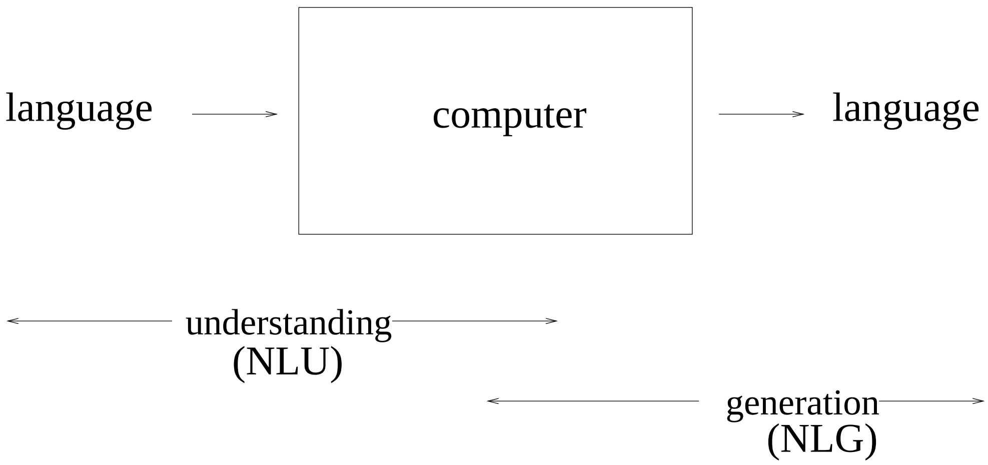
I. Background and Motivation

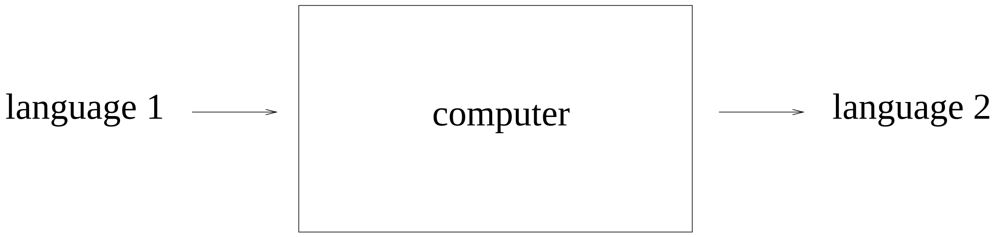
Natural Language Processing (NLP)

Goal: computers using natural language as input/output



Why NLP?

- Lots of information is in natural language format
- Lots of users want to communicate in natural language



Applications include:

- speech recognition
- text summarization
- machine translation
- user interfaces

Why NLP? (cont.)

- Wide variety of problems: linguistic, mathematical, psychological aspects
- Hard: “AI-complete”

Statistical NLP

Goal: Infer language properties from (annotated?) (text?) samples

Draws on probability, statistics, information theory, machine learning

Two threads (often intertwined; not everyone distinguishes!):

- statistical *models* — language assumed generated by a statistical source
- statistical *methods* — no assumption on language source; sample statistics used to make decisions

Statistical Models Example: PCFG’s

Probabilistic Context-Free Grammars (PCFG’s): strings generated by randomly picking rules according to their probabilities

(1.0)	S	→	$NP\ VP$	(.75)	VP	→	rocked
(1.0)	NP	→	the tutorial	(.25)	VP	→	bombed

$$
\begin{aligned}
P(\text{“the tut. rocked”}) &= \underbrace{1}_{S\rightarrow NP\ VP} \times \underbrace{1}_{NP\rightarrow \text{the tut.}} \times \underbrace{.75}_{VP\rightarrow \text{rocked}} \\
&= .75
\end{aligned}
$$

Statistical Methods Example: WSD

Word sense disambiguation *(WSD)*: find correct word sense from context

“They put money in the $\underbrace{\text{bank}}_{\text{savings? river?}}$ ”

A statistical solution [Lesk 86]: estimate the likelihood of $\langle\text{savings bank}\rangle$ co-occurring with “money” from entries in a *machine-readable dictionary*

Why Statistical NLP?

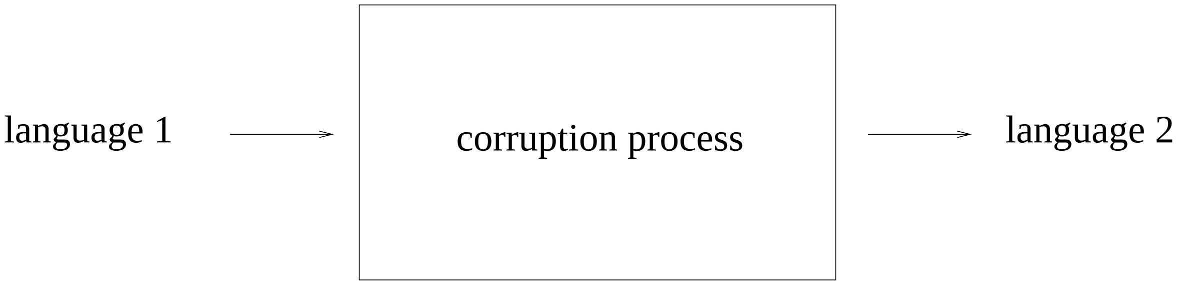
- Statistical models allow degrees of uncertainty (not just “grammatical/ungrammatical”)
  - ▷ confidence can be assessed (helps combine knowledge sources)
  - ▷ models can be iteratively trained/updated
- Statistical methods reduce the *knowledge acquisition bottleneck*
  - ▷ transfer to new domains is easier

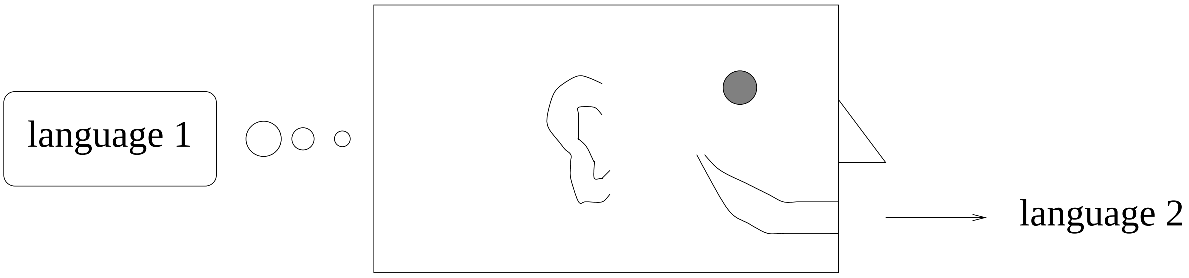
But statistical approaches were (are) not universally accepted ...

A Brief History

The 40's and 50's: statistical NLP popular

- Harris, Firth: empirical linguistics ("You shall know a word by the company it keeps" [Firth 57])
- Shannon, Weaver: cryptographic notions, the *noisy channel model*





A Brief History (cont.)

Late 50's–80's: statistical NLP in disfavor

> "It is fair to assume that neither sentence
>
> (1) *Colorless green ideas sleep furiously*
>
> nor
>
> (2) *Furiously sleep ideas green colorless*
>
> ... has ever occurred .... Hence, in any statistical model ... these sentences will be ruled out on identical grounds as equally "remote" from English. Yet (1), though nonsensical, is grammatical, while (2) is not." [Chomsky 1957]

A Brief History (cont.)

The 80's – present: statistical NLP once again mainstream

- revived by IBM: influenced by speech recognition
- Confluence with interest in machine learning
- nowadays,

  > "no one can profess to be a computational linguist without a passing knowledge of statistical methods .... anyone who cannot at least use the terminology persuasively risks being mistaken for kitchen help at the ACL banquet." [Abney 96]

The “Opposite” of Statistical NLP?

Some draw contrasts with knowledge-based methods, higher-level processes, linguistics...

- Chomsky
- “I don’t believe in this statistics stuff”
- “that’s not learning, that’s statistics”
- “AI-NLP ...is going nowhere fast”
- “Every time I fire a linguist, my performance goes up”

Statistics Complements Other Approaches

- Knowledge-based models can be converted to stochastic versions
  - ▷ CFG's $\longrightarrow$ PCFG's
  - ▷ statistical semantics, discourse models [Miller 96]
- Statistical methods can make use of knowledge bases (don't confuse methods and models)
  - ▷ WSD using dictionaries