North American Computational Linguistics Olympiad @ Illinois

Julia Hockenmaier
juliahmr@illinois.edu
3324 Siebel Center
Today’s meeting

What is Computational Linguistics?

What is NACLO?

How can you prepare for NACLO?
What is Computational linguistics?

Computational linguists develop theories about language that are precise and formal enough to be implemented on a computer.

Natural language engineers implement systems that solve a particular linguistic task (without necessarily trying to explain anything about language)
What would we need to build HAL?

• **Speech Recognizer and Synthesizer**
  to process audio signal and generate speech

• **Tokenizer/Segmenter**
  to break audio signal into words

• **Lexicon and Grammar**
  to know the structure and meaning of words and sentences

• **Syntactic/Semantic Parser**
  to obtain the structure and meaning of what was said

• **Natural Language Generation**
  to formulate replies

• **Discourse Model**
  to keep track of what the conversation is about
Current Machine Translation

Google Translate
Becoming a computational linguist

Undergraduate degrees in computer science and/or linguistics (or electrical engineering for speech processing)

Recommended coursework:
- Probability, Linear Algebra, Logic
- Programming
- Theoretical Linguistics
  (phonetics/phonology, morphology, syntax, semantics)
- Theoretical Computer Science
  (Formal languages, algorithms)
- Natural language processing, AI, machine learning
- Foreign language(s) are helpful, not mandatory
The different areas of linguistics

- **How does language work?**
  (core linguistics)
- **How do people learn and process language?**
  (psycholinguistics)
- **Where in the brain is language located?**
  (neurolinguistics)
- **How do languages change over time?**
  (historical linguistics)
- **How does language express identity/social status?**
  (sociolinguistics)
- **How can you teach foreign languages?**
  (applied linguistics)
How does language work?

What sounds are used in human speech?
(phonetics)

How do languages use and combine sounds?
(phonology)

How do languages form words?
(morphology)

How do languages form sentences?
(syntax)

How do languages convey meaning in sentences?
(semantics)

How do people use language to communicate?
(pragmatics)
What is grammar?
What is grammar?

Grammar formalisms
(= linguists’ programming languages)
A precise way to define and describe the structure of sentences.

Specific grammars
(= linguists’ programs)
Implementations (in a particular formalism) for a particular language (English, Chinese,....)
Why is NLP hard?
Overgeneration

English

Undergeneration

John saw Mary.
I ate sushi with tuna.

Did you go there?

I want you to go there.

I ate the cake that John had made for me yesterday.

John made but Mary just bought some cake.

Did you went there?

.....

John Mary saw.
with tuna sushi ate I.
“I made her duck”

What does this sentence mean?
- “duck”: noun or verb?
- “make”: “cook X” or “cause X to do Y”?
- “her”: “for her” or “belonging to her”?

Other examples of ambiguity:
- “I eat sushi with tuna” or “I eat sushi with chopsticks”
- “I saw the man with the telescope on the hill”
- “John saw Jim. He was drinking coffee.”
NACLO
Open Round (February 4, 2009)
  - Open to all middle and high school students
  - 3 hours long
  - Problems include "practice" and "theory" questions;
    Practice parts are worth about 60% of the score,
    Theory parts are worth about 40%.
  - Strong contestants will advance to the second round.

Invitational Round (March 11, 2009)
  - Harder problems.
  - 5 hours long
  - National winners will be eligible to participate
    in the international competition (ILO).
Problem types

- **Translation problems**: Given a set of sentences in language X and their English translations, translate new sentences in X into English (or vice versa).

- **Number problems**: Identify the number system of language X (number terms, basis 10 or something else?) from a set of sentences in X which describe basic arithmetic facts.

- **Calendar systems**: given sentences in X that refer to dates, identify the calendar that was used.
Problem types

- **Writing systems**: decipher a new writing system (and write something in it yourself)

- **Phonological problems**: what is the relationship between the writing system of a language and its sounds?

- **Formal problems**: identify the logical rules that underly linguistic phenomena (how do we get from “Mary ate an apple” to “An apple was eaten by Mary”?)

- **Computational problems**: identify a precise procedure to solve a particular linguistic task that you could program a computer to do

- **Other types**: kinship systems, associating sentences with images,…
NACLO @ Illinois

Open Round:
February 4, 2009, 8:30am-12pm.
Siebel Center for Computer Science, 210 N. Goodwin Ave,
Urbana, IL 61801. (Room TBC)

Organizers:
Roxana Girju and Julia Hockenmaier
naclo@cs.uiuc.edu

Website:
http://nlp.cs.uiuc.edu/NACLO/
What you need to bring

- Blank paper,
- Black nonerasable pens,
- Pencils (for scratch work)

- The use of computers, cell phones, pagers, calculators, watches with calculators, PDAs, etc. is not allowed!
Submitting solutions

- Use a **black nonerasable pen**, 
- Ensure handwriting is legible  
  *(Judges will grade only legible solutions)!*
- Write your **name & registration number on each page** of the booklet
- Write solutions in the appropriate spaces in the booklet
- You can use **extra blank paper**, but:
  - Only use one side of each sheet
  - Don’t write answers to more than one question on same sheet  
  *(The answers will be split by problem number and [scanned in and] shipped to graders around the world.)*
- You can solve the given problems in any order
- You should try to solve as many problems as possible.
- You may get partial credit for providing incomplete solutions or specific ideas for solving a problem - so write those down as well!
Preparing for NACLO
Translation problems
Quechua sentences
1. Antukaq chakranpiqa t'ikashanmi papa.
2. Siskuq chakranpiqa wiñashanmi sara.
3. Siskuq chakranpiqa rurushansi kiwña.
4. Antukaq chakranpiqa t'ikashanchá kiwña.
5. Siskuq chakranpiqa wiñashansí sara.
6. Antukaq chakranpiqa wiñashanchá papa.

Translations in random order
A. Potatoes may be growing in Antuka's field.
B. Barley may be flowering in Antuka's field.
C. Corn is growing in Sisku's field.
D. I've heard corn is growing in Sisku's field.
E. I've heard barley is yielding fruit in Sisku's field.
F. Potatoes are flowering in Antuka's field.
Now, translate the following Quechua sentences:

7. Istuchaq chakranpiqa t'ikashansi sara.
8. Sawinaq chakranpiqa wiñashanchá kiwña.
10. Kusiq chakranpiqa t'ikashanchá papa.
11. Inashuq chakranpiqa rurushansi kiwña.
Solution

F 1. Antukaq chakranpiqa t'ikashanmi papa.
   Gloss: Potatoes are flowering in Antuka's field.

C 2. Siskuq chakranpiqa wiñashanmi sara.
   Gloss: Corn is growing in Sisku's field.

E 3. Siskuq chakranpiqa rurushansi kiwña.
   Gloss: I've heard barley is yielding fruit in Sisku's field.

B 4. Antukaq chakranpiqa t'ikashanchá kiwña.
   Gloss: Barley may be flowering in Antuka's field.

D 5. Siskuq chakranpiqa wiñashansí sara.
   Gloss: I've heard corn is growing in Sisku's field.

A 6. Antukaq chakranpiqa wiñashanchá papa.
   Gloss: Potatoes may be growing in Antuka's field.
Sentence Translations

7. Istuchaq chakranpiqa t'ikashansi sara.
   Gloss: I hear corn is flowering in Istucha's field.

8. Sawinaq chakranpiqa wiñashanchá kiwña.
   Gloss: Barley may be growing in Sawina's field.

   Gloss: Potatoes are yielding fruit in Tumasa's field.

10. Kusiq chakranpiqa t'ikashanchá papa.
    Gloss: Potatoes may be flowering in Kusi's field.

11. Inashuq chakranpiqa rurushansi kiwña.
    Gloss: I hear barley is yielding fruit in Inashu's field.
Garden path sentences
The horse raced past the barn fell.

Is this sentence grammatical?
What about this one?

The car driven into the river sank.
Garden path sentences

People process language **incrementally** (one word at a time).

When we hear or read sentences, there are many **processing effects** that influence how easily we understand them.
Don’t bother coming early.

Take the turkey out at five to four.

I got canned peaches.

All Americans need to buy a house is a large amount of money.

Melanie is pretty busy.
Your task

Come up with two new garden path sentences.

What makes a garden path sentence harder to process?
See you at NACLO @ Illinois!

Open Round:
  February 4, 2009, 8:30am-12pm.
  Siebel Center for Computer Science, 210 N. Goodwin Ave, Urbana, IL 61801. (Room TBC)

What you need to bring:
  Black nonerasable pens,
  Blank paper, pencils (for scratch work)

Organizers:
  Roxana Girju and Julia Hockenmaier (naclo@cs.uiuc.edu)
  http://nlp.cs.uiuc.edu/NACLO/