



HAL
open science

Désambiguïisation des adjectifs à emploi nominal et des adverbes

Dimitra Kazantzi, Rania Voskaki

► **To cite this version:**

Dimitra Kazantzi, Rania Voskaki. Désambiguïisation des adjectifs à emploi nominal et des adverbes. 6ème Colloque international de linguistique grecque, 18-21 septembre 2003, Rethymno, Grèce, 2004, Greece. 10pp. hal-00621518

HAL Id: hal-00621518

<https://hal.science/hal-00621518v1>

Submitted on 3 Oct 2011

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.

ΕΞΑΛΕΙΨΗ ΑΜΦΙΣΗΜΙΩΝ ΣΕ ΟΥΣΙΑΣΤΙΚΟΠΟΙΗΜΕΝΑ ΕΠΙΘΕΤΑ ΚΑΙ ΕΠΙΡΡΗΜΑΤΑ

Ράνια Βοσκάκη
Δήμητρα Καζαντζή

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ
ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

Abstract

During the process of enrichment of an automatic processing system in Modern Greek, we confronted the problem of ambiguity between adverbs and adjectives with nominal use. The object of this study is the retrieval of these ambiguities on the one hand, and the efforts to eliminate them on the other hand. We were based on the theoretical principles of Z. S. Harris as well as on the studies of Maurice Gross, who proposed the complete and systematic description of each language, aiming at the automatic text processing. In particular, the basis was a corpus of 6.000 simple adverbs and a corpus of ~2.000 adjectives with nominal use. The verification of the analysis performed, was realised by UNITEX, an automatic processing system that was developed in LADL (Laboratoire d'Automatique et Documentaire Linguistique) of the University of Marne-la-Vallee. During the research, we have come to the following conclusions:

- There is ambiguity when an adjective with nominal use in neuter gender is found in the nominative, accusative and vocative of plural. For example:

1.0. *καθημερινά/ιδιαίτερα (daily/specially)*, adverb

1.1. *καθημερινά/ιδιαίτερα (everyday clothes/private lessons)*, adjective with nominal use

Φοράει τα καθημερινά της (She is wearing her everyday clothes)

Παραδίδει ιδιαίτερα (He offers private lessons)

Έρχεται καθημερινά και με βλέπει (He comes on a daily basis and meets me)

Ενδιαφέρεται ιδιαίτερα για τη λογοτεχνία (He is specially interested in literature)

In order to resolve such ambiguities, there are the following solutions:

- finite state automata,
- creation of syntactical-semantic analysers,
- classes of objects.

Λέξεις - κλειδιά

αμφισημίες, ουσιαστικοποιημένα επίθετα, επιρρήματα, πεπερασμένα αυτόματα, λεξικό-γραμματική, τάξεις αντικειμένων, μετασχηματιστική γραμματική, συστήματα αυτόματης ανάλυσης, εγκιβωτισμένο γράφημα, μεταβιβαστές, πολυλεκτικά σύνθετα.

1. Εισαγωγή

Αντικείμενο της παρούσας μελέτης είναι ο εντοπισμός των αμφισημιών που παρατηρούνται ανάμεσα στα ουσιαστικοποιημένα επίθετα και στα επιρρήματα της Νέας Ελληνικής, με σκοπό την αναγνώρισή τους από συστήματα αυτόματης ανάλυσης φυσικών γλωσσών. Εντάσσεται στο γενικότερο πλαίσιο γλωσσολογικής ανάλυσης της Νέας Ελληνικής με απώτερο στόχο την αυστηρά τυποποιημένη περιγραφή της.

Συγκεκριμένα, η έρευνα βασίστηκε στις θεωρητικές αρχές της υπολογιστικής γλωσσολογίας, που τέθηκαν από τον Z. S. Harris (:1976) κατά τη δεκαετία του '50 καθώς και στην τυποποίηση και επεξεργασία των μοντέλων του από το μαθητή του M. Gross. Πρόκειται για την εφαρμογή των προτεινόμενων μοντέλων στις φυσικές γλώσσες και στην ανάπτυξη μιας γενικής θεωρίας για τους κωδικούς, που θεμελιώνει την αλγεβρική θεωρία των πεπερασμένων

αυτομάτων και των τυποποιημένων γλωσσών. Σύμφωνα με την παραπάνω θεωρία, είναι εφικτό να συνδυαστεί ο συντακτικός πλούτος μιας φυσικής γλώσσας με την ακρίβεια μιας τυποποιημένης σημασιολογίας, η οποία καθίσταται επεξεργάσιμη μέσω αλγεβρικών τεχνικών.

Επιπλέον, υιοθετήθηκε ο φορμαλισμός που παρέχεται μέσω του μηχανισμού του Λεξικού-Γραμματική, έτσι όπως τον πρότεινε ο M. Gross (:1977) και εφαρμόστηκε στο LADL (Laboratoire d'Automatique et Documentaire Linguistique). Χρησιμοποιήθηκαν, παράλληλα, οι διαδικασίες αυτοματισμού πεπερασμένων καταστάσεων (Silberztein:1993,1999), που υλοποιούνται από συστήματα αυτόματης ανάλυσης κειμένων.

Η έρευνα στηρίχθηκε σε σώματα κειμένων αποτελούμενα από: 40.000 επίθετα, 6.000 απλά επιρρήματα, 2.000 ουσιαστικοποιημένα επίθετα και η επαλήθευση των αποτελεσμάτων των εφαρμογών που δημιουργήθηκαν έγινε σ' ένα σώμα κειμένων 280 MB (~45.000.000 λέξεων), που παραχωρήθηκε από τον καθηγητή Franz Guenther του Πανεπιστημίου του Μονάχου (Center for Information & Language Processing, Ludwig-Maximilian Universitat Munchen).

Ως εργαλείο ελέγχου και επαλήθευσης των αναλύσεων που πραγματοποιήθηκαν, χρησιμοποιήθηκε το UNITEX, ένα σύστημα αυτόματης ανάλυσης που αναπτύχθηκε στο IGM (Institut Gaspard Monge) του Πανεπιστημίου της Marne-la-Vallee.

2. Ουσιαστικοποιημένα Επίθετα

Ως ουσιαστικοποιημένα επίθετα χαρακτηρίζονται τα επίθετα που επέχουν στην πρόταση θέση ουσιαστικού. Τα επίθετα αυτά λειτουργούν ως κεφαλή της ονοματικής φράσης, αντικαθιστώντας το προσδιοριζόμενο ουσιαστικό. Προέρχονται, συνήθως, από μη μονολεκτικές λεξικές μονάδες ή σύνθετα, έτσι όπως τα ορίζει η Α. Αναστασιάδη-Συμεωνίδη (1986: σ.133). Πρόκειται για ένα επίθετο και ένα ουσιαστικό, το οποίο συχνά παραλείπεται, είτε επειδή εννοείται, είτε επειδή έχει πλέον αντικατασταθεί πλήρως από το επίθετο. Στην περίπτωση αυτή έχουν αποκτήσει μια ελλειπτική χρήση σε συγκεκριμένες σημασίες. Επιπλέον, το επίθετο συμφωνεί ως προς το γένος και τον αριθμό με το ουσιαστικό που έχει παραλειφθεί. Σε άλλες περιπτώσεις, πρόκειται για τις ελλειπτικές χρήσεις ενός επιθέτου που δηλώνει ένα ουσιαστικό, το οποίο έχει ήδη αναφερθεί στο λόγο. Αυτό σημαίνει ότι ο συγκεκριμένος τρόπος ουσιαστικοποίησης του επιθέτου είναι συχνά παραγωγικός, με την έννοια ότι εύκολα μπορούν να δημιουργηθούν νέα ουσιαστικοποιημένα επίθετα. Για παράδειγμα:

αμοιβαία (A') κεφάλαια (N), (N) → αμοιβαία (N)

ιδιαίτερα (A) μαθήματα (N), (N) → ιδιαίτερα (N)

λαϊκά (A) τραγούδια (N), (N) → λαϊκά (N)

αντικειμενικά (A) κριτήρια (N), (N) → αντικειμενικά (N)

τοξικά (A) απόβλητα (N), (N) → τοξικά (N)

Υπάρχουν, όμως, και ορισμένα ουσιαστικοποιημένα επίθετα που δε συνιστούν λεξικές φράσεις με κάποιο προσδιοριζόμενο ουσιαστικό, αλλά εκφράζονται συγχρονικά μόνο ως ουσιαστικά (Α. Anastassiadis, Α. Efthimiou, Α. Fliatouras: 2002). Ενδεικτικά είναι τα παρακάτω παραδείγματα: *γλυκά, ζωντανά, κ.α.*

Ακόμη, μεγάλος αριθμός επιθέτων, κυρίως ουδετέρου γένους, στέκονται στο λόγο εντελώς μόνα τους όταν χρησιμοποιούνται ως αφηρημένα ουσιαστικά. Π.χ.

Πολλά άδικα βλέπει κανείς στον κόσμο.

Από τα παραπάνω, καθίσταται προφανές ότι τα όρια ανάμεσα στις γραμματικές κατηγορίες των λέξεων είναι ρευστά και ότι η θέση τους μέσα στο συγκείμενο είναι αυτή που συχνά καθορίζει το χαρακτηρισμό τους σε μορφολογικό επίπεδο.

3. Αμφισημίες

Οι αμφισημίες, ως εγγενή στοιχεία της φυσικής γλώσσας, συναντώνται σε όλα τα επίπεδα μελέτης της φυσικής γλώσσας: στο μορφολογικό, συντακτικό, σημασιολογικό, ορθογραφικό. Αυτές, τις περισσότερες φορές, δε γίνονται αντιληπτές από το φυσικό ομιλητή μιας γλώσσας, αφού διαθέτει τη διαισθητική γνώση της δομής και της λειτουργίας της γλώσσας του, γίνονται όμως αντιληπτές παρατηρώντας τα αποτελέσματα μιας αυτόματης μορφολογικής ανάλυσης σε ηλεκτρονικά σώματα κειμένων.

Η αμφισημία εντοπίζεται στον τύπο της ονομαστικής, αιτιατικής και κλητικής του ουδετέρου αρκετών ουσιαστικοποιημένων επιθέτων στον πληθυντικό αριθμό, ο οποίος συμπίπτει με τα επιρρήματα σε *-a*. Π.χ. *εισαγωγικά, κοινωνικά*. Σε κάποια μάλιστα επίθετα που σχηματίζουν το θηλυκό σε *-a*, στους παραπάνω τύπους προστίθεται και ο τύπος της ονομαστικής, αιτιατικής και κλητικής του θηλυκού στον ενικό, ο οποίος αντιστοιχεί με τον τύπο των επιρρημάτων. Π.χ. *δεξιά, πλούσια*.

Στην παρούσα έρευνα αναζητήθηκαν μέσω των συστημάτων αυτόματης ανάλυσης (INTEX, UNITEXⁱⁱ) επιρρήματα και ουσιαστικοποιημένα επίθετα και διαπιστώθηκε ότι και στις δύο περιπτώσεις δίνονται τα ίδια ακριβώς αποτελέσματα. Αυτό σημαίνει ότι τα συστήματα αυτά δεν είναι ικανά να διακρίνουν, σε μορφολογικό επίπεδο, τη διαφορά ανάμεσα στα δύο μέρη του λόγου επίρρημα και ουσιαστικοποιημένο επίθετο, γεγονός που οφείλεται στο ότι το λεξικό που χρησιμοποιούν διαχειρίζεται τη γνώση σε επίπεδο λέξης (Γ. Ορφανός, Π. Γάκης, Α. Ιορδανίδου: 2000). Ενδεικτικά παρατίθενται δύο φράσεις από τα σώματα κειμένων που υποβλήθηκαν σε επεξεργασία, όπου η λέξη *αντικειμενικά* είναι αμφίσημη για τον υπολογιστή:

*1.α. Στόχος του υπουργείου είναι η αύξηση του φόρου που θα καταβάλουν το 1998 όσοι υπάρχουν στα **αντικειμενικά** να κομανθεί στα επίπεδα του 20% κατά μέσον όρο*

*1.β. Η αυτοψία μπορεί, να μη διενεργείται, μετά από αιτιολογημένη απόφαση του προϊσταμένου της Δ.Ο.Υ. στις περιπτώσεις που λόγοι **αντικειμενικά** δυσχερείς συνηγορούν προς αυτό*

Είναι προφανές, για ένα φυσικό ομιλητή της Ελληνικής γλώσσας, ότι στο πρώτο απόσπασμα η λέξη *αντικειμενικά* είναι ένα επίθετο σε θέση ουσιαστικού, ενώ στο δεύτερο λειτουργεί ως επίρρημα που προσδιορίζει το επίθετο *δυσχερείς*. Κατά συνέπεια, η συγκεκριμένη αμφισημία είναι τεχνητή και όχι πραγματική (Laporte-Monceaux:1999), εφόσον αυτή υφίσταται για τον

υπολογιστή και όχι για τους φυσικούς ομιλητές. Όπως παρατηρείται κατά την αυτόματη ανάλυση, η αμφισημία εντοπίζεται μετά από την εφαρμογή του συνόλου των λεξικών και δεν επιλύεται χωρίς τη συντακτική ανάλυση του κειμένου. Αυτό συμβαίνει γιατί οι λεξικές μονάδες δεν είναι οντότητες απομονωμένες, κλειστές στον εαυτό τους, αλλά παίρνουν αξία και αποκτούν σημασία μέσα από το συγκεκριμένο στο οποίο εμφανίζονται (M. Mathieu-Colas, C. Girardin: 1997). Μ' αυτό τον τρόπο αναδεικνύεται η φράση ως ελάχιστη συντακτικο-σημασιολογική μονάδα. Επομένως, θα πρέπει να εξεταστούν οι συντακτικές ιδιότητες κάθε λέξης, δηλαδή το συγκεκριμένο στο οποίο συναντάται και η σημασία που παίρνει μέσα σ' αυτό.

4. Περιοριστικοί κανόνες εμφάνισης επιρρημάτων

Όσον αφορά τα επιρρήματα, κρίθηκε αρχικά απαραίτητο να διερευνηθούν περιοριστικοί κανόνες εμφάνισής τους μέσα στα σώματα κειμένων. Λαμβάνοντας υπόψη τη βασική συντακτική λειτουργία των επιρρημάτων (Φιλιππάκη:1998) ως προσδιορισμοί ρήματος, επιθέτου, δευτέρου ή και τρίτου επιρρηματος, ουσιαστικού και αριθμητικού ή ποσοδεικτικού, έγινε η καταγραφή εκείνων των συντακτικών δομών που δημιουργούν προβλήματα αμφισημίας σε σχέση με τα ουσιαστικοποιημένα επίθετα κατά την αυτόματη ανάλυση.

Το αποτέλεσμα της έρευνας ήταν οι παρακάτω περιορισμοί, που σε καμία περίπτωση δεν είναι εξαντλητικοί, αλλά εξαλείφουν σε ικανοποιητικό βαθμό τις υπάρχουσες αμφισημίες.

1. Ρήμα + Επίρρημα + (Προσδιοριστής + Eⁱⁱⁱ) + Αριθμητικό/ Ποσοδεικτικό

Η συγκεκριμένη ακολουθία αφορά επιρρήματα που λειτουργούν επιτακτικά στα αριθμητικά και ποσοδεικτικά που προσδιορίζουν. Π.χ.:

...η Ευρωπαϊκή Ένωση, η οποία είναι αδιαφιλονίκητα πρώτη στην οικονομική συμβολή...

2. Προσδιοριστής + Επίρρημα + Επίθετο + (Όνομα + E)

Σ' αυτή την περίπτωση καταγράφεται η δυνατότητα ένα επίρρημα να προσδιορίζει ένα επίθετο που προηγείται ή όχι ενός ονόματος. Οι κατηγορίες επιρρημάτων που μπορούν να προηγηθούν είναι κυρίως τροπικά και λιγότερο συχνά χρονικά. Π.χ.

...υποχρεωτικά πρέπει να αποβάλει τον ανεκτικό περιεκτικό χαρακτήρα της ...

3. Ρήμα + Επίρρημα + Προσδιοριστής + (Επίθετο + E) + Όνομα

Εν προκειμένω εξετάζεται η δομή που περιλαμβάνει επίρρημα το οποίο λειτουργεί ως προσδιορισμός ρήματος και παρεμβάλλεται ανάμεσα στο ρήμα και το όνομα. Π.χ.

...η ενιαία πράξη συνδέει στενά τον στόχο της αγοράς...

4. (Σύνδεσμος +E) + Επίρρημα + Ρήμα

Ένα επίρρημα ως προσδιορισμός του ρήματος είναι δυνατόν να εμφανίζεται όχι μόνο μετά αλλά και πριν από αυτό. Π.χ.

...εάν αποδεδειγμένα δεν λαμβάνει όμοιο επίδομα ...

5. Ρήμα + Επίρρημα + Πρόθεση + (Προσδιοριστής + E) + (Επίθετο + E) + Όνομα

Πρόκειται για τις περιπτώσεις εκείνες των επιρρημάτων, που είναι δυνατόν να ακολουθούνται από προθέσεις ή προθετικές φράσεις. Οι προθέσεις αυτές εξαρτώνται από τα εκάστοτε επιρρήματα. Π.χ.

...συνδέονται άρρηκτα με τον σταθερό στόχο της βελτίωσης

6. Ρήμα + (Επίρρημα + Ε) + (Επίρρημα + Ε) + (Επίρρημα + Ε) (Πρόθεση + Ε) + (Όνομα + Ε)

Η συγκεκριμένη ακολουθία συμπεριλαμβάνει τις περιπτώσεις που ακολουθούν ένα, δύο ή και τρία επιρρήματα μετά από το ρήμα της φράσης. Πρόκειται για τροπικά, τοπικά, χρονικά και ποσοτικά επιρρήματα. Π.χ.

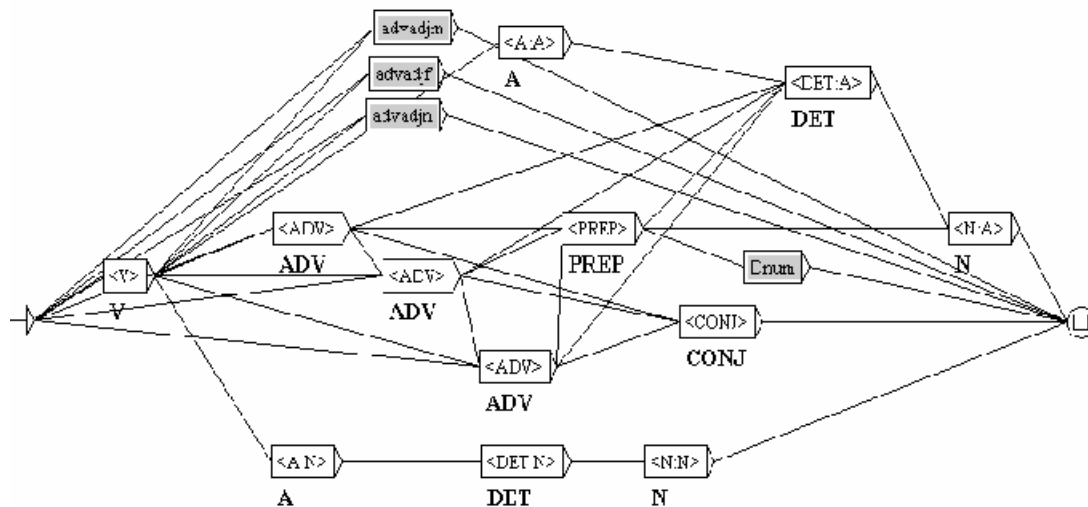
... μεγάλο ποσοστό υλοψηφίων τα πήγε πάρα πολύ καλά στις εξετάσεις.....

Οι περιορισμοί αυτοί κρίθηκε απαραίτητο να αναπαρασταθούν μέσω διαδικασιών αυτοματισμού πεπερασμένων καταστάσεων, προκειμένου να εισαχθούν στα υπάρχοντα συστήματα αυτόματης ανάλυσης (INTEX, UNITEX) και να επαληθευθούν μέσα στα διαθέσιμα σώματα κειμένων. Μετά την ολοκλήρωση της διαδικασίας αυτής, και εφόσον αποδείχθηκε μέσα από τα σώματα κειμένων ότι ισχύουν οι συγκεκριμένοι περιορισμοί, ενσωματώθηκαν στις γραμματικές των παραπάνω συστημάτων.

5. Διαδικασίες αυτοματισμού πεπερασμένων καταστάσεων

Οι διαδικασίες αυτοματισμού πεπερασμένων καταστάσεων αποτελούνται από ένα πεπερασμένο σύνολο καταστάσεων (από όπου προέρχεται και ο προσδιορισμός *πεπερασμένο*), ένα σύνολο συμβόλων και ένα σύνολο μεταβάσεων από κατάσταση σε κατάσταση. Η μετάβαση γίνεται με βάση κάποια σύμβολα εισόδου και τη συνάρτηση μετάβασης. Πρόκειται για «γραμμικά γραφήματα με αρχή, μέση και τέλος και χρησιμοποιούνται ως εργαλεία αναπαράστασης γλωσσικών φαινομένων σε διάφορα επίπεδα περιγραφής φυσικών γλωσσών: ορθογραφικές παραλλαγές, φωνολογική αναπαράσταση, αναπαράσταση κλιτικών παραδειγμάτων, κατάρτιση τοπικών γραμματικών, άρση λανθασμένων επιλογών και μετάφραση δίγλωσσων πληροφοριών» (Ε. Σκλαβούνου: 1994).

Για τη δημιουργία και την επαλήθευση των πεπερασμένων αυτομάτων που κατασκευάστηκαν, χρησιμοποιήθηκαν τα συστήματα αυτόματης ανάλυσης INTEX (<http://laseldi.univ-fcomte.fr/intex>) που κατασκευάστηκε από τον Max Silberztein και UNITEX (<http://www-igm.univ-mlv.fr/~unitex/manuel.html>) του ερευνητή Sebastien Paumier. Τα συστήματα αυτά, που είναι σχετικά απλά στη λειτουργία τους, χρησιμοποιούνται για τον αυτόματο εντοπισμό και την αναγνώριση συντακτικο-σημασιολογικών μονάδων και πολυλεκτικών ακολουθιών μέσα σε μεγάλα σώματα κειμένων, την άρση των λεξικών και συντακτικών αμφισημιών, τη δημιουργία συμφραστικών πινάκων, τη στατιστική επεξεργασία των γλωσσικών πληροφοριών καθώς και για την ημι- αυτόματη μετάφραση.



Αυτό είναι ένα από τα πεπερασμένα αυτόματα, που δημιουργήθηκαν για την αναγνώριση των επιρρημάτων. Κρίθηκε σκόπιμο να ενσωματωθούν σ' αυτό τα εξής εγκιβωτισμένα γραφήματα: το *advadjm*, που αφορά τα επίθετα αρσενικού γένους, το *advadjf* για τα επίθετα θηλυκού γένους και το *advadjn* για τα επίθετα ουδετέρου γένους. Το *Dnum* περιλαμβάνει τους αριθμούς. Με το *V* συμβολίζονται τα ρήματα, με το *ADV* τα επιρρήματα, με το *PREP* οι προθέσεις, με το *CONJ* οι σύνδεσμοι και με το *DET* οι προσδιοριστές. Αξίζει να σημειωθεί το γεγονός ότι για τον συμβολισμό χρησιμοποιήθηκαν λατινικοί χαρακτήρες, δεδομένου ότι αποτελούν κοινούς συμβολισμούς για όλες τις εξεταζόμενες γλώσσες στα συστήματα αυτόματης ανάλυσης.

Εφαρμόζοντας το παραπάνω πεπερασμένο αυτόματο, παράγεται ο εξής πίνακας συμφοραζομένων:

ου παρέχουν οι λύσεις της Connectix δεν V είναι ADV ιδιαίτερα A αξιόλογη, ειδικά ο παιχνίδι, η αεροπορία των Terrans δεν V είναι ADV ιδιαίτερα A αξιόμαχη.{S} Επίσης. {S} Αυτή η εσωτερική διαφοροποίηση V είναι ADV ιδιαίτερα A θεαματική σε πο. {S} Ειδικά αυτά που μπορούν να πετούν, V είναι ADV ιδιαίτερα A χρήσιμα, καθώς ενώ 50% των ατόμων με HP(+) γαστρίτιδα V είναι ADV εντελώς A ασυμπτωματικά.{S} Θεοδωρόπουλο αυτά ακριβώς τα θέματα. {S} V Θεωρείται ADV εξαιρετικά A ικανός αφού Μια κάρτα πολλών χρήσεων που V θα φανεί ADV αναμφισβήτητη A χρήσιμη. {S} Να V αφουγκράζονται ADV καθημερινά DET τον N παλμό του λαού. {S} Η Ζάκονθος θα συνεχίσει την αύξηση ADV πάνω PREP από Dnum 12% αυξάνοντας το μερίδιό της στην αγορά. {S} Ελπίζω ότι DET το ADV ιδιαίτερα A καλό N παράδειγμα της Πορτογαλίας θα ας λόγους, αλλά και από τις πράξεις, και ADV κυρίως PREP από DET τα N αποτελέσματα των πράξεών της. {S} Η Γενική Συνέλευση V συνεδριάζει ADV έκτακτα CONJ όταν το ζητήσει

6. Λεξικό-Γραμματική

Για την επιτυχή αναγνώριση των ουσιαστικοποιημένων επιθέτων διαπιστώθηκε μέσω της έρευνας που πραγματοποιήθηκε, ότι είναι πρακτικά ανέφικτο να εντοπιστούν περιοριστικοί κανόνες που να δίνουν ικανοποιητικά αποτελέσματα κατά την εφαρμογή πεπερασμένων αυτομάτων. Για το λόγο αυτό, όπως προαναφέρθηκε, υιοθετήθηκε ο φορμαλισμός του Λεξικού-Γραμματική, που προτάθηκε από τον M. Gross (:1977) και αποτελεί εφαρμογή της Μετασχηματιστικής Γραμματικής του Z. S. Harris (:1976). Για τη δημιουργία του πίνακα Λεξικού-Γραμματικής εξετάζεται καθένα ουσιαστικοποιημένο επίθετο μέσα στο συγκεκριμένο μιας στοιχειώδους πρότασης. Παράλληλα, αναζητούνται και δηλώνονται μέσα στον πίνακα όλοι οι εν δυνάμει μετασχηματισμοί της.

Το λεξικό-γραμματική αποτελείται από πίνακες excel με τη μορφή δυαδικών μητρών, των οποίων οι οριζόντιες γραμμές συμπληρώνονται από τα λήμματα, ενώ στις κάθετες αναγράφονται οι συντακτικές τους ιδιότητες. Σε κάθε τετράγωνο της μήτρας σημειώνεται ένα «+» ή ένα «-» προκειμένου να δηλωθεί αν η δεδομένη ιδιότητα αποδίδεται στο συγκεκριμένο λήμμα.

Ο πίνακας που ακολουθεί, αποτελεί ένα μέρος του Λεξικού-Γραμματικής των ουσιαστικοποιημένων επιθέτων που συναντώνται στη βασική συντακτική δομή *NO V AEN*. Στις δύο πρώτες κολώνες γίνεται αναφορά στα πιθανά υποκείμενα (αν δηλαδή πρόκειται για ανθρώπινο υποκείμενο ή όχι). Στην τρίτη κολώνα αναγράφονται τα πιθανά ρήματα που μπορεί να προηγούνται του συγκεκριμένου ουσιαστικοποιημένου επιθέτου, ενώ στην τέταρτη κολώνα καταχωρείται το υπό εξέταση ουσιαστικοποιημένο επίθετο. Αξίζει να σημειωθεί το γεγονός ότι στα ανθρώπινα υποκείμενα συμπεριλαμβάνονται επιπλέον περιληπτικά ουσιαστικά που χαρακτηρίζονται ως έμψυχα λόγω της μετωνυμίας (π.χ. *εταιρεία*). Στις στήλες που ακολουθούν περιγράφονται αναλυτικά όλοι οι δυνατοί μετασχηματισμοί της βασικής συντακτικής δομής των εν λόγω ουσιαστικοποιημένων. Παρατηρείται ότι στην προκειμένη περίπτωση οι μετασχηματισμοί αυτοί είναι 9.

N0 V AEN

| N0hum | N0-hum | V | AEN | N0 V DET=: ένας, μία, ένα AEN | N0 V DET=: ο, η, το AEN | N0 V A AEN | N0 V ADV AEN | N0 V DET=: ένας, μία, ένα (ADV + E) A AEN | N0 V DET=: ο, η, το (ADV + E) A AEN | N0 V ADV DET=: ο, η, το A AEN | N0 V ADV DET=: ένας, μία, ένα A AEN | N0 V (ADV + E) A AEN |
|-------|--------|---------------|--------------|-------------------------------|-------------------------|------------|--------------|---|-------------------------------------|-------------------------------|-------------------------------------|----------------------|
| + | - | διαθέτω | ακίνητο | + | + | + | + | + | + | + | + | + |
| + | - | διαχειρίζομαι | ακίνητο | + | + | + | + | + | + | + | + | + |
| + | - | έχω | ακίνητο | + | + | + | + | + | + | + | + | + |
| + | - | αγοράζω | αμοιβαία | - | + | + | + | - | + | + | - | + |
| + | - | πουλάω | αμοιβαία | - | + | + | + | - | + | + | - | + |
| + | - | πουλώ | αμοιβαία | - | + | + | + | - | + | + | - | + |
| + | - | κερνώ | γλυκό | + | + | + | + | + | + | + | + | + |
| + | - | φτιάχνω | γλυκό | + | + | + | + | + | + | + | + | + |
| + | - | ανοίγω | εισαγωγικά | - | + | - | - | - | - | - | - | - |
| + | - | βάζω | εισαγωγικά | - | + | - | + | - | - | - | - | - |
| + | - | κλείνω | εισαγωγικά | - | + | - | - | - | - | - | - | - |
| + | - | ετοιμάζω | ιδιαίτερο | + | + | + | + | + | + | + | + | + |
| + | - | κάνω | ιδιαίτερο | - | + | + | + | + | - | + | - | + |
| + | - | παραδίδω | ιδιαίτερο | - | + | + | + | - | - | + | - | + |
| + | - | ακούω | λαϊκά | - | - | + | + | - | + | + | - | + |
| + | - | τραγουδώ | λαϊκά | - | - | + | + | - | + | + | - | + |
| + | - | φορώ | χειμωνιάτικα | - | + | + | + | - | + | + | - | + |

Γενικά, στον συγκεκριμένο πίνακα λεξικού-γραμματικής παρατηρείται ότι το υποκείμενο είναι πάντα ανθρώπινο. Μεταξύ των άλλων, συμπεριλήφθησαν οι ακόλουθες συντακτικές δομές:

N₀ V AEN → *Ο Νίκος ακούει λαϊκά*

N₀ V DET=: ένας, μία, ένα AEN → *Η Μαρία φτιάχνει ένα γλυκό*

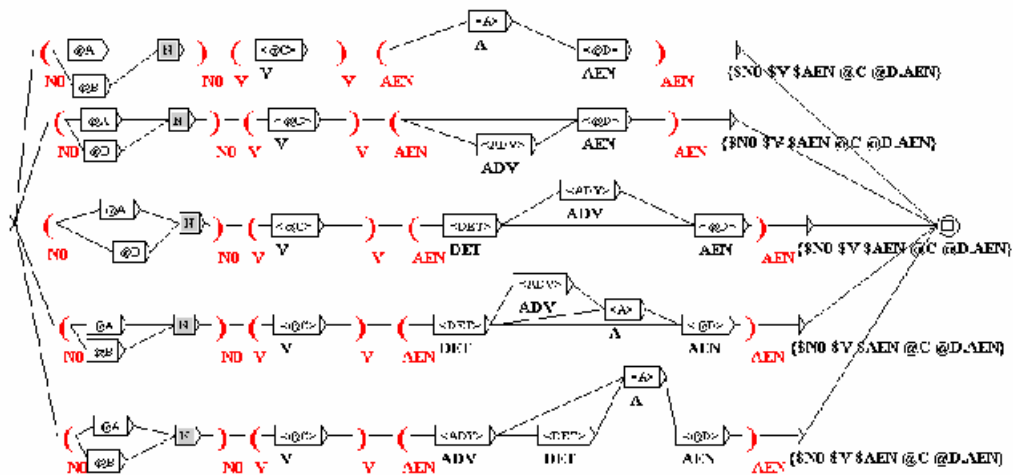
N₀ V DET=: ο, η, το A AEN → *Ο κ. Παλαδόπουλος διαθέτει τα ακριβότερα ακίνητα*

N₀ V A AEN → *Η εταιρεία αγοράζει εγχώρια αμοιβαία*

Παρόμοιοι συντακτικο-σημασιολογικοί πίνακες απαιτείται να δημιουργηθούν για κάθε διαφορετική συντακτική δομή μέσα στην οποία εμφανίζονται ουσιαστικοποιημένα επίθετα. Με αυτό τον τρόπο καθίσταται εφικτή η επαρκής εξάλειψη των αμφισημιών ανάμεσα στα ουσιαστικοποιημένα επίθετα και τα επιρρήματα, εφόσον εξετάζονται μέσα στο εν δυνάμει συγκεκριμένο.

Προκειμένου να εφαρμοστεί ο παραπάνω πίνακας σ' ένα αυτόματο σύστημα ανάλυσης και να επαληθευθούν τα παραγόμενα αποτελέσματα, απαιτείται η δημιουργία του παρακάτω βασικού πεπερασμένου αυτομάτου (πβλ. manual UNITEX και manual INTEX).

Το βασικό πεπερασμένο αυτόματο, που ακολουθεί, τυποποιεί τις καταγεγραμμένες ιδιότητες του πίνακα. Γενικά, το βασικό πεπερασμένο αυτόματο ενός πίνακα Λεξικού-Γραμματικής χρησιμοποιείται για να μεταφράσει τις καταγεγραμμένες ιδιότητες του πίνακα, δεδομένου ότι μετατρέπει τον πίνακα excel σε επιμέρους πεπερασμένα αυτόματα ένα για κάθε σειρά. Αυτό του είδους το αυτόματο διαφοροποιείται από τα απλά πεπερασμένα αυτόματα για τα οποία έγινε λόγος εκτενώς παραπάνω, εφόσον αναφέρεται σε μήτρες του πίνακα.



Στο παραπάνω βασικό πεπερασμένο αυτόματο, το @ συμβολίζει το σύνολο της αναγραφόμενης κάθε φορά κολώνας. Έτσι για παράδειγμα το @A παραπέμπει στην στήλη A που περιέχει τα έμψυχα υποκείμενα. Οι κόκκινες παρενθέσεις δηλώνουν τα υποσύνολα των ιδιοτήτων. Το N_0 αφορά το υποκείμενο, το V αφορά το ρήμα της εκάστοτε φράσης και το AEN^{iv} αφορά το ουσιαστικοποιημένο επίθετο. Στο τέλος κάθε κόμβου δηλώνεται η σειρά των όρων και δίνεται ο χαρακτηρισμός AEN.

Εφαρμόζοντας τον πίνακα λεξικού-γραμματικής σε συνδυασμό με το βασικό πεπερασμένο αυτόματο, παράγονται τα παρακάτω αποτελέσματα:

- {S}Η δημοσκόπηση απέδειξε ότι [DET οι N νέοι V ακούν ADV κυρίως A ελαφρά AEN λαϊκά {oi νέοι ακούν κυρίως ελαφρά λαϊκά.ακούω λαϊκά.AEN}](#)
- ότι οι νέοι ακούν κυρίως ελαφρά λαϊκά. {S}[DET Ο N συγγραφέας V βάζει AEN εισαγωγικά {Ο συγγραφέας βάζει εισαγωγικά.βάζω εισαγωγικά.AEN}](#)
- κείμενα όταν παραθέτει αναφορές τρίτων. {S}[DET Ο N Νίκος V φτιάχνει DET τα A καλύτερα AEN γλυκά{Ο Νίκος φτιάχνει τα καλύτερα γλυκά.φτιάχνω γλυκό.AEN}](#)
- ά. {S}[DET Ο N Γιάννης V φόρεσε ADV αμέσως DET τα A βαριά AEN χειμωνιάτικα {Ο Γιάννης φόρεσε αμέσως τα βαριά χειμωνιάτικα. φορώ χειμωνιάτικα.AEN}](#)
- ης φόρεσε αμέσως τα βαριά χειμωνιάτικα. {S}[DET Η Ν τράπεζα V πουλάει ADV περισσότερο A εγχώρια AEN αμοιβαία {Η τράπεζα πουλάει περισσότερο εγχώρια αμοιβαία. πουλάω αμοιβαία.AEN}](#)

7. Συμπεράσματα-Προοπτικές

Στην πορεία της έρευνας και στην προσπάθειά να οριστούν οι γενικοί περιοριστικοί κανόνες που παρουσιάστηκαν, διαπιστώθηκε ότι δεν είναι εφικτή η δημιουργία πεπερασμένων αυτομάτων όταν τα επιρρήματα βρίσκονται στην αρχή πρότασης. Αυτό συμβαίνει λόγω της δισεπίλυτης, προς το παρόν, αμφισημίας που δημιουργείται ανάμεσα στα ουσιαστικοποιημένα επίθετα και τα επιρρήματα, όταν αυτά συναντώνται σε τίτλους κειμένων ή σε αγγελίες. Ενδεικτικά αναφέρονται τα ακόλουθα παραδείγματα:

Ιδιαίτερα παραδίδονται σε μαθητές Λυκείου → *ιδιαίτερα*, *N*

Ιδιαίτερα δύσκολη αποδείχθηκε η μετακίνηση των ταξιδιωτών → *ιδιαίτερα*, *ADV*

Επίσης, επιβεβαιώθηκε το γεγονός ότι δεν είναι εφικτή η δημιουργία πεπερασμένων αυτομάτων για τα ουσιαστικοποιημένα επίθετα, δεδομένου ότι δεν υπάρχουν αρκετοί περιοριστικοί κανόνες, ώστε να επιλύονται ικανοποιητικά οι υπάρχουσες αμφισημίες. Για το λόγο αυτό, επιλέχθηκε η δημιουργία πινάκων λεξικού-γραμματικής για τα ουσιαστικοποιημένα επίθετα. Για την πλήρη, όμως, αναγνώριση των επιρρημάτων και των ουσιαστικοποιημένων επιθέτων απαιτείται η ολοκλήρωση του ηλεκτρονικού λεξικού των σύνθετων επιρρημάτων και του αντίστοιχου ηλεκτρονικού λεξικού των ουσιαστικοποιημένων επιθέτων. Επιπλέον, απώτερος στόχος είναι η ολοκλήρωση των πινάκων Λεξικού-Γραμματικής για το σύνολο των ουσιαστικοποιημένων επιθέτων της Νέας Ελληνικής.

Στη συνέχεια και εφόσον ολοκληρωθούν τα παραπάνω στάδια, θα προβούμε στην κατηγοριοποίηση των ουσιαστικοποιημένων επιθέτων σύμφωνα με το μοντέλο των Τάξεων Αντικειμένων με σκοπό την επίλυση των αμφισημιών που εξακολουθούν να παραμένουν. Η συγκεκριμένη κατηγοριοποίηση γίνεται με βάση μια σειρά σημασιολογικών και συντακτικών κριτηρίων. Πιο συγκεκριμένα, θα αναζητηθούν οι γενικοί επιλογείς (*opérateurs généraux*) σύμφωνα με τους οποίους τα κατηγορήματα κατατάσσονται σε σημασιολογικές ομάδες, όπως: ανθρώπινο, ζώο, φυτό, συγκεκριμένο, αφηρημένο, γεγονός, τοπικό. Έπειτα, με τη βοήθεια ειδικών επιλογέων (*opérateurs appropriés*) η κάθε σημασιολογική ομάδα διαιρείται σε ειδικές τάξεις (Ζ. Γαβριηλίδου: 1999). Κατά τον G. Gross (:1991) η σημασιολογική τάξη στην οποία ανήκει κάθε κατηγορήμα είναι αυτή που καθορίζει το σχήμα των ορισμάτων του και τις συντακτικές του ιδιότητες. Η τάξη, στην οποία ανήκει κάθε κατηγορήμα και τα ορίσματα που δέχεται, θα πρέπει να σημειώνονται στο μορφολογικό λεξικό, μαζί με άλλες πληροφορίες, όπως είναι τα συνώνυμα, τα αντώνυμα, η μετάφραση του σε άλλες γλώσσες.

Σημειώσεις

ⁱ Οι συμβολισμοί που παρατίθενται μέσα στις παρενθέσεις δηλώνουν αντίστοιχα: το *A* τα επίθετα, το *N* τα ουσιαστικά.

ⁱⁱ Για τα συστήματα αυτά γίνεται εκτενής αναφορά παρακάτω.

ⁱⁱⁱ Το σύμβολο *E* χρησιμοποιείται για να δηλωθεί το κενό.

^{iv} Τα αρχικά *AEN* σημαίνουν *Adjectif a Emploi Nominal*. Δεδομένου ότι οι συμβολισμοί των όρων στα συστήματα αυτόματης ανάλυσης αναγράφονται με λατινικούς χαρακτήρες, επιλέχθηκε η γαλλική αντιστοιχία του ουσιαστικοποιημένου επιθέτου.

Βιβλιογραφία

- Anastassiadis-Symeonidis, Anna. Efthimiou, Aggeliki. Fliatouras, Asimakis. 2002. "Conversion or ellipsis? Evidence from Modern Greek", Paper read at th Autumn LAGB Meeting, September 2002, Manchester.
- Anastassiadis-Symeonidis, Anna. Kyriakopoulou, Tita. Sklavounou, Elsa. Thilikos, Iassonas. Voskaki, Rania. 2000 "A system for analysing texts in Modern Greek: representing and solving ambiguities", in *Proceedings of COMLEX 2000, Workshop on Computational Lexicography and Multimedia Dictionaries*, 22-23 September 2000, Kato Achaia, Greece, Organized by WCL, Department of Electrical & Computer Engineering University of Patras, Patras 2000. 113-116.
- Gross Gaston. 1991. "Classes d'objets et traitement automatique des langues". 26th Colloquium of Linguistics (Poznan).
- Gross, Maurice. 1977. *Grammaire transformationelle du francais. Syntaxe du nom*. Vol. 2, Paris: Cantilene.
- Harris, Zellig Sabbetai. 1976. *Notes du cours de syntaxe*. trad. M. Gross, Paris: Le Seuil.
- Holton, David. Mackridge, Peter. Φιλippyάκη-Warburton, Eιρήνη. 1999. *Γραμματική της Ελληνικής Γλώσσας*, Εκδόσεις Πατάκη Αθήνα.
- Laporte, Eric. Monceaux, Anne. 1998-1999. "Elimination of lexical ambiguities by grammars. The ELAG system", *Linguistic? Investigations XXII*, Amsterdam-Philadelphie, Benjamins. 341-367.
- Mathieu-Colas, Michel. Girardin, Chantal. 1997. "Classes d'objets et traitement automatique de la synonymie". *Actes des VIIemes journees du LIPN*.
- Paumier, Sebastien. 2002. *Manual UNITEX*, <http://www-igm.univ-mlv.fr/~unitex/manuel.html>.
- Silberztein, Max. 1993. *Dictionnaires electroniques et analyse automatique de texts. Le systeme INTEX*, sous la direction de Maurice Gross, Collection Informatique Linguistique, Masson, Paris.
- Silberztein, Max. 1999. *Finite State Transducers and the processing of natural languages*, Tutorial Notes, INTEX, electronic Edition.
- Silberztein, Max. 2000. *Manual INTEX*, <http://laseldi.univ-fcomte.fr/intex>.
- Skavounou, Elsa. 1997. *Etude Comparee de la Nominalisation des adjectifs en Grec Moderne et en Francais*, These de Doctorat. Paris: Paris 8.
- Voyatzi, Stavroula. 2002. *Description syntactico-semantic des adverbs en vue d'un systeme d'analyse automatique des texts du grec moderne. Le cas des adverbs simples*, μεταπτυχιακή εργασία, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Θεσσαλονίκη.
- Αναστασιάδη-Συμεωνίδη, Άννα. 1986. *Η νεολογία στην κοινή νεοελληνική*, Επιστημονική Επετηρίδα της Φιλοσοφικής Σχολής, Θεσσαλονίκη.
- Βοσκάκη, Ράνια. Θηλυκός, Ιάσοντας. 2003. «Αναπαράσταση των ουσιαστικοποιημένων επιθέτων της Νέας Ελληνικής με διαδικασίες αυτοματισμού πεπερασμένων καταστάσεων», *Μελέτες για την ελληνική γλώσσα - Πρακτικά της 23ης ετήσιας συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Α.Π.Θ.*, 17-19 Μαΐου 2002, Θεσσαλονίκη. 814-825.

-
- Γαβριηλίδου, Ζωή. 1999. «Πολυλεκτικά παραθετικά σύνθετα: διδακτική προσέγγιση», *Μελέτες για την ελληνική γλώσσα - Πρακτικά της 19^{ης} ετήσιας συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Α.Π.Θ.*, 25-28 Απριλίου 1998, Θεσσαλονίκη. 66-80.
- Γάκης, Παναγιώτης, Ιορδανίδου, Άννα, Ορφανός, Γιώργος. 2000. «Μορφοσυντακτική ασάφεια της Νέας Ελληνικής: η περίπτωση επιθέτου-ουσιαστικού-επιρρήματος», *Μελέτες για την ελληνική γλώσσα - Πρακτικά της 20^{ης} ετήσιας συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Α.Π.Θ.*, 23-25 Απριλίου 1999, Θεσσαλονίκη. 373-383.
- Σκλαβούνου, Έλσα. 1999. «Ηλεκτρονικό Μορφολογικό Λεξικό Επιθέτων και Ουσιαστικών της Νέας Ελληνικής», *Μελέτες για την ελληνική γλώσσα - Πρακτικά της 19^{ης} ετήσιας συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Α.Π.Θ.*, 23-25 Απριλίου 1998, Θεσσαλονίκη 1999. 461-474.
- Σκλαβούνου, Έλσα. 1994. «Ένα παράδειγμα λεξιλογικής αναπαράστασης και επεξεργασίας με πεπερασμένα αυτόματα», *Μελέτες για την ελληνική γλώσσα - Πρακτικά της 15^{ης} ετήσιας συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Α.Π.Θ.*, 11-14 Μαΐου, Τιμητική προσφορά στον Καθ. Μ. Σετάτο, Θεσσαλονίκη 1994. 811-820.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.