

# North American Computational Linguistics Olympiad

Lori Levin

Language Technologies Institute

Carnegie Mellon University

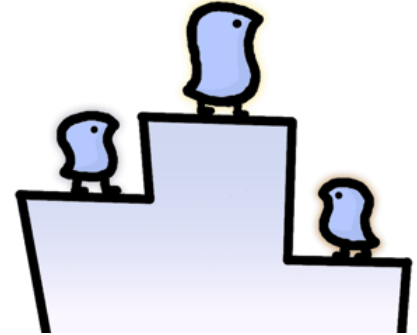
# Web sites

- North American Computational Linguistics Olympiad
  - <http://www.naclo.cs.cmu.edu>
- International Linguistics Olympiad
  - <http://www.ioling.org>
- Videos about NACLO on You Tube
  - <http://www.youtube.com/watch?v=82rbhy4Xjbs>
  - [http://youtu.be/ao2tX3\\_qakU](http://youtu.be/ao2tX3_qakU)



# IOL

International Linguistics  
Olympiad



<http://www.ioling.org>

# Fascinating, addictive language puzzles

## Tajik







дуусти хуби ҳамсояи сумо a good friend of your neighbor

ҳамсояи дуусти хуби сумо a neighbor of your good friend

ҳамсояи хуби дуусти сумо a good neighbor of your friend

*Problem by Adriana Solovyova.*

## Luvian

- |  |  |
|--|--|
| 1.  | 4.  |
| 2.  | 5.  |
| 3.  | 6.  |

*Original problem by A. N. Zhurinskij. English adaptation by Valentin Vydrin.*

## Swahili

- |                                  |                                       |
|----------------------------------|---------------------------------------|
| ___ tarehe tatu Disemba jumamosi | a. Monday, October 5 <sup>th</sup>    |
| ___ tarehe pili Aprili jumanne   | b. Wednesday, October 5 <sup>th</sup> |
| ___ tarehe nne Aprili jumanne    | c. Sunday, October 5 <sup>th</sup>    |
| ___ tarehe tano Oktoba jumapili  | d. Tuesday, April 2 <sup>nd</sup>     |
| ___ tarehe tano Oktoba jumatatu  | e. Tuesday, April 4 <sup>th</sup>     |
| ___ tarehe tano Oktoba jumatano  | f. Saturday, December 3 <sup>rd</sup> |

*Original problem by A. N. Zhurinskij. English adaptation by Valentin Vydrin.*

# Linguistics and NACLO

- Languages have structure that you can discover.

дуустн хубн хамсоай сумо a good friend of your neighbor

хамсоай дуустн хубн сумо a neighbor of your good friend

хамсоай хубн дуустн сумо a good neighbor of your friend

*Problem by Adriana Solovyova.*



# NACLO: what?

- A nation-wide pencil and paper contest with no pre-requisites.
  - Free too!
- Problems about human language and computation.
  - Easy problems:
    - everyone has a good time and learns something about human language and/or computation
  - Hard problems:
    - identify the students who are most skilled at seeing patterns and structure
    - assemble a team that can win an international competition

# NACLO: when and where

- Open competition
  - January 31, 2013
- Invitational competition
  - Mid March
- You can participate at your school or at a university host site.

# NACLO: Who

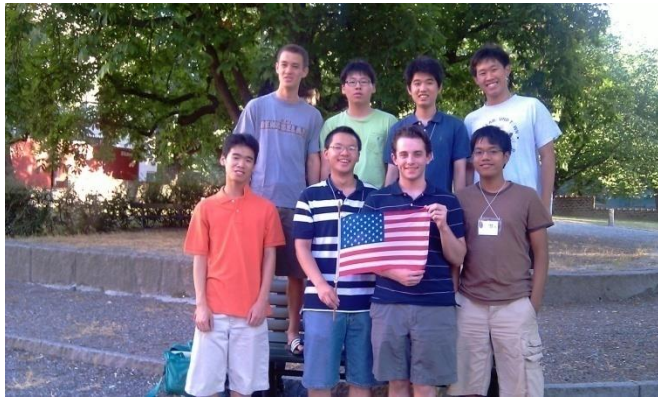
- Any student who is not in college yet
- Mostly high school
- Some middle school
- Winners have been 13 to 18 years old



# IOL International Linguistics Olympiad

<http://www.ioling.org>



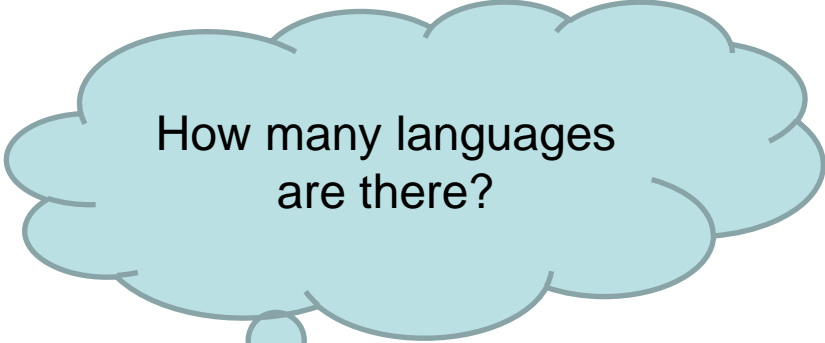


# NACLO and IOL

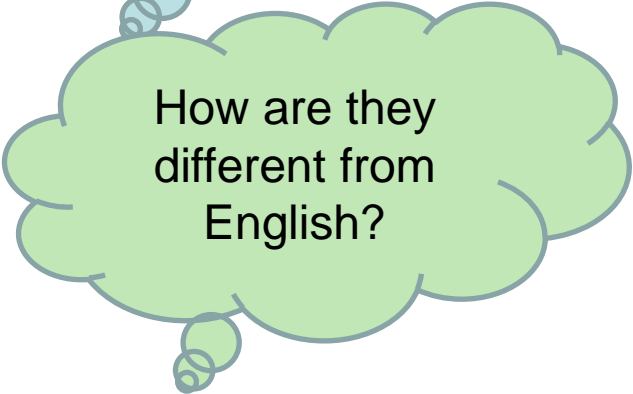
- The top eight students in the invitational competition are invited to the International Linguistics Olympiad (IOL).
  - IOL is in the UK in 2013
- IOL has individual and team competition (teams of four).
- Each country may send two teams.
  - Around 30 countries participate
- The US has participated in six IOLs.
- Canada has participated in two IOLs.

# What is Linguistics?

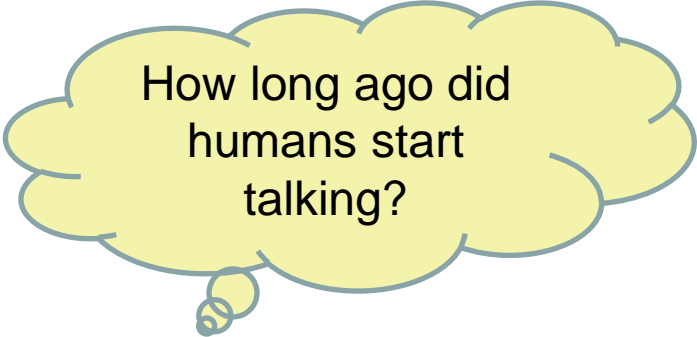
- The study of human language
  - As opposed to the study of human languages



How many languages are there?



How are they different from English?



How long ago did humans start talking?



What is grammar?

# Questions that Linguists ask

- What parts of the brain are used for producing and comprehending language?
- How do languages change?
- Why do languages change?
- How does language correlate with social factors?
  - E.g., Jocks and Burnouts (Eckert)
- What do human languages have in common?
- How are human languages different from animal communication?
- How is it that a baby can learn his/her first language perfectly, but an adult cannot learn a second language perfectly?

# Why linguistics?

- Human language is central to human communication and social interaction.
- Human language is a property of the human mind.
- You can practice discovering patterns and structure.
- You can practice scientific reasoning (forming hypotheses and knowing which data support them).

**Language Technologies**

**Computational Linguistics**

**Natural Language Processing**

# What is NLP?

- “Natural language processing is the technology for dealing with our most ubiquitous product: human language...”
  - Chris Manning and Dan Jurafsky: <http://www.nlp-class.org/>
  - “ubiquitous” means it’s everywhere

# We produce language to talk to

- Machines
  - Search queries
  - Siri
  - Telephone dialogue systems
- Each other
  - “human language .... in emails, web pages, tweets, product descriptions, newspaper stories, social media, and scientific articles, in thousands of languages and varieties.”
    - Chris Manning and Dan Jurafsky: <http://www.nlp-class.org/>
- Each other mediated by a machine that translates



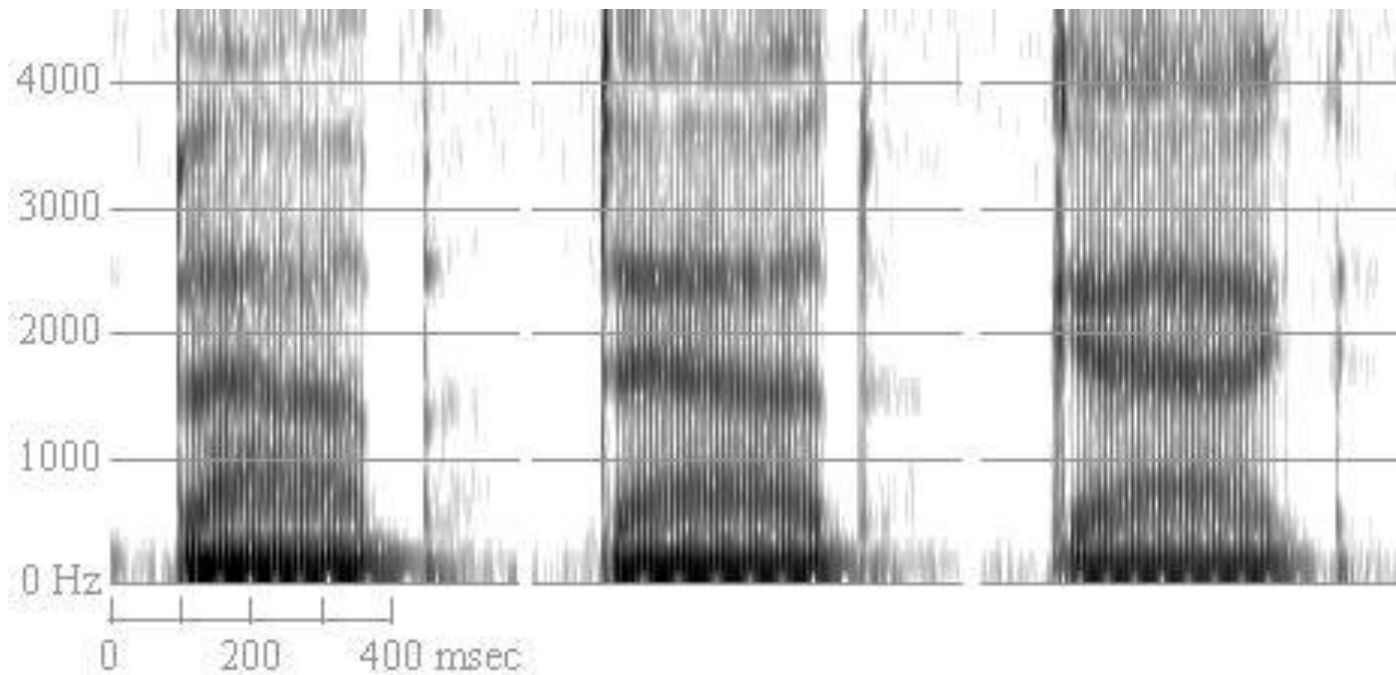
# You use Natural Language Processing every day

- Search queries
- Spell check
- Grammar check
- Spam detection
- Telephone dialogue systems
- Siri and similar things
- Google pops up ads that are related to the email you are typing.

But what **IS** NLP?

What frequencies are present in sound waves when you speak?

**bab, dad, gag**



# Where are the words?

## 世界人权宣言

联合国大会一九四八年十二月十日第217A(III)号决议通过并颁布

1948年12月10日，联合国大会通过并颁布《世界人权宣言》。这一具有历史意义的《宣言》颁布后，大会要求所有会员国广为宣传，并且“不分国家或领土的政治地位，主要在各级学校和其他教育机构加以传播、展示、阅读和阐述。”《宣言》全文如下：

## 序言

On December 10, 1948 the General Assembly of the United Nations adopted and proclaimed the Universal Declaration of Human Rights the full text of which appears in the following pages. Following this historic act the Assembly called upon all Member countries to publicize the text of the Declaration and to cause it to be disseminated, displayed, read and expounded principally in schools and other educational institutions, with out distinction based on the political status of countries or territories.

Where are the words?

**theyouthevent**

# Where are the words?

- There are no spaces in spoken language, so every spoken language is like Chinese writing:
  - How to recognize speech
  - How to wreck a nice beach
  - How to wreck an ice peach

# Where are the morphemes?

İnsan hakları evrensel beyannamesi

Önsöz

İnsanlık ailesinin bütün üyelerinde bulunan haysiyetin ve bunların eşit ve devir kabul etmez haklarının tanınması hususunun, hürriyetin, adaletin ve dünya barışının temeli olmasına,

İnsan haklarının tanınmaması ve hor görülmesinin insanlık vicdanını isyana sevkeden vahşiliklere sebep olmuş bulunmasına, dehşetten ve yoksulluktan kurtulmuş insanların, içinde söz ve inanma hürriyetlerine sahip olacakları bir dünyanın kurulması en yüksek amaçları oralak ilan edilmiş bulunmasına,

On December 10, 1948 the General Assembly of the United Nations adopted and proclaimed the Universal Declaration of Human Rights the full text of which appears in the following pages. Following this historic act the Assembly called upon all Member countries to publicize the text of the Declaration and to cause it to be disseminated, displayed, read and expounded principally in schools and other educational institutions, without distinction based on the political status of countries or territories.

# Which words are these?

n dcmbrr 10, 1948 th gnrl ssmbly f th ntd ntns dptd nd prclmd th nvrsl dclrtin f hmn rghts th fill txt f whch pprs n th fillwng pgs. fillwng ths hstrc ct th ssmbly clld pn ll mmbrr cntrs t pblcz th txt f th dclrtin nd t cs t t b dssmntd, dsplyd, rd nd xpndd princply n schls nd thr dctnl nstttns, wtht dstnctn bsd n th pltcl stts f cntrs r trrtrs.



# Ambiguity in English

- IRAQI HEAD SEEKS ARMS
- KIDS MAKE NUTRITIOUS SNACKS
- BRITISH LEFT WAFFLES ON FALKLAND ISLANDS
- STOLEN PAINTING FOUND BY TREE



**Careers**

**Humanitarian**

**Industry**

**Government**

**Academic**

**Education**



## Careers

Search engines

Natural language voice interfaces

Talking to machines

## Industry

Facebook

Twitter

Google

Yahoo

Reuters

General Motors

Microsoft

Amazon

Summarization

because there is more information than people can attend to

Sentiment detection

Did people like the product or movie?

Machine Translation

Translate from one language to another

# Careers

Machine Translation

Speech recognition

Summarization and  
information extraction

Detection of sentiment and  
deception

National Security:

There is more information  
than human analysts can  
attend to.

# Government

# Careers

Computer Assisted Language Learning  
Automatically detect errors

Automated grading of essays  
Educational Testing Service

Analysis of educational dialogue  
The way you interact affects the way you learn

# Education

# Careers

Work at a university

Train the next generation

Do research on unsolved problems in Natural Language Processing

# Academic