

C SC 620
Advanced Topics in Natural
Language Processing

Sandiway Fong

Student Presentations Today

- **Ben Wing:** *Cross-Linguistic Discovery of Semantic Regularity*. Peters, W. et al.

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 1: Translation, W. Weaver

- A War Anecdote - Language Invariants
 - “I was amazed to discover that, for (apparently) widely varying languages, the basic logical structures have important common features” - [citing Reichenbach]
 - Modern comment:
 - In linguistic theory, we have the Universal Grammar (UG) hypothesis. The idea there is a dedicated language faculty of the human brain. Core language principles and constraints are the same, languages exhibit parametric differences with respect to word order, sentence constructions, etc.

Paper 1: Translation, W. Weaver

- Translation and Computers
 - Weaver:
 - Semantic difficulties - multiple meanings
 - Domain restrictions - scientific literature (less semantic difficulty).
 - Rough translation - “inelegant (but intelligible)” still worthwhile
 - Wiener:
 - Basic English is the reverse of mechanical and throws upon such words as *get* a burden which is much greater than most words carry
 - Weaver:
 - Multiple meanings on *get* yes but a limited number of two word combinations *get up, get over, get back*
 - 2000 words => 4 million two word combinations - not formidable to a “modern” (1947) computer

Paper 1: Translation, W. Weaver

- **Translation and Computers**

- A translation procedure that does little more than handle a one-to-one correspondence of words cannot hope to be useful for problems of literary translation [23rd Psalm vs. technical material], in which style is important, and in which the problems of idiom, multiple meanings, etc., are frequent.

- **The Future of Computer Translation**

Look at four topics...

- Meaning and Context
 - What is the window size N to determine whether *fast* means *rapid* or *motionless*?
 - Statistical semantic character of language
 - Microcontext to settle difficult cases of ambiguity

Paper 1: Translation, W. Weaver

- **The Future of Computer Translation**
 - Language and Logic
 - Theorem [McCulloch & Pitts, 1943]: A [computer] constructed with regenerative loops of a certain formal character is capable of deducing any legitimate conclusion from a finite set of premises
 - Implies
 - “... as written language is an expression of logical character, this theorem assures one that the problem is at least formally solvable”
 - Translation and Cryptography
 - Citing Shannon’s work, he asks: “If we have useful methods for solving almost any cryptographic problem, may it not be that with proper interpretation we already have useful methods for translation?”

Paper 1: Translation, W. Weaver

- **The Future of Computer Translation**
 - Language and Invariants
 - Most promising approach ...
 - “Thus may it be true that the way to translate from Chinese to Arabic, or from Russian to Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication - the real but as yet undiscovered universal language - and then re-emerge by whatever particular route is convenient.”
 - Modern comment:
 - Interlingua or proto-language?