

C SC 620  
Advanced Topics in Natural  
Language Processing

3/11

Lecture 15

# Machine Translation

- *Readings in Machine Translation*, Eds. Nirenburg, S. *et al.* MIT Press 2003.
- Part 1: Historical Perspective
- Reading list:
  - *Introduction*. Nirenburg, S.
  - 1. *Translation*. Weaver, W.
  - 3. *The Mechanical Determination of Meaning*. Reifer, E.
  - 5. *A Framework for Syntactic Translation*. Yngve, V.
  - **6. *The Present Status of Automatic Translation of Languages*. Bar-Hillel, Y.**

# Machine Translation: Next Readings

- *Readings in Machine Translation*, Eds. Nirenburg, S. *et al.* MIT Press 2003.
- Part 1: Historical Perspective
- 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.

# Papers available

- On shelf (improperly) marked for LING 696G in Linguistics (Douglass) opposite front office

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.2 Unreasonableness of Aiming at Fully Automatic High Quality Translation (FAHQT)
  - Misplaced optimism from first years
    - Large number of problems were readily solved
    - Output of machine-simulated translations were often of a form which an intelligent and expert reader could make good sense and use of
  - Not sufficiently realized
    - Gap between such output and high quality translation was still enormous
    - Problems solved were the simplest ones, whereas the “few” remaining problems were the harder ones

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.2 Unreasonableness of Aiming at Fully Automatic High Quality Translation (FAHQT)
  - Most groups seem to have realized that FAHQT will not be attained in the near future
  - Consequence 1: keep trying
    - hope that the pursuit of this aim will yield interesting theoretical insights which will justify this endeavor, whether or not these insights will ever be exploited for some practical purpose
  - Consequence 2: try for something easier with a better chance of attainability in the near future

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.2 Unreasonableness of Aiming at Fully Automatic High Quality Translation (FAHQT)
  - Those who are interested in MT as a primarily practical device must realize that full automation is incompatible with high quality
  - Sacrifice quality, or
  - Reduce self-sufficiency of the machine output
  - Post-editing => computer-aided translation (CAT)

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.3 Commercial Partly Mechanized, High Quality Translation Attainable in the Near Future
  - Cost-benefit tradeoffs
  - Problem 1: Input
    - 0.25 to 0.5c/word typing
    - 1 to 3c/word cost of human Russian-to-English translation



# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.3 Commercial Partly Mechanized, High Quality Translation Attainable in the Near Future
  - Problem 2: A concerted effort will have to be made by a pretty large group in order to prepare the necessary dictionaries
    - Not straightforward
  - Modern Note:
    - (Free or readily available) high-quality lexical resources are still hard to come by even today

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.3 Commercial Partly Mechanized, High Quality Translation Attainable in the Near Future
  - Problem 3: Determining the optimal division of labor between human and machine
    - Easy for human, hard for machine
      - Example: period as end-marker or other purpose
  - Problem 4: Source language forms stored in dictionary as fully-inflected forms or canonical forms

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.4 Compromising in the Wrong Direction
  - Since we cannot have 100% automatic HQ translation, let us be satisfied with a machine output which is complete and unique, i.e. a smooth text of the kind you will get from a human translator but which has less than 100% chance of being correct - “95%”

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.4 Compromising in the Wrong Direction
  - Implementation 1: Print the most frequently target-language counterpart of a source-language word whose ambiguity has not been resolved
    - Requires large scale statistical studies
  - Implementation 2: Work with syntactical and semantical rules of analysis with a degree of validity of no more than 95%, so long as this degree is sufficient to insure uniqueness and smoothness of the translation
    - Esthetically appealing but ...
    - Wrong and dangerous - can reader detect mistranslations?

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.4 Compromising in the Wrong Direction
  - No need to compromise in the direction of reducing the reliability of the machine output
  - Fail-safe output
    - Provide post-editor with all possible help (alternatives to select from)
- Modern Note:
  - No MT system gives a rating of how confident it is in the translation

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.5 A Critique of the Overestimation of Statistics and the “Empirical Approach”
  - Warning against overestimating the impact of statistical information on the problem of MT and related questions
- Modern Note:
  - Statistical MT and other applications have been very popular in the past decade or so ...
    - Large corpora available on-line
    - Cheap CPU power
    - Perceived failure of symbolic approaches

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.5 A Critique of the Overestimation of Statistics and the “Empirical Approach”
  - “I believe this overestimation is a remnant of the time, seven or eight years ago, when many people thought that the statistical theory of communication would solve many, if not all, of the problems of communication”
  - Much valuable time spent on gathering statistics
  - Not every statistic on linguistic matters is automatically of importance for MT

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 1.5 A Critique of the Overestimation of Statistics and the “Empirical Approach”
  - Adherents of the “Empirical Approach”
    - Distrustful of existing grammar books and dictionaries
      - Most existing grammar books are normative
      - Translation dictionaries out-of-date
    - Regard it as necessary to establish from scratch grammatical rules
      - Not justified it’s any faster than modifying existing sources
    - Through human analysis of a large enough corpus of source-language material, constantly improving upon the formulation of these rules by constantly enlarging this corpus



# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 2. Critical Survey of the Achievements of the Particular MT Research Groups
  - 2.1.1 The Seattle Group (U of Washington, E. Reifler)
    - Low quality of output
    - Word-by-word translation plus
      - Word-order
      - Reducing syntactical and lexical ambiguities
    - Unbelievably optimistic claims
      - Compounding: “found moreover that only three matching procedures and four matching steps are necessary to deal effectively with any of these ten types of compounds of any language in which they occur”
      - “it will not be very long before the remaining linguistic problems in machine translation will be solved for a number of important languages”

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- Other tidbits
  - Interlingua: artificial mediating language
    - $n$  languages,  $2n$  programs - reduction from  $n(n-1)$
  - Interlingua: a real language
    - $n$  languages,  $2(n-1)$  programs
  - Artificial interlingua
    - Logical, unambiguous
    - Assumption that translation from a natural language into a logical one is somehow simpler than translation from one natural language into another is unwarranted

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- Other tidbits
  - Idea of a completely symmetrical  $n$ -ary dictionary, each entry consisting of exactly  $n$  words, one for each of the  $n$  languages concerned, is wholly unrealistic
    - Interlingual thesaurus
  - Assume  $L_1 \rightarrow L_2, L_2 \rightarrow L_3$ , how much better would  $L_1 \rightarrow L_3$  be compared to  $L_1 \rightarrow L_2 \rightarrow L_3$  and would it be cost-effective?

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 3. Conclusion
  - FAHQT not a reasonable goal, not even for scientific texts
  - Human translator is often obliged to make intelligent use of extra-linguistic knowledge which sometimes has to be of considerable breadth and depth.
  - Without this knowledge he would often be in no position to resolve semantic ambiguities
  - At present no way of constructing machines with such a knowledge is known, nor of writing programs which will ensure intelligent use of this knowledge
  - Modern Note: still true today ...

# Paper 6: The Present Status of Automatic Translation of Languages. Y. Bar-Hillel

- 3. Conclusion
  - For the preparation of practical MT programs, great linguistic sophistication seems to be neither requisite nor even especially helpful at the present state of the art
  - Basic linguistic research is of great important as such, and its support should preferably not be based on the pretense that it will lead to an improvement of MT techniques
  - It is likely that far-reaching illumination of the human factor in translation will not be achieved without an enormous amount of such basic research, but this is a very long-range affair that should preferably be kept separate from immediate goals