

Language

A brief overview

Pavel Logačev

What is Language?

- For our purposes: Languages are systems of arbitrary symbols and grammatical rules for the purpose of communicating ideas.
 - Arbitrary symbols
 - Rule-based system for combining those symbols
- ‘Language’ is the human capacity for acquiring and using one or several languages.

Arbitrariness in Language

Writing systems are arbitrary, but not really part of the language as such:

- hotel [English, Latin script]
- отель [Russian, Cyrillic script]
- होटल [Hindi, Devanagari]
- 飯店 [Mandarin, Traditional Chinese]

Arbitrariness in Language

Word-form to meaning mapping is arbitrary:

Turkish	Russian
saray 'palace'	сарай 'barn'
English	German
gift 'gift, present'	Gift 'poison'



Many more at <http://edl.ecml.at/LanguageFun/Sameworddifferentmeaning>

Arbitrariness in Language

Plural marking is arbitrary:

English

chin-**s**
(suffix)

Piro (Peru)

ka-zigi
(prefix)

German

cene(-ler)
(optional suffix)

Word order is (fairly) arbitrary:

English

around the car
(preposition)

Turkish

arabanın etrafında
(postposition)

German

um das Auto herum
(circumposition)

Linguistic Competence I

Which three of these are (fairly old) Turkish words?

talö

tark

līs

ğül

tal

rkat

līs

thal

Linguistic Competence I

Invalid Nonwords

*thal

*lɪs

*göl

*rkat

*talö

Valid Nonwords

tal

lis

tark

Linguistic Competence II

Which of these sentences are “*proper English*”?

- (1) The bird likes John.
- (2) Bird like John.
- (3) The blind woman saw the lake.
- (4) John hates with sunglasses.
- (5) Colorless green ideas sleep furiously.
- (6) Furiously sleep ideas green colorless.
- (7) Who did you see Peter and?

Linguistic Competence II

Which of these sentences are “*proper English*”?

- (1) The bird likes John.
- (2) *Bird like John.
- (3) #The blind woman saw the lake.
- (4) *John hates with sunglasses.
- (5) Colorless green ideas sleep furiously.
- (6) *Furiously sleep ideas green colorless.
- (7) Who did you see Peter and $_i$?

→ Although Language clearly serves to communicate meaning, some of its rules are purely about form.

Linguistic Competence

Someone who knows a language, knows:

- The sound inventory of that language [**phonetics**]
→ e.g., **/th/* in Turkish
- The sound structure of that language [**phonology**]
→ e.g., **/^rk/* in Turkish
- The structure of the words of that language [**morphology**]
→ e.g., **ler-futbolcu* in Turkish
- The structure of the sentences of that language [**syntax**]
- The meaning of previously unencountered sentences in that language [**semantics**]

Competence vs. Performance

Which of these sentences are “*proper English*”?

- (8) The horse raced past the barn fell.
- (9) Since Jay always jogs a mile and a half seems like no distance to him.
- (10) The key to the cabinets were unfortunately lost.
- (11) The mouse that the cat that the dog saw chased ate the cheese.
- (12) More people have been to Russia than I’ve been.

Competence vs. Performance

(8) The horse raced past the barn fell.

(9) Since Jay always jogs a mile and a half seems like no distance to him.

(10) *The key to the cabinets were unfortunately lost.

(11) The mouse that the cat that the dog saw chased ate the cheese.

(12) More people have been to Russia than I've been.

Competence: The implicit knowledge of a language's grammatical constraints.

→ **(Theoretical) Linguistics**

Performance: The set of processes that apply that knowledge in real-time.

→ **Psycholinguistics**

Competence vs. Performance

- Performance limitations may render a grammatical sentence unacceptable, and may also lead to grammaticality illusions.
- Performance may also lead to this, while competence helps us determine what the problem



nice girl
@christapeterso

Why do they have a sign saying that Schiff is currently following the rules



Manu Raju @mkraju · Nov 21

GOP adds new sign during the break

[Show this thread](#)



Phrase Structure

- So how we can represent those mysterious rules?
- The simplest way is in the form of **phrase structure rules**.

English

$S \rightarrow NP VP$	$VP \rightarrow Vi$	$VP \rightarrow Vt NP$	$NP \rightarrow Det N$
$Det \rightarrow the$	$N \rightarrow barber$	$Vi \rightarrow sleeps$	$Vt \rightarrow kissed$

Turkish

... $VP \rightarrow NP Vt$... $NP \rightarrow N$...

Phrase Structure

- **Simple phrase structure rules** miss many important generalizations:
 - Subject-verb agreement in number
 - Determiner-noun agreement in number and gender
- Solution: complex symbol categories
$$[S] \rightarrow [NP, \text{num: } \theta] [VP, \text{num: } \theta]$$
$$[NP] \rightarrow [Det, \text{num: } \theta, \text{gen: } \beta] [NP, \text{num: } \theta, \text{gen: } \beta]$$
- Such representations are used in contemporary syntactic frameworks, such as HPSG and LFG.

Some Key Properties of Language

- Finite number of symbols with an arbitrary relationship between form and meaning
- Finite number of rules
- Potential for creativity
- Displacement property

Creativity and Recursion I

- We acquire language in finite time, and yet each language has a potentially infinite number of sentences.

For example:

Anne laughed.

John mentioned that Anne laughed.

Peter said that John mentioned that Anne laughed.

Peter said that ... John mentioned that Anne laughed.

Creativity and Recursion I

Another example:

Anne laughed.

Anne laughed on Sunday.

Anne laughed on Sunday in the park.

Anne laughed on Sunday in the park by the bridge.

Anne laughed on Sunday in the park by the bridge over the lake

Creativity and Recursion I

Recursion: A recursive definition uses a category in parts of the definition of that category. (Produces ‘nesting’.)

Example:

[Peter said that [... John mentioned that [Anne laughed]]].

Phrase-structure rules:

$S \rightarrow NP Vi$	$NP \rightarrow Peter$	$Vt \rightarrow said$
$S \rightarrow NP Vt S'$	$NP \rightarrow John$	$Vt \rightarrow mentioned$
$S' \rightarrow that S$	$NP \rightarrow Anne$	$Vi \rightarrow said$

Creativity and Recursion II

- Some languages, like German, allow for recursion at the word-level:

Donau

Donauschiffahrt

Donauschiffahrtskapitän

Donauschiffahrtskapitänswitwe

Donauschiffahrtskapitänswitwenrente

Donauschiffahrtskapitänswitwenrentenfondsverwalter...

Creativity and Recursion II



(floor grinding machine rentals)

Language and Communication in Animals

- Language:
 - auditory (speech)
 - visual (sign language)
 - Animal communication:
 - auditory (e.g., monkeys, dogs)
 - visual (e.g., dogs, birds)
 - olfactory (e.g., pheromones in ants)
 - electric (e.g., electric eels)
 - seismic (e.g., frogs)
- No **displacement** property

Animals and Language

- Are humans really unique in our ability to use Language?
- Key properties: displacement and recursion (as a proxy for syntax)

Animals and Language

- Chimps and gorillas
 - use tools; sociable, able to learn and use symbols
 - are capable of deception
 - appear to have a concept of past and future
 - [primates may even have a vocal tract suitable for speech](#)
- Primates were taught American Sign Language:
 - Washoe: “First Signs of Washoe” [\[pt. 1\]](#) [\[pt. 2\]](#)
 - Nim Chimpsky, and many others

Animals and Language

Washoe (chimp)

- raised around humans
- taught ASL early
- vocabulary tests
- evidence of generalization
- used signs appropriately, and formed new expressions (e.g., “METAL CUP DRINK” for thermos)
- status of grammar unclear
- displacement property unclear

1965	Washoe is born in the wild
1966, June	Is brought to Nevada and begins training
1966, December	Has acquired her first 4 signs
1967, April	Signs her first combinations
1967, July	Has acquired her first 13 signs
1968, April	Has acquired her first 34 signs
1969, June	Has acquired 85 signs; end of first 3 years of training
1970	Is sent to the Institute for Primate Studies in Norman, Oklahoma
1975	Is reported to have 160 signs
2007	Is reported to know ~350 signs

Animals and Language

Koko (gorilla)

- environment similar to Washoe
- ~600 signs
- compounds like “elephant doll” for Pinocchio, “eye hat” for mask
- lied on occasion
- [\[signing video\]](#)
- documentary [\[youtube\]](#)

Me	‘What did you do to Penny?’
Koko	“Bite.”
Me	‘You admit it?’
Koko	‘Sorry bite scratch.’
At this point I showed Koko the mark on my hand	
Koko	‘Wrong bite.’
Me	“Why bite?”
Koko	“Because mad.’
Me	“Why mad?”
Koko	‘Don’t know.’

Animals and Language

- Data from early experiments very impressionistic, and by today's standards somewhat sparse.
- Later experiments used “lexigram keyboards” to avoid ambiguity.
- So far, evidence of displacement and recursion is sparse.
- There is evidence for what *might* be called syntax, though (fixed word order).

Linguistic Diversity

How many languages are there?

- That depends on what constitutes “a language”.
- “Mutual intelligibility” is a key criterion for being the dialects of the same language.

Linguistic Diversity

Some problematic examples:

- Hindi & Urdu (formerly: Hindustani)
- Mandarin & Cantonese
- Serbian & Croatian (formerly: Serbo-Croatian)

*"A language is a dialect with an army and a navy."
(Max Weinreich)*

Linguistic Diversity

- Most sources estimate the number of currently spoken languages at approximately 5000-7000.
- The [Langscape](#) database maps ~6400 languages worldwide. (@University of Maryland)
- The [World Atlas of Language Structures](#) has information on ~2679 languages. (@MPI for Evol. Anthropol.)

Primitive Languages?

- On occasion, some languages have been called “primitive”, by some. What could be the reasons?

Primitive Languages?

- On occasion, some languages have been called “primitive”, by some. What could be the reasons?
 - The fact that the language sounds different?
 - The fact that the language “lacks words”?
(e.g., colors, numerals)
 - Insufficient knowledge of the language?
- All known languages can express almost(?) any idea, but they may use different ways to do it.

Example: Past Tense

English

(1a/b) The water is/was cold.

Indonesian [\[WALS\]](#)

(2) *Air itu dingin.*

Water that cold

'The water is/was cold.'

Yagua [\[WALS, cf. table 1\]](#)

Ngiti (Congo)

(3a) *ma màkpěná*

'I whistled' (recent past)

(3b) *ma mákpěná*

'I whistled' (intermediate past)

(3c) *ma makpéna*

'I will whistle' (near future)

(3d) *ma makpénà*

'I used to whistle' (past habitual)

- Can some languages express more than others?

Language and Thought

- Do we think in Language? Is “thinking” just talking to yourself?

Language and Thought

- Do we think in Language? Is “thinking” just talking to yourself?
 - Watson (1924): “[...] Yes ... *‘thinking’ is largely sub-vocal talking.*”
 - Smith et al. (1947): Injected participant with paralytic *curare* - paralysis of pharyngeal muscles.
 - Participant could perceive, remember, and answer questions with not-yet-paralyzed muscles.
- ... so maybe you don't need muscles for talking to yourself ...

Language and Thought

- “Brother John”: Prominent case of an epileptic whose episodes prevented him from speaking or writing coherently, but could perform various non-linguistic tasks, including travel. (e.g. Traxler, 2011)
- Patients with Williams syndrome suffer from mental retardation, but no apparent linguistic limitations.
 - Double dissociation: Thought is not language.

Linguistic Determinism

- Sapir-Whorf Hypothesis: *Our perception and way of thinking is determined by the language we speak.*
- Partially based on Boas' (1911) observation of 4 words for 'snow' in an Eskimo language. (No, not 100).
- Implies that Eskimo-speakers 'carve up the world' differently from others. - Better snow-perception?

Linguistic Determinism

- Problem 1: There are only two words for snow in those languages (roughly: 'snow' and 'snowflake'); other words are derived or inflected forms.
- Problem 2: The argument is circular.
- Problem 3: Speakers of Pirahã can match sets with the same number of objects (e.g., 3 spoons and 3 forks), in spite of having no numerals (only quantities such as 'few', 'many').

Linguistic Relativism

- Revised Hypothesis: *Our perception and way of thinking is influenced by the language we speak, a little bit.*
- Winawer et al. (2007): Color-matching task with different shades of blue (diff. colors in Russian).
 - Russian-speakers were faster at matching two instances of the same class.
 - No effect for English speakers.



Goluboy



Siniy