

Computational Linguistics Forensic Linguistics A Bird's Eye View

Mohammad Haji-Abdolhosseini
LING 101X: Introduction to the Study of Linguistics
Jan. 23, 2007

Computational Linguistics (CL)

- A.k.a. Natural Language Processing (NLP), Language Engineering
- Study of making computers use human languages for specific tasks.
 - Conversational Agents
 - Machine Translation
 - Automatic Speech Recognition/Synthesis
 - Automatic Summarization
 - Text Categorization
 - Information Retrieval
 - Text/Dialogue Segmentation
 - Intelligent Tutoring Systems
 - ...



1/23/2007

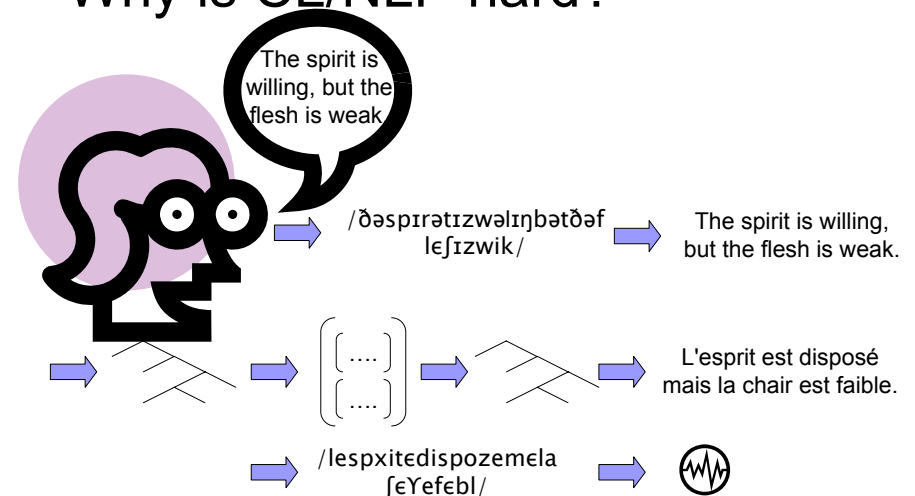
Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

2

The Beginnings

- The Turing Test (Allen Turing, 1950)
 - Question: Can machines think?
 - Question: What is intelligence?
 - A machine is intelligent if it can fool a human interrogator into thinking that it is human.
 - No machine has ever come close to passing the Turing test.

Why is CL/NLP hard?



1/23/2007

Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

4

1/23/2007

Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

3

Some CL/NLP Problems

- **Speech Recognition:**
 - People's voices and accents vary greatly.
 - Many words/phrases sound similar, e.g., *he's, his; hour, our; striker, strike her.*
- **Part-of-Speech Tagging:**
 - Many words belong to multiple parts of speech.
 - *Fruit flies like a banana.*
 - *Fruit/NN flies/NN like/V a/DT banana/NN ./.*
 - *Fruit/NN flies/V like/JJ a/DT banana/NN ./.*

1/23/2007

Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

5

Some CL/NLP Problems

- **Syntactic Parsing:**
 - Even though grammar rules are finite, the combinatorial possibilities of rules make an extremely complex system.
- **Semantic Analysis:**
 - What is the best way to represent meaning?
 - How much knowledge is needed for semantic representation?

1/23/2007

Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

6

Some CL/NLP Problems

- **Dialogue Agents:**
 - How are human dialogues structured?
 - How do we know what to say next?
- **Automatic Summarization:**
 - How do we find more salient sentences in a text that are representative of the content?
- **Text Data Mining:**
 - How do we discover new knowledge from large collections of text?

1/23/2007

Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

7

Career Opportunities in CL/NLP

- Grammar Writer
- Knowledge Engineer
- Computational Lexicographer
- Language Engineer
- Academic Positions



1/23/2007

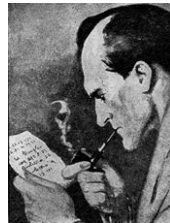
Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

8

Forensic Linguistics

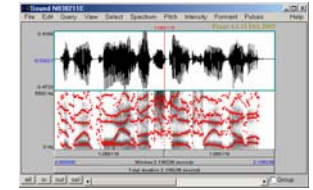
- The practice of finding information from linguistic data useful to law enforcement or for presentation at court.

- Forensic Phonetics
- Authorship Analysis
- Statement Analysis



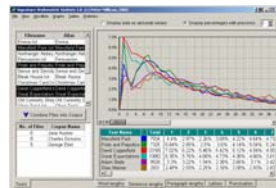
Forensic Phonetics

- Speaker Identification
 - Who said it?
- Transcription
 - What did the speaker say?
- Authentication of recordings
 - Is the recording genuine or has it been tampered with?
- Language or Accent Identification
 - Where is the speaker from?



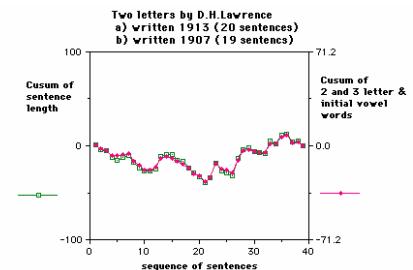
Authorship Analysis

- Authorship Attribution
 - Ransom letters
 - Hate mail
 - Any other document
- Similarity Analysis; Stylometry
 - Detecting plagiarism
- Authorship Profiling
 - Determining author's cohort



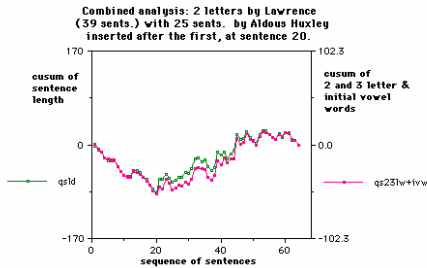
Authorship Analysis

- Finding author's linguistic "fingerprint."
 - Assumption: Each author has a set of subconscious writing habits that are constant over time.



Authorship Analysis

- Finding author's linguistic "fingerprint."
 - Assumption: Each author has a set of subconscious writing habits that are constant over time.



Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

1/23/2007

13

Authorship Profiling

- The Anthrax Scare:
 - Uses "-" in dates
 - Uses block letters
 - Addresses slanted
 - Quotes Russian style, Ns resemble Russian I
 - Fakes being Russian
 - Many clues pointing to the author being an American Scientist.



<http://www.anthraxinvestigation.com/AnthraxFoster.pdf>

Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

1/23/2007

14

Statement Analysis

- Detecting deception
 - Word definitions
 - *You know, I'm trying to be as honest as possible.*
 - Grammar
 - *A man pointed a gun at me and told me to get in the car.*
 - Research
 - Things people typically say when they lie.
 - *I swear to God. I only had two drinks. There were three men that attacked me.*

<http://www.statementanalysis.com/>



Thank You!

Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

1/23/2007

15

Computational Linguistics/Forensic
Linguistics: A Bird's Eye View

1/23/2007

16