”I’M SORRY DAVE, I’M AFRAID I CAN’T DO THAT”

Natural Language Processing on the Eve of 2001

Lillian Lee

(some of this material comes from a joint AAAI tutorial with John Lafferty)

Where Are The Flying Cars?

According to science fiction, the future has talking machines.

Metropolis (1926): “false Maria”

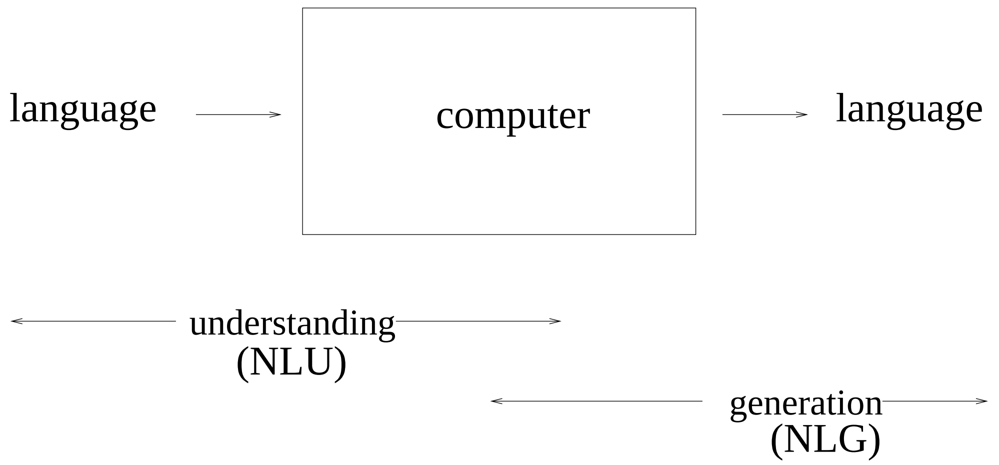
Star Wars: Episode IV: C3PO (Maria’s influence?)

2001: A Space Odyssey (1968): HAL (the HAL-9000)

Dave:	Open the pod bay doors, HAL.
HAL:	I’m sorry Dave, I’m afraid I can’t do that.

Natural Language Processing (NLP)

Goal: computers using natural language as input and/or output



NLU example: convert an utterance into a sequence of computer instructions.

NLG example: produce a summary of a patient's records.

NLP is Useful

Task	Input	Output
summarization	“document(s)” (CNN broadcasts)	summary
machine translation	signal in language 1	signal in language 2
question answering	query	answer to query
$\longrightarrow$ information retrieval	query	relevant documents
user interfaces	command in natural language	computer instructions

“Now we’re betting the company on these natural interface technologies”

– Bill Gates, 1997

NLP is Challenging

It is often said that NLP is "AI-complete":

> All the difficult problems in artificial intelligence manifest themselves in NLP problems.

This idea dates back at least to the Turing Test:

> "The question and answer method seems to be suitable for introducing almost any one of the fields of human endeavour that we wish to include" [Turing, "Computing Machinery and Intelligence", 1950]

Why is NLP hard?

- "Doesn't Microsoft do that already?"
- Ad from the 70's or 80's (source: S. Shieber): the problem has already been solved ...

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

Ambiguity

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

What can we infer about the computer?

1. (*) It understands you as well as your mother understands you

2. It understands (that) you like your mother

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

1 and 3: Does this mean well, or poorly?

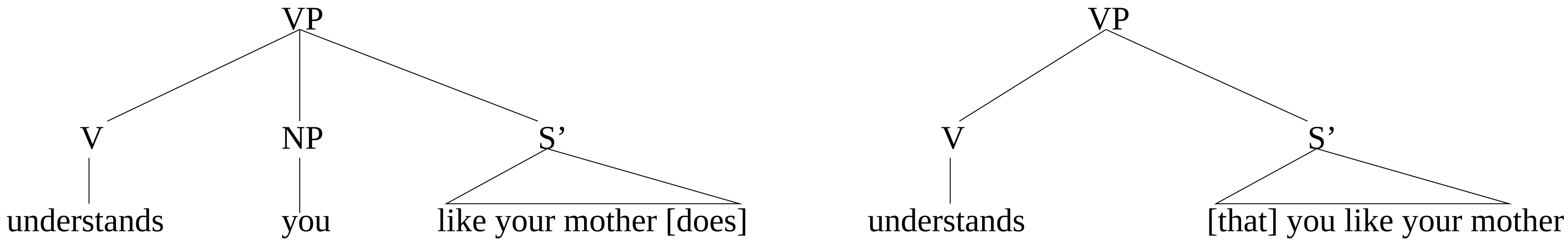
Ambiguity at Many Levels

At the acoustic level (*speech recognition*):

1. "... a computer that understands you like your mother"
2. "... a computer that understands your lie cured mother"

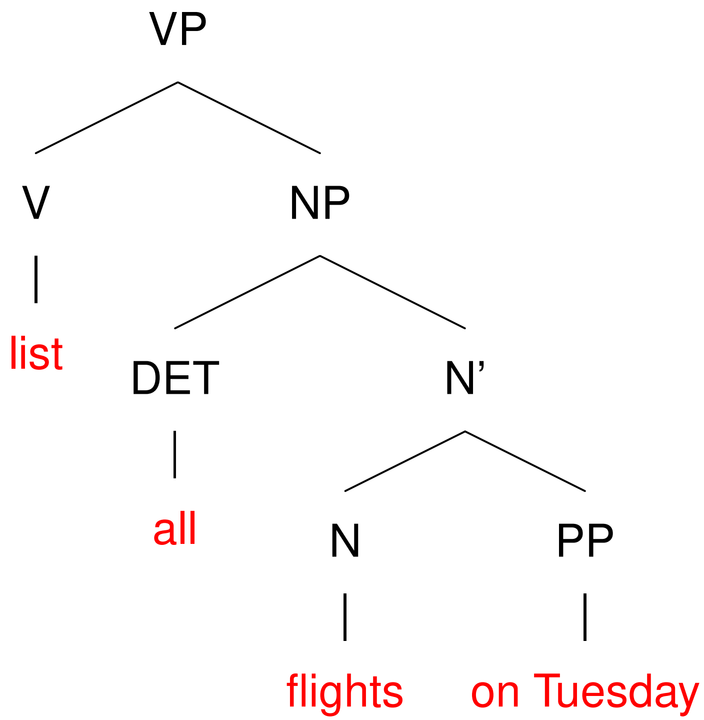
Ambiguity at Many Levels (cont.)

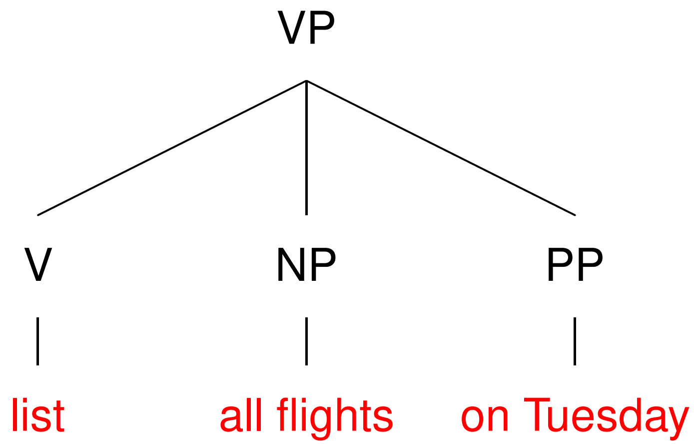
At the syntactic (structural) level:



Different structures lead to different interpretations.

More Syntactic Ambiguity





Ambiguity at Many Levels (cont.)

At the semantic (meaning) level:

mother = ? (OED)

A female parent

A cask or vat used in vinegar-making

This is an instance of word sense ambiguity.

More Word Sense Ambiguity

- They put money in the <u>bank</u>

  = buried in mud?

- “L’avocat general”

  = “the general avocado”?

- (*) “I saw her duck with a telescope”

Ambiguity at Many Levels (cont.)

At the discourse (multiple-clause) level:

1. Alice says they've built a computer that understands you like your mother.
2. But she ...

    2a. ... doesn't know any details.

    2b. ... doesn't understand me at all.

This is an instance of anaphora, where "she" co-refers to some other discourse entity.

I’m Afraid I Can’t Do That

The task seems so difficult! What resources do we need?

1. Knowledge about language
2. Knowledge about the world

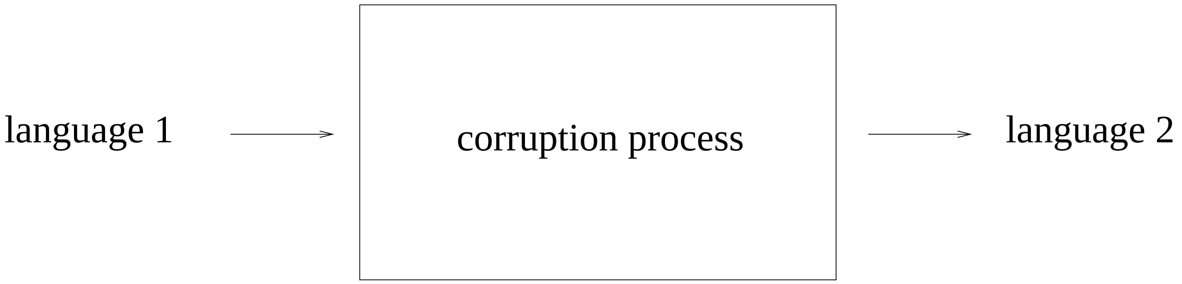
A Crazy Idea

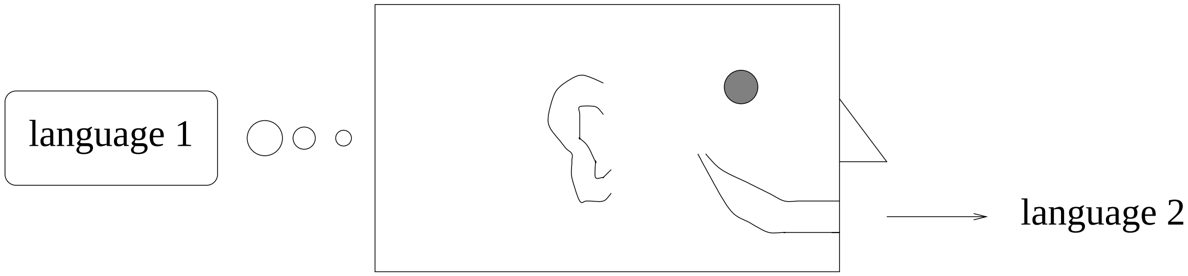
"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages."

C.E. Shannon, A Mathematical Theory of Communication,
The Bell System Technical Journal, July 1948.

Reductio ad Absurdum

The *noisy channel model*:





Actually, a Great Idea

Computers can **learn** such models.

Statistical NLP: Infer language properties from (annotated?) (text?) samples.

- Helps ease the knowledge acquisition bottleneck.

Probabilities are Realistic

"It's hard to recognize speech"

vs.

"It's hard to wreck a nice beach"

Which is more likely? (both are *grammatical*)

Applications: speech recognition, handwriting recognition, spelling correction, ...

General problem in statistical NLP: density estimation

P("I saw her duck [with a telescope]" $\rightarrow$ verb attachment)

P("L'avocat general" $\rightarrow$ "the general avocado")

No, Really, It's a Crazy Idea

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]

Who Are You Calling Crazy?

- “I don’t believe in this statistics stuff”
- “that’s not learning, that’s statistics”

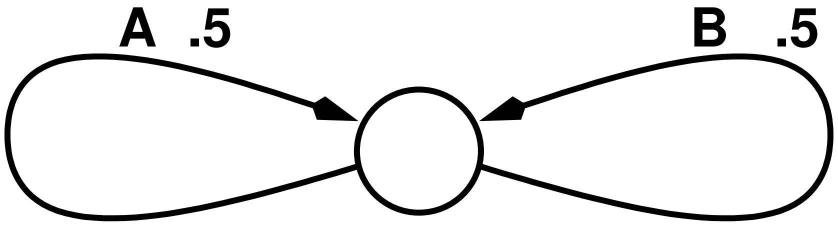
- Knowledge-intensive NLP “is going nowhere fast”
- “Every time I fire a linguist, my performance goes up”

A Computational Model of Language

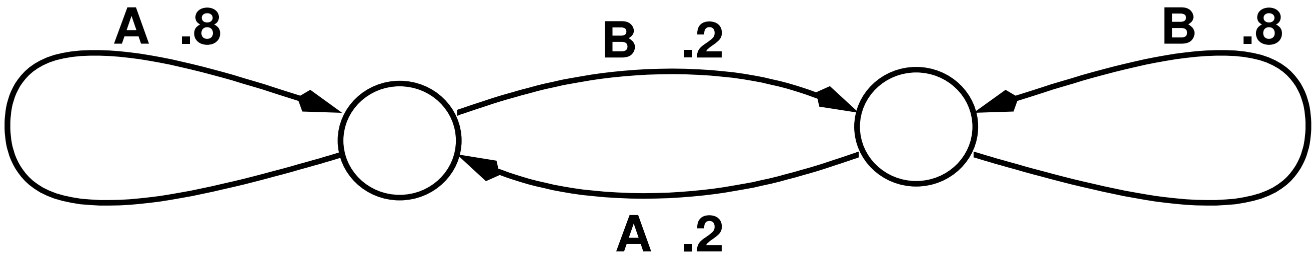
A useful conceptual and practical device: *coin-flipping models*

- A sentence is generated by a randomized algorithm
  - ▷ The generator can be in one of several "states"
  - ▷ Flip coins to choose the next state.
  - ▷ Flip other coins to decide which letter or word to output
- Shannon: *"The states will correspond to the "residue of influence" from preceding letters"*

Coin-Flipping Models



A B B A B B A A B A B A B B A A B A B A B



A A A A A A B B B B B B A A A A B B B B B

Markov Approximations to English

From Shannon's original paper:

1. *Zero-order approximation:*

   XFOML RXKXRJFFUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD

2. *First-order approximation:*

   OCRO HLI RGWR NWIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH RBL

3. Second-order approximation:

   ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

Markov Approximations (cont.)

From Shannon's original paper

4. *Third-order approximation:*

   IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTABIN IS REGOACTIONA OF CRE

Markov random field with 1000 "features," no underlying "machine" (Della Pietra et. al, 1997):

WAS REASER IN THERE TO WILL WAS BY HOMES THING BE RELOVERATED THER WHICH CONISTS AT FORES ANDITING WITH PROVERAL THE CHESTRAING FOR HAVE TO INTRALLY OF QUT DIVERAL THIS OFFECT INATEVER THIFER CONSTRANDED STATER VILL MENTTERING AND OF IN VERATE OF TO

Word-Based Approximations

1. *First-order approximation:*

   REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE T

2. *Second-order approximation:*

   THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED

Shannon’s comment:

“*It would be interesting if further approximations could be constructed, but the labor involved becomes enormous at the next stage.*”

No Myths...Only a Beginning

"The linguistic content of our program thus far is scant indeed. It is limited to one set of rules for analyzing a string of characters into a string of words, and another set of rules for analyzing a string of words into a string of sentences. Doubtless even these can be recast in terms of some information theoretic objective function. But *it is not our intention to ignore linguistics, neither to replace it. Rather, we hope to enfold it in the embrace of a secure probabilistic framework so that the two together may draw strength from one another and guide us to better natural language processing systems* in general and to better machine translation systems in particular."

— *The Mathematics of Statistical Machine Translation*
[Brown, Della Pietra, Della Pietra, and Mercer, 1993]

NLP is Cross-Disciplinary

Excellent opportunities for interdisciplinary work.

- **Linguistics**: models of language
  - ▷ emphasizes 100% accuracy (*competence*)
- **Psychology**: models of cognitive processes
  - ▷ emphasizes biological/cognitive plausibility
- **Mathematics and statistics**: properties of models
  - ▷ emphasizes formal aspects

On the whole, NLP tends to be applications-oriented: 95% is OK; models need be neither biologically plausible nor mathematically satisifying.