

The Future of Semantics: arm-chair theorizing or data fetishism?

panellists:

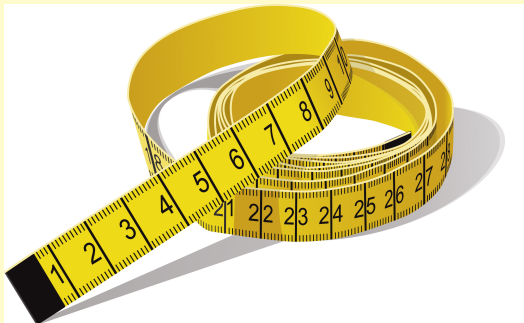
Martin Stokhof, Angelika Kratzer,
Matthew Stone, Noah Goodman

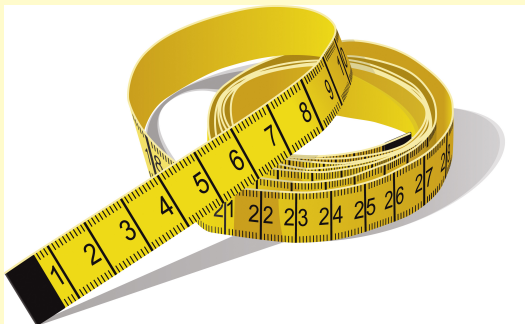
chair:

Jelle Zuidema

Amsterdam Colloquium & SMART Cognitive Science







$$\forall \wedge \implies \forall \exists \models \diamond \neg$$

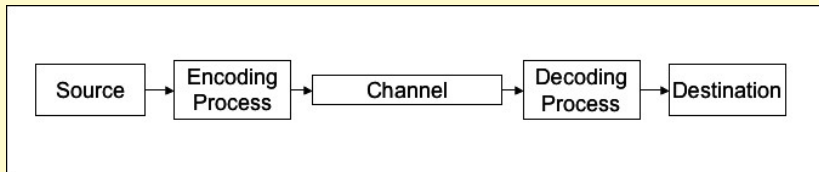
$$\int_a^b p \sim \mathcal{N} \frac{\delta X}{\delta t} \Sigma \Pi \approx P(H|D) = \frac{P(H)P(D|H)}{P(D)}$$

Claude Shannon, 1916-2001

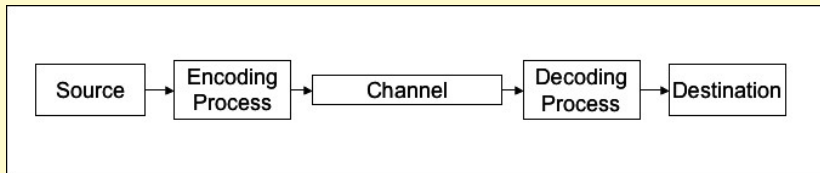


(Shannon with his selflearning mouse Theseus; he also built rocket-powered flying disks, the Ultimate Machine, ...)

Shannon's noisy channel model (1948)



Shannon's noisy channel model (1948)



$$C = \lim_{T \rightarrow \infty} \frac{\log_2 N(T)}{T} \quad (1)$$

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (2)$$

Shannon's information theory

(Weaver, 1949)

- ▶ Three levels of analysis:

Shannon's information theory

(Weaver, 1949)

- ▶ Three levels of analysis:
 - ▶ Technical level

Shannon's information theory

(Weaver, 1949)

- ▶ Three levels of analysis:
 - ▶ Technical level
 - ▶ Semantic level

Shannon's information theory

(Weaver, 1949)

- ▶ Three levels of analysis:
 - ▶ Technical level
 - ▶ Semantic level
 - ▶ Effectiveness level

Shannon's information theory

(Weaver, 1949)

- ▶ Three levels of analysis:
 - ▶ Technical level
 - ▶ Semantic level
 - ▶ Effectiveness level
- ▶ Shannon deals with the first only.

Shannon's influence

- ▶ Shannon's 1948 paper has been extraordinary influential in many fields.
- ▶ In language modelling, his application of ngram-models was instantly popular. Still widely used as language models (in the narrowest sense): to assign probabilities to sequences.
- ▶ Provoked Noam Chomsky to demonstrate the inadequacy of Markov models for describing syntactic structure.
- ▶ Established the need for probabilistic models of language (although Chomsky c.s. made them unpopular in linguistics for a while).

Shannon (1956): “Information theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press...

Shannon (1956): “Information theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press... Our fellow scientists in many different fields, attracted by the fanfare and by the new avenues opened to scientific analysis, are using these ideas in their own problems. Applications are being made to biology, psychology, linguistics, fundamental physics, economics, the theory of organization, and many others...

Shannon (1956): “Information theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press... Our fellow scientists in many different fields, attracted by the fanfare and by the new avenues opened to scientific analysis, are using these ideas in their own problems. Applications are being made to biology, psychology, linguistics, fundamental physics, economics, the theory of organization, and many others... Although this wave of popularity is certainly pleasant and exciting for those of us working in the field, it carries at the same time an element of danger...

Shannon (1956): “Information theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press... Our fellow scientists in many different fields, attracted by the fanfare and by the new avenues opened to scientific analysis, are using these ideas in their own problems. Applications are being made to biology, psychology, linguistics, fundamental physics, economics, the theory of organization, and many others... Although this wave of popularity is certainly pleasant and exciting for those of us working in the field, it carries at the same time an element of danger... It will be all too easy for our somewhat artificial prosperity to collapse overnight when it is realized that the use of a few exciting words like information, entropy, redundancy, do not solve all our problems.”

- ▶ Pierce and Carroll's ALPAC report (1956) persuaded the U.S. government to stop funding (statistical) machine translation (MT) research.

- ▶ Pierce and Carroll's ALPAC report (1956) persuaded the U.S. government to stop funding (statistical) machine translation (MT) research.
- ▶ “Certainly, language is second to no phenomenon in importance. And the tools of computational linguistics are considerably less costly than the multibillion-volt accelerators of particle physics. The new linguistics presents an attractive as well as an extremely important challenge.”

- ▶ Pierce and Carroll's ALPAC report (1956) persuaded the U.S. government to stop funding (statistical) machine translation (MT) research.
- ▶ “Certainly, language is second to no phenomenon in importance. And the tools of computational linguistics are considerably less costly than the multibillion-volt accelerators of particle physics. The new linguistics presents an attractive as well as an extremely important challenge.”
- ▶ Machine translation is field for “mad inventors or untrustworthy engineers”

- ▶ Pierce and Carroll's ALPAC report (1956) persuaded the U.S. government to stop funding (statistical) machine translation (MT) research.
- ▶ “Certainly, language is second to no phenomenon in importance. And the tools of computational linguistics are considerably less costly than the multibillion-volt accelerators of particle physics. The new linguistics presents an attractive as well as an extremely important challenge.”
- ▶ Machine translation is field for “mad inventors or untrustworthy engineers”
- ▶ Pierce (1959, JASA): “We are safe in asserting that speech recognition is attractive to money. The attraction is perhaps similar to the attraction of schemes for turning water into gasoline, extracting gold from the sea, curing cancer, or going to the moon. One doesn't attract thoughtlessly given dollars by means of schemes for cutting the cost of soap by 10%. To sell suckers, one uses deceit and offers glamour.”

Resurgence of statistical and corpus-based methods

Early 1980s Speech Recognition

Resurgence of statistical and corpus-based methods

Early 1980s Speech Recognition

Late 1980s Machine Translation

Resurgence of statistical and corpus-based methods

Early 1980s Speech Recognition

Late 1980s Machine Translation

Early 1990s Lexical Semantics

Resurgence of statistical and corpus-based methods

Early 1980s Speech Recognition

Late 1980s Machine Translation

Early 1990s Lexical Semantics

Late 1990s Constituency Structure

Resurgence of statistical and corpus-based methods

- Early 1980s Speech Recognition
- Late 1980s Machine Translation
- Early 1990s Lexical Semantics
- Late 1990s Constituency Structure
- Early 2000s Dependency Structure

Resurgence of statistical and corpus-based methods

Early 1980s Speech Recognition

Late 1980s Machine Translation

Early 1990s Lexical Semantics

Late 1990s Constituency Structure

Early 2000s Dependency Structure

Late 2000s Topic Models, Semantic Role Labelling

Resurgence of statistical and corpus-based methods

Early 1980s Speech Recognition

Late 1980s Machine Translation

Early 1990s Lexical Semantics

Late 1990s Constituency Structure

Early 2000s Dependency Structure

Late 2000s Topic Models, Semantic Role Labelling

Early 2010s Compositional Semantics?

So?

So?



So?



Armchair Theorizing or Data Fetishism?

- ▶ Is there anything to worry about in logical semantics?
- ▶ Should we all convert to designing probabilistic models based on corpus-data?

Armchair Theorizing or Data Fetishism?

- ▶ Is there anything to worry about in logical semantics?
- ▶ Should we all convert to designing probabilistic models based on corpus-data?

Panel:

- ▶ Martin Stokhof (ILLC, Universit of Amsterdam)
- ▶ Angelika Kratzer (University of Massachusetts at Amherst)
- ▶ Noah Goodman (Stanford University)
- ▶ Matthew Stone (Rutgers University)