What is Linguistics?

Matilde Marcolli
CS101: Mathematical and Computational Linguistics

Winter 2015
• **Linguistics** is the scientific study of language

- What is Language? (langage, lenguaje, ...)
- What is a Language? (lange, lengua,...)

Similar to ‘What is Life?’ or ‘What is an organism?’ in biology

• *natural* language
  as opposed to artificial (formal, programming, ...) languages

• The point of view we will focus on:
  Language is a kind of **Structure**

  - It can be approached mathematically and computationally, like many other kinds of structures
  - The main purpose of mathematics is the understanding of structures
Language Families
- Niger-Congo (1,532)
- Austronesian (1,257)
- Trans New Guinea (477)
- Sino-Tibetan (449)
- Indo-European (439)
- Afro-Asiatic (374)
- Nilo-Saharan (205)
- Oto-Manguean (177)
- Austro-Asiatic (169)
- Tai-Kadai (92)
- Dravidian (85)
- Creole (82)
- Tupian (76)
- Mayan (69)
- Altaic (66)
- Uto-Aztecan (61)
- Arawakan (59)
- Torricelli (56)
- Sepik (55)
- Quechuan (46)
- Na-Dene (46)
- Algic (44)
- Hmong-Mien (38)
- Uralic (37)
- North Caucasian (34)
- Penutian (33)
- Macro-Ge (32)
- Ramu-Lower Sepik (32)
- Carib (31)
- Panoan (28)
- Khoisan (27)
- Salishan (26)
- Tucanoan (25)
- Isolated Languages (75)
The Indo-European Language Family: Phylogenetic Tree (Graph)
• How are different languages related? What does it mean that they come in families?

• How do languages evolve in time? Phylogenetics, Historical Linguistics, Etymology

• How does the process of language acquisition work? (Neuroscience)

• Semiotic viewpoint (mathematical theory of communication)

• Discrete versus Continuum (probabilistic methods, versus discrete structures)

• Descriptive or Predictive? to be predictive, a science needs good mathematical models
Levels of Structure

A language exists at many different levels of structure

- **Phonology**: *sound* structures, building blocks (phonemes), phonetics (physical aspects), autosegmental phonology (tone systems), feature geometry (generative theory), optimality theory (neural networks)

- **Morphology**: *words* (roots, affixes, stress), building blocks (morphemes), morphological typology (use of morphemes), lexicology, word formation, paradigms (words associated to same lexeme, conjugation, declension), morphosyntax
• **Syntax:** \( \sigma\nu\nu\tau\alpha\xi\nu\varsigma = \text{ordering together} \) *sentences* (principles, rules, and processes by which sentences are formed), generative grammar (i-language, grammaticality), transformational grammar, principles and parameters (government and binding), generalized phrase structure grammar, head-driven phrase structure grammar, lexical functional grammar

• **Semantics:** *meaning* (how communication of meaning happens through signifiers like words, sentences), homonymy, synonymy, cognitive linguistics, philosophy of language, truth values (formal logic, propositional calculus), computational semantics
An analogy

Physics looks very different at different scales:

- General Relativity and Cosmology ($\geq 10^{10}$ m)
- Classical Physics ($\sim 1$ m)
- Quantum Physics ($\leq 10^{-10}$ m)
- Quantum Gravity ($10^{-35}$ m)

Despite dreams of a Unified Theory, we deal with different mathematical models for different levels of structure
Similarly, we view language at different “scales”:

- units of sound
- words
- sentences
- global meaning

We expect to be dealing with different mathematical structures and different models at these various different levels.
Phonetics

- Certain **speech-sounds** are specific to certain languages (clicks in Khoisan languages)
- Minimum known: 11 phonemes in Rotokas (East Papua)
- Maximum known: 141 phonemes in !Kung (Khoisan group, Namibia)
- **Consonants** (complete or partial closure of vocal tract): two coordinates
  - **place**: labial, coronal, dorsal, radical, glottal
  - **manner**: nasal, stop, sibilant fricative, non-sibilant fricative, approximant, flap or tap, trill, lateral fricative, lateral approximant, lateral flap
- **Vowels** (open vocal tract): close/mid/open
- **Other**: Semivowels, approximants, diphthongs, fricative vowels
- **Production modes**: pulmonic egressive (most languages) + other mechanisms (ejectives, clicks, implosives)
Articulatory Phonetics

- Phonemes classified according to vocal organs

<table>
<thead>
<tr>
<th>Place</th>
<th>Adjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>lips</td>
<td>labial</td>
</tr>
<tr>
<td>teeth</td>
<td>dental</td>
</tr>
<tr>
<td>alveolar</td>
<td>alveolar</td>
</tr>
<tr>
<td>ridge</td>
<td>palatal</td>
</tr>
<tr>
<td>palate</td>
<td>velar</td>
</tr>
<tr>
<td>velum</td>
<td>uvular</td>
</tr>
<tr>
<td>larynx</td>
<td>laryngeal</td>
</tr>
<tr>
<td>glottis</td>
<td>glottal</td>
</tr>
</tbody>
</table>

Nasal cavity

lip

palate

velum

alveolar ridge

lip

teeth

tongue tip

Tongue (raised)

Tongue (rest position)

larynx

epiglottis

glottis

(trace between vocal folds)

uvula

trachea
International Phonetic Alphabet (IPA)

- universal phonetic alphabet for all world languages

IPA chart of pulmonic consonants:

<table>
<thead>
<tr>
<th>Consonants (pulmonic)</th>
<th>Labial</th>
<th>Labiodental</th>
<th>Coronal</th>
<th>Dorsal</th>
<th>Radical</th>
<th>Laryngeal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td>η</td>
<td>η</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Plosive</td>
<td>p b</td>
<td>t d</td>
<td>θ  δ</td>
<td>c j</td>
<td>k g</td>
<td>q g</td>
</tr>
<tr>
<td>Fricative</td>
<td>φ β</td>
<td>f v</td>
<td>θ  δ</td>
<td>s z</td>
<td>e z</td>
<td>ç j x y</td>
</tr>
<tr>
<td>Approximant</td>
<td>u j</td>
<td>ɹ r</td>
<td></td>
<td>j u</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tap, flap</td>
<td>v r</td>
<td>ɹ r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>B r</td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>Lateral fricative</td>
<td></td>
<td>ɹ l̞ ɹ ɾ</td>
<td>ɾ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approximant</td>
<td></td>
<td>ɹ l̞ ɹ ɾ</td>
<td>ɾ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral flap</td>
<td></td>
<td>ɹ l̞ ɹ ɾ</td>
<td>ɾ</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured ɾ. Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.
## IPA chart of non-pulmonic consonants

### CONSONANTS (NON-PULMONIC)

<table>
<thead>
<tr>
<th>Anterior click releases (require posterior stops)</th>
<th>Voiced implosives</th>
<th>Ejectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʘ Bilabial fricated</td>
<td>ɓ Bilabial</td>
<td>’ Examples:</td>
</tr>
<tr>
<td>ǀ Laminal alveolar fricated (“dental”)</td>
<td>ɗ Dental or alveolar</td>
<td>ǀp’ Bilabial</td>
</tr>
<tr>
<td>ǃ Apical (post)alveolar abrupt (“retroflex”)</td>
<td>ʄ Palatal</td>
<td>ǀt’ Dental or alveolar</td>
</tr>
<tr>
<td>ǂ Laminal postalveolar abrupt (“palatal”)</td>
<td>ʅ Velar</td>
<td>ǀk’ Velar</td>
</tr>
<tr>
<td>ǁ Lateral alveolar fricated (“lateral”)</td>
<td>ʠ Uvular</td>
<td>ǀs’ Alveolar fricative</td>
</tr>
</tbody>
</table>
IPA vowel chart

Vowels at right & left of bullets are rounded & unrounded.
Example: Vowels in English

English vowels & major diphthongs, with keywords in b__t

dashed arrows for diphthongs
Tone systems, tonal languages (tonemes)

- Mandarin is the most widely spoken tonal language

- A large number (up to 70%) of world languages are tonal

- *Register* tone systems (Niger-Congo languages): tones distinguished by relative pitch

- *Contour* tone systems (Sino-Tibetan languages): tones distinguished by shape;

- combined system (Cantonese) both shape and pitch

- grammatical information may be encoded in the tone
### IPA diacritics and suprasegmentals

<table>
<thead>
<tr>
<th>SUPRASEGMENTALS</th>
<th>TONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary stress</td>
<td>Level tones</td>
</tr>
<tr>
<td>Extra stress</td>
<td>é ˥ Top</td>
</tr>
<tr>
<td>Secondary stress</td>
<td>é ˥ High</td>
</tr>
<tr>
<td>Long</td>
<td>é ˥ Mid</td>
</tr>
<tr>
<td>Half-long</td>
<td>ë ˥ Low</td>
</tr>
<tr>
<td>Short</td>
<td>ë ˥ Bottom</td>
</tr>
<tr>
<td>Extra-short</td>
<td>é ˥ Tone terracing</td>
</tr>
<tr>
<td>Syllable break</td>
<td>↑ Upstep</td>
</tr>
<tr>
<td>(no break)</td>
<td>↓ Downstep</td>
</tr>
</tbody>
</table>

INTONATION:
- Minor (foot) break
- Major (intonation) break
- Global rise
- Global fall
Other suprasegmental features (long-distance effects)

- **Metaphony**: general phenomena of assimilation of one type of vowel by another within a word
- **Vowel harmony** (progressive or regressive)

\[ V_aCV_bCV_bC \mapsto V_aCV_aCV_aC \]

vowel of type \( V_a \) assimilates vowels of type \( V_b \) across intervening consonant segments
(Turkish, Hungarian, Korean)
Phonological models of these phonetic structures

*Autosegmental Phonology* for tone systems, vowel harmony, ...

- Describe the segments of the sound structure by multiple tears, each a $+/−$ binary variable describing one feature
- Factoring phonemes into a product of *features*
  - Segmental tier: $±$ sonorant, $±$ continuant, $±$ voice, ...
  - Timing tier: lengths of segments (long/short vowels)
  - Stress tier: distribution of stress, main stress, ...
  - Tone tier: $±$ high/low pitch, ...

Feature Geometry

- Autosegmental structures as trees, geometric operations on trees

Optimality theory in phonology

- Generative phonology aims at a predictive model of sound systems in natural languages
- Change and constraints
- phonological changes as a replication process
- *faithfulness constraints* tend to maintain the output like the input (along some of the structural dimensions)
- *markedness constraints* forbid occurrences of certain outputs
- *ranking* of constraints
- *candidate set* of outputs given an input and the ranked constraints
- *optimality* by smallest number of violations