

What is NLP?

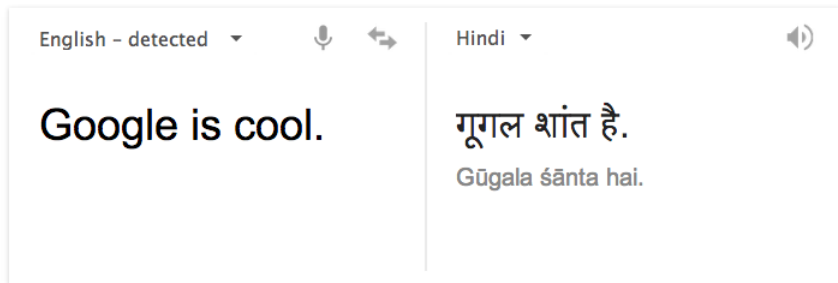
Fundamental and Scientific Goal

Deep understanding of *broad* language

Engineering Goal

Design, implement, and test systems that process natural languages for practical applications

Goals can be very ambitious: Good quality translation



[Open in Google Translate](#)

Well, even humans have made blunders

Pepsi Chinese blunder

“Come alive with the Pepsi Generation”, when translated into Chinese meant, “Pepsi brings your relatives back from the dead.”

Well, even humans have made blunders

Pepsi Chinese blunder

“Come alive with the Pepsi Generation”, when translated into Chinese meant, “Pepsi brings your relatives back from the dead.”

KFC's Chinese blunder

KFC's slogan, “Finger lickin' good”, when translated into Chinese meant “We'll eat your fingers off.”

Well, even humans ...



And Goals Can be Practical: Auto Completion

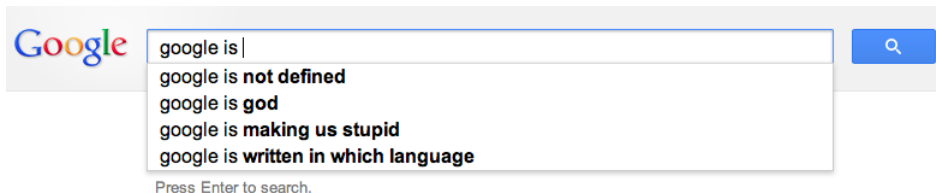
Search bar containing the text "wold cup 2014". To the right of the text is a microphone icon and a blue search button with a magnifying glass icon.

Navigation tabs: **Web**, Images, News, Videos, Maps, More ▾, Search tools.

About 86,50,00,000 results (0.25 seconds)

Showing results for **world cup 2014**
Search instead for **wold cup 2014**

And Goals can be Practical: Search Engines



And Goals can be Practical: Information Extraction

New York Times Co. named **Russell T. Lewis**, 45, **president and general manager** of its flagship **New York Times newspaper**, responsible for all business-side activities. He was **executive vice president and deputy general manager**. He succeeds **Lance R. Primis**, who in September was named **president and chief operating officer** of **the parent**.

Person	Company	Post	State
Russell T. Lewis	New York Times newspaper	president and general manager	start
Russell T. Lewis	New York Times newspaper	executive vice president	end
Lance R. Primis	New York Times Co.	president and CEO	start

Computer science class fails to notice their TA was actually an AI chatbot



by NAPIER LOPEZ — 9 weeks ago in SHAREABLES



And Goals can be Practical: Domain-specific Chatbots

Jill wasn't very good for the first few weeks after she started in January, often giving odd and irrelevant answers. Her responses were posted in a forum that wasn't visible to students.

"Initially her answers weren't good enough because she would get stuck on keywords," said Lalith Polepeddi, one of the graduate students who co-developed the virtual TA. "For example, a student asked about organizing a meet-up to go over video lessons with others, and Jill gave an answer referencing a textbook that could supplement the video lessons — same keywords — but different context. So we learned from mistakes like this one, and gradually made Jill smarter."

After some tinkering by the research team, Jill found her groove and soon was answering questions with 97 percent certainty. When she did, the human TAs would upload her responses to the students. By the end of March, Jill didn't need any assistance: She

1 wrote the class directly if she was 97 percent positive her answer was correct.

¹<http://www.news.gatech.edu/2016/05/09/artificial-intelligence-course-creates-ai-teaching-assist>

And Goals can be Practical: Sentiment Analysis



- Spam detection
- Machine Translation services on the Web
- Text Summarization
- ...

- Spam detection
- Machine Translation services on the Web
- Text Summarization
- ...

Natural Language Technology not yet perfect

But still good enough for several useful applications



Language Technology

making good progress

mostly solved

Spam detection

Let's go to Agra! ✓

Buy V1AGRA ... ✗

Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

Sentiment analysis

Best roast chicken in San Francisco! 👍

The waiter ignored us for 20 minutes. 👎

Coreference resolution

Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)

I need new batteries for my *mouse*.

Parsing

I can see Alcatraz from the window!

Machine translation (MT)

第13届上海国际电影节开幕...

The 13th Shanghai International Film Festival...

Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30

Party
May 27
add

still really hard

Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose

Economy is good

Dialog

Where is Citizen Kane playing in SF?

Castro Theatre at 7:30. Do you want a ticket?

Why is NLP hard?

Lexical Ambiguity

- *Will Will will Will's will?*

Why is NLP hard?

Lexical Ambiguity

- *Will Will will Will's will?*
- *Rose rose to put rose roes on her rows of roses.*

Why is NLP hard?

Lexical Ambiguity

- *Will Will will Will's will?*
- *Rose rose to put rose roes on her rows of roses.*
- *Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo.*

Why is NLP hard?

Lexical Ambiguity

- *Will Will will Will's will?*
- *Rose rose to put rose roes on her rows of roses.*
- *Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo.*
→ Buffaloes from Buffalo, NY, whom buffaloes from Buffalo bully, bully buffaloes from Buffalo.

Why is NLP hard?

Language ambiguity: Structural

- *The man saw the boy with the binoculars.*
- *Flying planes can be dangerous.*
- *Hole found in the room wall; police are looking into it.*

Language imprecision and vagueness

Why is NLP hard?

Language ambiguity: Structural

- *The man saw the boy with the binoculars.*
- *Flying planes can be dangerous.*
- *Hole found in the room wall; police are looking into it.*

Language imprecision and vagueness

- *It is very warm here.*

Why is NLP hard?

Language ambiguity: Structural

- *The man saw the boy with the binoculars.*
- *Flying planes can be dangerous.*
- *Hole found in the room wall; police are looking into it.*

Language imprecision and vagueness

- *It is very warm here.*
- *Q: Did your mother call your aunt last night?*
A: I'm sure she must have.

But that's the fun part of it

Why is the teacher wearing sun-glasses?

...

But that's the fun part of it

Why is the teacher wearing sun-glasses?

...

Because the class is so **bright**.

News Headlines

- Hospitals Are Sued by 7 Foot Doctors
- Stolen Painting Found by Tree
- Teacher Strikes Idle Kids

Ambiguity is pervasive

- Find at least 5 meanings of this sentence:
 - ▶ I made her duck

Ambiguity is pervasive

- Find at least 5 meanings of this sentence:
 - ▶ I made her duck
- I cooked duck for her
- I cooked duck belonging to her
- I created the (artificial) duck, she owns
- I caused her to quickly lower her head or body
- I waved my magic wand and turned her into a duck

Ambiguity is pervasive

Syntactic Category

- 'Duck' can be a noun or verb
- 'her' can be a possessive ('of her') or dative ('for her') pronoun

Ambiguity is pervasive

Syntactic Category

- 'Duck' can be a noun or verb
- 'her' can be a possessive ('of her') or dative ('for her') pronoun

Word Meaning

- 'make' can mean 'create' or 'cook'

Ambiguity is pervasive

Grammar

make can be

- **Transitive:** (verb with a noun direct object)
- **Ditransitive:** (verb has 2 noun objects)
- **Action-transitive:** (verb has a direct object + verb)

Ambiguity is pervasive

Grammar

make can be

- **Transitive:** (verb with a noun direct object)
- **Ditransitive:** (verb has 2 noun objects)
- **Action-transitive:** (verb has a direct object + verb)

Phonetics

- I'm eight or duck
- I'm aid her duck

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**
- I saw the man on the hill with the telescope. **5 parses**

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**
- I saw the man on the hill with the telescope. **5 parses**
- I saw the man on the hill in Texas with the telescope. **14 parses**

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**
- I saw the man on the hill with the telescope. **5 parses**
- I saw the man on the hill in Texas with the telescope. **14 parses**
- I saw the man on the hill in Texas with the telescope at noon. **42 parses**

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**
- I saw the man on the hill with the telescope. **5 parses**
- I saw the man on the hill in Texas with the telescope. **14 parses**
- I saw the man on the hill in Texas with the telescope at noon. **42 parses**
- I saw the man on the hill in Texas with the telescope at noon on Monday.
132 parses

Why is Language Ambiguous?

- The goal in the production and comprehension of natural language is *efficient* communication.

Why is Language Ambiguous?

- The goal in the production and comprehension of natural language is *efficient* communication.
- Allowing resolvable ambiguity
 - ▶ permits shorter linguistic expressions

Why is Language Ambiguous?

- The goal in the production and comprehension of natural language is *efficient* communication.
- Allowing resolvable ambiguity
 - ▶ permits shorter linguistic expressions
 - ▶ avoids language being overly complex

Why is Language Ambiguous?

- The goal in the production and comprehension of natural language is *efficient* communication.
- Allowing resolvable ambiguity
 - ▶ permits shorter linguistic expressions
 - ▶ avoids language being overly complex
- Language relies on people's ability to use their knowledge and inference abilities to properly resolve ambiguities

Natural Languages vs. Computer Languages

- Ambiguity is the primary difference between natural and computer languages.

Natural Languages vs. Computer Languages

- Ambiguity is the primary difference between natural and computer languages.
- Formal programming languages are designed to be unambiguous

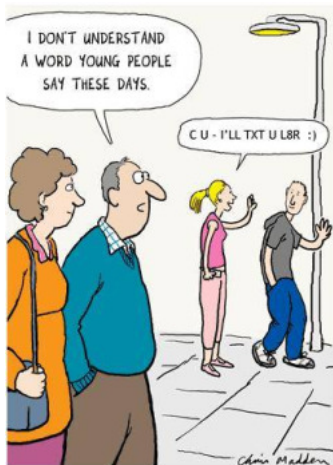
Natural Languages vs. Computer Languages

- Ambiguity is the primary difference between natural and computer languages.
- Formal programming languages are designed to be unambiguous
 - ▶ *Formal programming languages can be defined by a grammar that produces a unique parse for each sentence in the language.*

Natural Languages vs. Computer Languages

- Ambiguity is the primary difference between natural and computer languages.
- Formal programming languages are designed to be unambiguous
 - ▶ *Formal programming languages can be defined by a grammar that produces a unique parse for each sentence in the language.*
- Programming languages are also designed for efficient (deterministic) parsing.

Why else is NLP hard?



Why else is NLP hard?

Why else is NLP hard?

Non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either

Why else is NLP hard?

Non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either

Segmentation Issues

the New York-New Haven Railroad

Why else is NLP hard?

Non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either

Segmentation Issues

the New York-New Haven Railroad

the [New] [York-New] [Haven] [Railroad]

Why else is NLP hard?

Non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either

Segmentation Issues

the New York-New Haven Railroad

the [New] [York-New] [Haven] [Railroad]

the [New York]-[New Haven] [Railroad]

Why else is NLP hard?

Idioms

- dark horse
- Ball in your court
- Burn the midnight oil

Why else is NLP hard?

Idioms

- dark horse
- Ball in your court
- Burn the midnight oil

neologisms

- unfriend
- retweet
- Google/Skype/photoshop

Why is NLP hard?

New Senses of a word

- That's *sick* dude!
- Giants

Why is NLP hard?

New Senses of a word

- That's *sick* dude!
- Giants ... *multinationals, conglomerates, manufacturers*

Why is NLP hard?

New Senses of a word

- That's *sick* dude!
- Giants ... *multinationals, conglomerates, manufacturers*

Tricky Entity Names

- Where is *A Bug's Life* playing ...
- *Let It Be* was recorded ...

What we do in NLP?

Tools Required

- Knowledge about language
- Knowledge about the world
- A way to combine knowledge resources

What we do in NLP?

Tools Required

- Knowledge about language
- Knowledge about the world
- A way to combine knowledge resources

How is it generally done?

- Probabilistic models built from language data
 - ▶ $P(\text{"maison"} \rightarrow \text{"house"})$ is high

What we do in NLP?

Tools Required

- Knowledge about language
- Knowledge about the world
- A way to combine knowledge resources

How is it generally done?

- Probabilistic models built from language data
 - ▶ $P(\text{"maison"} \rightarrow \text{"house"})$ is high
 - ▶ $P(\text{I saw a van}) > P(\text{eyes awe of an})$
- Extracting rough text features does half the job.