Evolution of Language'14 UvA/FGW BA course

L3 - The study of language and its complexity

All human beings are born free and equal in dignity and rights

Intersection

Coordination

Coordination

Predication

Modification

Quantification

Language (1)

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

Alle mensen worden vrij en gelijk in waardigheid en rechten geboren. Zij zijn begiftigd met verstand en geweten, en behoren zich jegens elkander in een geest van broederschap te gedragen.

Alle Menschen sind frei und gleich an Würde und Rechten geboren. Sie sind mit Vernunft und Gewissen begabt und sollen einander im Geiste der Brüderlichkeit begegnen.

Tous les êtres humains naissent libres et égaux en dignité et en droits. Ils sont doués de raison et de conscience et doivent agir les uns envers les autres dans un esprit de fraternité.

Evri man en mere olketa born frii en ikwol lo digniti en raits blo olketa. Olketa evriwan olketa garem maeni fo tingting en olketa sapos fo treatim isada wittim spirit blo bradahood.

Solomons Pidgin

Language relations (1)

- The universal declaration of human rights example is deliberately complex
- There are more systematic ways of looking at differences
  - Core vocabulary
    - Everyday words that tend to be stable over time
  - Systematic correspondences
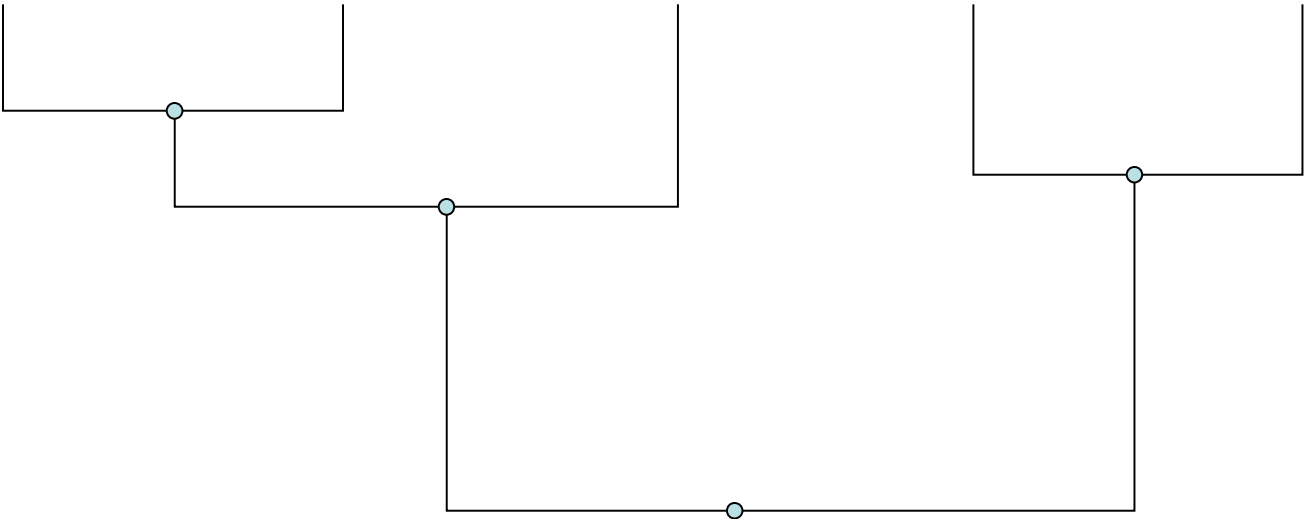   - Sounds that are reliably different between languages

Related languages

vaːdər   faːtər   faːðə   padre   paðɾe

patər

Language relations (2)

- Languages have similarities
- Four possible reasons for relatedness
  - Descend from the same ancestral language
  - They are influenced by each other
  - There are functional reasons
  - There are cognitive reasons

- English is atypical: descends from Germanic languages, but heavily influenced by French

What is a language?

- "a language is a dialect with an army and a navy"
  - Max Weinreich (?), Louis-Hubert Lyautey (?)

- Different languages are mutually unintelligible

  - Dialects of the same language are mutually intelligible
  - But how do we define "intelligible"?

How many languages are there?

- ~6000
  - Many of which only have very few speakers
    - One-over-*f* or Zipf distribution
  - And many of which are disappearing
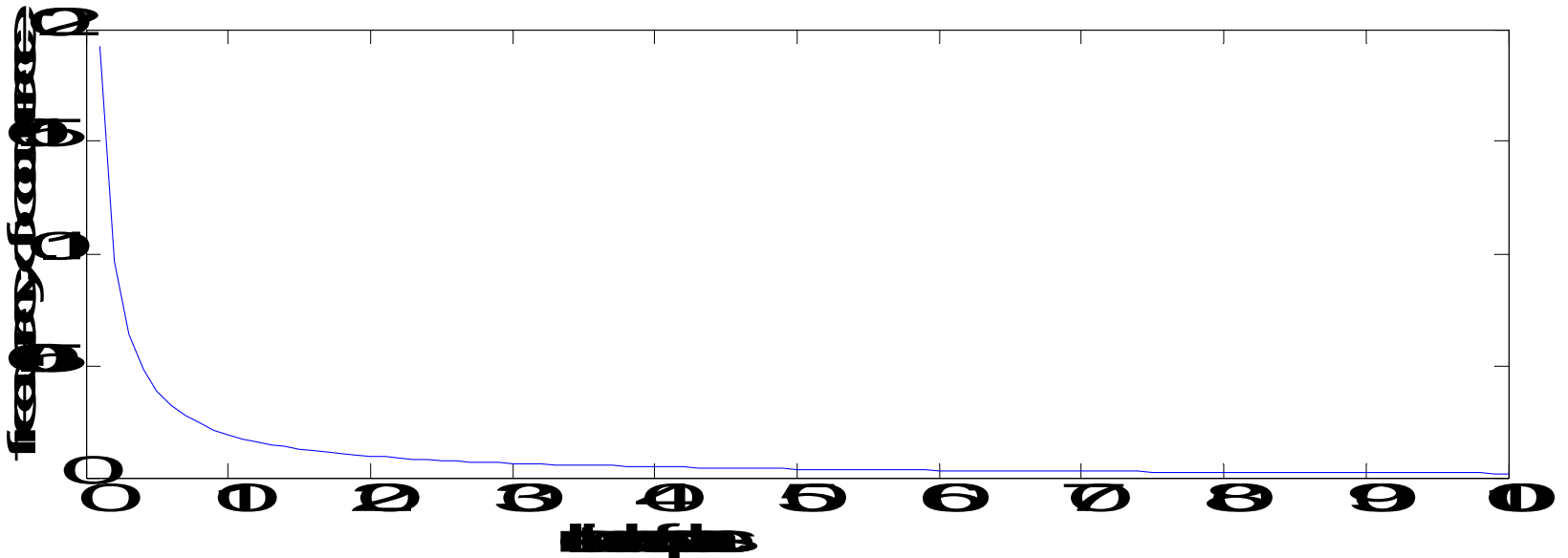
Language families

- Languages can be grouped along historical relations
  - But not all end up in one group
- If there was 1 proto-language the information is lost
  - And it is unlikely there ever was only one language
- There are many language families
  - >100 families
  - ~90 isolated languages

Cavalli-Sforza, 2001
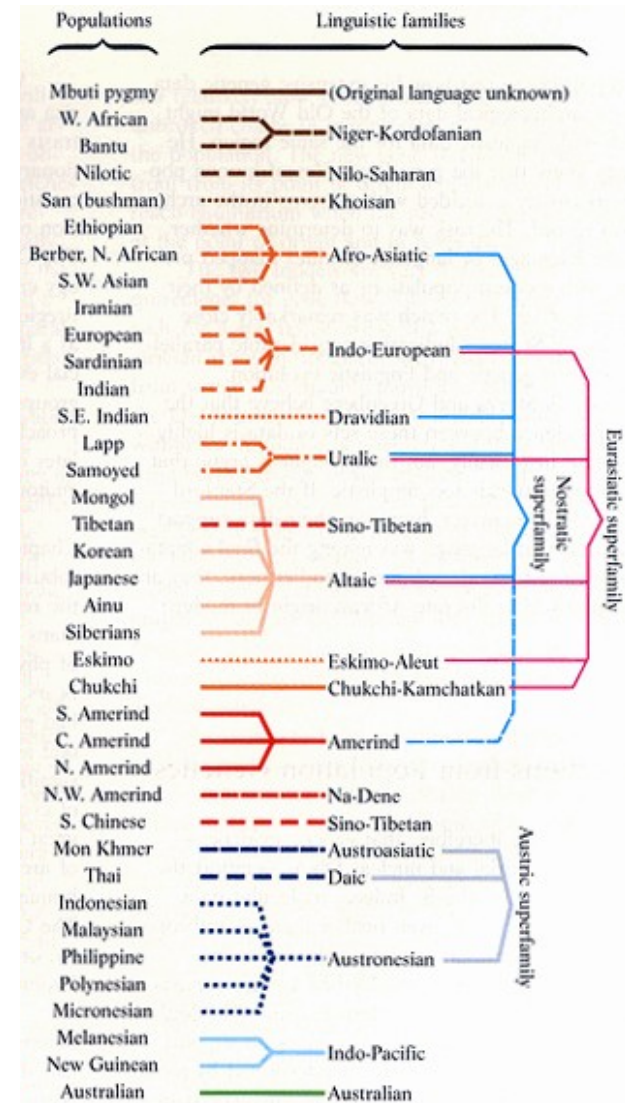
Linguistic diversity

- In order to understand language universals, we must understand linguistic diversity
  - We must understand what are the constants in the diversity
  - This was first articulated by Greenberg ca. 1963

UNIVERSALS OF LANGUAGE

SECOND EDITION

EDITED BY
JOSEPH H. GREENBERG

AII MIT PRESS CLASSIC

Language (2)

Điều 1: Tất cả mọi người sinh ra đều được tự do và bình đẳng về nhân phẩm và quyền. Mọi con người đều được tạo hoá ban cho lý trí và lương tâm và cần phải đối xử với nhau trong tình bằng hữu.

Vietnamese

Adesahi tsuo ɔ, a bɔ mɛ nɛ nɔ fɛɛ nɔ e ye e he, nɛ nɔ tsuaa nɔsɔ ngɛ odehe si himi kɛ he blɔhi a blɔ fa mi. A bɔ mɛ kɛ nɔ́ se kɔmi kɛ he nule juɛmi, nɛ e hia kaa nɔ fɛɛ nɔ nɛ e na nyɛmi suɔmi kɛ ha nɔ tsuaa nɔ.

Dangme

Himmaka' nittakookano hattak yokasht toksalicha'nikat ki'yo. Hattak mǫ́makat ittíllawwi bíyyi'kacha nanna mǫ́maka̱ ittibaachaffa'hitok.

Chickasaw

Inuit tamarmik inunngorput nammineersinnaassuseqarlutik assigiimmillu ataqqinassuseqarlutillu pisinnaatitaaffeqarlutik. Solaqassusermik tarnillu nalunngissusianik pilersugaapput, imminnullu iliorfigeqatigiittariaqaraluarput qatanngutigiittut peqatigiinnerup anersaavani.

Inuktitut (Greenlandic)

Phonetics

- The study of how signals are produced
  - Language-independent
- Places of articulation
- Manners of articulation
- Use of vocal folds
- Airstream mechanisms

- Parallel in sign language

labial

alveolar

palatal

velar

glottal
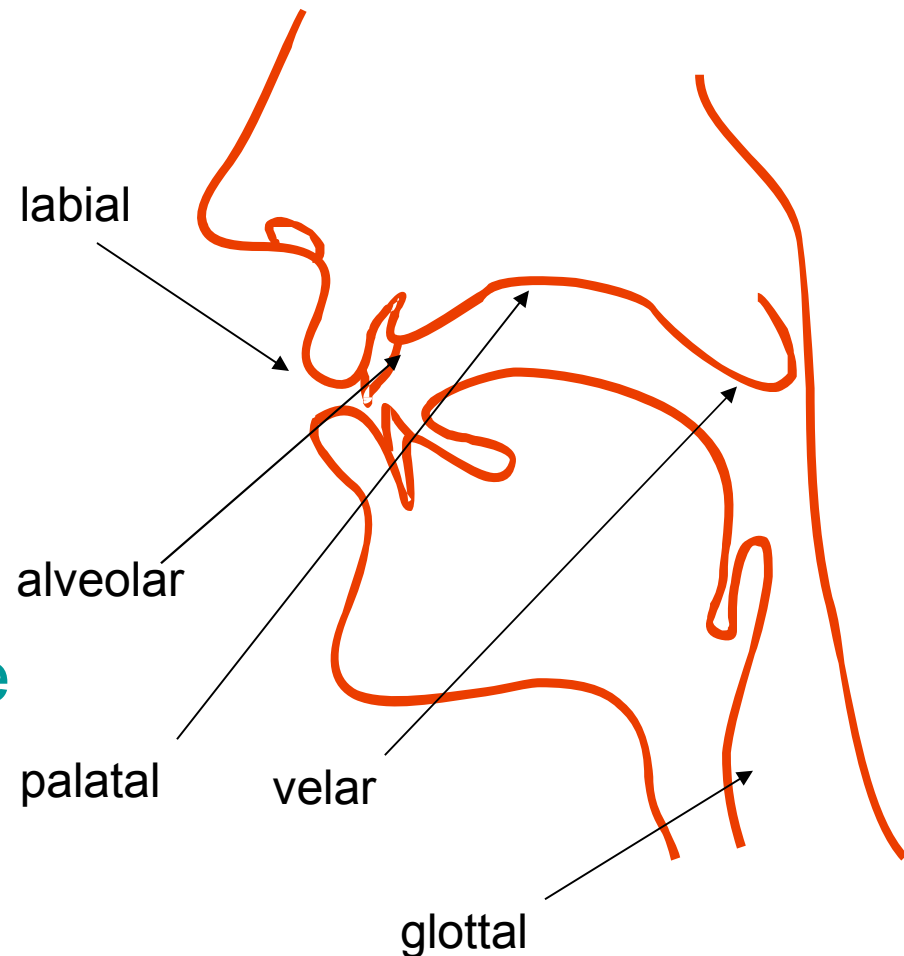
Phonology

- How sounds are used in language
  - Minimal pairs
    - beet – boat
    - bat – bet
    - right – light
  - Allophones
    - pit – spit (in homage to Pullum 1989)
    - (Dutch) uil – lui
  - But:
    - Russian: łuk (onion) – lʲuk (porthole)
    - Czech: ra:t ([he] loves) – řa:t (order)

Examples (1)

- Voicing (Hindi)
  - ta:n (musical tone)
  - tʰa:n (a bale of cloth)
  - da:n (donation)
  - dʰa:n (paddy)

Examples (2)

- Air stream (and place) Montana Salish
  - tsáqʷəlʃ (western larch)
  - ts' áɬt (it's cold)
  - tɬ' áq' (hot)
- Clicks (!Xóõ)

Tone

- All spoken languages use intonation (variation in pitch)
  - This is often learned (note foreign accents)
- Some languages use it systematically to distinguish meaning
  - Tone languages
  - Chinese, Vietnamese, many African languages (Dangme), many American languages (Navaho)
  - Norwegian, Swedish, Limburgian dialects

Tone (cantonese)

- Mā – hemp
- Mă – horse
- Mâ – scold
- Mà – interrogative
- Má - mother

Phonotactics

- The way sounds are combined into words
  - English, Dutch are rather complex
    - "Strength" = CCCVCC
    - "Schraalst" = CCCVCCC
    - But: "prtskvna" = Georgian "to peel"
  - Japanese is simpler: CVN
    - MacDonalds = Macudonaradu
- Phonotactics is language-dependent
  - English: spin - *zbin
  - But Polish: Zbigniew is OK.

Universals?

- Languages have vowels and consonants
  - But what about signed languages?

- Certain vowel systems are near-universal
  - But there are good non-cognitive explanations for most phonological phenomena

- Syllables have a sonorant as their nucleus
  - But: Tashlhiyt Berber

Table 3. Corpus for test 2

gis	inside	rar	give back !
ls	put !	flt	leave it !
ns	spend a night !	ssnd	churn !
ks	pasture !	kst	pasture it !
fk	give !	fkt	give it
mnSk	how many	tkti	shehas thought
tsti	she has filtered	txznt	you sg. have stored
tsqsat	she asked her	irgl	he has locked
tnda	it f. has been churned	tldi	she has shot
tasa	liver	afud	knee

Puech & Louali 1999

Core aspects of language

- Morphology and syntax are somehow more central to linguistics than other aspects of language
  - Phonology is also much studied
- Perhaps they are more typical
  - Combinatorial, learned, arbitrary
- Phonetics is more shared with animals
- Semantics, Pragmatics, social use is more about the function of language than the form/the computational mechanisms

Colorless green ideas…

- "Colorless green ideas sleep furiously"
  - A sentence constructed by Noam Chomsky to illustrate that nonsensical sentences can nevertheless be grammatical
  - It can (and will) be read with normal sentence intonation
  - Viz. "Furiously sleep ideas green colorless"
- To show "statistical theories don't work"
  - All bigram probabilities are almost zero

Colorless green ideas…

- Chomsky draws the conclusion that syntax (grammar) is autonomous and that there must be highly language-specific learning and pre-wired knowledge "Universal grammar"
- But note that the sentence is generally judged by highly educated people
  - How would illiterate people judge such sentences?
  - Probably they would show better performance on the grammatical sentence, too
  - But perhaps this can be explained from more sophisticated statistical learning?

Colorless green ideas

- Perhaps statistical classification learning and derivation of patterns may explain the difference in performance, too
  - Very much like Goldberg's construction learning

Wild ideas

Bees sleep peacefully

Colorless liquid

Wild plants

Ideas exist

Odorless liquid

Green plants

Unicorns exist

Bees buzz peacefully

Odorless green liquid

Unicorns sleep

Bees buzz furiously
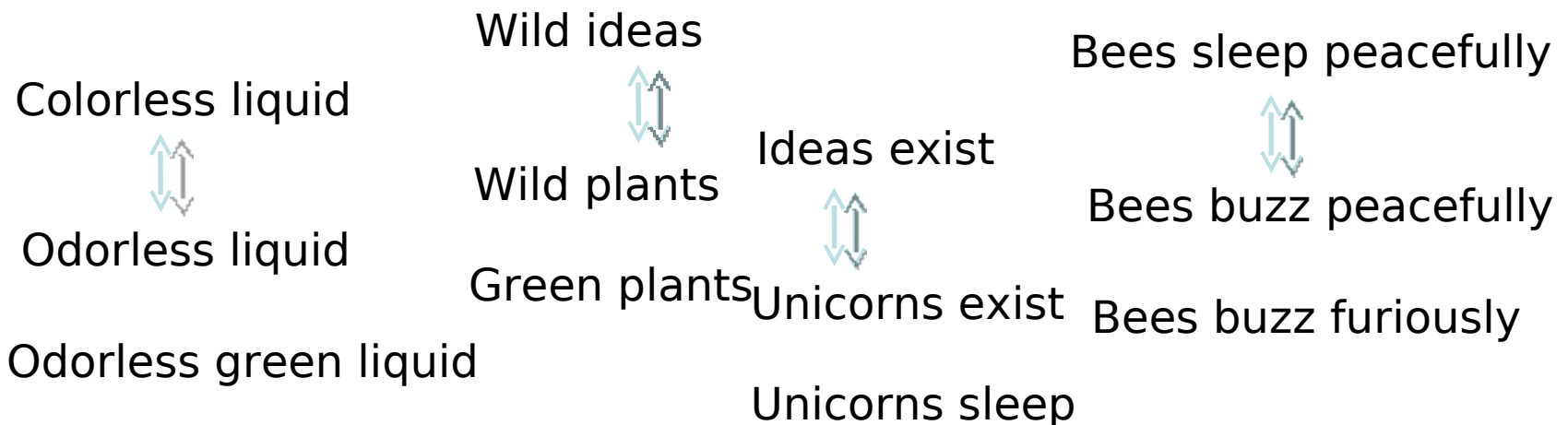
By noticing certain words re-occur in similar contexts, we can derive patterns statistically – but although this is statistical learning it may be human-specific

Morphology (1)

- Morphemes are the smallest meaningful elements of a language
  - inflectional morphology (grammatical function)
    - walk – walks – walked – walking
  - Derivational morphology (lexical function)
    - establish – establishment – establishmentarian – establishmentarianism – antiestablishmentarianism

Morphological diversity (1)

- Not all languages express the same things with morphology
  - Future: French j'irai English I will go
  - Cases: Ivan gave Olga the book
    Russian: Ivan dal knigu Ol'ge
    Ivan dal Ol'ge knigu = Ivan gave Olga a book

Morphological diversity (2)

- Bahing transitive verb morphology
  - Ca. 75 verb endings

	1s	**1di**	**1de**	**1pi**	**1pe**	**2s**	**2d**	**2p**	**3s**	**3d**	**3p**
1s						Σ-na Σ-tana	Σ-nasi Σ-ntanasi	Σ-nami Σ-ntanani	Σ-^{N}a Σ-to^N	Σ-^{N}asi Σ-to^{N}si	Σ-^{N}ami Σ-to^{N}mi
1di									Σ-sa Σ-tasa	Σ-sa Σ-tasa	Σ-sa Σ-tasa
1de						Σ-^{ʔ}a Σ-tana	Σ-^{ʔ}a Σ-tanasi	Σ-^{ʔ}a Σ-tanani	Σ-su Σ-tasu	Σ-su Σ-tasu	Σ-sumi Σ-tasume
1pi									Σ-ia Σ-taia	Σ-ia Σ-taiasi	Σ-iami Σ-taiami
1pe						Σ-^{ʔ}a Σ-tak ·	Σ-^{ʔ}a Σ-tak	Σ-^{ʔ}a Σ-tak mi	Σ-ka Σ-tak	Σ-ka Σ-tak	Σ-kami Σ-tak me
2s	Σ-i Σ-pti		Σ-si Σ-tasi		Σ-ki Σ-tami/-taki				Σ-∅ Σ-pti	Σ-i(?)/-esi Σ-ptisi	Σ-umi Σ-ptimi
2d	Σ-isi Σ-tasi		Σ-si Σ-tasi		Σ-kini/-kimi Σ-takini				Σ-sa Σ-tasi	Σ-sa Σ-tasi	Σ-sami Σ-ntani
2p	Σ-ini Σ-tini		Σ-sini Σ-tasini		Σ-kimi Σ-ntanime				Σ-ni Σ-ntani	Σ-ni Σ-ntanisi	Σ-ami Σ-ntanimi
3s	Σ-i Σ-ti	Σ-so Σ-taso	Σ-si Σ-tasi	Σ-so Σ-taso	Σ-ki Σ-taki	Σ-e Σ-te	Σ-si(?) Σ-taso	Σ-ni Σ-ntani	Σ-wa Σ-pta	Σ-se Σ-ptasi	Σ-me S-ptami
3d	Σ-isi Σ-tisi	Σ-sosi Σ-tasosi/-taso	Σ-si Σ-tasi	Σ-sosi Σ-tasosi	Σ-kisi Σ-tasimi(?)	Σ-esi/-si Σ-tesi	Σ-si(?) Σ-tasosi	Σ-nisi Σ-ntanisi	Σ-se Σ-tase	Σ-se Σ-tase	Σ-se/-mesi S-tasemi
3p	Σ-imi Σ-timi	Σ-somi Σ-tasomi/-taso	Σ-simi Σ-tasi	Σ-somi Σ-tasomi	Σ-kimi Σ-takimi/-taki	Σ-emi Σ-temi	Σ-sumi(?) Σ-tasomi	Σ-nimi Σ-ntanimi	Σ-me Σ-mtame	Σ-me Σ-mtamesi/-mtame	Σ-me/-mesi Σ-mtamemi

Patient (column headers) · Agent (row labels)

Morphology (2)

- ## Languages use morphology differently
  - Chinese, Vietnames: one morpheme – one word (isolating languages)

  Khi tôi đến nhà bạn tôi, chúng tôi bắt đầu làm bài.
  when I come house friend I PLURAL I begin do lesson
  'When I came to my friend's house, we began to do lessons.'

  - Turkish, Hungarian: multiple morphemes per word, but morphemes can be distinguished (agglutinating)
  - Russian, Latin: morphemes mix (fusional)

	Singular	Plural
Nominative	adam	adam-lar
Accusative	adam-ı	adam-lar-ı
Genitive	adam-ın	adam-lar-ın
Dative	adam-a	adam-lar-a
Locative	adam-da	adam-lar-da
Ablative	adam-dan	adam-lar-dan

Ia	Singular	Plural
Nominative	stol	stol-y
Accusative	stol	stol-y
Genitive	stol-a	stol-ov
Dative	stol-u	stol-am
Instrumental	stol-om	stol-ami
Prepositional	stol-e	stol-ax

Morphology (3)

- Polysynthetic languages (Chukchi)
  - Many morphemes are combined into long words
  - These would be sentences in other languages

təmeyŋəlevtəpəɣtərkən = I have a fierce headache

tə-meyŋə-levtə-pəɣt-ərkən
t = first person singular subject
meyŋ = big
levt = head
pəɣt = ache
rkən = imperfect aspect

Morphological universals?

- More animate patients of an action are more morphologically marked
  - I hit the stone
  - The stone hit me
  - I was hit by the stone
- Many similar universals exist, but one can understand them as functional (almost information-theoretical) adaptations
  - Cf. Piantadosi
  - Comrie 1981

Syntax (1)

- The study of how sentences are formed out of words

- The focus of much of general linguistics
  - Most linguistics papers in TiCS are on syntax
  - Because it is important in converting complex meaning into signals
  - But also because this is the most active process in English…

Syntax (2)

- Words appear in phrases
  - The structure of sentences is phrase structure
  - Absolute position is usually unimportant
  - Position relative to other components counts
  - Structure is hierarchical

Position 3: subject

The$_1$ old$_2$ carpet$_3$ was$_4$ dirty$_5$

John$_1$ met$_2$ Mary$_3$ last$_4$ Tuesday$_5$

Position 3: direct object

```
          S
        /   \
      NP     VP
     /|\    / \
The old carpet was dirty
```

```
        S
       / \
      /   VP
    NP  V /|\
   /   / NP  PP
John met Mary last Tuesday
```
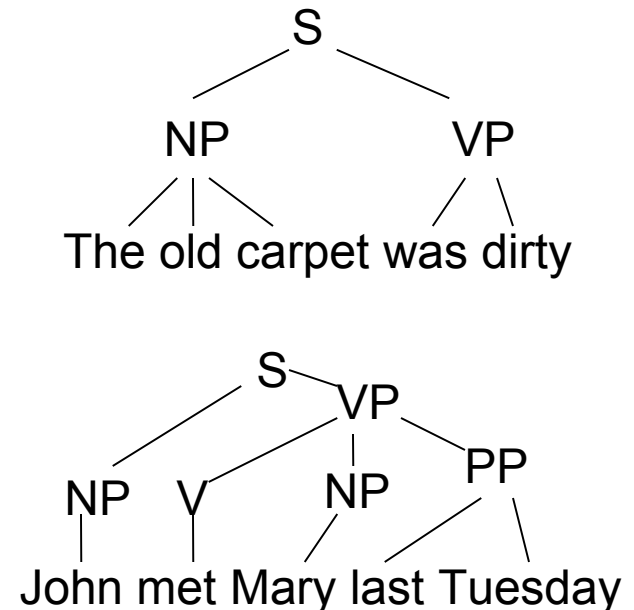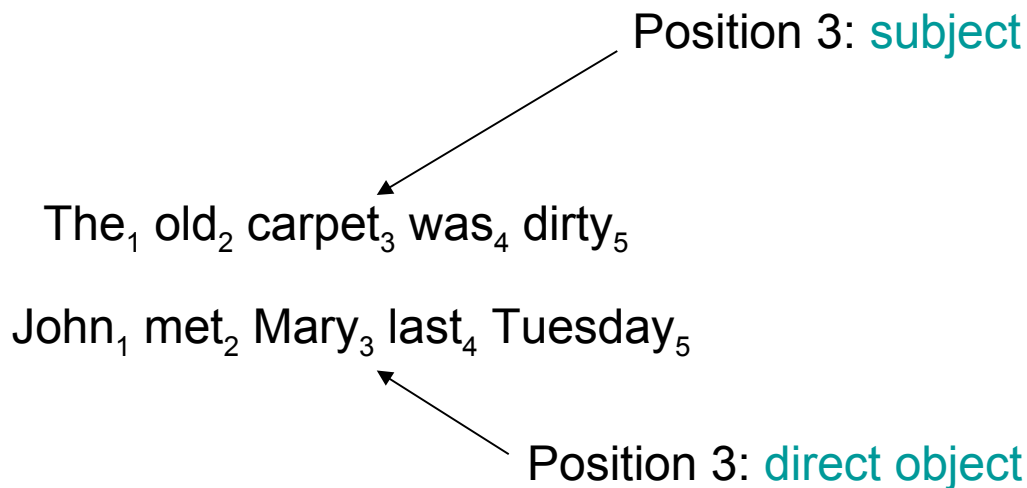
Syntax (3)

- Syntactical structure can be recursive
  - Phrases can be embedded in other phrases

- John saw Mary.
- Alice said that John saw Mary.
- Bob thought that Alice said that John saw Mary.
- Etc…

Syntactic universals?

- Languages have recursion
  - Not Pirahã(?)

- All languages use phrase structure
  - Probably true
  - Not necessary for a communication system (computer protocols)
  - Perhaps explainable from older cognitive constraints?

- All languages have phrases and sentences
  - "Yesterdays consumption of the sandwich by Bart"
  - "Bart ate the sandwich yesterday"
  - Less easily explained from general cognition

Word order universals (1)

- "The farmer killed the Duckling" SVO

- "Hasan öküzü aldı"
  Hasan Ox Bought *Turkish*: SOV

- "Lladdodd y ddraig y dyn"
  Killed the dragon the man *Welsh*: VSO

- "Nahita ny mpianatra ny vehivavy"
  saw the student the woman
  Malagasy: VOS

Word order universals (2)

World atlas of linguistic structures
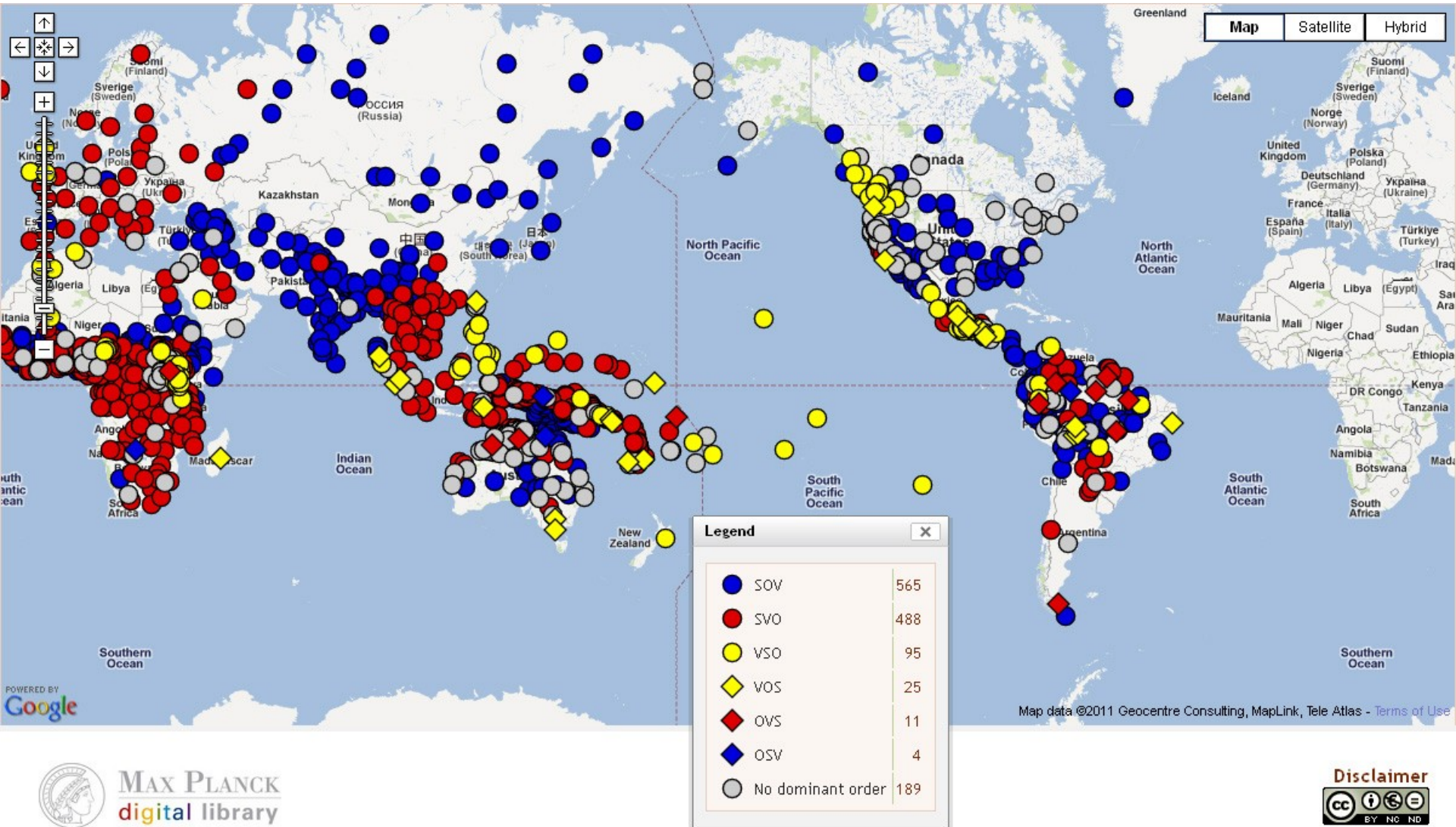
Word order universals (3)

- "Toto yahsɨye kamara"
  man grabbed jaguar *Hixkaryana*: OVS

- It was thought that object first languages did not exist, but they are just very rare

Implicational universals

- Does one type of word order predict another?
  - VO -> prepositions
  - OV -> postpositions

- But even to such universals there are exceptions
  - Although there are strong tendencies
- Functional explanation?
  - Head first versus head last?
- Historical explanation?
  - Adpositions derive from verbs

Semantics (1)

- The study of syntax is often formal
  - Only interested in form
  - A sentence is either grammatical or not

- Semantics studies the relation between form and meaning
  - On all levels of language

Semantics (2)

- How do different languages divide up meaning spaces?
  - Color terms
  - Space: here, there vs. aquí, ahí, allá
  - Taste: sour/bitter vs. acide/amer vs. zuur/bitter/wrang
  - Animals/Plants

Semantics (3)

- The relation between words in sentences and meaning
  - The direct relation is already not simple
    - "The old man the boat"
  - But semantics likes to focus on even more subtle ones:
    - It's raining
    - It isn't dry

Semantic universals?

- There appear to be tendencies of how sensory spaces are named
  - E.g. Color Spaces (Berlin & Kay 1969)
  - But this might be due to properties of vision
  - And they are tendencies
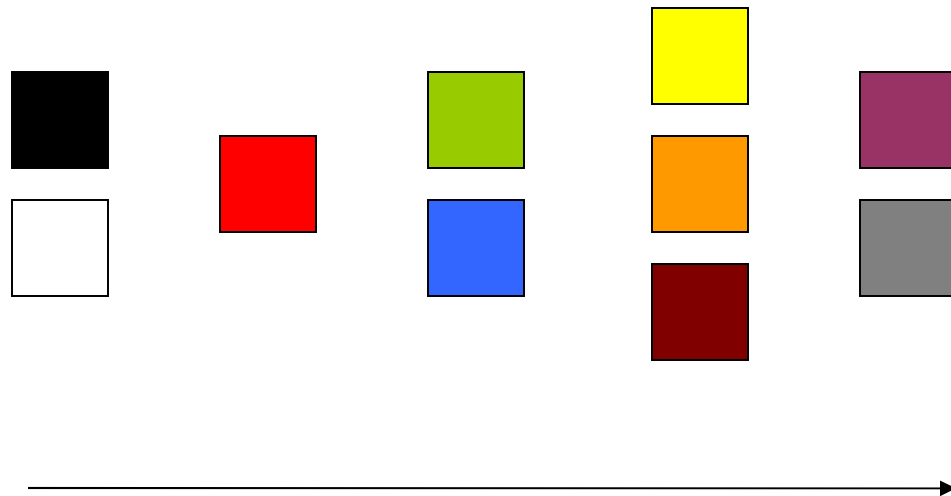
Pragmatics (1)

- Language is often used indirectly

- And almost always in context

- Pragmatics investigates why and how language is used the way it is

Pragmatics (2)

- Generally, people do not exactly say what they mean
  - "Could you open the window?" – "Yes"
  - "Open the window!"

Pragmatics (3)

- People are very good at providing the right information and the right amount
  - "Could you open the small window?" → there is more than one window

  - A: "Where is the Anne Frank House?"
    B: "Follow the tram line and turn right after the church."

- Determining what to say is a very difficult task
  - Not easily solved in e.g. computer dialog systems

Pragmatics differences

- French and Dutch both have polite and familiar forms of address:
  - Vous/tu
  - U/jij

- But in French using "tu" to the plumber would be unacceptable, whereas in Dutch using "U" indicates there is a problem
 - Details depend on age differences etc.

Sociolinguistics (1)

- Whereas pragmatics investigates how language is used between individuals, sociolinguistics focuses on language in larger groups

- Gender
- Socio-economic class
- Ethnic group
- Age group
- Regional variation

Sociolinguistics (2)

- Different accents are appropriate in different situations
  - Cockney vs. RP
  - Perception of regional accents differs
    - Brittain: no especially negative attitudes
    - NL: considered backward
  - Often there is a confusion between the social status of a group and the status of the language they speak
    - Poor = ugly, backward, wrong, primitive, tough, free
    - Rich = snobbish, civilized, important

Sociolinguistics and complexity

- Lupyan and Dale (2010) show an inverse relation between language complexity and population size
  - They relate this to the number of second language speakers (higher for large languages)
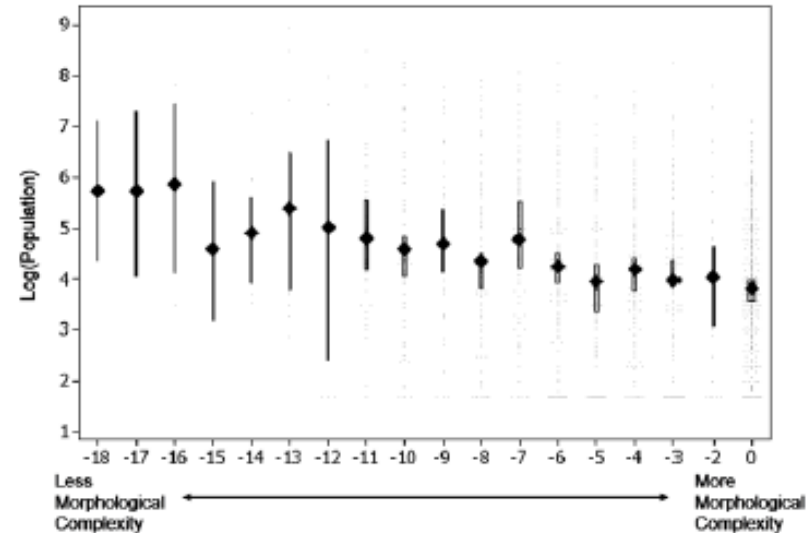
Figure 3. Languages spoken by more people have simpler inflectional morphology. X-axis scores represent a measure of lexical devices compared to the use of inflectional morphology. Filled symbols represent population means for languages with a given complexity score; bars show 95% confidence intervals of the median. Bar width is proportional to sample size for each score.
doi:10.1371/journal.pone.0008559.g003

Sociolinguistics and evolution

- People are more favorably disposed towards people who speak the same
  - "Kin recognition"
  - If languages are sufficiently complex, mastery is a costly signal of group membership
  - Languages may be "excessively" complex because of reasons of group selection
  - Also: this is less effective in larger groups

Conclusion

- A longish introduction to the issues and terminology of linguistics
  - In order to aid independent appreciation of the literature

Some misconceptions

- There are "incorrect" languages
- There are primitive languages
- People only speak one language
- Languages are uniform
- "That's not a language, it's just a dialect"
- Written language is spoken language

Goals of linguistics (1)

- Promoting "correct" language usage

- Establishing a standard
  - Historically important
    - Sanskrit, Greek, Latin etc.
  - This is still an important task of "language academies" in some countries
  - And in first language education
  - Also, standards are established for languages with recent official status

Goals of linguistics (2)

- Describing languages
  - An important part of modern linguistics
- Describing "field" languages
  - Complete grammars
  - E.g. Franz Boas
- Describing "large" languages
  - Focus on detail
- Description for second language learning

Goals of linguistics (3)

- Understanding language history
  - Through comparison of variants
- Focus of much 19th century work
  - The Grimm brothers
- Nowadays often combined with archaeology and DNA-studies
- Important for understanding diachronic processes
  - how and why does language change?

Goals of linguistics (4)

- Explaining language
  - E. g. Chomsky
- Why are languages the way they are?
- Often with a focus on synchronic processes
  - Phenomena in a language such as:
    - I ask – he asks – I asked
    - John saw Mary – Whom did John see?
  - But these are not necessarily exactly the same as cognitive processes

Goals of linguistics (5)

- Understanding language cognition
- How does language work in the brain?
- Focus of modern psycholinguistics, cognitive linguistics
  - E.g. Tomasello
- Language acquisition studies
  - E. g. Kuhl

Goals of linguistics (6)

- Most linguistic work addresses 2-5
  - And sometimes a bit of 1

- We are mostly interested in 4 and 5
  - The cognitive questions

Nature versus nurture (1)

- Old debate in the study of humans
  - With focus changing periodically

- Is behavior innate (genetically determined) or learned (culturally determined)?
  - Of course: both
  - But which is most important?
  - And what is the nature of nature?

Nature versus nurture (2)

- A bewildering variety of languages

- Nevertheless: all humans can learn all languages
  - And no other animal can
  - What is special about humans?

Nature versus nurture (3)

- This debate did not really preoccupy 19th century linguists
  - Interested in description, history

- But for cognitive linguistics it becomes important
  - Especially since Chomsky

Nature versus nurture (5)

- Arguments for specialized nature:

- Poverty of the stimulus
  - Infants have insufficient information to learn their languages
  - Or not?

- Deep similarities between languages
  - Universals
  - Principles and Parameters
  - Processing mechanisms

Nature versus nurture (4)

- What is the nature of nature?
  - Is it a highly specialized language acquisition device?
  - Chomsky, Pinker

  - Or is it a further development of domain-general cognitive mechanisms?
  - Tomasello, Christiansen

How to gather data (1)

- We can look at the whole complex thing at once
  - Descriptive linguistics

- We can look at isolated linguistic phenomena
  - Grammaticality judgments of constructed sentences
  - General linguistics

- We can look at the simplest responses to input
  - ERP, EEG, reaction times, eye tracking etc.
  - Psycholinguistics

How to gather data (2)

- We can rely on intuitions of *one* speaker
    - Happens often in descriptive linguistics, because there is only one speaker

- Or we can rely on *large numbers* of speakers
    - Corpus data

http://www.circassianworld.com/new/images/stories/Articles/TevfikE_82.jpg
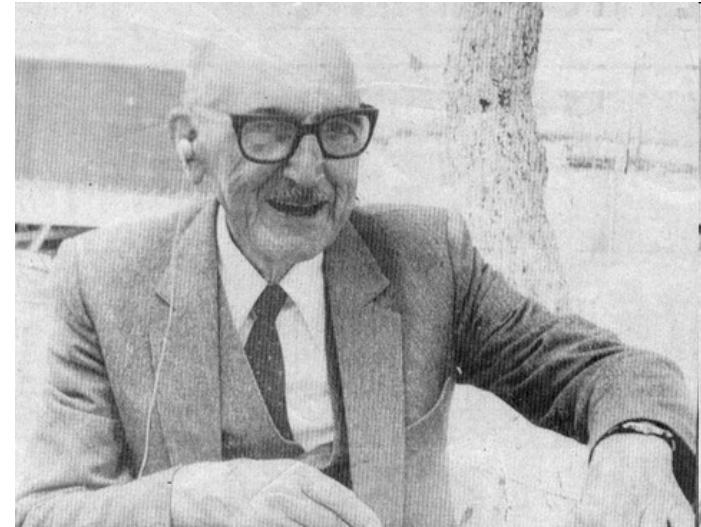
Tevfik Evenç, last speaker of Ybykh

How to gather data (3)

- We can study multiple languages
  - Look for universals (Greenberg 1963)

- But Chomsky proposed all languages are similar, because of Universal Grammar
  - There has been a period in which many linguists studied only one language
  - But even UG linguists nowadays study multliple languages