

Language & Technology

Lecture 1: What is Language Technology?

Alëna Aksënova & Aniello De Santo

Stony Brook University
alena.aksenova@stonybrook.edu
aniello.desanto@stonybrook.edu

Why Language Technology?

Because language is important!

- ▶ **communication system**
transfer message, convey information
- ▶ **cultural artefact**
signal group membership
(dialects, sociolects)
- ▶ **uniquely human cognitive ability**
language in the brain

Why Language Technology?

Because language is important!

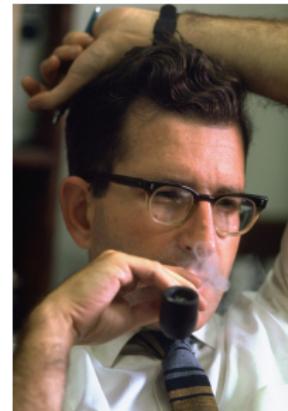
- ▶ communication system
transfer message, convey information
- ▶ cultural artefact
signal group membership
(dialects, sociolects)
- ▶ uniquely human cognitive ability
language in the brain



Why Language Technology?

Because language is important!

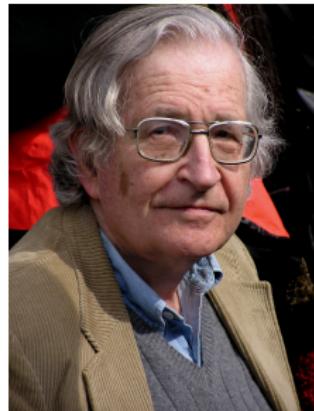
- ▶ communication system
transfer message, convey information
- ▶ cultural artefact
signal group membership
(dialects, sociolects)
- ▶ uniquely human cognitive ability
language in the brain



Why Language Technology?

Because language is important!

- ▶ communication system
transfer message, convey information
- ▶ cultural artefact
signal group membership
(dialects, sociolects)
- ▶ uniquely human cognitive ability
language in the brain



So What is the Right Notion of Language?

Communication? Culture? Cognition?

- ▶ **Linguists** care about cognition and rules.
 - ▶ How is language computed in the brain?
 - ▶ How is it acquired?
 - ▶ What grammatical rules do we have in our head?
- ▶ **Non-linguists** care about communication and culture.
 - ▶ What is that person trying to tell me?
 - ▶ Is this text clearly written?
 - ▶ What is the original meaning of this proper name?
 - ▶ How does that person talk?
grammar nazis, complaints about youth slang, AAVE

Language for Language Technology

- ▶ Most users of language technology are not linguists.
Their expectations about proper language use must be met.
- ▶ Linguistic insights are indispensable for language technology.
You cannot emulate what you do not understand!
- ▶ Therefore all facets of language matter.
But sometimes, some facets matter more than others.

Language Technology in Daily Life

- ▶ search engines
Google, DuckDuckGo, Altavista
- ▶ voice recognition
Siri, Dragon Naturally Speaking, Youtube Captions
- ▶ optical character recognition (OCR) software
Adobe Acrobat, OmniPage, Tesseract
- ▶ text-to-speech (TTS) software
Siri, Natural Reader, eSpeak
- ▶ machine translation
Google translate, Babelfish, Moses
- ▶ dialog systems/chatbots
online help desk, Cleverbot

How Well Does it Work?

Let's check the output of multiple Google translate steps:

English ⇒ Chinese ⇒ Macedonian ⇒ French ⇒ ⋯ ⇒ English



A Few Caveats

- ▶ Translating a translation is always sub-par
real world example: German dubs for anime based on French dubs
- ▶ Back-translation never yields the original,
even when done by very skilled humans.
- ▶ Literary texts are hard to translate.
- ▶ This might just be a hoax — but if so, it's a good one.

Evaluation

Evaluation

- ▶ meaning of song largely lost, but individual lines mostly fine
- ▶ odd phrasing
 - the wind is howling storm
- ▶ wrong lexical choices
 - you cannot do [put?] it back in
- ▶ incomplete sentences
 - let us very angry
- ▶ clear grammar mistakes
 - the fear is that once guided me; it runs perfect woman
- ▶ curiously absent: subject-verb agreement mistakes

Summary

surprisingly good, but still tons of mistakes for 225 word text

Another Example

If one examines precapitalist deappropriation, one is faced with a choice: either accept the semantic paradigm of discourse or conclude that the raison d'etre of the artist is deconstruction. Thus, postdialectic theory implies that reality may be used to entrench class divisions. If the semantic paradigm of discourse holds, we have to choose between deconstructive subtextual theory and Lacanist obscurity.

Another Example

*If one examines precapitalist deappropriation, one is faced with a choice: either accept the semantic paradigm of discourse or conclude that the *raison d'être* of the artist is deconstruction. Thus, postdialectic theory implies that reality may be used to entrench class divisions. If the semantic paradigm of discourse holds, we have to choose between deconstructive subtextual theory and Lacanist obscurity.*

Therefore, in Charmed, Spelling examines the cultural paradigm of narrative; in Beverly Hills 90210, although, he analyses the semantic paradigm of discourse. Sartre uses the term 'the cultural paradigm of narrative' to denote not narrative, but neonarrative.

Another Example

*If one examines precapitalist deappropriation, one is faced with a choice: either accept the semantic paradigm of discourse or conclude that the *raison d'être* of the artist is deconstruction. Thus, postdialectic theory implies that reality may be used to entrench class divisions. If the semantic paradigm of discourse holds, we have to choose between deconstructive subtextual theory and Lacanist obscurity.*

Therefore, in Charmed, Spelling examines the cultural paradigm of narrative; in Beverly Hills 90210, although, he analyses the semantic paradigm of discourse. Sartre uses the term 'the cultural paradigm of narrative' to denote not narrative, but neonarrative.

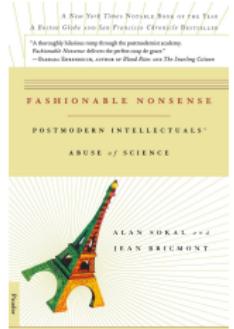
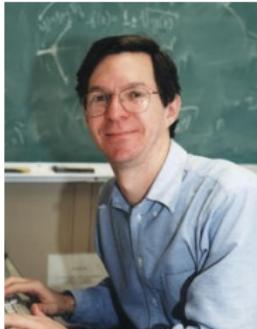
Author: The Postmodernism Generator

<http://www.elsewhere.org/pomo/>

What the \$#%@% was That?

Motivation

- ▶ mock deliberate obscurity of postmodern intellectuals
- ▶ inspired by **Sokal hoax**



- ▶ no grammar mistakes
- ▶ much longer sentences, more complex
- ▶ sentences seem connected (though they aren't)
- ▶ total gibberish (which is kind of the point)
- ▶ behind the scenes: old, simplistic technology (basically Mad Libs on steroids)

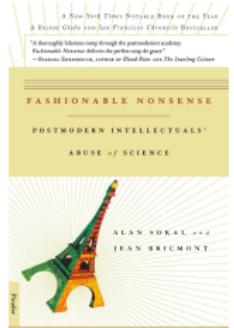
Summary

Looks impressive, but actually very simplistic and limited.

What the \$#%@% was That?

Motivation

- ▶ mock deliberate obscurity of postmodern intellectuals
- ▶ inspired by **Sokal hoax**



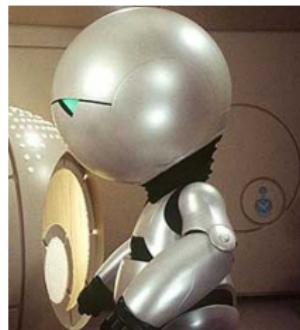
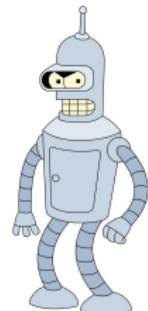
- ▶ no grammar mistakes
- ▶ much longer sentences, more complex
- ▶ sentences seem connected (though they aren't)
- ▶ total gibberish (which is kind of the point)
- ▶ behind the scenes: old, simplistic technology
(basically Mad Libs on steroids)

Summary

Looks impressive, but actually very simplistic and limited.

Language Technology in Fiction

Fiction is full of talking robots, androids, AIs.



Robots' Narrow Range of Language Competence



Robots' Narrow Range of Language Competence



Robots' Narrow Range of Language Competence



completely human



Robots' Narrow Range of Language Competence



completely human



Robots' Narrow Range of Language Competence



completely human



perfect but weird voice



Robots' Narrow Range of Language Competence



completely human



perfect but weird voice



Robots' Narrow Range of Language Competence



completely human



perfect but weird voice



Arnold

Moral of the Story

- ▶ Even the most primitive sci-fi killer robots have greater linguistic competence than present-day technology.
- ▶ Only a few quirks:
 - ▶ robotic voice,
 - ▶ no contractions,
 - ▶ stilted and overly elaborate style,
 - ▶ literal interpretation of metaphors
(if it makes for a funny scene).
- ▶ But perfect grammar and efficient, context-appropriate communication.

Why We Don't Have Talking Robots Yet

- ▶ Humans cannot fathom what might be hard about language because we learn it naturally, like walking.
- ▶ But both are incredibly complex and difficult.
(we are still excited by robots than can walk up stairs)
- ▶ Using language involves several skills, none of them are trivial.

1.	segmenting sound waves into sounds	doable
2.	combining sounds into words	doable
3.	analyzing how words connect to each other in the sentence	hard
4.	inferring the meaning of sentences	hard
5.	determining the contribution of a sentence to the current discourse	very hard
6.	translating the discourse into instructions and information about the real world	nigh impossible

Why We Don't Have Talking Robots Yet

- ▶ Humans cannot fathom what might be hard about language because we learn it naturally, like walking.
- ▶ But both are incredibly complex and difficult.
(we are still excited by robots than can walk up stairs)
- ▶ Using language involves several skills, none of them are trivial.

- | | | |
|----|-------------------------------------------------------------------------------------|------------------------|
| 1. | segmenting sound waves into sounds | doable |
| 2. | combining sounds into words | doable |
| 3. | analyzing how words connect to each other
in the sentence | hard |
| 4. | inferring the meaning of sentences | hard |
| 5. | determining the contribution of a sentence
to the current discourse | very hard |
| 6. | translating the discourse into instructions
and information about the real world | nigh impossible |

Pitfalls of Language: World Knowledge

- | | | | |
|-----|----|----------------------------------|----------|
| (1) | a. | Bill and Mary have many friends. | plural |
| | b. | Bill has many friends. | singular |
| (2) | a. | Bill and Mary met. | plural |
| | b. | ?? Bill met. | singular |

Hypothesis

In sentences of the form *subject met*, the subject must be plural.

- | | | | |
|-----|----|-------------------------------------------------------------|----------|
| (3) | a. | The newly assembled board has five members. | singular |
| | b. | The newly assembled board met for the first time yesterday. | |

Revised Hypothesis

In sentences of the form *subject met*, the subject must refer to multiple individuals/objects.

Pitfalls of Language: World Knowledge

- | | | | |
|-----|----|----------------------------------|----------|
| (1) | a. | Bill and Mary have many friends. | plural |
| | b. | Bill has many friends. | singular |
| (2) | a. | Bill and Mary met. | plural |
| | b. | ?? Bill met. | singular |

Hypothesis

In sentences of the form *subject met*, the subject must be plural.

- | | | | |
|-----|----|-------------------------------------------------------------|----------|
| (3) | a. | The newly assembled board has five members. | singular |
| | b. | The newly assembled board met for the first time yesterday. | |

Revised Hypothesis

In sentences of the form *subject met*, the subject must refer to multiple individuals/objects.

Pitfalls of Language: World Knowledge

- | | | | |
|-----|----|----------------------------------|----------|
| (1) | a. | Bill and Mary have many friends. | plural |
| | b. | Bill has many friends. | singular |
| (2) | a. | Bill and Mary met. | plural |
| | b. | ?? Bill met. | singular |

Hypothesis

In sentences of the form *subject met*, the subject must be plural.

- | | | | |
|-----|----|-------------------------------------------------------------|----------|
| (3) | a. | The newly assembled board has five members. | singular |
| | b. | The newly assembled board met for the first time yesterday. | |

Revised Hypothesis

In sentences of the form *subject met*, the subject must refer to multiple individuals/objects.

Pitfalls of Language: World Knowledge

- | | | | |
|-----|----|----------------------------------|----------|
| (1) | a. | Bill and Mary have many friends. | plural |
| | b. | Bill has many friends. | singular |
| (2) | a. | Bill and Mary met. | plural |
| | b. | ?? Bill met. | singular |

Hypothesis

In sentences of the form *subject met*, the subject must be plural.

- | | | | |
|-----|----|-------------------------------------------------------------|----------|
| (3) | a. | The newly assembled board has five members. | singular |
| | b. | The newly assembled board met for the first time yesterday. | |

Revised Hypothesis

In sentences of the form *subject met*, the subject must refer to multiple individuals/objects.

Pitfalls of Language: World Knowledge [cont.]

- (4) a. The board has five members and met for the first time yesterday.
- b. ?? The board has four wheels and met for the first time yesterday.

Need for World Knowledge

A board with four wheels is a skateboard, not a committee
⇒ *board* does not refer to multiple individuals here

- (5) The board met yesterday even though it only has one member right now.

Not All World Knowledge Matters

The committee has only one member and *board* thus does not refer to multiple individuals. But the sentence is still fine.

Pitfalls of Language: World Knowledge [cont.]

- (4) a. The board has five members and met for the first time yesterday.
- b. ?? The board has four wheels and met for the first time yesterday.

Need for World Knowledge

A board with four wheels is a skateboard, not a committee
⇒ *board* does not refer to multiple individuals here

- (5) The board met yesterday even though it only has one member right now.

Not All World Knowledge Matters

The committee has only one member and *board* thus does not refer to multiple individuals. But the sentence is still fine.

Pitfalls of Language: World Knowledge [cont.]

- (4) a. The board has five members and met for the first time yesterday.
- b. ?? The board has four wheels and met for the first time yesterday.

Need for World Knowledge

A board with four wheels is a skateboard, not a committee
⇒ *board* does not refer to multiple individuals here

- (5) The board met yesterday even though it only has one member right now.

Not All World Knowledge Matters

The committee has only one member and *board* thus does not refer to multiple individuals. But the sentence is still fine.

Pitfalls of Language: World Knowledge [cont.]

- (4) a. The board has five members and met for the first time yesterday.
- b. ?? The board has four wheels and met for the first time yesterday.

Need for World Knowledge

A board with four wheels is a skateboard, not a committee
⇒ *board* does not refer to multiple individuals here

- (5) The board met yesterday even though it only has one member right now.

Not All World Knowledge Matters

The committee has only one member and *board* thus does not refer to multiple individuals. But the sentence is still fine.

Pitfalls of Language: Form and Meaning I

(6) I know two students handed in every homework.

(7) I know two students who handed in every homework.

The Puzzle

- ▶ Why can a sentence have two interpretations?
- ▶ Why is one interpretation preferred?
- ▶ Why does the word *who* forbid the second interpretation?

Pitfalls of Language: Form and Meaning I

(6) I know two students handed in every homework.

Meaning 1: Two specific students handed in every homework.

Meaning 2: Every homework was handed in by at least two students.

(7) I know two students who handed in every homework.

The Puzzle

- ▶ Why can a sentence have two interpretations?
- ▶ Why is one interpretation preferred?
- ▶ Why does the word *who* forbid the second interpretation?

Pitfalls of Language: Form and Meaning I

(6) I know two students handed in every homework.

Meaning 1: Two specific students handed in every homework.

Meaning 2: Every homework was handed in by at least two students.

(7) I know two students who handed in every homework.

Meaning 1: still the same

Meaning 2: no longer available

The Puzzle

- ▶ Why can a sentence have two interpretations?
- ▶ Why is one interpretation preferred?
- ▶ Why does the word *who* forbid the second interpretation?

Pitfalls of Language: Form and Meaning I

(6) I know two students handed in every homework.

Meaning 1: Two specific students handed in every homework.

Meaning 2: Every homework was handed in by at least two students.

(7) I know two students who handed in every homework.

Meaning 1: still the same

Meaning 2: no longer available

The Puzzle

- ▶ Why can a sentence have two interpretations?
- ▶ Why is one interpretation preferred?
- ▶ Why does the word *who* forbid the second interpretation?

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself/himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself/himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself/himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself/himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him/himself.

Pitfalls of Language: Form and Meaning II

The interpretation of pronouns is also very tricky.

- (8) a. Hugo likes himself.
 b. Hugo likes him.
 c. Hugo knows that Bill likes him/himself.
- (9) a. Hugo introduced himself to Bill.
 b. Hugo introduced Bill to himself/himself.
- (10) Hugo looked at his contemporaries — less clever than
 him/himself — and saw them outstrip him.

Pitfalls of Language: Form and Meaning III

Even contractions can have an effect on meaning.

(11) Who do you want to leave?

(12) Who do you wanna leave?

Pitfalls of Language: Form and Meaning III

Even contractions can have an effect on meaning.

- (11) Who do you want to leave?

Answer 1: I want to leave Bill.

Answer 2: I want Bill to leave.

- (12) Who do you wanna leave?

Pitfalls of Language: Form and Meaning III

Even contractions can have an effect on meaning.

- (11) Who do you want to leave?

Answer 1: I want to leave Bill.

Answer 2: I want Bill to leave.

- (12) Who do you wanna leave?

Answer 1: still possible

Answer 2: impossible

Pitfalls of Language: Form and Meaning IV

Adjectives usually modify nouns.

- (13) *retired physicist* = *physicist who is retired*

But sometimes they can only modify a subpart.

- (14) a. *nuclear physicist* ≠ *physicist who is nuclear*
b. *nuclear physicist* = *a scientist working in nuclear physics*

And sometimes they can do both.

- (15) a. *radical feminist* = *a feminist who is radical*
b. *radical feminist* = *an advocate of radical feminism*

Pitfalls of Language: Form and Meaning IV

Adjectives usually modify nouns.

- (13) retired physicist = physicist who is retired

But sometimes they can only modify a subpart.

- (14) a. nuclear physicist \neq physicist who is nuclear
b. nuclear physicist = a scientist working in nuclear physics

And sometimes they can do both.

- (15) a. radical feminist = a feminist who is radical
b. radical feminist = an advocate of radical feminism

Pitfalls of Language: Form and Meaning IV

Adjectives usually modify nouns.

- (13) retired physicist = physicist who is retired

But sometimes they can only modify a subpart.

- (14) a. nuclear physicist \neq physicist who is nuclear
b. nuclear physicist = a scientist working in nuclear physics

And sometimes they can do both.

- (15) a. radical feminist = a feminist who is radical
b. radical feminist = an advocate of radical feminism

Pitfalls of Language: Form and Meaning IV

Adjectives usually modify nouns.

- (13) retired physicist = physicist who is retired

But sometimes they can only modify a subpart.

- (14) a. nuclear physicist \neq physicist who is nuclear
b. nuclear physicist = a scientist working in nuclear physics

And sometimes they can do both.

- (15) a. radical feminist = a feminist who is radical
b. radical feminist = an advocate of radical feminism

Remark

Even here, world knowledge
can play a big role!

Pitfalls of Language: Form and Meaning IV

Adjectives usually modify nouns.

- (13) retired physicist = physicist who is retired

But sometimes they can only modify a subpart.

- (14) a. nuclear physicist \neq physicist who is nuclear
b. nuclear physicist = a scientist working in nuclear physics

And sometimes they can do both.

- (15) a. radical feminist = a feminist who is radical
b. radical feminist = an advocate of radical feminism

Remark

Even here, world knowledge
can play a big role!



Firestorm, nuclear physicist

Pitfalls of Language: Form and Meaning V

Word order also affects meaning, but not consistently.

- (16) a. The house is old and run down
= The house is run down and old
- b. an old shabby house
= a shabby old house
- (17) a. The house is small and expensive
= The house is expensive and small
- b. a small expensive house
≠ an expensive small house

Pitfalls of Language: Just Form

Even when the meaning of a phrase is clear,
it can still be ungrammatical.

- (18) a. a rusty car
 - b. a car that is rusty
- (19) a. a former president
 - b. ?? a president that is former

And the presence of a single word can make a sentence
ungrammatical, even if it fine in very similar sentences.

- (20) a. Who do you think Bill likes?
 - b. Who do you think likes Bill?
- (21) a. Who do you think that Bill likes?
 - b. ?? Who do you think that likes Bill?

Pitfalls of Language: Just Form

Even when the meaning of a phrase is clear,
it can still be ungrammatical.

- (18) a. a rusty car
 - b. a car that is rusty
- (19) a. a former president
 - b. ?? a president that is former

And the presence of a single word can make a sentence
ungrammatical, even if it fine in very similar sentences.

- (20) a. Who do you think Bill likes?
 - b. Who do you think likes Bill?
- (21) a. Who do you think that Bill likes?
 - b. ?? Who do you think that likes Bill?

Pitfalls of Language: Comprehensability

Many sentences are perfectly grammatical but hard to understand.
Garden path sentences are a prime example.

- (22) a. The horse raced past the barn fell.
b. The cotton clothing is made of grows in Mississippi.
c. Until the police arrest the drug dealers control the street.

Another case is **center embedding** being harder than
right embedding.

- (23) a. The cheese that the mouse that the cat chased ate was rotten.
b. The cheese was rotten that the mouse ate that the cat chased.

Why is this the case?

How can we prevent a machine from producing difficult sentences?

Pitfalls of Language: Comprehensability

Many sentences are perfectly grammatical but hard to understand.
Garden path sentences are a prime example.

- (22) a. The horse raced past the barn fell.
b. The cotton clothing is made of grows in Mississippi.
c. Until the police arrest the drug dealers control the street.

Another case is **center embedding** being harder than
right embedding.

- (23) a. The cheese that the mouse that the cat chased ate was rotten.
b. The cheese was rotten that the mouse ate that the cat chased.

Why is this the case?

How can we prevent a machine from producing difficult sentences?

Pitfalls of Language: Comprehensability

Many sentences are perfectly grammatical but hard to understand.
Garden path sentences are a prime example.

- (22) a. The horse raced past the barn fell.
b. The cotton clothing is made of grows in Mississippi.
c. Until the police arrest the drug dealers control the street.

Another case is **center embedding** being harder than
right embedding.

- (23) a. The cheese that the mouse that the cat chased ate was rotten.
b. The cheese was rotten that the mouse ate that the cat chased.

Why is this the case?

How can we prevent a machine from producing difficult sentences?

Pitfalls of Language: Comprehensability

Many sentences are perfectly grammatical but hard to understand.
Garden path sentences are a prime example.

- (22) a. The horse **that was** raced past the barn fell.
b. The cotton clothing is made of grows in Mississippi.
c. Until the police arrest the drug dealers control the street.

Another case is **center embedding** being harder than
right embedding.

- (23) a. **The cheese that the mouse that the cat chased ate was rotten.**
b. **The cheese was rotten that the mouse ate that the cat chased.**

Why is this the case?

How can we prevent a machine from producing difficult sentences?

Pitfalls of Language: Comprehensability

Many sentences are perfectly grammatical but hard to understand.
Garden path sentences are a prime example.

- (22) a. The horse **that was** raced past the barn fell.
b. The cotton **that** clothing is made of grows in Mississippi.
c. Until the police arrest the drug dealers control the street.

Another case is **center embedding** being harder than
right embedding.

- (23) a. **The cheese that the mouse that the cat chased ate was rotten.**
b. **The cheese was rotten that the mouse ate that the cat chased.**

Why is this the case?

How can we prevent a machine from producing difficult sentences?

Pitfalls of Language: Comprehensability

Many sentences are perfectly grammatical but hard to understand.
Garden path sentences are a prime example.

- (22) a. The horse **that was** raced past the barn fell.
b. The cotton **that** clothing is made of grows in Mississippi.
c. Until the police arrest, the drug dealers control the street.

Another case is **center embedding** being harder than
right embedding.

- (23) a. **The cheese that the mouse that the cat chased ate was rotten.**
b. **The cheese was rotten that the mouse ate that the cat chased.**

Why is this the case?

How can we prevent a machine from producing difficult sentences?

Pitfalls of Language: Comprehensability

Many sentences are perfectly grammatical but hard to understand.
Garden path sentences are a prime example.

- (22) a. The horse **that was** raced past the barn fell.
b. The cotton **that** clothing is made of grows in Mississippi.
c. Until the police arrest, the drug dealers control the street.

Another case is **center embedding** being harder than
right embedding.

- (23) a. **The cheese that the mouse that the cat chased ate was rotten.**
b. **The cheese was rotten that the mouse ate that the cat chased.**

Why is this the case?

How can we prevent a machine from producing difficult sentences?

The Human Mystery

- ▶ We all have complete mastery of these rules even though
 - ▶ we were never told about them (not even in grammar courses)
 - ▶ we have no conscious knowledge of them.
- ▶ Humans don't learn language, we acquire it naturally, without explicit instruction.
- ▶ Even a five year old is better at language than computers.

The Big Question of Linguistics

Why are humans so good at language?

What does their knowledge look like, and how is it acquired?

The Human Mystery

- ▶ We all have complete mastery of these rules even though
 - ▶ we were never told about them (not even in grammar courses)
 - ▶ we have no conscious knowledge of them.
- ▶ Humans don't learn language, we acquire it naturally, without explicit instruction.
- ▶ Even a five year old is better at language than computers.

The Big Question of Linguistics

Why are humans so good at language?

What does their knowledge look like, and how is it acquired?

Two Technological Challenges

- ▶ Native-like performance would require an incredibly complex model. lexicon, grammar, meaning, world knowledge
- ▶ This isn't feasible for any task that needs to be done fast and efficiently:
 - ▶ spell checking 1000 page document in few seconds
 - ▶ answering thousands of online translation requests per second
 - ▶ tracking millions of interconnected tweets
 - ▶ speech technology on a slow phone with limited battery
- ▶ Quality comes at the cost of resources.

The Dual Challenge of Language Technology

- ▶ Industrial applications: simple models that work well enough
Natural Language Processing (NLP)
- ▶ Native-like performance: complex, resource-demanding models
Computational Linguistics

Two Technological Challenges

- ▶ Native-like performance would require an incredibly complex model. lexicon, grammar, meaning, world knowledge
- ▶ This isn't feasible for any task that needs to be done fast and efficiently:
 - ▶ spell checking 1000 page document in few seconds
 - ▶ answering thousands of online translation requests per second
 - ▶ tracking millions of interconnected tweets
 - ▶ speech technology on a slow phone with limited battery
- ▶ Quality comes at the cost of resources.

The Dual Challenge of Language Technology

- ▶ Industrial applications: simple models that work well enough
Natural Language Processing (NLP)
- ▶ Native-like performance: complex, resource-demanding models
Computational Linguistics

- ▶ **Twilight of Simple Models and Shallow Methods**

We will still see improvements in real-world applications for the next few years, but overall quality will plateau due to diminishing returns.

- ▶ **Rise of Deep Methods**

There will be a surge in interest in more complex models because they can handle much more complex tasks.

- ▶ **Linguistics at the Forefront**

In order to make deep methods efficient, we need a very good understanding of the object being modeled. For this, **linguistic know-how will be indispensable**.