

Η συμβολή της μορφολογικής γεννήτριας στην υπολογιστική γλωσσολογία

Ευάγγελος Χ. Παπακίτσος, Γεωργία Παπαθανασοπούλου* & Αικατερίνη Αλέξη*
ΑΣΠΑΙΤΕ, *Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

ABSTRACT

Natural Language Generation is the area of Computational Linguistics that deals with the production of text in natural language, based on a system of machine representation in digital form. The core of this area consists of a morphological generator, which is the software module that creates words. The designing of this module depends on the morphotactics of the processed language. In this paper the implementation of a generator is briefly described in relation to Modern Greek, focusing on issues of particular interest as regards its functionality.

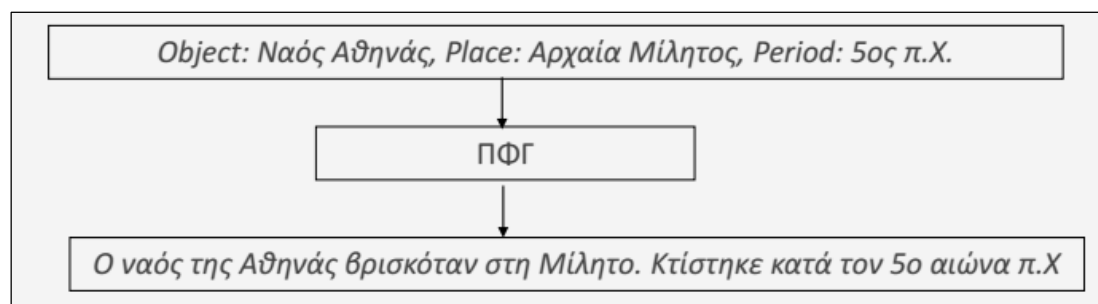
ΛΕΞΕΙΣ-ΚΛΕΙΔΙΑ: μορφολογική γεννήτρια, παραγωγή φυσικής γλώσσας, υπολογιστική γλωσσολογία

1. Εισαγωγή

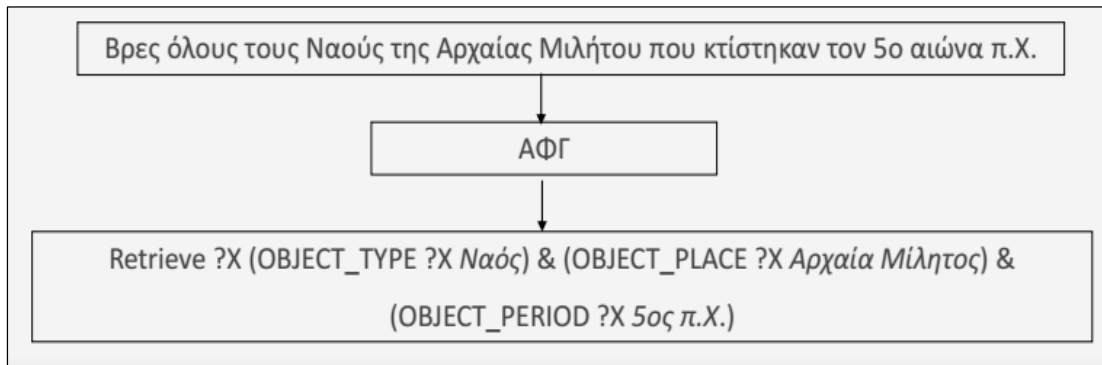
Παραγωγή Φυσικής Γλώσσας (εφεξής ΠΦΓ) είναι η διαδικασία της αυτόματης δημιουργίας κειμένων φυσικής γλώσσας έτσι ώστε να ανταποκρίνονται σε συγκεκριμένους επικοινωνιακούς στόχους (βλ. McDonald 1992). Βασίζεται σε μεθόδους και τεχνικές από τις περιοχές της τεχνητής νοημοσύνης, της επεξεργασίας φυσικής γλώσσας, της αναπαράστασης και διαχείρισης γνώσεων και της υπολογιστικής γλωσσολογίας.

Ένα πληροφοριακό σύστημα ΠΦΓ μετατρέπει σε κείμενο φυσικής γλώσσας μια ψηφιακή αναπαράσταση του ηλεκτρονικού υπολογιστή που βασίζεται σε μια αντίστοιχη τυπική αναπαράσταση της φυσικής γλώσσας, όπως π.χ. η απεικόνιση δεδομένων μιας βάσης εκθεμάτων μουσείων που βασίζεται σε περιγραφές των εκθεμάτων σε φυσική γλώσσα (βλ. Σχήμα 1).

Σχήμα 1. Απεικόνιση δεδομένων μιας βάσης εκθεμάτων μουσείων με βάση περιγραφές των εκθεμάτων σε φυσική γλώσσα



Λειτουργικά, η ΠΦΓ μέσω ηλεκτρονικού υπολογιστή είναι η αντίθετη διαδικασία της Ανάλυσης Φυσικής Γλώσσας (εφεξής ΑΦΓ). Δηλαδή, ΑΦΓ είναι η απεικόνιση κειμένων σε μια αναπαράσταση «κατανοητή» από τον υπολογιστή, όπως π.χ. η απεικόνιση μιας ερώτησης φυσικής γλώσσας σε μια ερώτηση βάσης δεδομένων (βλ. Σχήμα 2).

Σχήμα 2. Απεικόνιση μιας ερώτησης φυσικής γλώσσας σε μια ερώτηση βάσης δεδομένων

Επομένως, κατά την ΠΦΓ το πληροφοριακό σύστημα χρειάζεται να λάβει αποφάσεις για το πώς να μετατρέψει μια ψηφιακά κωδικοποιημένη ιδέα σε λέξεις. Απλά παραδείγματα τέτοιων συστημάτων είναι αυτά που δημιουργούν φόρμες επιστολών. Συνήθως δεν περιλαμβάνουν κανόνες γραμματικής, αλλά μπορούν να δημιουργήσουν μία επιστολή προς έναν καταναλωτή, που αναφέρει π.χ. ότι μια πιστωτική κάρτα εξάντλησε το όριό της. Τα πιο πολύπλοκα συστήματα επιχειρούν τη δυναμική δημιουργία κειμένου προκειμένου να ανταποκριθούν σε έναν επικοινωνιακό στόχο. Όπως και σε άλλους τομείς της επεξεργασίας φυσικής γλώσσας, αυτό μπορεί να γίνει είτε χρησιμοποιώντας μοντέλα της γλώσσας (γραμματικές), είτε χρησιμοποιώντας στατιστικά μοντέλα που προέρχονται από την ανάλυση φυσικών γραπτών κειμένων. Παραδείγματα τέτοιων εφαρμογών είναι τα ακόλουθα:

- Συστήματα που παράγουν κείμενο περίληψης από δεδομένα σε βάσεις δεδομένων ή από σύνολα δεδομένων (data-to-text). Αυτά τα συστήματα συνήθως εκτελούν την ανάλυση δεδομένων, καθώς και την παραγωγή κειμένου, ενώ από εμπορική άποψη είναι οι πιο επιτυχημένες εφαρμογές. Η σχετική θεματολογία περιλαμβάνει κείμενο μετεωρολογικής πρόβλεψης από τα στοιχεία του καιρού (βλ. Goldberg et al. 1994, Reiter et al. 2005, Belz 2008, Stripada et al. 2014), σύνοψη οικονομικών και επιχειρηματικών δεδομένων όπως η ανάλυση μεγάλων ποσοτήτων δεδομένων πωλήσεων λιανικής (βλ. Anand & Kahn 1992) και περιλήψεις ιατρικών δεδομένων από ηλεκτρονικούς ιατρικούς φακέλους (βλ. Law et al. 2005, Harris 2008, Portet et al. 2009).
- Συστήματα υποβοήθησης του συγγραφικού έργου. Τέτοια εργαλεία μιμούνται το σενάριο γραψίματος με αποκοπή και επικόλληση (cut-and-paste), στο οποίο ένας συγγραφέας σχηματίζει το κείμενό του από τα διάφορα αποτελέσματα αναζήτησης κάνοντας έτσι τη διαδικασία πιο αποτελεσματική και αποδοτική (βλ. Sauper & Barzilay 2009, Galitsky et al. 2012, Galitsky 2013).
- Αυτόματη σύνθεση ομιλίας ή αλλιώς σύνθεση φωνής από κείμενο (text-to-speech). Πρόκειται για την αυτόματη μετατροπή ηλεκτρονικών κειμένων σε φωνή μέσω υπολογιστή. Η φωνή είναι βασικό υποστηρικτικό εργαλείο για την υποκατάσταση της γραπτής ή γενικότερα οπτικής πληροφορίας για ειδικές ομάδες ατόμων, όπως άτομα με προβλήματα όρασης (τυφλοί, μερικώς βλέποντες), ηλικιωμένους, άτομα με δυσλεξία, άτομα που δυσκολεύονται να χειριστούν έντυπα μέσα, άτομα που δεν γνωρίζουν αρκετά την γλώσσα (π.χ. μετανάστες), καθώς και σε κάθε άλλη περίπτωση όπου η ακουστική πληροφορία είναι πιο λειτουργική από την οπτική.
- Μηχανική ή αυτόματη μετάφραση, δηλαδή την αυτόματη μετάφραση κειμένου από μια γλώσσα σε μια άλλη μέσω υπολογιστή. Είναι μια από τις πρώτες και

πιο δύσκολες εφαρμογές της υπολογιστικής γλωσσολογίας η οποία αντλεί γνώσεις από πολλούς υποκλάδους της. Χρησιμοποιούνται διαφορετικές τεχνικές για παραγωγή μετάφρασης και παρουσιάζει έντονο ακαδημαϊκό και εμπορικό ενδιαφέρον (βλ. Παπαθανασοπούλου 2015).

Τα τυπικά στάδια της ΠΦΓ, όπως αυτά προτείνονται από τους Dale & Reiter (2000) είναι:

- α) *Προσδιορισμός περιεχομένου*, όπου καθορίζεται ποιες πληροφορίες θα περιέχονται στο κείμενο,
- β) *Διάρθρωση εγγράφου*, όπου αποφασίζεται η συνολική οργάνωση των πληροφοριών που παρατίθενται,
- γ) *Συνάθροιση*, όπου πραγματοποιείται η συγχώνευση παρόμοιων προτάσεων για τη βελτίωση της αναγνωσιμότητας και της φυσικότητας του κειμένου,
- δ) *Λεκτική επιλογή*, όπου γίνεται η τοποθέτηση των κατάλληλων λέξεων στις αντίστοιχες έννοιες,
- ε) *Δημιουργία αναφορικών εκφράσεων*, όπου δημιουργούνται οι αναφορικές εκφράσεις που προσδιορίζουν αντικείμενα και περιφέρειες (π.χ. βόρειο-ανατολικός άνεμος). Το στάδιο αυτό περιλαμβάνει επίσης αποφάσεις για αντωνυμίες και άλλους τύπους αναφορικού λόγου.
- στ) *Πραγμάτωση*, όπου δημιουργείται το ίδιο το κείμενο, που θα πρέπει να είναι σωστό σύμφωνα με τους κανόνες του συντακτικού, της μορφολογίας και της ορθογραφίας.

Επιπλέον των παραπάνω, προσφάτως η έρευνα στις εφαρμογές ΠΦΓ έχει επικεντρωθεί και στον σημαντικό ρόλο του επικοινωνιακού στόχου, που αποκαλείται επίσης και *επικοινωνιακή πρόθεση* (communicative intention), είτε πρόκειται π.χ. για ένα σύστημα διαλόγου, είτε για ένα σύστημα μηχανικής μετάφρασης. Ο λόγος είναι ότι, σε αντίθεση με ένα σύστημα ΑΦΓ, τα διαδραστικά συστήματα ΠΦΓ συνθέτουν το δικό τους επικοινωνιακό περιεχόμενο (βλ. Stent & Bangalore 2014, Horacek & Zock 2015).

Στον πυρήνα των συστημάτων ΠΦΓ, ειδικά στο στάδιο της λεκτικής επιλογής, βρίσκεται η υπολογιστική μονάδα χειρισμού των μορφοφωολογικών στοιχείων, που θα αποκαλείται εφεξής *μορφολογική γεννήτρια*. Η μορφολογική γεννήτρια έχει ιδιαίτερη βαρύτητα σε γλώσσες πλούσιας μορφολογίας, όπως η νέα ελληνική. Η περιγραφή του αλγοριθμικού χειρισμού των μορφοφωολογικών φαινομένων της νέας ελληνικής θα ακολουθήσει στην επόμενη ενότητα, μετά από μια σύντομη αναφορά σε αντίστοιχα εργαλεία λογισμικού.

2. Μορφολογική γεννήτρια

Η αυτόνομη ύπαρξη μορφολογικής γεννήτριας έχει νόημα μόνο στα πλαίσια λογισμικού για πειραματική χρήση. Έτσι, η δυνατότητα σύνθεσης λέξεων μπορεί να βρίσκεται σε λογισμικό επεξεργασίας φυσικής γλώσσας που πραγματοποιεί και ΑΦΓ. Για τη νέα ελληνική, μερικά υπολογιστικά συστήματα με τέτοια δυνατότητα είναι τα εξής:

- Ο επεξεργαστής ΑΘΗΝΑ (βλ. Ralli & Galiotou 1987, 1991, Γαλιώτου 1991), υλοποιημένος σε γλώσσα προγραμματισμού Turbo-Prolog, που αποτελείται από λεξικό, κανόνες και λειτουργεί σαν πεπερασμένο αυτόματο (finite state automaton). Έχει δοκιμαστεί σε κείμενα 10.000 λέξεων και προβλέπει την αναγνώριση νεολογισμών. Ο επεξεργαστής ΑΘΗΝΑ βασίζεται στο θεωρητικό μοντέλο της Γενετικής Λεξικής Μορφολογίας και κατέδειξε την εφαρμοσιμότητά του στη νέα ελληνική.
- Οι επεξεργαστές διεπίπεδης μορφολογίας (βλ. Μαρκόπουλος 1998, Sgarbas et al. 1995).

- Το ελληνικό σύστημα ΠΦΓ που βασίζεται στο αντίστοιχο μοντέλο λογισμικού IDAS για την Αγγλική γλώσσα (βλ. Karamanis 2000).
- Το ΣΟΜΕ (βλ. Παπακίτσος 2000), με διακριτή μονάδα μορφολογικής γεννήτριας (βλ. Αλέξη 1997), υλοποιημένο σε γλώσσα προγραμματισμού Turbo-Pascal, που αποτελείται από λεξικό μορφημάτων και βασίζεται στα αλγοριθμικά μοντέλα της λειτουργικής αποσύνθεσης (βλ. Allen et al. 1987, Sproat 1992, Παπακίτσος 2014: 36-37) και της οκνηρής αναγνώρισης (βλ. Dura 1994, Παπακίτσος 2014: 41-42). Έχει δοκιμαστεί σε κείμενα 1.880.000 λέξεων (ελληνικό μέρος του ECI corpus) και η ακρίβεια ανάλυσης ξεπέρασε το 98% σε αντιπροσωπευτικό δείγμα ορθογραφικά σωστών λέξεων. Ο αυτόματος μορφολογικός αναλυτής του ΣΟΜΕ έχει υλοποιηθεί και σε γλώσσα Visual C (βλ. Ντάφλου 2013).
- Ένα μεγάλης κλίμακας μορφολογικό λεξικό της ελληνικής, που αναπτύχθηκε στο ΕΚΕΦΕ «Δημόκριτος» και χρησιμοποιήθηκε για την ανάπτυξη ενός λημματοποιητή και ενός μορφολογικού αναλυτή που χρησιμοποιήθηκαν σε διάφορες εφαρμογές επεξεργασίας φυσικής γλώσσας για τα ελληνικά, όπως π.χ. εξόρυξη και φιλτράρισμα πληροφορίας (βλ. Petasis et al. 2001).
- Το ενσωματωμένο σύστημα μορφολογικής επεξεργασίας του ελληνικού Wordnet (βλ. Galiotou et al. 2001, Γρηγοριάδου κ.ά. 2001, Kornilakis et al. 2003, 2004, Grigoriadou et al. 2004), που βασίζεται στο ΣΟΜΕ.
- Ένα πειραματικό σύστημα που βασίζεται στο πρότυπο του λογισμικού Linguistica (Goldsmith 2001) και επικεντρώνεται στην τεχνική ανεπίβλεπτης εκμάθησης μορφολογίας (unsupervised morphology learning technique), όπου επιχειρείται η υπολογιστική επεξεργασία συγκεκριμένων φαινομένων σχηματισμού λέξεων της Νέας Ελληνικής (όπως επιθηματοποίησης και σύνθεσης) μέσω της χρήσης διαφόρων κειμενοσωμάτων, χωρίς όμως ιδιαίτερη επιτυχία (βλ. Karasimos & Petropoulou 2010).

Η αρχιτεκτονική της μορφολογικής γεννήτριας εξαρτάται από τη μορφολογική τυπολογία και τη μορφοτακτική της υποκείμενης γλώσσας. Η νέα ελληνική, ως γλώσσα συνενωτικής (concatenative) μορφολογίας, χρειάζεται μια αρχιτεκτονική στην οποία οι γλωσσικές διεργασίες της κλίσης, της παραγωγής και της σύνθεσης να αντιμετωπίζονται από διαφορετική μονάδα του λογισμικού λόγω των διαφορετικών φαινομένων που εμφανίζονται κατά περίπτωση. Ειδικά στην περίπτωση της παραγωγής, εδώ εξετάζεται η διεργασία της προσφυματοποίησης (affixation), γιατί οι άλλες περιπτώσεις παραγωγής π.χ. με ετεροίωση ή μετάπλαση (βλ. Ράλλη 2011: 140-146) δεν αντιμετωπίζονται καλύτερα αλγοριθμικά, αλλά μέσω δομών δεδομένων.

Η κλίση αντιμετωπίζεται αλγοριθμικά με τετριμμένο τρόπο. Το προγραμματιστικό βάρος δίνεται στις δομές δεδομένων. Δηλαδή, δημιουργούνται σύνθετες λίστες θεμάτων και λίστες καταλήξεων (ο όρος *κατάληξη*, όπως και *κλιτικό επίθημα*, απαιτούν περαιτέρω συζήτηση). Οι λίστες θεμάτων περιέχουν τα συγγενή μεταξύ τους θέματα, όπως στις περιπτώσεις των ανισοσύλλαβων ουσιαστικών (π.χ. *ψαρ-*, *ψαραδ-*) και των ρηματικών θεμάτων (π.χ. *γραφ-*, *γραψ-*). Αυτά τα θέματα συνοδεύονται από δείκτες που παραπέμπουν στις συνδεδεμένες με αυτά λίστες καταλήξεων. Έτσι, με την επιλογή ενός θέματος και με τη βοήθεια του σχετικού δείκτη σχηματίζονται όλοι οι κλιτοί τύποι του λεξήματος. Οι λεπτομέρειες της υλοποίησης και οι θεωρητικές προεκτάσεις τους αξίζουν μελλοντικά ιδιαίτερης παρουσίασης.

Η σύνθεση αντιμετωπίζεται, κατά ένα μέρος, με παρόμοιο τρόπο με την κλίση μέσω δομών δεδομένων, αλλά και αλγοριθμικά. Τα χαλαρά πολυλεκτικά σύνθετα (βλ. Ράλλη 2011: 199-203), αλλά και αυτά με πρώτο συνθετικό θέμα ουσιαστικό ή ρήμα, αντιμετωπίζονται κυρίως με δείκτες που συνδέουν τα δύο συνθετικά μεταξύ τους. Τα

σύνθετα με πρώτο συνθετικό επίθετο ή επίρρημα μπορούν να αντιμετωπιστούν όπως η προηγούμενη κατηγορία (μέσω δομών δεδομένων, δηλαδή με λίστες και δείκτες) ή μέσω θεματικών ιδιοτήτων όπως η κατηγορία του θέματος (ουσιαστικό, επίθετο, ρήμα, επίρρημα). Στη δεύτερη περίπτωση, ο έλεγχος γίνεται αλγοριθμικά με την εξέταση της συμβατότητας μεταξύ των κατηγοριών των θεμάτων. Γενικά, ο αλγοριθμικός τρόπος αναγνωρίζει τους νεολογισμούς, αλλά προκαλεί υπερπαραγωγή αντιγραμματικών λέξεων. Τέλος, εδώ αντιμετωπίζονται τα μορφοφωνολογικά φαινόμενα όπως η διατήρησης ή μη του συνδετικού φωνήεντος (-ο-) π.χ. *αερ-άμυνα*, *αερ-ο-διάδρομος*, αλλά και γενικότερα η συμπεριφορά του φωνηεντικού χαρακτήρα του πρώτου συνθετικού όπως στα: *αρχι-* + *επίσκοπος* = *αρχιεπίσκοπος*, *αρχι-* + *άγγελος* = *αρχάγγελος*.

Οι μορφοφωνολογικές μεταβολές κωδικοποιούνται με τον τρόπο που φαίνεται αργότερα στην προσφύματοποίηση, χωρίς αυτό να σημαίνει ότι μπορούμε πάντα να αποφύγουμε την παραγωγή αντιγραμματικών τύπων όπως **αρχιεπίσκοπος*, **αρχιάγγελος*. Προφανώς, τα παραπάνω αποτελέσματα της μορφολογικής γεννήτριας σχετίζονται και με τον τρόπο που θα γίνει η διαχείριση των αχώριστων μορίων και της λεξικοποίησης γενικότερα.

Η διεργασία της προσφύματοποίησης χωρίζεται σε δύο διακριτές περιπτώσεις, την επιθηματοποίηση (suffixation) και την προθηματοποίηση (prefixation). Στην πρώτη ακολουθείται διαχείριση είτε μέσω δομών δεδομένων, είτε αλγοριθμικά μέσω ιδιοτήτων. Με τον πρώτο τρόπο, η ρίζα συνδέεται μέσω δείκτη με λίστα επιτρεπτών επιθημάτων όπως στα: *γραφ-+-ημ-*= *γράφημα* και *γραφ-+-ικ-*= *γραφικός*. Με το δεύτερο τρόπο, η επιθηματοποίηση γίνεται αλγοριθμικά με έλεγχο της ιδιότητας της κατηγορίας όπως: *γραφ-[ΡΗΜΑ]* + *-ικ-[ΡΗΜΑ/ΟΥΣ >ΕΠΙΘ]* = *γραφικός* [ΕΠΙΘ]. Έτσι βεβαίως δεν αποφεύγεται η υπερπαραγωγή αντιγραμματικών τύπων όπως *μαυρ-ίζω*, **μαυρ-ιάζω*, **μαυρ-άζω*, **μαυρ-ώνω*, **μαυρ-εύω* κ.λπ. Αντίστοιχα γίνεται η διαχείριση της προθηματοποίησης, δηλαδή μέσω δομών δεδομένων, σε έκταση ανάλογη του τρόπου κωδικοποίησης της λεξικοποίησης και των αχώριστων μορίων, όπως προηγουμένως, αλλά και γενικότερα των προθημάτων.

Κρίσιμη υπομονάδα της μορφολογικής γεννήτριας και απαραίτητη στις περισσότερες από τις προηγούμενες περιπτώσεις είναι αυτή της διαχείρισης των μορφοφωνολογικών μεταβολών. Εδώ υπάρχουν δύο δυνατότητες υλοποίησης. Στην πρώτη κωδικοποιούνται τα φωνητικά χαρακτηριστικά των φωνηέντων (βλ. Πίνακα 1) και των συμφώνων (βλ. Πίνακα 2). Κατόπιν, κωδικοποιούνται οι μεταβολές τους μέσω των κανόνων ενός θεωρητικού μοντέλου, όπως π.χ. της γενετικής φωνολογίας (βλ. Φιλιππάκη-Warburton 1992: 58-63, 88-90). Το αποτέλεσμα προκύπτει αλγοριθμικά από τη σύγκριση και την αναζήτηση της νέας μορφής που είναι συμβατή με την προκύπτουσα δέσμη φωνητικών χαρακτηριστικών.

Πίνακας 1. Φωνητικά χαρακτηριστικά των φωνηέντων

| | α | ε | η | ι | ο | υ | ω |
|------------------|---|---|---|---|---|---|---|
| εμπρόσθιο | - | + | + | + | - | + | - |
| υψηλό | - | - | + | + | - | + | - |
| χαμηλό | + | - | - | - | - | - | - |

Πίνακας 2. Φωνητικά χαρακτηριστικά των συμφώνων

| | β | γ | δ | ζ | θ | κ | λ | μ | ν | π | ρ | σ | τ | φ | χ |
|------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| συμφωνικό | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| φωνηεντικό | - | - | - | - | - | - | + | - | - | - | + | - | - | - | - |
| αντηχητικό | - | - | - | - | - | - | + | + | + | - | + | - | - | - | - |
| έρρινο | - | - | - | - | - | - | - | + | + | - | - | - | - | - | - |
| ηχηρό | + | + | + | + | - | - | + | + | + | - | + | - | - | - | - |
| διαρκές | + | + | + | + | + | - | - | - | - | - | - | + | - | + | + |
| εμπρόσθιο | + | - | + | + | + | - | + | + | + | + | + | + | + | + | - |
| κορονικό | - | - | + | + | + | - | + | - | + | - | + | + | + | - | - |
| οπίσθιο | - | + | - | - | - | + | - | - | - | - | - | - | - | - | + |
| υψηλό | - | + | - | - | - | + | - | - | - | - | - | - | - | - | + |
| πλευρικό | - | - | - | - | - | - | + | - | - | - | - | - | - | - | - |
| συριστικό | - | - | - | + | - | - | - | - | - | - | - | + | - | - | - |

Στη δεύτερη δυνατότητα υλοποίησης κωδικοποιούνται απευθείας οι μεταβολές των γειτονικών γραφημάτων, όπως αυτές προκύπτουν από το συνδυασμό τους (βλ. Πίνακα 3, για ένα μέρος των μεταβολών). Στον Πίνακα 3, στην επικεφαλίδα κάθε στήλης εκτός της πρώτης (οριζόντια) εμφανίζεται το δεύτερο στη σειρά γράφημα. Στην πρώτη στήλη του Πίνακα 3 και στο πρώτο κελί της κάθε γραμμής του (κατακόρυφα) εμφανίζεται το πρώτο στη σειρά γράφημα, ενώ στο αντίστοιχο κελί του καρτεσιανού συνδυασμού τους εμφανίζεται ο μετασχηματισμός του πρώτου γραφήματος όταν ακολουθεί το δεύτερο. Η παύλα (-) σηματοδοτεί την αποβολή του πρώτου γραφήματος. Ξεχωριστή καταχώριση γίνεται για τα δασυνόμενα φωνήεντα ώστε να αντιμετωπιστούν μεταβολές με ιστορικές καταβολές. (Δεν εμφανίζονται στον Πίνακα 3).

Πίνακας 3. Πίνακας μεταβολών γραφημάτων

| | α | β | γ | δ | ε | ζ | η | θ | ι | κ | λ | μ | ν |
|----|--------|---|---|---|--------|---|--------|---|--------|---|---|---|---|
| -α | α, -αν | α | α | α | α, -αν | α | α, -αν | α | α, -αν | α | α | α | α |
| -ε | ε,- | ε | ε | ε | - | ε | ε,- | ε | ε,- | ε | ε | ε | ε |
| -ι | ι,- | ι | ι | ι | ι,- | ι | ι,- | ι | ι,- | ι | ι | ι | ι |
| -κ | ξ | κ | κ | κ | ξ | κ | ξ | κ | ξ | κ | κ | κ | κ |
| -ν | ν | μ | γ | ν | ν | ν | ν | ν | ν | γ | λ | μ | ν |
| -ο | ο,- | ο | ο | ο | ο,- | ο | ο,- | ο | ο,- | ο | ο | ο | ο |
| -σ | σ | σ | σ | σ | σ | σ | σ | σ | σ | σ | σ | σ | σ |
| -υ | υ | υ | υ | υ | υ | υ | υ | υ | υ | υ | υ | υ | υ |

Εδώ εμφανίζεται ξανά το δίλημμα του τρόπου υλοποίησης των γλωσσικών φαινομένων: θα υλοποιηθεί το αίτιο της μεταβολής (πρώτος τρόπος) ή το αποτέλεσμα του (δεύτερος τρόπος); Για μεν τους γλωσσολόγους περισσότερο ενδιαφέρον έχει ο πρώτος τρόπος υλοποίησης, με τις δέσμες φωνητικών χαρακτηριστικών. Για τους γλωσσομηχανικούς προτιμότερος είναι ο δεύτερος τρόπος υλοποίησης, γιατί έτσι αποφεύγεται η αυξημένη υπολογιστική πολυπλοκότητα, η οποία είναι μετρήσιμη με διάφορες μεθόδους που ονομάζονται *μετρικές λογισμικού* (software metrics). Η αυξημένη υπολογιστική πολυπλοκότητα οδηγεί σε ανεπιθύμητα σφάλματα που μπορεί να είναι εξαιρετικά δύσκολο να εντοπιστούν προτού διορθωθούν.

3. Συμπεράσματα

Στην παρούσα εργασία παρουσιάστηκαν τα γενικά χαρακτηριστικά των πληροφοριακών συστημάτων ΠΦΓ, καθώς και μερικές δημοφιλείς εφαρμογές τέτοιων

συστημάτων. Ο πυρήνας λογισμικού ενός συστήματος ΠΦΓ υποστηρίχθηκε ότι είναι η μορφολογική γεννήτρια, η σχεδίαση της οποίας εξαρτάται από το είδος της φυσικής γλώσσας που διαχειρίζεται. Ως παράδειγμα υλοποίησης παρουσιάστηκε μια μορφολογική γεννήτρια της νέας ελληνικής που δημιουργήθηκε στα πλαίσια ενός συστήματος ολοκληρωμένης μορφολογικής επεξεργασίας (ΣΟΜΕ). Η νέα ελληνική ως γλώσσα συνενωτικής μορφολογίας και δεδομένης της μορφοτακτικής της χρειάζεται ξεχωριστή υπολογιστική αντιμετώπιση καθεμίας από τις τρεις μορφολογικές διεργασίες, της κλίσης, της παραγωγής/προσφύματοποίησης και της σύνθεσης. Για κάθε περίπτωση διεργασίας παρουσιάστηκαν παραδείγματα του τρόπου κωδικοποίησης των αντίστοιχων γλωσσικών φαινομένων και κάποιων συνεπειών στα αποτελέσματα της μορφολογικής γεννήτριας, ανάλογα με τον τρόπο κωδικοποίησης. Βασικό κριτήριο επιλογής του τρόπου αυτού στα πληροφοριακά συστήματα είναι η επίτευξη αυξημένης παραγωγής ορθών γραμματικά αποτελεσμάτων και μειωμένης υπολογιστικής πολυπλοκότητας.

Βιβλιογραφία

- Αλέξη, Α. 1997. Υλοποίηση γεννήτριας μορφολογικού επεξεργαστή και αξιολόγηση σχετικών αλγορίθμων. Πτυχιακή εργασία, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Τμήμα Πληροφορικής & Τηλεπικοινωνιών.
- Allen, J., Hunnicutt, M.S., Klatt, D. 1987. *From Text to Speech: The MITalk System*. Cambridge: Cambridge University Press.
- Anand, T. & Kahn, G. 1992. Making sense of gigabytes: A system for knowledge-based market analysis. In P. Klahr & A.F. Scott (eds), *Innovative Applications of Artificial Intelligence 4. Proceedings of the IAAI-92 Conference*. Menlo Park, CA: AAAI Press, 57-70.
- Belz, A. 2008. Automatic generation of weather forecast texts using comprehensive probabilistic generation-space models. *Natural Language Engineering* 14, 431-455.
- Dale, R. & Reiter, E. 2000. *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.
- Dura, E. 1994. Lexicon and lazy word parsing. *Proceedings of the International Conference Language Engineering on the Information Highway, 26-30 September 1994, Santorini, Greece*. Athens: ILSP.
- Γαλιώτου, Ε. 1991. Η τεχνητή νοημοσύνη στην επεξεργασία φυσικής γλώσσας. Ανακοίνωση στην Επιστημονική Ημερίδα Τεχνητής Νοημοσύνης, Αθήνα, 23 Φεβρουαρίου 1991, Ελληνική Εταιρεία Τεχνητής Νοημοσύνης.
- Galiotou, E., Giannopoulou, G., Grigoriadou, M., Ralli, A., Brewster, C., Arhakis, A., Papakitsos, E., Pantelidou, A. 2001. Semantic tests and supporting tools for the Greek Wordnet. Paper presented at the NAACL-2001 Workshop on "WordNet and other Lexical Resources", Carnegie Mellon University, Pittsburgh, USA.
- Galitsky, B. 2013. A web mining tool for assistance with creative writing. *Advances in information retrieval. Lecture Notes in Computer Science* 7814, 828-831.
- Galitsky, B., de la Rosa, J. L., Dobrocsi, G. 2012. Inferring the semantic properties of sentences by mining syntactic parse trees. *Data & Knowledge Engineering* 81-82, 21-45.
- Goldberg, E., Driedger, N., Kittredge, R. 1994. Using natural-language processing to produce weather forecasts. *IEEE Expert* 9 (2), 45-53.
- Goldsmith, J. 2001. Unsupervised learning of the morphology of a natural language. *Computational Linguistics* 27 (2), 153-198.
- Γρηγοριάδου, Μ., Γαλιώτου, Ε., Παντελίδου, Α., Παπακίτσος, Ε. 2001. Υπολογιστικά εργαλεία υποστήριξης του ελληνικού Wordnet. *Πρακτικά ημερίδας «Λεξικογραφικές Βάσεις Δεδομένων – Ηλεκτρονικοί Γλωσσολογικοί Πόροι» (στα πλαίσια της 22ης Συνάντησης Γλωσσολογίας του ΑΠΘ)*, 33-40.
- Grigoriadou, M., Kornilakis, H., Galiotou, E., Stamou, S., Papakitsos, E. 2004. The software infrastructure for the development and validation of the Greek Wordnet. *Journal of Information Science and Technology* 7 (1-2), 89-105.

- Harris, M.D. 2008. Building a large-scale commercial NLG system for an EMR. *Proceedings of the Fifth International Natural Language Generation Conference*. Association for Computational Linguistics Stroudsburg, PA, 157-160.
- Horacek, H. & Zock, M. 2015. *New Concepts in Natural Language Generation: Planning, Realization and Systems*. New York: Bloomsbury.
- Karamanis, N. 2000. Greek nominal morphology for a classification-based natural language generation system. *Proceedings of the 14th International Symposium on Theoretical and Applied Linguistics, 20-22 April 2000, Department of Theoretical and Applied Linguistics, School of English, Aristotle University of Thessaloniki*. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.20.221&rep=rep1&type=pdf>.
- Karasimos, A. & Petropoulou, E. 2010. A crash test with Linguistica in Modern Greek: The case of derivational affixes and bound stems. *Proceedings of the International Conference on Language Resources and Evaluation LREC 2010*, 182-189.
- Kornilakis, H., Grigoriadou, M., Galiotou, E., Papakitsos, E. 2003. Aligning, annotating and lemmatizing a corpus for the validation of Balkan Wordnets. Paper given at the Workshop on Balkan Language Resources and Tools, November 2003, Thessaloniki.
- Kornilakis, H., Grigoriadou, M., Galiotou, E., Papakitsos, E. 2004. Using a lemmatizer to support the development and validation of the Greek WordNet. Paper given at the Second Global Wordnet Conference, January 2004, Brno, Czech Republic.
- Law, A., Freer, Y., Hunter, J., Logie, R., McIntosh, N., Quinn, J. 2005. A comparison of graphical and textual presentations of time series data to support medical decision making in the neonatal intensive care unit. *Journal of Clinical Monitoring and Computing* 19 (3), 183-194.
- Μαρκόπουλος, Γ. 1998. Υπολογιστική επεξεργασία του νεοελληνικού ονόματος. Διδακτορική διατριβή, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.
- McDonald, D. D. 1992. Natural-language generation. In S. C. Shapiro (ed.), *Encyclopedia of Artificial Intelligence*. New York: Wiley, 983-997.
- Ντάφλου, Ο. 2013. Αυτόματος μορφολογικός αναλυτής της νέας ελληνικής σε γλώσσα προγραμματισμού Visual C: Βελτίωση αλγορίθμων. Διπλωματική εργασία, Διεπιστημονικό Διαπανεπιστημιακό ΠΜΣ «Τεχνογλωσσία», Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών & Εθνικό Μετσόβιο Πολυτεχνείο.
- Παπαθανασοπούλου, Γ. 2015. Σχεδιασμός και υλοποίηση βάσης δεδομένων για οργάνωση διπλωματικών εργασιών του Δ.Π.Μ.Σ Τεχνογλωσσία. Διπλωματική εργασία, Διεπιστημονικό Διαπανεπιστημιακό ΠΜΣ «Τεχνογλωσσία», Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών & Εθνικό Μετσόβιο Πολυτεχνείο.
- Παπακίτσος, Ε.Χ. 2000. Συμβολή στη μορφολογική επεξεργασία της νέας ελληνικής: Λειτουργική αποσύνθεση – Καρτεσιανό ηλεκτρονικό λεξικό. Διδακτορική διατριβή, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.
- Παπακίτσος, Ε.Χ. 2014. *Γλωσσική τεχνολογία λογισμικού: II. Πραγμάτωση*. Αθήνα: Εθνική Βιβλιοθήκη της Ελλάδος.
- Petasis, G., Karkaletsis, V., Farmakiotou, D., Androutsopoulos, I., Spyropoulos, C.D. 2001. A Greek morphological lexicon and its exploitation by a Greek controlled language checker. In Y. Manolopoulos and S. Evripidou (eds), *Proceedings of the 8th Panhellenic Conference in Informatics*. Berlin: Springer, 80-89.
- Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., Sykes, C. 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence* 173 (7-8), 789-816.
- Ράλλη, Α. 2011. *Μορφολογία*. Τέταρτη Έκδοση. Αθήνα: Πατάκης.
- Ralli, A. & Galiotou, E. 1987. A morphological processor for Modern Greek. *Proceedings of the 3rd European Meeting of the Association for Computational Linguistics*. Copenhagen: ACL, 26-31
- Ralli, A. & Galiotou, E. 1991. Affixation in Modern Greek: A computational treatment. Paper given at the EURISCON '91 Conference.
- Reiter, E., Sripada, S., Hunter, J., Yu, J., Davy, I. 2005. Choosing words in computer-generated weather forecasts. *Artificial Intelligence* 167, 137-169.

- Sauper, C. & Barzilay, R. 2009. Automatically generating Wikipedia articles: A structure-aware approach. *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, Suntec, Singapore, 2-7 August 2009*, 208-216. Available at: <http://www.aclweb.org/anthology/P09-1024>.
- Sgarbas, K., Fakotakis, N., Kokkinakis, G. 1995. A PC-KIMMO-based morphological description of Modern Greek. *Literary and Linguistic Computing* 10 (3), 189-201.
- Sproat, R.W. 1992. *Morphology and Computation*. Cambridge, Mass.: The MIT Press.
- Sripada, S., Burnett, N., Turner, R., Mastin, J., Evans, D. 2014. Generating a case study: NLG meeting weather industry demand for quality and quantity of textual weather forecasts. *Proceedings of INLG 2014*, 1-5
- Stent, A. & Bangalore, S. (eds) 2014. *Natural Language Generation in Interactive Systems*. Cambridge: Cambridge University Press.
- Φιλίππáκη-Warburton, E. 1992. *Εισαγωγή στη θεωρητική γλωσσολογία*. Αθήνα: Νεφέλη.