Topics in Computational Linguistics

Week 3: Linguistic Essentials and Mathematical Foundations

Shu-Kai Hsieh

Lab of Ontologies, Language Processing and e-Humanities
GIL, National Taiwan University

March 14, 2014

Language, Cognition, and Computation

- **logical construct and reasoning**
- **probabilistic phenomena**
- **contextual usages**

What Math that Linguists Should Know?

Partee [2]'s foci in 1990.

Example

Set theory (relations and functions), Logic and formal systems (axiomatization and model theory), Algebra (Groups, Lattices, Boolean and Heyting Algebras), Automata.

Manning and Schütze [1]'s foci in 1999.

Example

Calculus, **Vector and Matrices**, **Statistics and Probability**, Information Theory.

Elementary Probability Theory

- Probability spaces
- Conditional probability and independence
- Bayes' theorem
- Random variables
- Expectation
- Distribution
- Bayesian statistics

Lab[1]: Zipf's Law

Elementary Information Theory

- The notion of Entropy
- Joint entropy and conditional entropy
- Mutual information
- Relative entropy (or Kullback-Leibler divergence)
- Cross entropy/perplexity: the relation to language
- The entropy of English

What Linguistics that non-linguists Should Know?

- Generativism is (just) ONE important page in linguistic history.
- Ambiguity is the key to linguistic complexity.
- Language Analytics from different perspectives (formal, cognitive-functional, computational), at different levels (phonetic-phonoligical, morpho-syntactic, semantic-pragmatic, historical-typological, socio-cultural, psycho-neural,...), by different theories (you don't want me to list them here...)
- Language Resources (dictionary, lexicon and corpus).

Charles Fillmore, Founder of FrameNet, Dead at 84

February 18, 2014

We are sad to report that our esteemed colleague at ICSI, Charles J. Fillmore, died February 13 after a long struggle with cancer. He passed away peacefully at home, with his wife Lily Wong Fillmore at his side. A professor of linguistics at UC Berkeley for many years and head of the FrameNet Project at ICSI from 1997 until his death, he was one of the great figures in twentieth century linguistics, best known for his foundational work on case grammar, frame semantics, and construction grammar. We will not attempt to detail his enormous contributions to linguistics, but just say that everyone who was privileged to study or work with him will feel the loss of a gentle friend and wise mentor.

The UC Berkeley Linguistics Department has set up a Web page in his memory. Visit it at http://linguistics.berkeley.edu/charles-j-fillmore-1929-2014.

Read George Lakoff's tribute to Fillmore.

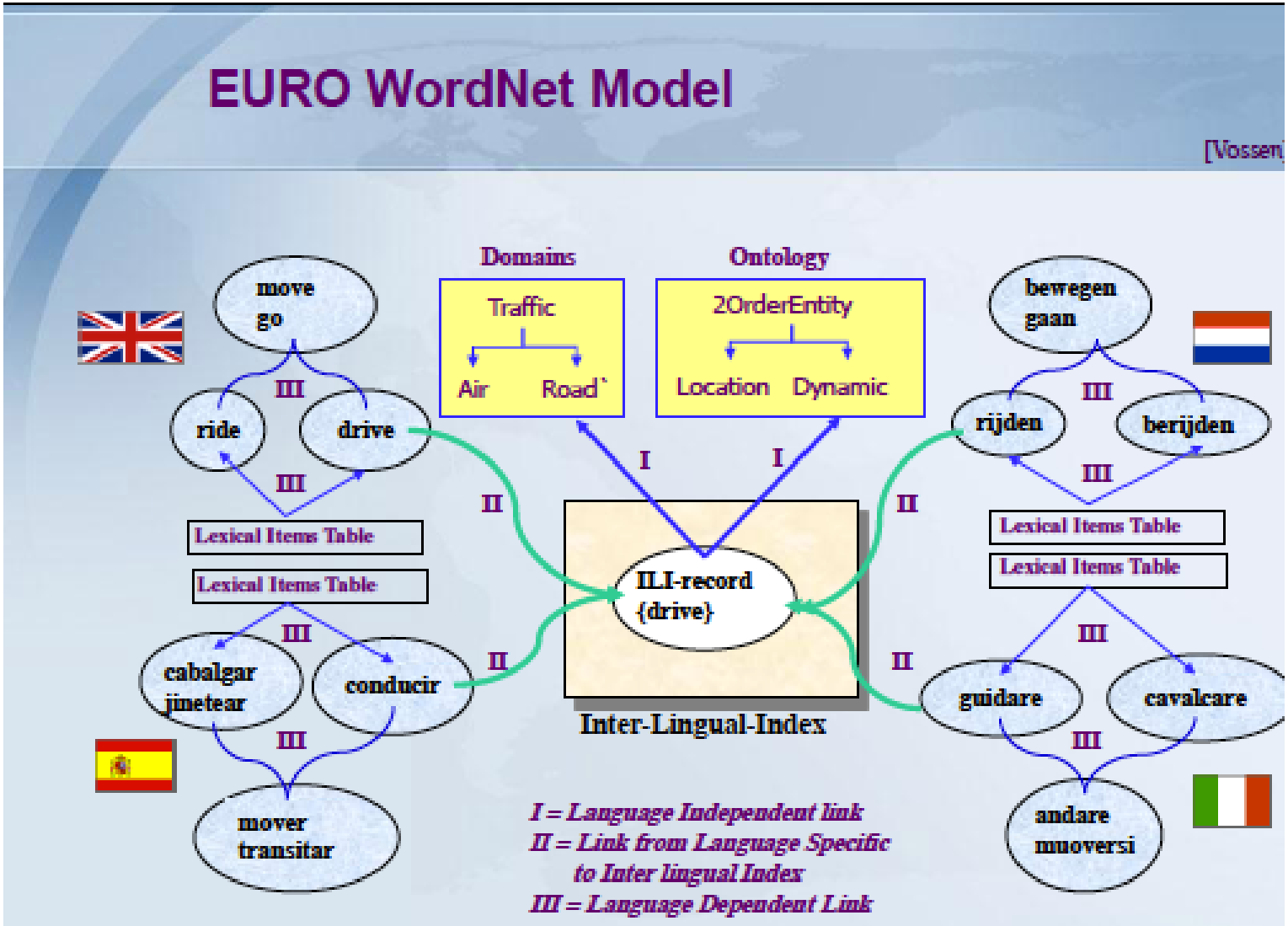
source:

`http://www.icsi.berkeley.edu/icsi/news/2014/02/charles-fillmore-dies-at-84`

Unrevealed Story of Words [1]

- Intricate interaction between writing and linguistic system.
  (For some reason *hanzi* is greatly ingrained in Chinese mind).
- Either ask-the-speaker or ask-the-linguist doesn't work.

Collaboration is IN



Homework (source: Dice book p78)

80% Take two short peices of aligned texts (in Chinese and English, respectively) and compute the relative frequencies of the *letters* and *Chinese characters* in the texts. Assume these are the true probabilities. What is the **entropy** of the distributions for each text?

BONUS Take another piece of English text and compute a second probability distribution over letters by the same method. What is the KL divergence between the two distributions? (You will need to 'smooth' the second distribution and replace any zero with a small quantity ϵ)

20% 預習 chapter 4 (Jurafsky and Martin)

 Christopher D Manning and Hinrich Schütze.

Foundations of statistical natural language processing.

MIT press, 1999.

 Barbara Partee, Alice Ter Meulen, and Robert Wall.

Mathematical methods in linguistics, volume 30.

Springer, 1990.