

# LING 520: Computational Analysis of English

Semester: FALL '16

Instructor: Sowmya Vajjala

Iowa State University, USA

25 August 2016

# Class outline

- ▶ Why is NLP hard?
- ▶ What are some common NLP tasks?
- ▶ Assignment 1 overview
- ▶ Programming project demos

Why is NLP hard?  
(what sort of issues pose problems for a computer?)

# Language is ambiguous

## Some ambiguous sentences

- ▶ Newspaper headlines
  - ▶ "Children make delicious snacks"
  - ▶ "Dead expected to rise"
  - ▶ "Republicans grill IRS chief over lost emails"
- ▶ Normal, grammatical sentences can be ambiguous too:
  - ▶ "I saw a man on a hill with a telescope."
  - ▶ "Look at the man with one eye"

We are not even talking about ambiguities involving speech or alternative interpretations due to stress/emphasis on some word.

# Some types of ambiguity

1. Lexical ambiguity: due to multiple meanings or senses of word usage  
e.g., He stood near the **bank**
2. Structural ambiguity: due to syntactic structure  
e.g., I saw the man on the hill with telescope.
3. Semantic ambiguity: more interpretations possible  
e.g., John and Mary are married (to each other? or to different people?)
4. Referential ambiguity  
e.g., She dropped the *plate* on the *table* and broke **it**
5. Ambiguity due to the use of non-literal language  
e.g., Time flies like an arrow

Good source to read more:

<http://cs.nyu.edu/faculty/davise/ai/ambiguity.html>

# Ambiguity for humans

this happened to me last night ...

- ▶ "ISU sends out alert as devastating weed spreads in Iowa" - read a local newspaper headline. What did you understand?

# Ambiguity for humans

this happened to me last night ...

- ▶ "ISU sends out alert as devastating weed spreads in Iowa" - read a local newspaper headline. What did you understand?
- ▶ My immediate reaction: Some new drug is circulating and ISU wants its students to be aware and not consume it.

# Ambiguity for humans

this happened to me last night ...

- ▶ "ISU sends out alert as devastating weed spreads in Iowa" - read a local newspaper headline. What did you understand?
- ▶ My immediate reaction: Some new drug is circulating and ISU wants its students to be aware and not consume it.
- ▶ However, the actual news is about some poisonous wild plant, and the alert was issued by ISU weed and crop specialists to farmers.

... so some humans also cannot disambiguate certain things.



# "common" knowledge for humans

Look at these two sentences:

Dog bit man.

Man bit dog.

- For a computer, both of them are linguistically the same. We know only the first one is "normal" English sentence (I hope!) because we have "world knowledge".

# Language is creative

Literary texts have their own language style: long sentences, neologisms, creative usage of words etc.

# Language can be complex to understand

Legal documents, Writing style of some authors, propaganda materials, etc.

# Language is Diverse

## Some Examples

- ▶ There are different types of text online: news, tweets, SMS, email, forum posts, speech transcripts etc.
- ▶ Each genre has some specific characteristics of its own
- ▶ NLP methods should generalize to all genres, and at the same time capture such specific characteristics
- ▶ Example: a machine translation model created by training examples from European parliament speeches should also be able to translate casual day to day conversations.

# Other Issues

- ▶ different text formats (pdf, doc, txt etc)
- ▶ spelling variations
- ▶ sarcasm
- ▶ slang
- ▶ using synonyms, paraphrases etc

.. and so on.

# Languages are many

.. and each language has its own special characteristics apart from similarities with other languages. NLP should handle both these aspects.

So, the summary is:

perfect NLP is hard to achieve because of all these issues that come up when we start using computers to analyze language!

# Question Time

- ▶ Apart from all these, what are some CALL specific problems for NLP?



# Question Time

- ▶ Apart from all these, what are some CALL specific problems for NLP?
  - ▶ One issue I can think of: all NLP tools usually work under the assumption that there is only one standard way of writing in a given language, and everyone writes grammatically correct sentences.

## Some NLP tasks

# Let us take a small text snippet

“I’m not big, I’m not fancy,” she said as she sat in a booth, looking out the window to Lincoln Way. “But I don’t mind.”

Chinese Homestyle Cooking’s owners, Tina and Chung Song, have run the restaurant for almost 20 years from the small building off the corner of Sheldon Avenue and Lincoln Way. But in late October, they’ll close their business when the lease on their building runs out.

Tina’s start in the restaurant business came in 1982, when she emigrated from Taiwan and began working as a waitress for her sister’s restaurants in Des Moines and Ankeny. While in the U.S., she got a call from Chung, who she had grown up with as a child in their home country.

Source: Ames Tribune (<http://goo.gl/zvx9Uw>)

1. When she says “I” in the first sentence, does she mean herself literally?

# Let us take a small text snippet

“I’m not big, I’m not fancy,” she said as she sat in a booth, looking out the window to Lincoln Way. “But I don’t mind.”

Chinese Homestyle Cooking’s owners, Tina and Chung Song, have run the restaurant for almost 20 years from the small building off the corner of Sheldon Avenue and Lincoln Way. But in late October, they’ll close their business when the lease on their building runs out.

Tina’s start in the restaurant business came in 1982, when she emigrated from Taiwan and began working as a waitress for her sister’s restaurants in Des Moines and Ankeny. While in the U.S., she got a call from Chung, who she had grown up with as a child in their home country.

Source: Ames Tribune (<http://goo.gl/zvx9Uw>)

1. When she says “I” in the first sentence, does she mean herself literally?
2. What is she referring to? When will we know what is she referring to?

# Let us take a small text snippet

“I’m not big, I’m not fancy,” she said as she sat in a booth, looking out the window to Lincoln Way. “But I don’t mind.”

Chinese Homestyle Cooking’s owners, Tina and Chung Song, have run the restaurant for almost 20 years from the small building off the corner of Sheldon Avenue and Lincoln Way. But in late October, they’ll close their business when the lease on their building runs out.

Tina’s start in the restaurant business came in 1982, when she emigrated from Taiwan and began working as a waitress for her sister’s restaurants in Des Moines and Ankeny. While in the U.S., she got a call from Chung, who she had grown up with as a child in their home country.

Source: Ames Tribune (<http://goo.gl/zvx9Uw>)

1. When she says "I" in the first sentence, does she mean herself literally?
2. What is she referring to? When will we know what is she referring to?
3. Who is "She"?

# Let us take a small text snippet

“I’m not big, I’m not fancy,” she said as she sat in a booth, looking out the window to Lincoln Way. “But I don’t mind.”

Chinese Homestyle Cooking’s owners, Tina and Chung Song, have run the restaurant for almost 20 years from the small building off the corner of Sheldon Avenue and Lincoln Way. But in late October, they’ll close their business when the lease on their building runs out.

Tina’s start in the restaurant business came in 1982, when she emigrated from Taiwan and began working as a waitress for her sister’s restaurants in Des Moines and Ankeny. While in the U.S., she got a call from Chung, who she had grown up with as a child in their home country.

Source: Ames Tribune (<http://goo.gl/zvx9Uw>)

1. When she says "I" in the first sentence, does she mean herself literally?
2. What is she referring to? When will we know what is she referring to?
3. Who is "She"?
4. What is "home country" in the last sentence?

# Let us take a small text snippet

“‘I’m not big, I’m not fancy,” she said as she sat in a booth, looking out the window to Lincoln Way. “But I don’t mind.”

Chinese Homestyle Cooking’s owners, Tina and Chung Song, have run the restaurant for almost 20 years from the small building off the corner of Sheldon Avenue and Lincoln Way. But in late October, they’ll close their business when the lease on their building runs out.

Tina’s start in the restaurant business came in 1982, when she emigrated from Taiwan and began working as a waitress for her sister’s restaurants in Des Moines and Ankeny. While in the U.S., she got a call from Chung, who she had grown up with as a child in their home country.

Source: Ames Tribune <http://goo.gl/zvx9Uw>

1. What is the main event of this text?

# Let us take a small text snippet

“‘I’m not big, I’m not fancy,” she said as she sat in a booth, looking out the window to Lincoln Way. “But I don’t mind.”

Chinese Homestyle Cooking’s owners, Tina and Chung Song, have run the restaurant for almost 20 years from the small building off the corner of Sheldon Avenue and Lincoln Way. But in late October, they’ll close their business when the lease on their building runs out.

Tina’s start in the restaurant business came in 1982, when she emigrated from Taiwan and began working as a waitress for her sister’s restaurants in Des Moines and Ankeny. While in the U.S., she got a call from Chung, who she had grown up with as a child in their home country.

Source: Ames Tribune <http://goo.gl/zvx9Uw>

1. What is the main event of this text?
2. What is "Chinese Homestyle Cooking" referring to?



# Let us take a small text snippet

“‘I’m not big, I’m not fancy,” she said as she sat in a booth, looking out the window to Lincoln Way. “But I don’t mind.”

Chinese Homestyle Cooking’s owners, Tina and Chung Song, have run the restaurant for almost 20 years from the small building off the corner of Sheldon Avenue and Lincoln Way. But in late October, they’ll close their business when the lease on their building runs out.

Tina’s start in the restaurant business came in 1982, when she emigrated from Taiwan and began working as a waitress for her sister’s restaurants in Des Moines and Ankeny. While in the U.S., she got a call from Chung, who she had grown up with as a child in their home country.

Source: Ames Tribune <http://goo.gl/zvx9Uw>

1. What is the main event of this text?
2. What is "Chinese Homestyle Cooking" referring to?
3. What is the relationship between "Chinese Homestyle cooking" and Tina?

# Let us take a small text snippet

“‘I’m not big, I’m not fancy,” she said as she sat in a booth, looking out the window to Lincoln Way. “But I don’t mind.”

Chinese Homestyle Cooking’s owners, Tina and Chung Song, have run the restaurant for almost 20 years from the small building off the corner of Sheldon Avenue and Lincoln Way. But in late October, they’ll close their business when the lease on their building runs out.

Tina’s start in the restaurant business came in 1982, when she emigrated from Taiwan and began working as a waitress for her sister’s restaurants in Des Moines and Ankeny. While in the U.S., she got a call from Chung, who she had grown up with as a child in their home country.

Source: Ames Tribune <http://goo.gl/zvx9Uw>

1. What is the main event of this text?
2. What is "Chinese Homestyle Cooking" referring to?
3. What is the relationship between "Chinese Homestyle cooking" and Tina?
4. Is Lincoln Way something related to President Lincoln?

# NLP tasks: Word Collocations and Concordances

- ▶ Task: compiling lists of words, or word sequences occurring in documents.
- ▶ Simplest and least ambiguous form of language processing.
- ▶ Can get to more advanced collocations beyond surface forms.
- ▶ Established methods already exist for collecting different kinds of collocations

# NLP tasks: Pattern Extraction

- ▶ Task: Extract the language patterns that exist in textual data.
- ▶ Regular expressions are very useful for this
- ▶ More advanced methods (which rely on machine learning) exist to extract unknown patterns from unstructured text documents.

# NLP tasks: POS Tagging

What is the big deal about automatic tagging?

- ▶ Task: Given a sequence of words, return the POS tags for each word.
- ▶ An example problem: What is the best tag for a word in a context?
  - ▶ I wish to cite this work.  
PRP/I VBP/wish TO/to VB/cite DT/this NN/work ./.
  - ▶ He has a wish.  
PRP/He VBZ/has DT/a NN/wish ./.
- ▶ Largely considered solved for English, but there are still issues if we go beyond typical newspaper language (e.g., tagging speech or tweets). Still an unsolved problem for several languages.

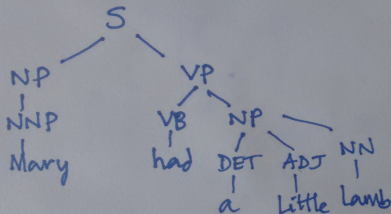
# NLP tasks: Parsing

- ▶ Task: Construct the syntactic structure of a given sentence.
- ▶ Two kinds of trees can be generated in NLP: Phrase structure tree (Constituency tree), Dependency tree
- ▶ PST: shows parse structure in terms of Noun Phrases, Verb Phrases, Prep. Phrases etc.
- ▶ Dependency Tree: shows relations between words in a sentence in terms of a pre-defined set of relations
- ▶ Very active area of current research, for multiple languages.
- ▶ Important note: POS tagging errors can carry over and affect parser efficiency.

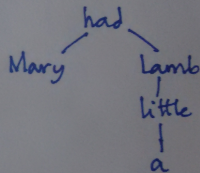
# NLP tasks: Parsing

Mary had a little Lamb.

(a) Phrase structure Tree



(b) Dependency tree



# NLP tasks: Word Sense Disambiguation

- ▶ Task: For words that can have multiple meanings, what is the right sense of the word in a given sentence?
- ▶ Example: "Let us go inside, it is cold" vs "I have cold and cough"
- ▶ Very important for applications such as machine translation, information retrieval
- ▶ Good progress for English WSD. One of the active areas of research in the field.



# NLP tasks: Named Entity Recognition

- ▶ Task: Identify and classify named entities (e.g., person names, organization names, locations etc.,)
- ▶ Application: Information extraction from text
- ▶ Some NER is domain specific (biomedical NER, financial NER etc)
- ▶ Current methods of NER: hand-crafted or automatically compiled lists + statistical machine learning models
- ▶ Active area of research for English and other languages.

# NLP tasks: Semantic Role Labeling

- ▶ SRL is all about doing a "semantic parse" of a sentence. The task here is to identify argument structure of a sentence and thematic roles of different entities.
- ▶ Example: (source: <http://www.cs.upc.edu/~srlconll/>)

The following sentence, taken from the PropBank corpus, exemplifies the annotation of semantic roles:

[<sub>A0</sub> He ] [<sub>AM-MOD</sub> would ] [<sub>AM-NEG</sub> n't ] [<sub>V</sub> **accept** ] [<sub>A1</sub> anything of value ] from [<sub>A2</sub> those he was writing about ] .

Here, the roles for the predicate **accept** (that is, the *roleset* of the predicate) are defined in the PropBank Frames scheme as:

**V:** verb  
**A0:** acceptor  
**A1:** thing accepted  
**A2:** accepted-from  
**A3:** attribute  
**AM-MOD:** modal  
**AM-NEG:** negation

- ▶ Active area of research. Still hard, but making progress.

# NLP tasks: Discourse Analysis

- ▶ Task: Given a text (more than one sentence), analyze the relationships between sentences, identify what pronouns refer to what nouns, how is the same entity referred in different ways (Barack Obama, Obama, The President and so on).
- ▶ Application: Text summarization, Question-Answering, Essay scoring etc.
- ▶ Hard problem, but active research topic and hence, making good progress

# NLP tasks: Text - Speech interface

- ▶ Text to speech conversion, Speech to text conversion
- ▶ Traditionally studied separately under Speech processing. Not a part of NLP courses generally.
- ▶ Difficult task, but lot of improvement in speech recognition happened in the recent past, partly due to "deep learning" methods
- ▶ You can easily incorporate speech recognition into your Python code now, without writing any speech related code yourself!

# NLP tasks: Entailment and Paraphrasing

- ▶ Tasks: Analyzing if the meaning of a sentence is entailed in another sentence, if both sentences are paraphrases of each other etc.,
- ▶ Uses: question answering, information extraction, summarization etc.
- ▶ Still very hard. Most of the research is with English.. so we don't know if that is even harder for several other languages.

# NLP tasks: Language Generation

- ▶ Task: Generate text automatically.
- ▶ Texts should be grammatically and semantically correct.  
Should be human like.
- ▶ One of the toughest problems in NLP.
- ▶ Example 1: Create weather reports, match summaries, reports etc. automatically (without human intervention!)
- ▶ arria.com is a NLG company that does this successfully for English.
- ▶ There are some software libraries that support the development of NLG systems for some languages currently.

# Question Time

Where in CALL applications are the following tasks relevant:

- ▶ Paraphrase identification
- ▶ Natural Language Generation
- ▶ Text to Speech conversion
- ▶ Speech to Text conversion
- ▶ Named Entity Recognition

# NLP: References, Resources etc

- ▶ Access to various publications:  
<http://aclweb.org/anthology/>  
(Almost all major publication venues in the field are open-access!)
- ▶ Information about resources for different languages: ACL Wiki  
<http://www.aclweb.org/aclwiki>
- ▶ Know about other NLP courses around the world etc:  
ACLWeb again



Assignment 1 description  
(Deadline: 10 September)  
On Blackboard as PDF

# Next Week

## ► To Do:

1. Start revising your programming concepts. Look at the notes you made in the previous course.
2. Go through Radev's course lectures: Week1 and last three lectures from Week 3.
3. Read Chapter 1 in Jurafsky and Martin book
4. Python users: You can start with Chapter 1 in the book if you want to do some stuff on your own.

## ► Next week:

1. Programming review and practice. Come prepared to work either on your laptops or lab machines.
2. Setting up Perl and Python programming environments on lab machines or personal laptops

## Programming Project - Demos