Language Use and Uncertainty

Khalil Sima'an

Institute for Logic, Language and Computation (ILLC)
University of Amsterdam, The Netherlands

Amsterdam – 22 November 2005

The dream of Artificial Intelligence

- Systems that extract information from textual or spoken media.
  Examples: inf retrieval, info extraction, data mining etc.
- Systems that transform text/speech to text/speech
  Examples: translation systems, summarization, dictation, reading etc.
- Systems that communicate with people through language.
  Examples: dialogue systems

"Language Understanding" is Crucial for Communication

"Language Understanding"? What plays a role?

Example system

A computer provides information about train schedules:

C: Good evening. How can I help you?

U: I want to travel to Utrecht. Eh... from Amsterdam tomorrow evening.

C: What time do you want to arrive in Utrecht?

U: I want to depart at around half eight.

C: There is a train at seven thirty six from Amsterdam CS, arriving at seven fifty six in Utrecht CS. Is that suitable for you?

⋮

What problems do we expect to face?

It's All About Expectations

A speech-driven routing system built for banking services (ABN-AMRO) over the telephone.

One session went as follows:

System Q: `Would you like information about mortgages, loans or investments?`

User A: `Yes.`

It's All About Expectations

A speech-driven routing system built for banking services (ABN-AMRO) over the telephone.

One session went as follows:

System Q: `Would you like information about mortgages, loans or investments?`

User A: `Yes.`

Premature Optimism

Reactions:

- Reaction of hardcore "computer scientists":
  "I thought this problem has been solved!"
  "Why don't you build a compiler for this?"
- McDonnel-Douglas ad in 1985 (Ref: S. Shieber, L. Lee):

  ``At last, a computer that understands you like your mother"

  Understands you like your mother?

Premature Optimism

Reactions:

- Reaction of hardcore "computer scientists":
  "I thought this problem has been solved!"
  "Why don't you build a compiler for this?"
- McDonnel-Douglas ad in 1985 (Ref: S. Shieber, L. Lee):

  ``At last, a computer that understands you like your mother"

Understands you like your mother?

Premature Optimism

Reactions:

- Reaction of hardcore "computer scientists":
  "I thought this problem has been solved!"
  "Why don't you build a compiler for this?"
- McDonnel-Douglas ad in 1985 (Ref: S. Shieber, L. Lee):

  ``At last, a computer that understands you like your mother"

  Understands you like your mother?

Human Perception: Ambiguity (1)

"A computer that understands you like your mother."

Written: A computer that understands

... you as well as your mother does.

... (that) you like your mother.

... you as well as it understands your mother.

Spoken: A computer that understands ...

... you *like your* mother.

... your *lie cured* mother.

Other examples of ambiguity

Word-sense: different meanings:

`west` *bank* `of the river` vs. `my savings in the` *bank*

`She ran up a big bill` vs. `She ran up a big hill.`

Part-of-speech: different categories:

`following` as `verb`, `adj`, or `noun`.

Sentence structure: `I saw the man with the telescope.`

`The telegraphy and telephony services are important.`

More Observations: Human Perception

Robustness: people can process "weird" utterances
"what did Sally whisper that she had secretly said" "what I do to me?"

Relative grammaticality: people see different levels of grammaticality:
"Those are the books you should read before talking about becomes difficult."
People disagree on how "grammatical" this utterance is.

Language change: language use changes over time
The word "following" is originally a verb, now as "adjective" and "noun".

More Observations: Human Perception

We are able to guess given a certain context:

Next word: fill in the blank ("puntje-puntje" exercises):

`I want to make a collect ...`

Spelling error: what is the most plausible word?

`I have been teading`

`"correction∈{leading,reading,feeding}`

Speech hazard: what is the most plausible utterance:

`I want to travel to`

`Almelo/Ermelo/marmalade/Elsloo.`

Human Perception: Summary

Human communication seems to hinge on the ability to

Resolve ambiguities

Accommodate weird/irregular/unexpected utterances

Guess/ predict/correct words/phrases in context

Adapt/learn new words, forms and uses of language

How to model language use with such properties?

What plays a role in language processing?

What influences language processing:

Language knowledge (words, structure?, ...)

World knowledge (`dogs bark vs. humans speak, yell, shout; eat-pizza vs. eat-bank??`)

Situation: Time, place and culture (e.g., Belgium/NL, city/village dweller, teenagers/old folks, level of education), etc.

Experience we "bet" based on expectations/experience

⋮

We operate under uncertainty (Input/Output) $\Rightarrow$ Probability

Situated+collective experience $\Rightarrow$ Data/learning/estimation

Will Formal Grammars Do?

Do we merely "compose words/phrases with words/phrases"?

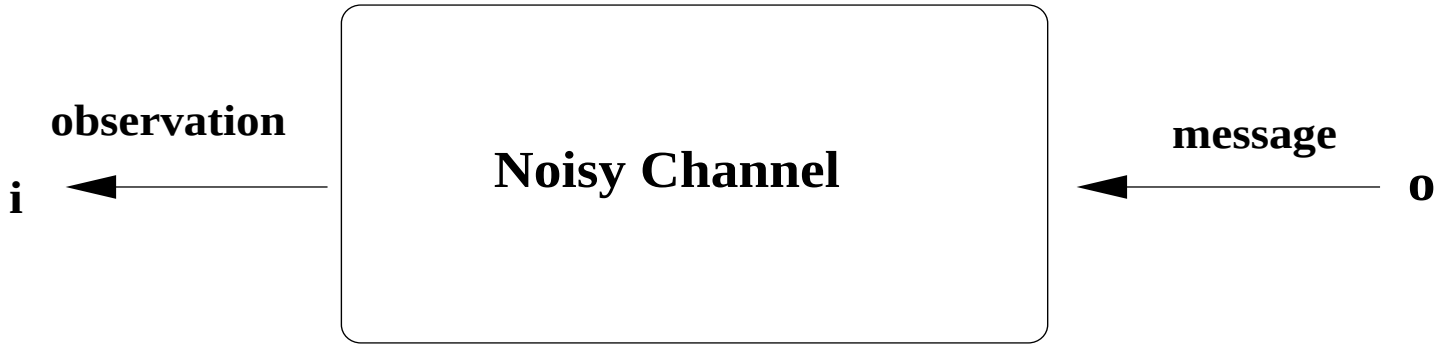
More evidence from psycholinguistics:

Memory effects: we "replay" from earlier experience.

Frequency effects: more frequent is preferred.

Combine formal grammars with learning!

Noisy-Channel Model (C. E. Shannon, 1948)



If we know a set Ω of candidate pairs, we may select for any given input i:

$$\arg \max_{\{o:\langle i,o\rangle \in \Omega\}} P(o \mid i) = \arg \max P(i|o)P(o)$$

Examples: Speech Recognition, Machine Translation, Syntactic Parsing, Correction

What's in a probability?

$$\arg\max P(i|o)P(o)$$

Define Ω: a language of $\langle i, o \rangle$ pairs
You may use formal, linguistic grammars, logic ...

Language Model: $P(o)$ is a model of language use
Higher probabilities imply more regular!

Task Model: $P(i|o)$ is a model of the task
How does o change into i?
The plausibility of a pair $\langle i, o \rangle$

Combination: $P(i|o)P(o)$
Independent models up to the definition of Ω

Where the probabilities come from?

Data: a sample of $\langle i, o \rangle$ pairs representative of language use that we care to model.

Data: what kind of data?

Learning: estimate the probabilities from Data
machine learning = statistical estimation

Variance/Bias
- what happens if Data changes?
- what happens when Data grows large?
- what happens when Data is too small?

Consistency: convergence in the limit to correct values

Human Perception

Questions:

- Does visual/musical perception work similar to language perception?
- How do we learn language?
- Where does grammar meet probability?
- Where does memory meet analogy?
- Where does plausibility meet similarity?
- Where does statistics meet learning?

Probabilistic models start where grammar and logic stop!