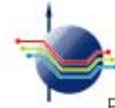


Ερευνητικό Κέντρο Αθηνά
Athena Research Center

Ερευνητικό Κέντρο Καινοτομίας στις Τεχνολογίες
της Πληροφορίας, των Επικοινωνιών, της Γνώσης

Research and Innovation Center in Information,
Communication and Knowledge Technologies



Ερευνητικό Κέντρο Αθηνά
Athena Research Center

Η εξέλιξη στα συστήματα Μηχανικής Μετάφρασης

Σοφιανόπουλος Σωκράτης

Ινστιτούτο Επεξεργασίας του Λόγου



Δομή παρουσίασης

- Τι είναι η Μηχανική Μετάφραση (Machine Translation)
- Ιστορική αναδρομή
- Είδη συστημάτων Μηχανικής Μετάφρασης
- Η Μηχανική Μετάφραση σήμερα
- Παράδειγμα: Το μεταφραστικό σύστημα METIS-II
- Το Μέλλον της Μηχανικής Μετάφρασης



Τι είναι η Μηχανική Μετάφραση

- Αυτόματη μετάφραση κειμένου από μια γλώσσα σε μια άλλη μέσω υπολογιστή
- Από τις πρώτες εφαρμογές της Υπολογιστικής Γλωσσολογίας
- Διαφορετικές τεχνικές για παραγωγή της μετάφρασης
- Μεγάλης σπουδαιότητας
- Ακαδημαϊκό και εμπορικό ενδιαφέρον
- Σύνθετο και απαιτητικό πρόβλημα
 - Χρόνος ανάπτυξης
 - Υπολογιστικό κόστος
 - Ποιότητα μετάφρασης
 - Κριτήρια αξιολόγησης



Ιστορική Αναδρομή

1954: Τα πρώτα βήματα

- Παρουσίαση του 1^{ου} συστήματος MM: Georgetown-IBM
- Ανταγωνισμός Ρωσίας – Η.Π.Α.

1966: Η απομυθοποίηση

- Έκθεση ALPAC: περιορισμός κονδυλίων
- Αναθεώρηση προσεγγίσεων

1970-80: Οι πρώτες επιτυχίες

- Διερεύνηση νέων τεχνικών
- Συστήματα νέας γενιάς: π.χ. SYSTRAN, METEO

1980-90: Αναθέρμανση ενδιαφέροντος

- Πολύγλωσσα συστήματα (EUROTRA: 9 γλώσσες της ΕΕ, 72 ζεύγη)
- Στατιστικά συστήματα MM
- Ανάπτυξη γλωσσικών πόρων (εργαλεία ανάλυσης, σώματα κειμένων)



Είδη συστημάτων Μηχανικής Μετάφρασης

A) Μηχανική Μετάφραση με κανόνες

B) Μηχανική Μετάφραση με στοχαστικές μεθόδους



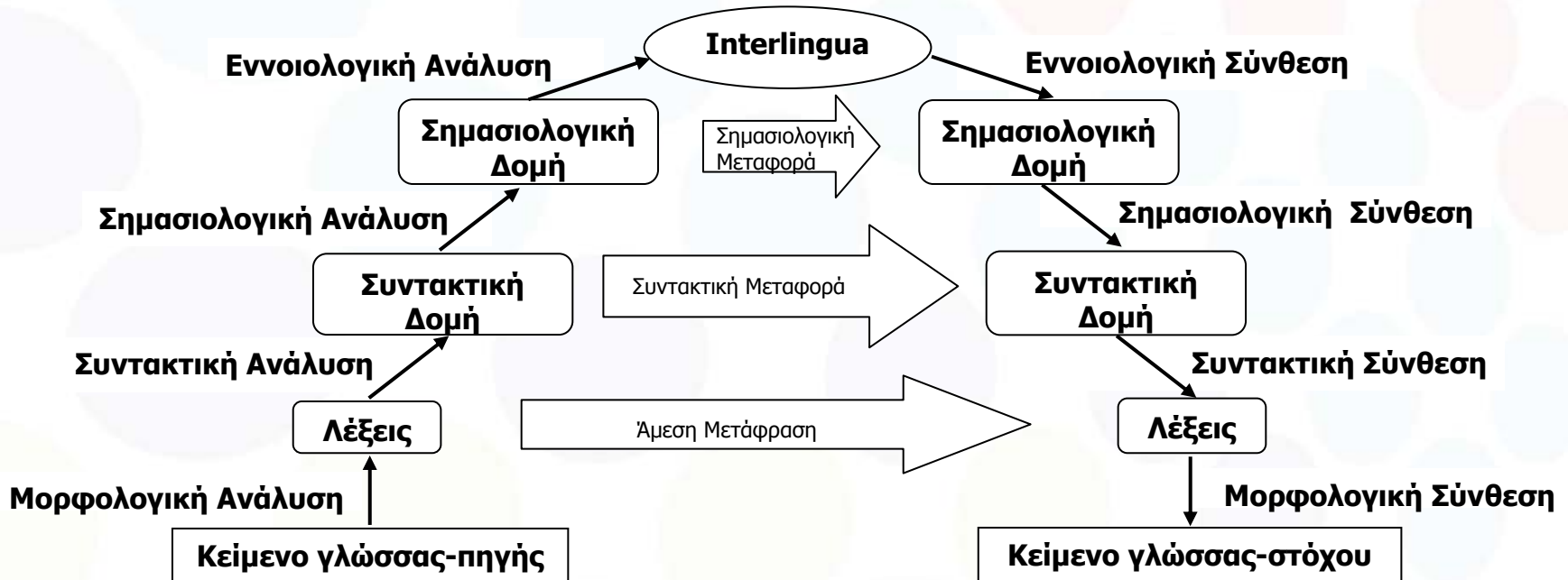
A) Μηχανική Μετάφραση με κανόνες

- Χρήση σύνθετων κανόνων για τη μετάφραση
- Η υλοποίηση εξαρτάται από το γλωσσικό ζεύγος
- + Υψηλή ποιότητα μετάφρασης
- + Αντιμέτωπιση σύνθετων μεταφραστικών φαινομένων
 - Ιδιαίτερα χρονοβόρα διαδικασία ανάπτυξης
 - Ανεξέλεγκτη αύξηση του πλήθους των κανόνων



Κατηγοριοποίηση Κανονιστικών Συστημάτων

- Ανάλογα με το επίπεδο ανάλυσης:
 - Συστήματα Άμεσης Μετάφρασης (Direct)
 - Συστήματα Μεταφοράς (Transfer)
 - Διαγλωσσικά Συστήματα (Interlingua)





B) Στοχαστικές Μέθοδοι

- Σύγχρονες και ελπιδοφόρες προσεγγίσεις
- Βασίζονται αποκλειστικά σε δεδομένα (συλλογές κειμένων)
- Αξιοποιούν τις τεχνολογικές εξελίξεις άλλων τομέων
- Βασικές προσεγγίσεις
 - Στατιστική Μηχανική Μετάφραση (Statistical)
 - Μετάφραση βάσει παραδειγμάτων (Example-based)
 - Υβριδικά συστήματα μετάφρασης (Hybrid)



B1) Στατιστική Μηχανική Μετάφραση

- Στατιστικά μοντέλα, των οποίων οι παράμετροι προκύπτουν από την ανάλυση παράλληλων σωμάτων κειμένων
- Αντιμετώπιση μετάφρασης ως υπολογιστικού προβλήματος
- + Γενική μέθοδος, εφαρμόσιμη σε οποιοδήποτε γλωσσικό ζεύγος
- + Πλήρης απουσία γλωσσικών κανόνων
 - Αδυναμία μοντελοποίησης σύνθετων γλωσσικών φαινομένων
 - Μη αξιοποίηση εργαλείων γλωσσικής ανάλυσης
 - Απαραίτητοι πόροι: δεν είναι διαθέσιμοι για όλες τις γλώσσες

B2) Μετάφραση βάσει παραδειγμάτων

- Προσομοίωση της λειτουργίας του ανθρώπου κατά τη μετάφραση
- Μετάφραση κατ' αναλογία με χρήση παράλληλων κειμένων με παραδείγματα μεταφράσεων
- Απλή μορφή τους: Μεταφραστικές μνήμες
- + Ικανοποιητική απόδοση σε ειδικές εφαρμογές (π.χ. μετάφραση καταλόγων)
- + Χρήση ως υποσυστήματα σε άλλα μεταφραστικά συστήματα
 - Αδυναμία μοντελοποίησης σύνθετων γλωσσικών φαινομένων
 - Απαραίτητοι πόροι: δεν είναι διαθέσιμοι για όλες τις γλώσσες
 - Απαιτούν μεγάλη κάλυψη των φαινομένων κάθε γλώσσας
 - Χειρότερα αποτελέσματα από τα στατιστικά συστήματα

B3) Υβριδικά Συστήματα

- Συνδυασμός τεχνικών από τις παραπάνω κατηγορίες συστημάτων
- Αξιοποίηση των προτερημάτων κάθε μεθόδου
- Βασίζονται κυρίως στα στατιστικά, εμπλουτίζοντάς τα με πόρους και μεθόδους από άλλες τεχνικές
 - Γραμματική, σημασιολογική και συντακτική πληροφορία
 - Στατιστική πληροφορία και συντακτικές φράσεις
- Συνδυασμός υπάρχόντων συστημάτων
 - Διαφορετική τεχνική για κάθε γλωσσικό ζεύγος
 - Παράλληλη λειτουργία: πολλά συστήματα για μια μετάφραση
 - Γραμμική λειτουργία: αξιοποίηση κάθε συστήματος σε συγκεκριμένη εργασία
 - Παράδειγμα: Ανάπτυξη του Google Translate

Η Μηχανική Μετάφραση σήμερα

- Εφαρμογές προσωπικών υπολογιστών (SYSTRAN)
- Μετάφραση στο διαδίκτυο (Babel fish, SYSTRAN, Google Translate, Microsoft Translator)
 - Δωρεάν, χωρίς μεγάλες απαιτήσεις στην ποιότητα
 - Η μετάφραση χρειάζεται διόρθωση
 - Συνεχώς αυξάνονται τα υποστηριζόμενα ζεύγη γλωσσών
 - Από τις ευρύτερα χρησιμοποιούμενες υπηρεσίες της Google
- Μετά από 50 χρόνια έρευνας, το αποτέλεσμα δεν είναι ικανοποιητικό
- Ενίσχυση έρευνας
 - Η.Π.Α.: Χρηματοδότηση για στρατιωτικούς σκοπούς
 - Ευρωπαϊκή Ένωση: Ετήσιο κόστος μεταφράσεων πάνω από \$1 δισεκατομμύριο

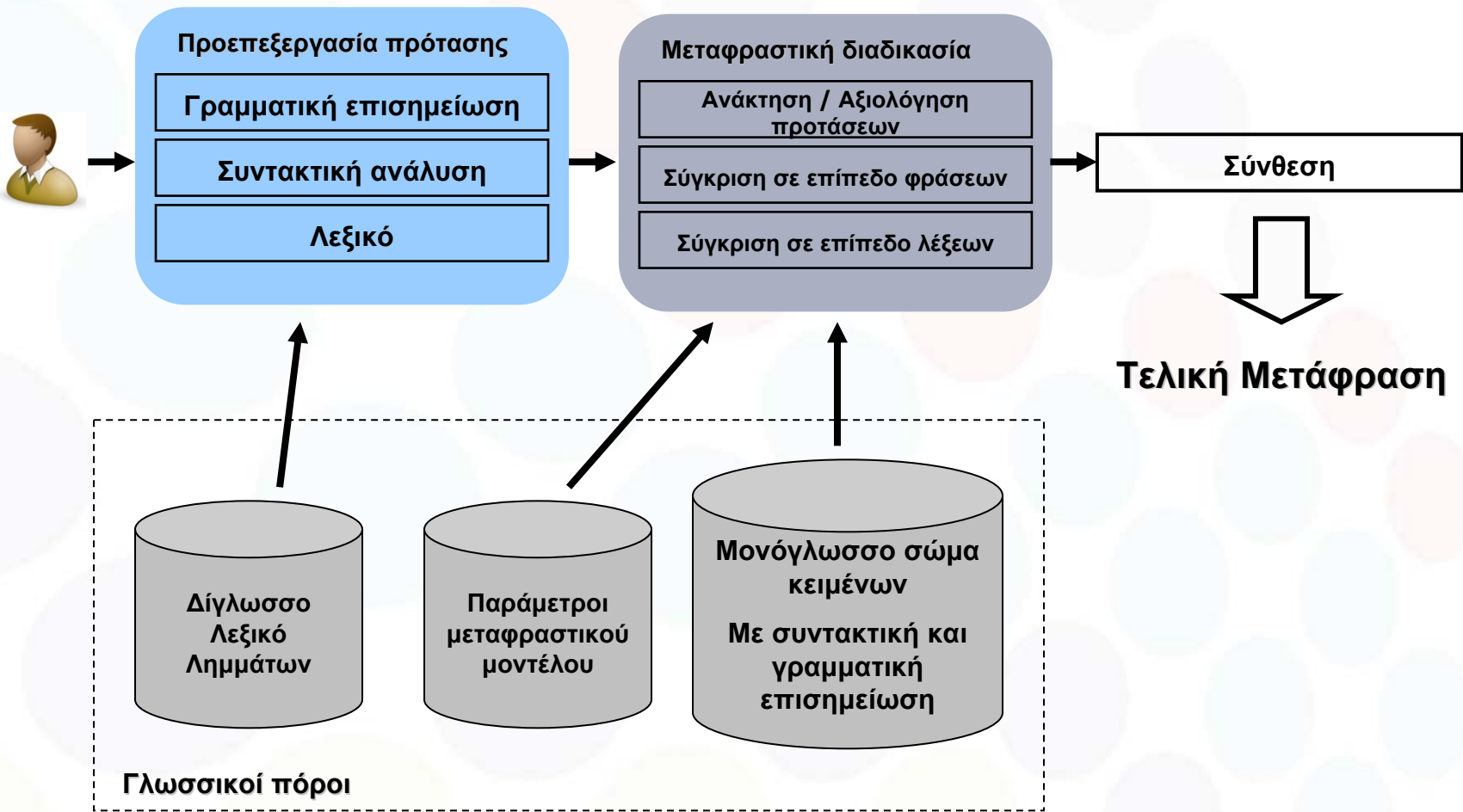


Το μεταφραστικό σύστημα METIS-II

- Αναπτύχθηκε από το Ι.Ε.Λ. 2004-2007 (FP6-IST-003768)
- Βασική ιδέα: Αυτόματη μετάφραση με τη χρήση **μονόγλωσσου** σώματος κειμένων της γλώσσας-στόχου
- **Υβριδικό** σύστημα: συνδυάζει διάφορες τεχνικές, κυρίως τεχνικές αναγνώρισης προτύπων (pattern recognition)
- **Αρθρωτή** (modular) αρχιτεκτονική, που επιτρέπει τον εύκολο χειρισμό διαφόρων γλωσσικών ζευγών
- **Παγιωμένη** γλωσσική τεχνολογία: εργαλεία λημματοποίησης, γραμματικής & συντακτικής επισημείωσης
- **Απουσία** κανόνων αντιστοίχισης
- Μετάφραση με χρήση συντακτικών φράσεων



Αρχιτεκτονική του συστήματος





Αξιολόγηση συστήματος

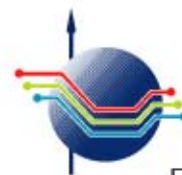
- Αν και πρόκειται για **πρωτότυπο**, το METIS-II έχει σημειώσει πολύ καλές επιδόσεις σε σύγκριση με εμπορικά συστήματα, για το ζεύγος Ελληνικά -> Αγγλικά

Κείμενο αξιολόγησης			
	BLEU	NIST	TER
SYSTRAN	0.3946	7.7041	37.258
METIS-II	0.4590	8.2496	41.058



Το Μέλλον της Μηχανικής Μετάφρασης

- Αδυναμία παραγωγής τέλειων μεταφράσεων
- Υπολογιστικά εργαλεία μετάφρασης: δίγλωσσα λεξικά, συστήματα διαχείρισης ορολογίας, μεταφραστικές μνήμες, μεταφραστικές πλατφόρμες
- Πρόσβαση σε διαγλωσσικές πληροφορίες
- Το διαδίκτυο ως τεράστιος γλωσσικός πόρος
- Υβριδικά συστήματα
- Αξιοποίηση τεχνολογικών εξελίξεων



Ερευνητικό Κέντρο Αθηνά
Athena Research Center

Ερευνητικό Κέντρο Καινοτομίας στις Τεχνολογίες
της Πληροφορίας, των Επικοινωνιών, της Γνώσης

Research and Innovation Center in Information,
Communication and Knowledge Technologies