

C SC 620
Advanced Topics in Natural
Language Processing

Lecture 13

3/4

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.

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- MT Linguistics
 - MT linguist (*vs. traditional linguist*)
 - Mostly concerned with differences in behavior between a given pair of languages
 - Need not adhere strictly to the results of scientific language research.
 - When they serve his purpose, he will consider them
 - He will ignore them when an arbitrary treatment of the language material better suits his purpose

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- MT Linguistics
 - MT linguist (*vs. traditional linguist*)
 - Practicality is a consideration of the highest order
 - First concern is source-target semantic agreement and intelligibility
 - Semantics: a poor relation of linguistics, re-directed to psychologists and philosophers

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- The Problem of Editing
 - Pre-editor
 - Works with the input language
 - Determines the intended nongrammatical meaning
 - Annotates input, resolving ambiguity, specifying which lexeme to pick
 - Post-editor
 - Works with the output language (only)
 - Selects the preferred translation based on output context

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- No Editor
 - Fully automatic
 - Or a pre-editor who “instructs the operator of the machine to press a special key, with the result that a mechanical memory selects only output equivalents characteristic of that branch of knowledge”

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- Compound Forms

- The **mechanical dissection of complexes** and their identification via the identification of their constituents means that practically no complex form, all of whose constituents are prolific and/or productive, needs to be coded into the mechanical memory. Only the prolific and productive constituents need be coded. The increase in the number of mechanical operations which such an arrangement implies will be amply compensated for by a reduction in the size of the memory
- Examples:
 - *sea-* in *seaside, seaboard, seaway*
 - *-s* in *seas, boards, ways*

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- Compound Forms
 - Three difficulties in extending this analysis
 - Meaning of a compound often cannot be inferred from its components
 - X-factor, letter or letter sequence could be part of the preceding as well as the following constituent
 - Example (Russian):
 - » Ryblollovu to a fisherman
 - » *Rybolovu to the tin of fishes
 - Extemporized, i.e. unpredictable, compounds
 - Examples:
 - » Holdability
 - » (German) Mitlgift with/poison dowry

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- The Mechanical Determination of Grammatical Meaning
 - Steps:
 - Meaning of each source form in isolation
 - Determination of semantic coincidences exhibited by syntactically correlated co-occurrences in the input text
 - Example (German) of grammatical meaning:
 - *den* (acc masc sg/dat pl) *Männern* (dat pl)
 - Example (German) of nongrammatical meaning:
 - Er bestand die Prüfung/he passed the exam
 - » bestand -> passed

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- The Mechanical Determination of Grammatical Meaning
 - Substantives that can also occur as proper names
 - Can only be resolved by pre-editor
 - Examples:
 - Bauer -> farmer
 - Gerber -> tanner
 - The “Pinpointing” of Composite Intended Meanings
 - Mongenetic vs. polygenetic meaning
 - Pinpointer and pinpointtee

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- Two Groups of Form Classes
 - Form Classes with a Very Large Membership
 - Substantives
 - Attributive adjectives
 - Principal verbs
 - Invariable attributive adjectives derived from substantives by suffix -er
 - Predicative adjectives
 - Adverbs of adjectival origin
 - Cardinal numbers

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- Two Groups of Form Classes
 - Form Classes with a Comparatively Very Small Membership
 - Determiners
 - Pro-substantives
 - Prepositions
 - Verbs that take predicate complements: auxiliaries etc.
 - Separated verb prefixes
 - Adverbs
 - Conjunctions
 - Interjections
 - Total membership: < 2000

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- Memory Systems
 - Large-Drum System
 - 4 units
 - Capital memory for substantives
 - Attribute adjective memory
 - Principal verb memory
 - Predicate adjective memory
 - Small-Drum System
 - Individual memory for each operational form class (10-15)
 - Memory sections
 - Memory equivalents of all low-frequency forms may be grouped according to the number of their component alphabetic and/or non-alphabetic minimal symbols
 - I.e. use N-symbol sections

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- Operational Form-Class Filter System
 - Steps:
 1. All free initial capital forms directed to capital memory
 2. Input of the initial letter of all other free forms activates the small-drum system
 3. All source forms which are members of small operational form classes are identified in processed in the small-drum system
 4. The moment a signal has been fed in which occurs in a sequence position not existing in the small-drum system, the latter is disconnected and the large-drum system is connected
 5. Forms thus rejected by the small-drum system are first directed to the capital memory

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- Operational Form-Class Filter System
 - Steps:
 6. All forms identified in the capital memory are processed there. Free source forms rejected by the capital memory are, in a fixed sequence, redirected to the other memories
 7. They are first directed to the attributive adjective memory
 8. Of forms not identified in 7, the pronominal forms are redirected to the small-drum system
 9. All other free forms rejected are directed to the principal verb memory
 - V + separable prefix processed by co-occurrence
 10. All forms rejected in 9 are redirected to the memory for predicate adjectives and adverbs of adjectival and numeral origin
 11. All source forms not identified so far are forwarded to the output side in their original symbols

Paper 3: The Mechanical Determination of Meaning. E. Reifler

- Conclusion
 - More details needed for pinpointers and pinpointees
 - But the operational form-class filtering system described here, together with the mechanical determination of the constituents of substantive compounds, amply demonstrate the feasibility of a mechanization of the work of a human pre-editor whose intervention had previously been held to be necessary. Nor does it appear from present indication that a human post-editor will be necessary