλουθες κατηγορίες:

- μονάδες απλές (ή μονολεκτικές), π.χ. *φυλακίζω* (= δεσμεύω τα κεφάλαιά μου σε επενδύσεις με μακροπρόθεσμη απόδοση)
- μονάδες σύνθετες (ή πολυλεκτικές) [1, 4], *άλμα επί κοντώ*, *μαύρο χρώμα*
- αρκτικόλεξα και ακρωνύμια [3], π.χ. *ΧΑΑ* (= Χρηματιστήριο Αξιών Αθηνών), και *ΝΕ.Χ.Α.* (και *ΝΕΧΑ*) (= Νέα Χρηματιστηριακή Αγορά)
- σύμβολα που απαντούν στο ειδικό λεξιλόγιο των οικονομικών όρων και αποτελούν έναν καινούργιο τύπο γλωσσικών μονάδων που χρησιμοποιούνται από συγκεκριμένες ομάδες χρηστών, π.χ. *T* (= η ημέρα κατάρτισης πράξης), *T+1* (= η επόμενη εργάσιμη ημέρα από την ημέρα πράξης), *T+2* (= η μεθεπόμενη εργάσιμη ημέρα από την ημέρα πράξης).

2.2.3 Πολυλεκτικές μονάδες

Όσον αφορά τις πολυλεκτικές μονάδες ειδικότερα, ο ορθογράφος πρόκειται να αναγνωρίζει και να ελέγχει γλωσσικούς σχηματισμούς των παρακάτω τύπων:

- επίθετο + ουσιαστικό, π.χ. *έγγραφη απόδειξη*, *εξωτερικό λάκτισμα*, *μαύρη ζώνη*, *ομαδική εκκίνηση*
- επίθετο/-α ή μετοχή/-ές + επίθετο/-α + ουσιαστικό σε ονομαστική (+ ουσιαστικό σε γενική), π.χ. *εγγυημένη ανανεούμενη πιστωτική διευκόλυνση*, *έντοκα γραμμάτια δημοσίου*
- ουσιαστικό + ουσιαστικό (και τα δύο σε ονομαστική πτώση), π.χ. *βιομηχανία κλειδί*, *σουτ βολίδα*
- ουσιαστικό + πρόθεση + ουσιαστικό σε αιτιατική ή γενική, π.χ. *εγγύηση με περιορισμούς*, *αντιστάθμιση έναντι του πληθωρισμού*, *μήκος της πίστας*, *σώμα με σώμα*

10 Τμήμα Γαλλικής Γλώσσας και Φιλολογίας, Φιλοσοφική Σχολή, Α.Π.Θ.

- ουσιαστικό + ουσιαστικό/-ά σε γενική πτώση (με ή χωρίς άρθρο), π.χ. *εγγύηση επιτοκίου, έγκριση ενάρξεως εργασιών, ναός της Σοφοκλέους*
- ουσιαστικό σε ονομαστική + σύμβολο (+ ουσιαστικό σε γενική πτώση), π.χ. *μετοχή Α' (κατηγορίας), διαιτητής 1^{ης} κατηγορίας* ή ουσιαστικό σε ονομαστική + ουσιαστικό σε γενική πτώση + σύμβολο, π.χ. *ομολογία κατηγορίας 2*
- ουσιαστικό + επίρρημα και το αντίστροφο, π.χ. *κίνηση όπισθεν, οιονεί αγορά*
- προθετικά σύνολα με επιρρηματική χρήση, π.χ. *άνευ συμμετοχής στα κέρδη, υπέρ το άρτιον* (= όταν εκδίδονται νέες μετοχές με τιμή διάθεσης μεγαλύτερη αυτής της ονομαστικής αξίας τους)
- αρκτικόλεξο + πολυλεκτική μονάδα, π.χ. *Δ.Τ. ακίνητης περιουσίας*
- επιρρηματικό συμπλήρωμα, π.χ. *ακριβώς στο χρόνο*.

2.2.3 Ημιπαγιωμένες εκφράσεις

Οι ημιπαγιωμένες εκφράσεις [3] αποτελούν μία ιδιαίτερη κατηγορία γλωσσικών τύπων, οι οποίοι δημιουργούν ποικίλους συνδυασμούς μέσα στο λόγο (π.χ. στη χρηματιστηριακή ορολογία συναντούμε τις εκφράσεις *παίζω αέρα* και *παίζω με αέρα*). Για το λόγο αυτό, είναι προτιμότερο να μην τις αναπαραστήσουμε με λεξικά αλλά με τοπικές γραμματικές, μέσω της διαδικασίας πεπερασμένων αυτομάτων [7]. Στο πλαίσιο της αυτόματης επεξεργασίας των φυσικών γλωσσών, οι τυποποιημένες γραμματικές αναπαριστούν με γράφους πεπερασμένων αυτομάτων ή μεταβιβαστών μορφολογικές και συντακτικές πληροφορίες. Με τα πεπερασμένα αυτόματα είναι δυνατή μόνο η αναγνώριση των φραστικών ακολουθιών μέσα σε ένα σώμα κειμένου, ενώ με τους μεταβιβαστές πεπερασμένων καταστάσεων παράγονται και νέες πληροφορίες (μορφολογικές, συντακτικές, σημασιολογικές) σχετικές με τις φραστικές ακολουθίες που έχουμε ήδη εισαγάγει [8].

Ένα παράδειγμα πεπερασμένου αυτόματου δίνεται στο τέλος του άρθρου, πριν τις βιβλιογραφικές παραπομπές.

2.3 Γενικότερες παρατηρήσεις στις μονάδες ειδικών λεξιλογίων

2.3.1 Μορφή

Από τη μέχρι τώρα μελέτη των λημμάτων του ειδικού λεξιλογίου, προκύπτουν ορισμένα βασικά χαρακτηριστικά για τις γλωσσικές μονάδες, τα οποία συνοψίζονται στα εξής:

(α) ένα μεγάλο μέρος των καταγεγραμμένων όρων αποτελούν μεταφραστικά δάνεια [2] π.χ. *σκιάδης εταιρ(ε)ία* (< shadow company), ενώ ένας αξιόλογος αριθμός μονάδων απαρτίζεται από μη προσαρμοσμένα δάνεια κυρίως από την αγγλοαμερικανική και τη γαλλική, π.χ. *λίτζι(ν)γκ, μουαγ(ι)έν*

(β) στο σύνολο των καταχωρισμένων μονολεκτικών (απλών) όρων, παρατηρήθηκαν και περιπτώσεις παράγωγων μονάδων που κατασκευάζονται συχνά με την προσθήκη επιθήματος σε δάνεια βάση, π.χ. *σορτάρω* (= πουλώ αέρα, δηλαδή μετοχές που δεν κατέχω, τον οποίο ελπίζω να "κλείσω" φτηνότερα, αποκομίζοντας το αντίστοιχο κέρδος), *σορτάκιας* (= ο πωλητής αέρα, δηλαδή μετοχών που δεν κατέχει, τον οποίο ελπίζει να "κλείσει" φτηνότερα, αποκομίζοντας το αντίστοιχο κέρδος)

(γ) σε αρκετές περιπτώσεις μονάδων έχει παρατηρηθεί ορθογραφική πολυτυπία, με ισοδύναμη χρήση όλων των γλωσσικών τύπων που αντιστοιχούν σε κάθε μονάδα, π.χ. *αντιντάμπι(ν)γκ / αντι-ντάμπι(ν)γκ*

(δ) λόγω του ότι βρισκόμαστε σε πεδίο επιστημονικού λεξιλογίου, παρουσιάζονται αρκετοί λόγοι τύποι τόσο σε επίπεδο μορφολογίας (καταλήξεις γενικής σε *-έως*, π.χ. *ομόλογο ολοκληρώσεως*) όσο και σε επίπεδο λεξιλογικών επιλογών (χρήση λόγιων προθέσεων, συνδέσμων κτλ., π.χ. *υπόλοιπο εις νέου*).

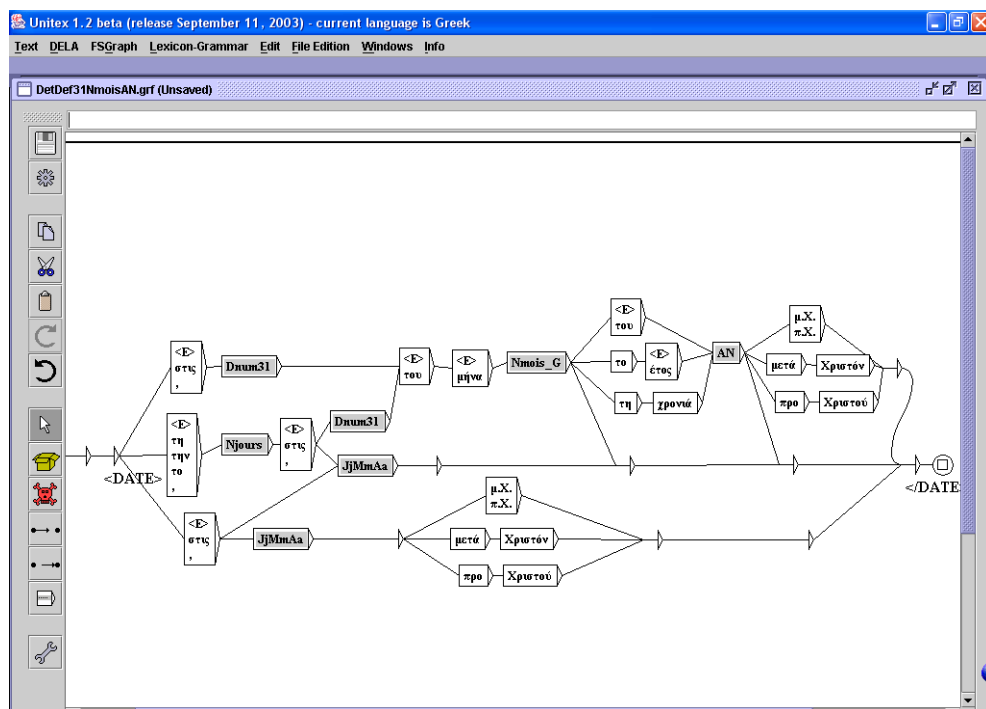
2.4 Πλεονεκτήματα αναγνώρισης και ελέγχου των ειδικών όρων

Η αναγνώριση και ο έλεγχος των όρων ειδικού λεξιλογίου από τον «Ελληνικό Ορθογράφο» αναβαθμίζει σημαντικά την αποτελεσματικότητα της λειτουργίας του, καθώς:

- ο αριθμός των μονάδων που θα αναγνωρίζονται και θα ελέγχονται από το διορθωτή αυξάνεται σημαντικά, ενώ αξίζει να σημειωθεί ότι πολλές από τις μονάδες αυτές δεν καταγράφονται σε κανένα άλλο οργανωμένο σύνολο γλωσσικών δεδομένων (λεξικά, βάσεις δεδομένων), με αποτέλεσμα ο «Ελληνικός Ορθογράφος» να καθίσταται μοναδικό εργαλείο γλωσσικής επεξεργασίας
- στοιχεία που ενδεχομένως θα σημαδεύονταν αν θεωρούνταν αυτόνομες μονάδες μέσα στο κείμενο (π.χ. λόγιες καταλήξεις) τώρα θα αντιμετωπίζονται ως μέρη ενός πολυλεκτικού σχηματισμού που θα αναγνωρίζεται ως ληματικός τύπος από τον ορθογραφικό ελεγκτή
- σε περίπτωση που μία πολυλεκτική μονάδα παρουσιάσει απόκλιση σε σχέση με τον παγιωμένο τύπο που χρησιμοποιείται στο αντίστοιχο ειδικό λεξιλόγιο, ο ορθογράφος θα προτείνει άμεση διόρθωση, ακόμη και αν πρόκειται για την αλλαγή μιας πτώσης ή μιας απλής κατάληξης σε κλιτό τύπο.

Είναι προφανές ότι με την προσθήκη της ειδικής ορολογίας στην αποθήκη δεδομένων που θα αναγνωρίζει ο «Ελληνικός Ορθογράφος» αυξάνεται η αξιοπιστία του ως εργαλείου γλωσσικής επεξεργασίας, ενώ παράλληλα επιτυγχάνεται ένας αρμονικός συνδυασμός στο

χειρισμό γενικού και ειδικών λεξιλογίων, πάντα προς όφελος του χρήστη ηλεκτρονικού υπολογιστή.



Πρόκειται για την αναπαράσταση των εκφράσεων που δηλώνουν την ημερομηνία.

Βιβλιογραφικές παραπομπές

- [1] Αναστασιάδη-Συμεωνίδη, Α. *Η Νεολογία στην Κοινή Νεοελληνική*: Επιστημονική Επετηρίδα της Φιλοσοφικής Σχολής του Α.Π.Θ., Παράρτημα 65, Θεσσαλονίκη, 1986.
- [2] Αναστασιάδη-Συμεωνίδη, Α. *Νεολογικός δανεισμός στη νεοελληνική. Δάνεια από τη γαλλική και την αγγλοαμερικανική - Μορφολογική ανάλυση*, Εστία, 1994.
- [3] Gaston, G. *Les expressions figées en français. Noms composés et autres locutions*, Ophrys, 1996.
- [4] Kyriacopoulou, T. - Mrabti, S. - Yannacopoulou, A. «Le dictionnaire électronique des noms composés en grec moderne», *Linguisticæ Investigations* 25:1, Amsterdam/Philadelphia, John Benjamins, 2002, pp. 7-28.
- [5] Kyriacopoulou, T. *L'analyse automatique des textes écrits : le cas du grec moderne*. University Studio Press, Θεσσαλονίκη, 2005.
- [6] Μήτσκα, Α. - Μιχαηλίδης, Ι. «Αυτόματη επεξεργασία αρκτικολέξων και ακρωνύμων νέας ελληνικής», *Πρακτικά της 24ης Ετήσιας Συνάντησης του Τομέα Γλωσσολογίας του Τμήματος Φιλολογίας του Α.Π.Θ.*, 2003, 482-493.
- [7] Paumier, S. *Unitex. Manuel d'utilisation*, Université de Marne-la-Vallée, 2002.

<http://www-igm.univ-mlv.fr/~unitex> .

[8] Silberztein, M. *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX*, sous la direction de Maurice Gross, Collection Informatique Linguistique, Masson, Paris, 1993.

[9] Silberztein, M. *INTEX and the processing of natural languages*, Tutorial, 1999. Notes: <http://www.nyu.edu/pages/linguistics/intex/> .

[10] Τζιάφα, Ε. *Ένα Σύστημα Αυτόματης Ανάλυσης Κειμένων της Νέας Ελληνικής: Μέθοδοι Μορφολογικής Περιγραφής και Ηλεκτρονικής Κωδικοποίησης των Όρων του Χρηματιστηρίου*, αδημοσίευτη διπλωματική εργασία, Θεσσαλονίκη, ΑΠΘ-Φιλοσοφική Σχολή-Τμήμα Γαλλικής Γλώσσας και Φιλολογίας-Σπουδαστήριο της Διδακτικής των Ζωντανών Γλωσσών, 2005.

Άννα Αναστασιάδη – Συμεωνίδη

Καθηγήτρια Γλωσσολογίας του Α.Π.Θ.

Ταχ. διεύθυνση: Α.Π.Θ. Τμήμα Φιλολογίας GR-54124, Θεσσαλονίκη

Τηλέφωνο & Τηλεμοιότυπο: +30 2310 997015

Ηλ-διεύθυνση: ansym@lit.auth.gr

Τίτα Κυριακοπούλου

Επίκουρη Καθηγήτρια του Α.Π.Θ. Ταχ.διεύθυνση: Α.Π.Θ. Τμήμα Γαλλικής Γλώσσας και Φιλολογίας GR-54124, Θεσσαλονίκη

Τηλέφωνο: +30 2310 997516

Ηλ-διεύθυνση: tita@frl.auth.gr

Γεωργία Νικολάου

Υποψήφια Διδάκτωρ, Τμήμα Φιλολογίας.

Ταχ.διεύθυνση: Α.Π.Θ. Τμήμα Γαλλικής Γλώσσας και Φιλολογίας GR-54124, Θεσσαλονίκη

Τηλέφωνο: +30 2310 997516

Ηλ-διεύθυνση: ngeorgia@lit.auth.gr

Άννα Παναγιώτου – Τριανταφυλλοπούλου

Αναπληρώτρια Καθηγήτρια Γλωσσολογίας του Πανεπιστημίου Κύπρου

Ταχ.διεύθυνση: Πανεπιστήμιο Κύπρου, Τμήμα Κλασικών Σπουδών και Φιλοσοφίας, Τ.Κ. 20537, CY-1678 Λευκωσία

Τηλέφωνο: +357 22892355

Τηλεμοιότυπο: +357 22756274

Ηλ-διεύθυνση: gppanay@ucy.ac.cy

Βασιλική Φούφη

Μεταπτυχιακή φοιτήτρια του Διατμηματικού Προγράμματος Μεταπτυχιακών Σπουδών του Α.Π.Θ.

Ταχ.διεύθυνση: Α.Π.Θ. Τμήμα Γαλλικής Γλώσσας και Φιλολογίας GR-54124, Θεσσαλονίκη

Τηλέφωνο: +30 2310 997516

Ηλ-διεύθυνση: foufivaso@hotmail.com