

ΟΡΟΛΟΓΙΑ ΚΑΙ ΑΥΤΟΜΑΤΗ ΜΕΤΑΦΡΑΣΗ

Ελένη Ευθυμίου

Περίληψη

Οι εφαρμογές επεξεργασίας φυσικής γλώσσας σε ΗΥ (NLP), διακρίνονται από ιδιαίτερη επιτυχία, όταν αφορούν θεματολογικά περιορισμένες υπογλώσσες. Το λεξιλογικό αλλά και το φραστικό περιεχόμενο μίας υπογλώσσας καλύπτεται, σε μεγάλο ποσοστό, από τα στοιχεία ορολογίας της. Στην επεξεργασία λοιπόν, με σκοπό την αυτόματη μετάφραση, η έγκαιρη αναγνώριση των στοιχείων ορολογίας, κυρίως των πολυλεξικών όρων, είναι καθοριστική.

Γιά την αυτόματη μεταφορά των όρων από τη μία γλώσσα στην άλλη κατά την μετάφραση, προτείνονται από την ερευνητική ομάδα ορολογίας του προγράμματος EUROTRA, δύο νέες μεθοδολογίες αναγνώρισης των όρων-φράσεων που βασίζονται στην χρήση κωδικών αριθμών αναγνώρισης (ΚΑΑ) των όρων, οι οποίοι είναι κοινοί για τις βάσεις δεδομένων ορολογίας, και στις δύο γλώσσες, που συμμετέχουν κάθε φορά στη μετάφραση. Οι μεθοδολογίες αυτές συνδέουν την βάση δεδομένων ορολογίας με τα αρχικά στάδια συντακτικής ανάλυσης της προς μετάφραση γλώσσας, με τρόπο ώστε να απλουστεύεται η διαδικασία μετάφρασης, να διευκολύνεται διαχειριστικά το σύστημα και να επιτυγχάνεται απόλυτη ταυτότητα εννοιών.

TERMINOLOGY AND MACHINE TRANSLATION

Eleni Efthimiou

Abstract

NLP applications, connected with specific sublanguages are considered especially successful. The lexical as well as the phrasal (for multiword units) component of a sublanguage is extensively derived by the terminology of this sublanguage. If our goal is Machine Translation, the early recognition of parts of sentences as belonging to terminology, especially in the case of multiword terms, is significant.

The treatment of terms in EUROTRA, was based on using 10-digit identification numbers for terms in the translation module, common to the term DBs of all languages involved in translation. This may activate the system's default transfer mechanisms from the source to the target language. For this purpose, two techniques were developed. Both connect the term DB with the early stages of syntactic analysis in the source language. As a result, the translation module is significantly simplified and coherence of the system as well as match of concepts are preserved.

Η αυτόματη ή μηχανική μετάφραση αποτελεί έναν χώρο συνεχούς έρευνας και προσπάθειας εφαρμογών λόγω της άμεσης χρησιμότητας και της αυξημένης ζήτησης στην διεθνή αγορά, αυτού του τύπου προϊόντων.

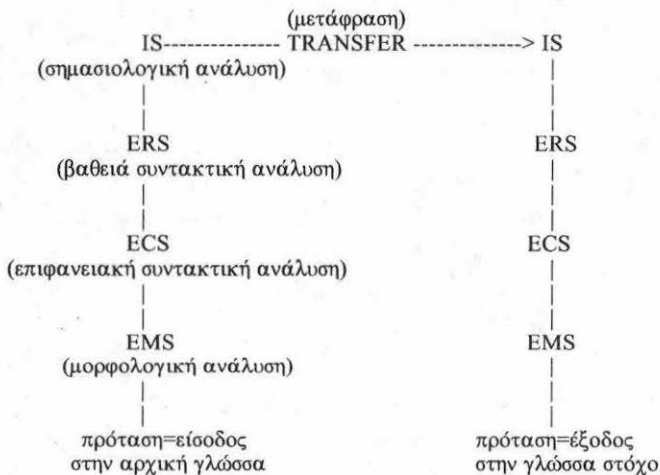
Είναι σημαντικό όμως για την καλύτερη προσέγγιση/αξιολόγηση αυτού του είδους εργαλείων, να γίνει αρχικά σαφής ο διαχωρισμός ανάμεσα στις περιπτώσεις μετάφρασης που χειρίζεται ένας φυσικός μεταφραστής και σε εκείνες που αναμένεται να χειριστεί ένα αυτόματο σύστημα επεξεργασίας. Αν σκεφθούμε την φυσική γλώσσα σαν βιολογικό σύστημα που χαρακτηρίζει τον άνθρωπο και απλουστευτικά την ονομάσουμε "Σύνολο Γ", τότε αν απομονώσουμε από αυτό το "υποσύνολο γ", των γραμματικών/συντακτικών φαινομένων και των στοιχείων του λεξικού που χρησιμοποιεί ένας περιορισμένος θεματικά χώρος για να καλύψει τις γλωσσικές ανάγκες του (έκφραση, πληροφορία, επικοινωνία κ.λ.π.), φθάνουμε στην έννοια της Υπογλώσσας (Sublanguage).

Η έννοια Υπογλώσσα θα κυριαρχήσει σε ό,τι θα ακολουθήσει, δεδομένου ότι οι πλέον επιτυχημένες εφαρμογές στον χώρο της Επεξεργασίας Φυσικής Γλώσσας, όπως αυτός είναι ευρύτερα γνωστός, αφορούν στην ανάλυση ή σύνθεση κειμένων ορισμένης θεματολογίας κάθε φορά, ώστε να γίνεται εφικτή η απομόνωση και συμμετοχή στα εργαλεία (modules) ανάλυσης/ σύνθεσης, στον μέγιστο δυνατό βαθμό αλγοριθμικής τους κάλυψης, των στοιχείων του λεξικού και της σύνταξης που αποτελούν την υπογλώσσα του συγκεκριμένου χώρου.

Όταν τα κείμενα μίας υπογλώσσας αποτελούν αντικείμενο αυτόματης μετάφρασης από μία αρχική γλώσσα σε μία γλώσσα στόχο, τα στάδια επεξεργασίας στον ΗΥ καθορίζονται συνήθως ως εξής:

- α) Στάδιο Ανάλυσης: περιλαμβάνει τα υποσύνολα της γραμματικής (σύνολο κανόνων συντακτικής αναγνώρισης φυσικής γλώσσας από τον ΗΥ) και του λεξικού που απαιτούνται για την αναγνώριση/ανάλυση του συνόλου των προτάσεων της προς μετάφραση, αρχικής γλώσσας.
- β) Στάδιο Μετάφρασης: επιτυγχάνει την ταύτιση των δομών της αρχικής γλώσσας με τις αντίστοιχες δομές στην γλώσσα στόχο και περιλαμβάνει το σύνολο των λημμάτων του απαιτούμενου δίγλωσσου λεξικού καθώς και τους κανόνες εκείνους που εξομαλύνουν τις δομικές διαφορές ανάμεσα στις δύο γλώσσες.
- γ) Στάδιο Σύνθεσης: περιλαμβάνει το υποσύνολο της γραμματικής και του λεξικού που απαιτούνται για την σύνθεση του κειμένου στην γλώσσα στόχο.

Τα στάδια αυτά τηρούνται στο σύστημα που ανέπτυξε το πρόγραμμα EUROTRA, στο οποίο ορίζονται τέσσερα επίπεδα μονογλωσσικής επεξεργασίας, όμοια για την ανάλυση και την σύνθεση, σύμφωνα με τον ακόλουθο πίνακα.



Από όσα ήδη αναφέραμε γίνεται σαφές, ότι το λεξικό αποτελεί σημαντικό στοιχείο και των τριών σταδίων επεξεργασίας. Και όταν μιλάμε για ένα λεξικό που αποτελεί μέρος ενός μηχανισμού (module) επεξεργασίας φυσικής γλώσσας, αναφερόμαστε, ουσιαστικά, σε μία βάση δεδομένων η οποία περιέχει τα λήμματα από όλες τις λέξεις που εμφανίζονται στο corpus των κειμένων που αφορούν την υπογλώσσα του εκάστοτε ενδιαφέροντός μας.

Τα στοιχεία της βάσης δεδομένων του λεξικού είναι οργανωμένα σύμφωνα με την συντακτική τους κατηγορία και με τέτοιο τρόπο ώστε να παρέχουν όλες τις λεξικές πληροφορίες που χρειάζεται το σύστημα των κανόνων της γραμματικής για να προχωρήσει στην επεξεργασία μίας πρότασης και να αποφασίσει αν θα ολοκληρωθεί η διαδικασία αναγνώρισης ή αν θα σταματήσει και θα απορριφθεί η προς ανάλυση είσοδος (=πρόταση).

Όμως δεν είναι μόνον η βάση δεδομένων των κοινών στοιχείων του λεξικού που συμμετέχει στις διαδικασίες επεξεργασίας των προτάσεων μίας υπογλώσσας. Πολύ σημαντική θέση σε μία υπογλώσσα κατέχουν τα στοιχεία ορολογίας που διέπουν την υπογλώσσα αυτή και ορίζουν τον χαρακτήρα του θεματικού της χώρου, καθώς και την διαφοροποίηση της από άλλες υπογλώσσες διαφορετικής θεματολογίας.

Τα στοιχεία ορολογίας μίας υπογλώσσας συγκεντρώνονται στην βάση δεδομένων ορολογίας που συμμετέχει παράλληλα με την βάση δεδομένων κοινών στοιχείων του λεξικού στην επεξεργασία των προς ανάλυση ή σύνθεση προτάσεων.

Αυτό όμως, που κυρίως διαφοροποιεί μία βάση δεδομένων ορολογίας από μία βάση δεδομένων κοινών στοιχείων του λεξικού, είναι ότι ενώ στην περίπτωση του κοινού λεξικού, στην βάση δεδομένων δεν περιέχεται ποτέ συντακτική κατηγορία μεγαλύτερη από την μηδενική κατηγορία που αντιστοιχεί στη λέξη, μία βάση δεδομένων ορολογίας μπορεί να περιέχει όχι μόνο μονολεκτικούς όρους, αλλά και όρους που αντιστοιχούν σε μεγαλύτερες συντακτικά, φραστικές κατηγορίες, χωρίς να είναι τυπικά δυνατό να περιορισθεί το είδος ή το μέγεθος των συντακτικών κατηγοριών που θα μπορούσαν, δυνάμει, να χαρακτηρισθούν όροι μίας υπογλώσσας.

Παράδειγμα αποτελούν οι ακόλουθοι όροι από την περιοχή των τηλεπικοινωνιών:

ΛΗΜΜΑ	ΣΥΝΤΑΚΤΙΚΗ ΚΑΤΗΓΟΡΙΑ	ΚΑΑ
δορυφόρος	ουσιαστικό	2000017656
δορυφορική		
ζεύξη	ονοματική φράση	2000017708
δορυφόρος με σταθεροποίηση		
τριών αξόνων	ονοματική φράση	2000022071

Το λήμμα "δορυφόρος" μπορεί να εμφανίζεται συγχρόνως σε μία βάση δεδομένων ορολογίας και σε μία βάση δεδομένων κοινών στοιχείων του λεξικού. Αυτό όμως, είναι αδύνατο στην περίπτωση φραστικών/πολυλεξικών κατηγοριών, όπως το παράδειγμα "δορυφορική ζεύξη", το οποίο μπορεί να συμπεριληφθεί, αποκλειστικά σε μία βάση δεδομένων ορολογίας.

Το γεγονός αυτό δημιουργεί κατ' αρχήν σημαντικά προβλήματα στην υπολογιστική επεξεργασία, όταν απαιτείται να ληφθεί υπόψη ο χαρακτηρισμός μίας φραστικής ενότητας σαν όρος.

Το δομικό δένδρο που ακολουθεί, με επαναλαμβανόμενη υπόταξη ΟΦ-συμπληρωμάτων σε Γενική Πτώση, μπορεί να αναλύει έναν όρο ("κέντρο περιοχής κάλυψης" στο παράδειγμά) καθώς και οποιαδήποτε ΟΦ με την ίδια δομή.



Από την άλλη πλευρά, θα ήταν ανεδαφικό να αγνοηθεί ο χαρακτηρισμός "όρος", μίας φραστικής ενότητας, σε μία συγκεκριμένη υπογλώσσα, ιδιαίτερα όταν τελικός σκοπός της επεξεργασίας είναι η αυτόματη μετάφραση.

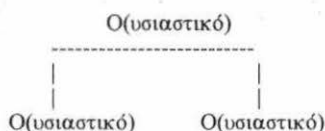
Αν σκεφθούμε τον ορισμό του όρου που δεν είναι άλλος από την σύμβαση/δέσμευση ένα κομμάτι λόγου να ταυτίζεται αποκλειστικά με μία και μόνο μία έννοια μέσα σε μία υπογλώσσα, έτσι ώστε να έχει ενεπρος-ένα απόλυτη εννοιολογική ταύτιση με τον αντίστοιχο όρο σε οποιαδήποτε φυσική γλώσσα, ανεξάρτητα από τα συντακτικά ή λεξιλογικά μέσα που χρησιμοποιούν διαφορετικές φυσικές γλώσσες για να εκφράσουν ίδιες έννοιες, τότε γίνεται κατανοητή η σπουδαιότητα της αναγνώρισης των όρων για την αυτόματη μετάφραση.

Αυτό τονίζεται ιδιαίτερα από το γεγονός ότι ήδη υπάρχουν γνωστές προσπάθειες συλλογής όρων από διάφορους θεματικούς χώρους, σε κεντρικές βάσεις δεδομένων, όπου η κάθε έννοια ταυτίζεται με έναν κωδικό αριθμό αναγνώρισης (ΚΑΑ) και κάθε γλώσσα μπορεί να προσθέτει τις δικές της αντιστοιχίες στις έννοιες που αντιπροσωπεύουν οι διάφοροι κωδικοί αριθμοί. Παράδειγμα αυτής της προσπάθειας αποτελεί η καταχώριση όρων στην βάση euroidicautom, όπου σε κάθε έννοια-όρο αντιστοιχεί ένας δεκαψηφίος αριθμός αναγνώρισης.

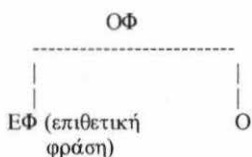
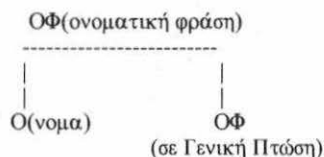
Στη συνέχεια, θα προσπαθήσουμε να δείξουμε πόσο είναι σημαντικό για την υπολογιστική επεξεργασία, όταν σκοπεύουμε στην αυτόματη μετάφραση, να εντοπίζεται ο ΚΑΑ ενός όρου στα πρώιμα στάδια επεξεργασίας, ώστε στο στάδιο μετάφρασης να καταλήγουμε σε μεταβίβαση του ΚΑΑ μόνο, απ την αρχική γλώσσα στην γλώσσα στόχο. Στη διαδικασία αυτή δεν συμμετέχει μετάφραση των φραστικών κατηγοριών ή των λεξιλογικών στοιχείων που απαρτίζουν τον όρο στις δύο γλώσσες.

Γιά τον σκοπό αυτό θα χρησιμοποιήσουμε δύο ενδεικτικά παραδείγματα περιπτώσεων μεταφραστικής/υπολογιστικής πολυπλοκότητας, χρησιμοποιώντας ορολογία δορυφορικής επικοινωνίας από τα Αγγλικά και Ελληνικά.

Στα Αγγλικά ο σχηματισμός νέων όρων, για έναν πολύ μεγάλο αριθμό περιπτώσεων, ακολουθεί το παραγωγικό σχήμα του σύνθετου ουσιαστικού, που αποτελείται από δύο ουσιαστικά τα οποία συνδιάζονται στο επίπεδο της λέξης για να δώσουν μία καινούρια λέξη, σύνθετη αυτή τη φορά, σύμφωνα με το ακόλουθο σχήμα.



Στη μετάφραση στα Ελληνικά, το σχήμα αυτό συναντά δύο δομικά αντίστοιχα σχήματα για την επιλογή ανάμεσα στα οποία, δεν υπάρχει δυνατή πρόβλεψη. Τα σχήματα αυτά είναι τα ακόλουθα:



Τα παραδείγματα που ακολουθούν δηλώνουν το πρόβλημα.

- ι) Αγγλικά: [O [O_satellite] [O_communication]
 Ελληνικά: [ΟΦ [ΕΦ_δορυφορική] [O_επικοινωνία]
- ιι) Αγγλικά: [O [O_ground] [O_station]
 Ελληνικά: [ΟΦ [O_σταθμός] [ΟΦ_GEN_εδάφους]

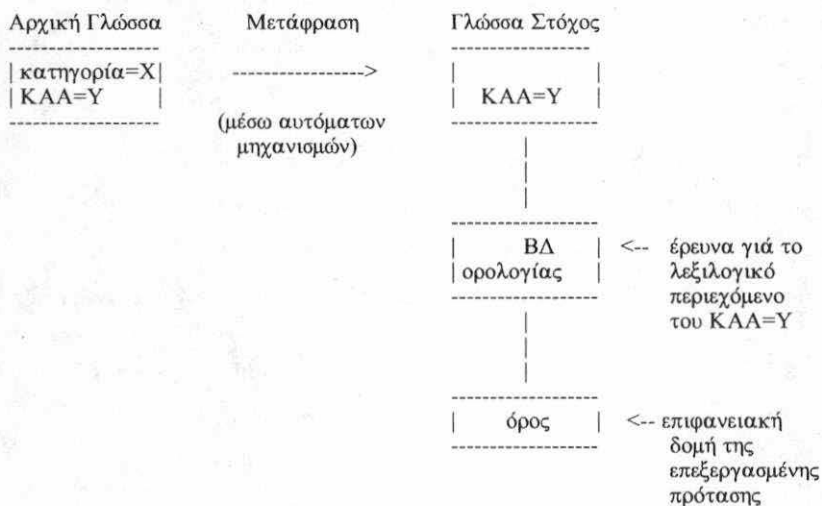
Είναι γεγονός, ότι ένας πολύ μεγάλος αριθμός τεχνικών όρων στα αγγλικά αλλά και σε γλώσσες που η δημιουργία σύνθετων λέξεων αποτελεί εξαιρετικά παραγωγική διαδικασία, όπως στα Γερμανικά και Ολλανδικά, ακολουθεί το παράδειγμα που μόλις αναφέραμε, και για τον σχηματισμό σύνθετων με περισσότερες από δύο λέξεις.

Οι υπολογιστικές δυσκολίες είναι εμφανείς. Είναι αναγκαίο να υπάρχει ένας ξεχωριστός κανόνας, στο στάδιο μετάφρασης, για κάθε περίπτωση με διαφορετικό λεξιλογικό περιεχόμενο, πράγμα που επιβραδύνει την επεξεργασία (χρειάζεται μία λίστα ειδικών κανόνων) και κάνει το σύστημα αντιοικονομικό και δυσκίνητο.

Αν, αντίθετα, οι σύνθετοι όροι αναγνωρίζονται σε κάποιο στάδιο της μονογλωσσικής ανάλυσης, πριν από το στάδιο μετάφρασης και συνοδεύονται από τον ΚΑΑ που τους αντιστοιχεί στην βάση δεδομένων ορολογίας, δεδομένου ότι ο αριθμός αυτός είναι κοινός για όλες τις γλώσσες, η διαδικασία μετάφρασης μπορεί να περιοριστεί στην αυτόματη μεταφορά του αριθμού αναγνώρισης από την μία γλώσσα στην άλλη.

Με αυτόν τον τρόπο υπεραπλουστεύεται η διαδικασία, το σύστημα μετάφρασης γίνεται πολύ πιο ευέλικτο και γρήγορο, καθώς επίσης επιτυγχάνεται εκατότοις-εκατό πιστότητα μετάφρασης αφού η διαδικασία της σύνθεσης στην γλώσσα στόχο αφορά απλή έρευνα στην βάση δεδομένων ορολογίας για τον εντοπισμό του λεξιλογικού περιεχομένου του όρου που έχει μεταφερθεί.

Η όλη διαδικασία παρουσιάζεται στον πίνακα που ακολουθεί:



Η διαδικασία που περιγράφεται εδώ, εντάσσεται απόλυτα σε ένα σύστημα αυτόματης μετάφρασης σαν αυτό του προγράμματος EUROTRA.

Το δεύτερο παράδειγμα μας αφορά στην περίπτωση εκείνη, όπου απαιτείται αλλαγή της συντακτικής κατηγορίας από γλώσσα σε γλώσσα, για να επιτευχθεί σωστή μετάφραση.

Εάν θελήσουμε, στο προηγούμενο παράδειγμα από τα Αγγλικά, να ακολουθήσουμε μία αναλυτική διαδικασία μετάφρασης, όπου η ανάλυση και μεταφορά από την αρχική γλώσσα στην γλώσσα στόχο φθάνει για όλα τα στοιχεία μέχρι τα φύλλα του δομικού δένδρου, δηλαδή τις λέξεις, και όπου οι σύνθετοι όροι μεταφράζονται όπως οι υπόλοιπες φραστικές ενότητες, τότε θα υποχρεωθούμε πάλι σε έναν εξαιρετικά μεγάλο αριθμό αυθαίρετων ειδικών κανόνων.

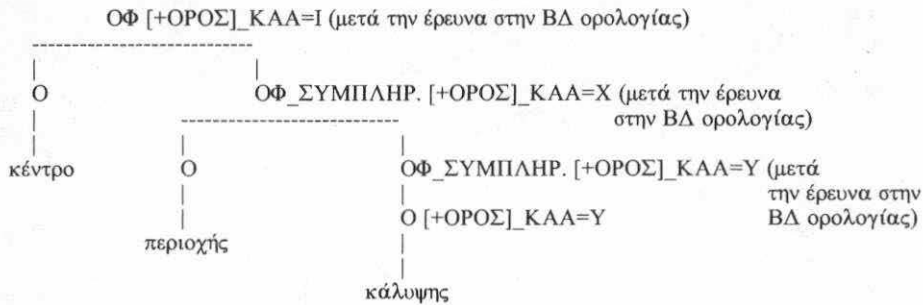
Στο σχήμα [N [N] [N]], όπως στο "satellite communication", η λέξη "satellite" χαρακτηρίζεται "ουσιαστικό" στα Αγγλικά. Για να μεταφραστεί στα Ελληνικά ο όρος, η συντακτική κατηγορία ουσιαστικό πρέπει να μετατραπεί στην συντακτική κατηγορία επίθετο. Επιβάλλεται λοιπόν ένας δομικός κανόνας μετάφρασης που ουσιαστικά σκοτώνει ό,τι έρχεται από την αρχική γλώσσα για να το αντικαταστήσει με ό,τι είναι αποδεκτό στη γλώσσα στόχο.

Στην περίπτωση αυτή, το πρώτο μέρος του κανόνα θα πρέπει να περιγράψει ακριβώς και μόνο την δομή που αφορά, ώστε να είναι δυνατός ο εντοπισμός της αλγοριθμικά, ενώ το δεύτερο μέρος του να χτίζει μία καινούρια δομή σύμφωνα με τα δομικά σχήματα της γλώσσας στόχου.

Για να γίνει καλύτερα κατανοητή αυτή η διαδικασία, παρατίθεται ένας σύνθετος κανόνας - δείγμας μετάφρασης από τα Γερμανικά στα Ελληνικά διάλυσης δομής συμπληρωματικής πρότασης με "ότι", και αντικατάστασης της με απαρεμφατική πρόταση.

```
rule8 = NP: {cat=np} [N: {cat=n, d_lu=vorschlag},  
    ANY: * {},  
    ~: {cat=s, role=arg2, stype=subord, dia=D, sNEG=N, index=I}  
    [V: {cat=v}, ANY1: * {} ]]  
=>  
NP <N, ANY, {cat=s, role=arg2, el_mood=sub, stype=subord,  
dia=D, neg=N, sTENSE='$nil', el_diath=act, el_mstense=untns, index=I}  
<V, ANY1> >.
```

Αλλά, δεδομένου ότι, οι λεξικές κατηγορίες επιδρούν σε ολόκληρη την συντακτική δομή που τις περιέχει, οι κανόνες διάλυσης της αρχικής δομής θα πρέπει να είναι τόσο περιεκτικοί, ώστε να περιλαμβάνουν τις λεξικές αλλά και τις φραστικές κατηγορίες που εξαρτώνται ή χαρακτηρίζονται από αυτές, όπως στις δομές που ακολουθούν:



Η συμπληρωματική τεχνική στην συντακτική αναγνώριση, βασίζεται στην χρήση γενικών σχημάτων (templates) όρων κάτω από τα οποία παρατίθενται λίστες με τις διάφορες λεξικές εμφανίσεις που συμφωνούν με το κάθε σχήμα. Στην περίπτωση αυτή, οι κανόνες περιγράφουν στο αριστερό τους μέρος, το περιεχόμενο των όρων σύμφωνα με το γενικό σχήμα, ενώ στο δεξί, δίνουν τον ΚΑΑ που αντιστοιχεί σε κάθε περίπτωση. Όπου το γενικό σχήμα των κανόνων είναι το ακόλουθο:

```

:bits:
temp1 = ($head1,$head2 < - > $head3)

where:
t1: ($head1=x,$head2=y < - > $head3=z),
t2: etc...

```

Αυτός ο μηχανισμός εφαρμόζεται στην διαδικασία ανάλυσης, στο πέρασμα από την επιφανειακή δομή στην βαθειά σημασιολογική δομή των προτάσεων. Εφαρμόζεται δε, σαν φίλτρο για τις περιπτώσεις εκείνες που δεν καλύπτονται από την εφαρμογή της συντακτικής αναγνώρισης.

Και στην περίπτωση των γενικών σχημάτων όρων, καταλήγουμε στο ζητούμενο αποτέλεσμα. Ο ΚΑΑ ενός όρου παρέχεται σαν χαρακτηριστικό της φραστικής κατηγορίας που τον απαρτίζει, στην διαδικασία ανάλυσης. Έτσι στο στάδιο μετάφρασης είναι δυνατή η ενεργοποίηση της διαδικασίας αυτόματης μεταφοράς, που απλουστεύει το σύστημα, σύμφωνα με την θέση που αναπτύχθηκε εδώ.

Σε ό,τι προηγήθηκε, θελήσαμε να δείξουμε πόσο σημαντική διαδικασία για την αυτόματη μετάφραση, είναι η αναγνώριση των πολυλεξικών όρων στα αρχικά στάδια ανάλυσης/επεξεργασίας, καθώς και η σύνδεση κάθε όρου με έναν ΚΑΑ, κοινό για όλες τις βάσεις δεδομένων ορολογίας των γλωσσών που συμμετέχουν σε μία εφαρμογή αυτόματης μετάφρασης με δεδομένη υπογλώσσα.

Στόχος της ερευνητικής προσπάθειας είναι να περιορισθεί η μετάφραση, κυρίως των πολυλεξικών όρων, στην μεταβίβαση των ΚΑΑ από την αρχική γλώσσα στην γλώσσα στόχο. Η επιτυχία αυτής της προσπάθειας εξασφαλίζει οικονομία στην διαχείριση, ταχύτητα εκτέλεσης του προγράμματος και απόλυτη μεταφραστική πιστότητα των εννοιών. Σημείο κλειδί για την επίτευξη του στόχου αυτού, αποτελεί η σύνδεση της διαδικασίας γραμματικής ανάλυσης με την βάση δεδομένων ορολογίας στην αρχική γλώσσα, βάσει ενός μηχανισμού αναζήτησης και με την προϋπόθεση, ότι η υπογλώσσα που μας απασχολεί κάθε φορά, έχει πλήρως αποκωδικοποιηθεί και στο δομικό και στο λεξιλογικό επίπεδο. Η

σπουδαιότητα αυτού του χειρισμού της ορολογίας, γίνεται σαφής, αν ληφθεί υπόψη το γεγονός, ότι, σε ένα πολύ μεγάλο ποσοστό, το λεξιλογικό, και για τους πολυλεξικούς όρους, το φραστικό περιεχόμενο μιας υπογλώσσας, αποτελείται από όρους της θεματολογίας της.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Chomsky, N., Lectures on Government and Binding: the Pisa Lectures, Foris Publications, 1981.
2. Chomsky, N., Some Concepts and Consequences of the Theory of Government and Binding, MIT Press, 1982.
3. Chomsky, N., Barriers, MIT Press, 1986.
4. Efthimiou, E., "A Surface-Syntax Based Approach to Terminology in NLP", paper presented in Eurotra annual meeting, Sept., 1992 (to appear).
5. Eurotra Reference Manual 7.0, Commission of the European Communities, Dec. 1990
6. Copeland, C., J. Durand, S Krauwer, B. Maegaard ed., The Eurotra Linguistic Specifications, in Studies in Machine Translation and Natural Language Processing, published by Commission of the European Communities, 1991.
7. Talbot, G. et al., "Interim Report on the Integration of Terms into the Eurotra System", Feb., 1992.
8. Talbot, G. et al., "Integration of Terms into ETS", paper presented in Eurotra annual meeting, Sept., 1992 (to appear).
9. Koopman, L., The Syntax of Verbs: From V-movement Rules in the Kru-languages to Universal Grammar, Unpublished Doctoral Dissertation, McGill University, 1983.
10. Jackendoff, R., X' Syntax: A Study of Phrase Structure, MIT Press, 1977.

Ελένη Ευθυμίου Δρ. Γλωσσολογίας Senior Researcher
ΙΕΛ-Ινστιτούτο Επεξεργασίας Λόγου
Μάργαρα 22, Ν. Ψυχικό, Αθήνα 115 25, τηλ.: 6712250
e-mail: eefthimiou@eurokom.ie