C SC 620
Advanced Topics in Natural Language Processing

Lecture 17

3/25

Reading List

- *Readings in Machine Translation*, Eds. Nirenburg, S. *et al*. MIT Press 2003.
- Reading list:
  - 12. Correlational Analysis and Mechanical Translation. Ceccato, S.
  - **13. Automatic Translation: Some Theoretical Aspects and the Design of a Translation System. Kulagina, O. and I. Mel'cuk**
  - 16. Automatic Translation and the Concept of Sublanguage. Lehrberger, J.
  - 17. The Proper Place of Men and Machines in Language Translation. Kay, M.

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- 1. The Place of Automatic Translation (AT) Among Problems of Wider Range
- Observation:
  - Too broad: quite naturally broken down into a number of simpler tasks which are to be solved autonomously (first)
  - Too narrow: quite naturally included into broader problems which dominate AT
- Presuppositions:
  - Knowledge of the language pairs
  - Understanding the context
  - Knowing how to accumulate translation experience to gradually raise the quality

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- 1.1 The Linguistic Problem
  - Knowledge of Language means ability to do
    - Analysis: T (text) -> M (meaning), and
    - Synthesis: M -> T
    - Notation for specifying meaning (*Semantic Language*)
      - Example (invariance of meaning under translation):
        - » We fulfilled your task easily
        - » What you had set us as a task was done by us with ease
        - » It was easy for us to fulfill your rask
        - » Fulfilling your task turned out to be easy for us

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- Broadness
  - The three AT tasks are also tasks of general linguistics, moreover cardinal problems of any serious theory of language
    - If linguistics had more or less complete solutions to offer here, only some minor (tech) problems would have to be solved to make practical AT possible (*Failure of linguistics*)
  - Also important for other applications of language information processing
    - e.g. information retrieval, automatic editing and abstracting (summarization), man-machine communication

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- Conclusion 1
  - Any serious progress in AT depends on progress in linguistics on the three tasks
  - Progress in linguistics possible only if linguistics is transformed on the basis of new approaches and conceptions, in close connection with mathematics

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- 1.2 The Gnostical Problem
  - Knowledge of language does not guarantee good translation. Knowledge of situational context also needed.
- A. Different Meanings Correspond to the Same Situation
  - The largest city of the USSR
  - The capital of the USSR

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- B. The Same Meaning Corresponds to Different Situations
  - To this purpose he used the book
  - To do this he made use of the book
  - Situations:
    - Read a book to get information or divert oneself
    - Put a book on a ream of sheets to prevent the wind from scattering them
    - Throw a book at a dog to drive the animal away

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- Knowledge of Situation needed
- (1) Multiple meanings, each of which refers to a certain situation, all of them different
  - Examples:
    - The box is in the pen (Bar-Hillel)
      - Pen: enclosure
      - Pen: writing instrument
    - Slow neutrons and protons (Bar-Hillel)
      - Wide and narrow scope for *slow*

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- Knowledge of Situation needed
- (1) Single meaning, unique situation (a knock at the door), language-particular
  - Example:
    - Come in! (Russian)
    - Forward! (Italian)

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- Conclusion 2
  - Progress in AT dependent on progress in the study of human thinking and cognition

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- 1.3 The Problem of Automating Researchers' Activity
- AT System:
  - Algorithms
    - T->M, M->T, M->S (situation), S->M
  - Data (for each language) - dynamic
    - Lexical
    - Syntactic
    - Stylistic
    - Distribution and functioning of all items in the whole range of possible contexts
    - Rules of correspondence between these items
    - Encyclopedia

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- Start with imperfect system
- Need to organize algorithms and data and have maintenance devices that accept man-made corrections and learn by itself
- Need systems to automatically collect and classify language data

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

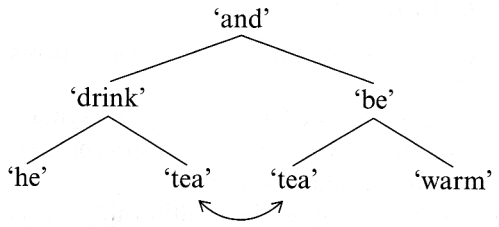
- Conclusion 3
  - Practical solution of AT depends on our ability to automate the scientific activities of humans

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- 2 Principal Components of an AT System
- 2.1 Analysis Algorithm
- 2.1.1 Lexico-morphological Analysis
  - "Morphs"
  - Word form -> Information (distribution and syntactic functions, semantic information)
- 2.1.2 Syntactic Analysis
  - Sentence -> syntactic tree(s)
  - Morphological ambiguities may be resolved here

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- 2.1.3 Semantic Analysis
  - Syntactic tree -> semantic structure (SEMS)
  - Possibly disambiguate syntactic trees here
  - Representation
    - Example: *He drinks warm tea*



Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- Synthesis
  - (Situation level excluded)
  - Replace semantic nodes
    - 1-to-1
    - Several nodes -> 1 node
    - 1 node -> several nodes
      - 'Rush along' -> very/great + fast + move
    - Syntactic node -> single/several semantic nodes
  - Semantic items to syntactic items
    - Success + great degree -> dramatic success
    - Staff -> staff [lab], personnel [hospital], crew [tank or ship], team [football], troupe [theater]

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- 2.2 Semantic Dictionary
- Text -> meaning: simplification
- Basic (English)
  - A few hundred items (plus technical items)
  - Other words must be expressible in Basic by means of non-ambiguous and readily understandable paraphrases
- Merge two Basics into one
  - Semantic Language, AT Interlingua
- Multiple stages: Russian of degree *N*

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

- 2.3 Synthesis Algorithm
  - (Exclude Semantic Synthesis)
  - Syntactic Synthesis
    - By Primitive Word Groups (PWG)
      - Head and dependents
      - Verb, noun, adjective and adverb groups
    - Assemble PWGs into Definitive (Terminal) Word Groups (DWG)
      - Look at master and place PWG
      - Finite verb, subject, object, circumstantial complements, adverb and nominal/infinitive complement groups
    - Arrange DWGs to ensure acceptable word order
      - Preference rules at work

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

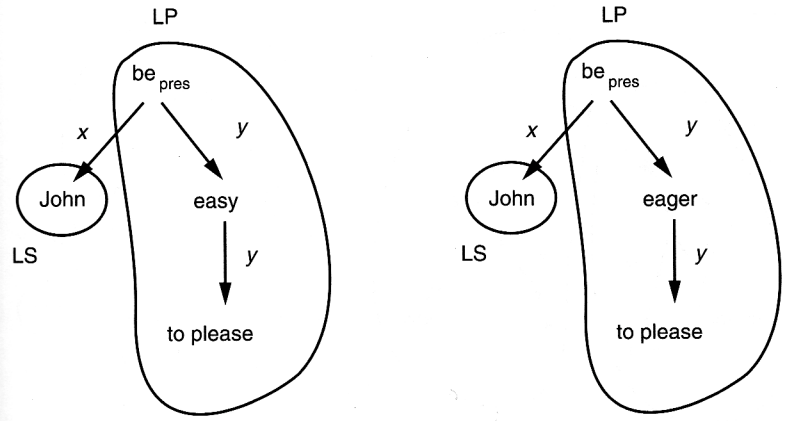


Figure 13.3
1_E) John is easy to please ⇒ (1_E)
2_E) John is eager to please ⇒ (2_E)

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

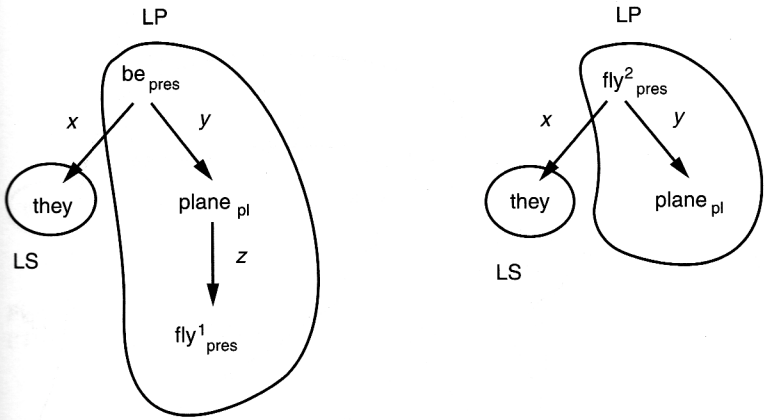


Figure 13.4
3_E) They are flying planes

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

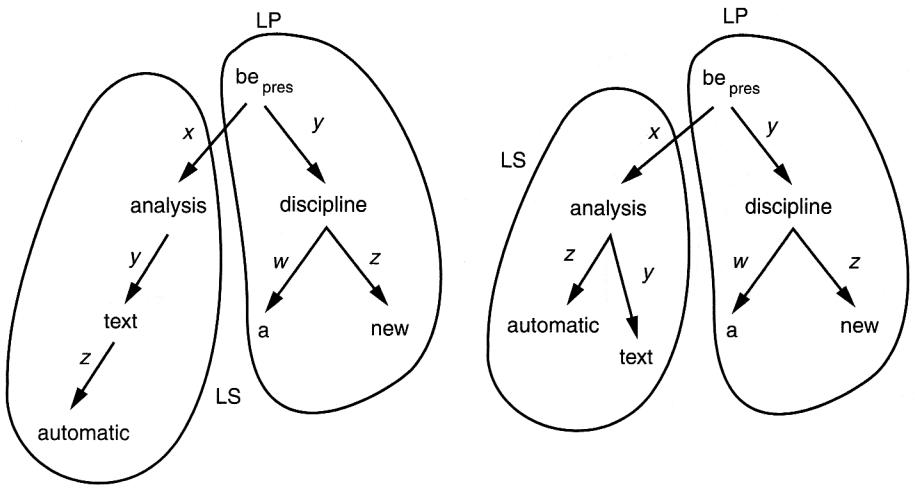


Figure 13.5
Automatic text analysis is a new discipline ($4'_E$) or ($4''_E$)

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

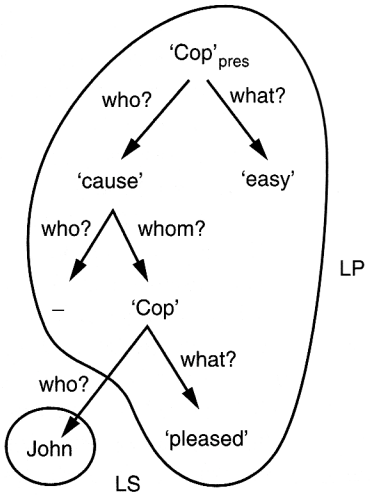


Figure 13.6
3. *Semantic analysis*

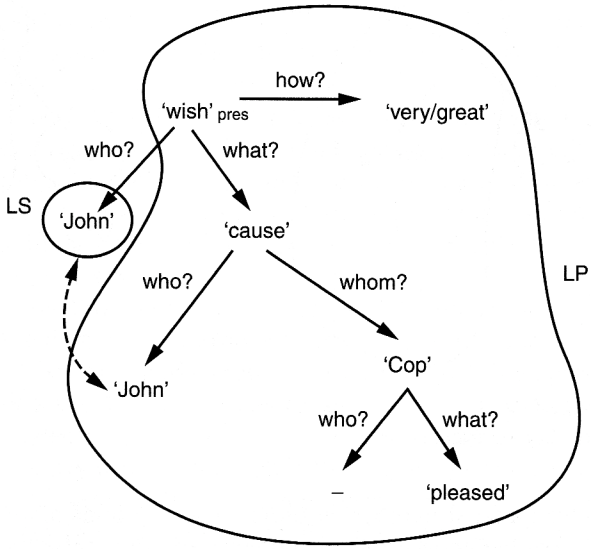


Figure 13.7
('To cause John be pleased is easy')

Figure 13.8
('John wishes very [much] that he (John) causes [someone] to be pleased')

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

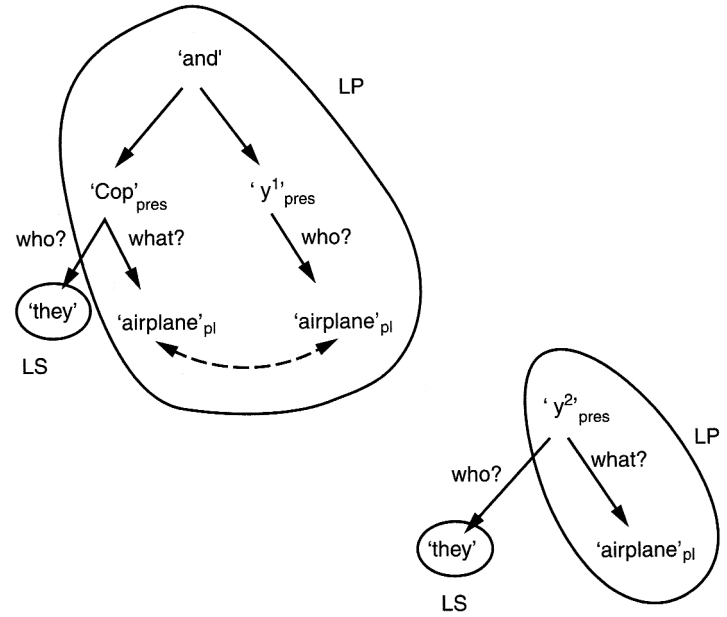


Figure 13.9
('They are airplanes, and [these] airplanes [are] flying')

Paper 13. Automatic Translation. Kulagina, O. and I. Mel'cuk

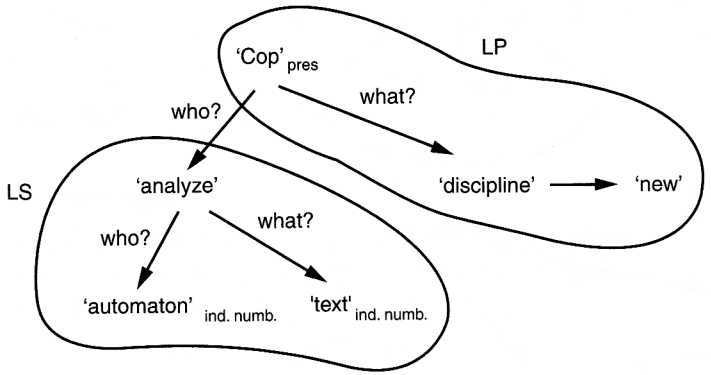


Figure 13.10
(4′): The syntactic tree is dropped by semantic analysis because of the semantic unacceptability of 'automatic text' (only devices, or actions and the like, can be 'automatic').

('Automaton[a] analyze[s] text[s]—is new discipline').