Variability of Languages in Time and Space

Lecture V
Phonology

- Consonant and Vowel Inventories
- Segmental Processes
- Syllables
- Suprasegmentals

Anja Nedoluzhko
Phonological Typology

• What kinds of phoneme inventories are found?
• Helps to analyze phonological data (*I’ve seen this before... Typically it goes like this...*)
• What is common/uncommon?
  – Odden 2013: „It is very difficult to refuse a claim of the form „X is more common than Y,” except if a very detailed numerical study is undertaken.“ (p. 207)
• Understand markedness
  – „X is marked“ (relative to Y): “[ʕ] is more marked than [q]” – “[q] is more marked than [k]”
• Most research on spoken languages
Phomenic Inventories Databases

- **SPA**, Stanford Phonology Archive (Vihman, 1974)
  - 196 languages,
  - phonological descriptions according to grammars
- **UPSID**, UCLA Phonological Segment Inventory Database (Maddieson 1984, 1997)
  - statistical survey, phonemic inventories,
  - 451 languages in the last version
  - [http://web.phonetik.uni-frankfurt.de/upsid.html](http://web.phonetik.uni-frankfurt.de/upsid.html)
- Phonemic inventories within **WALS** (Maddieson 2013)
  - statistical survey, phonemic inventories,
  - 564 languages
  - [http://wals.info/](http://wals.info/)
- **PHOIBLE** database (Moran et al. 2014, updated in 2019)
  - segment inventories of 1,672 languages
  - [https://phoible.org/](https://phoible.org/)
Phonemes

• The smallest distinct acoustic unit in a language
• Does not convey meaning but distinguishes meaning of larger units
• At the beginnings and ends of syllables vs. in the middle: consonants and vowels
  – pin, tin, kin, fin, thin, sin, shin
  – dim, din, ding, did, dig, dish
  – pin, pen, pan, pun, pain, pine, pawn
WALS: Consonant Inventories

small

moderately small

average

moderately large

large

6

15

19

25

34

...
WALS: Consonant Inventories

only 6 consonants ( /p, t, k, b, d, g/ )

Rotokas

Papua New Guinea

A language of Papua New Guinea

<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>FEEDBACK</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO 639-3</td>
<td>P00</td>
</tr>
<tr>
<td>Location</td>
<td>Bougainville Province, Central Bougainville district, central mountains. 28 villages.</td>
</tr>
<tr>
<td>Language Maps</td>
<td>Papua New Guinea, Map 13</td>
</tr>
<tr>
<td>Classification</td>
<td>North Bougainville, Rotokas</td>
</tr>
<tr>
<td>Dialects</td>
<td>Alta, Atillina, Pipiya.</td>
</tr>
<tr>
<td>Typology</td>
<td>SOV.</td>
</tr>
<tr>
<td>Language Resources</td>
<td>OLAC resources in and about Rotokas</td>
</tr>
<tr>
<td>Writing</td>
<td>Latin script (Latn).</td>
</tr>
</tbody>
</table>
WALS: Consonant Inventories

6 15 19 25 34 ...

- small
- moderately small
- average
- moderately large
- large

122 consonants
(mainly because it has a very large number of different click sounds with which a word may begin)

Click sounds? How to pronounce them?
https://www.youtube.com/watch?v=31zzMb3U0iY

spoken in Botswana
WALS: Consonant Inventories

http://wals.info/feature/1A#2/19.3/152.8
Typology for Consonant Inventories: Correlations

- Hypothesis (Lindblom - Maddieson, 1988): There is an overall relationship between the size of a consonant inventory and the kind of consonants it includes.

<table>
<thead>
<tr>
<th>C’s inventory size</th>
<th>special C’s</th>
<th>glottalized C’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>small</td>
<td>8.7%</td>
<td>8.7%</td>
</tr>
<tr>
<td>moderately small</td>
<td>13.1%</td>
<td>10.7%</td>
</tr>
<tr>
<td>average</td>
<td>22.1%</td>
<td>21.5%</td>
</tr>
<tr>
<td>moderately large</td>
<td>27.4%</td>
<td>39.3%</td>
</tr>
<tr>
<td>large</td>
<td>40.7%</td>
<td>66.7%</td>
</tr>
</tbody>
</table>

Languages with special consonants by consonant inventory size
Typology for Consonant Inventories

- Place of articulation: labial, alveolar, velar > alveopalatal (pol. *zioło*) > uvular, dental, retroflex (pol. *żaba*) > pharyngeal
- Place of articulation: Most languages have one laryngeal consonant (/h/ > /ʔ/ > /ɦ/)
- Manner of articulation: Stops > fricatives, nasals
- Most languages have at least one fricative (Klamath only /s/)
- Most languages have glides /w j/ – but in some languages, /w j/ do not contrast with high vowels
- Most languages have at least one nasal (some n. American languages lack them)
Vowel Quality Inventories

• Three scales
  – Front – Central – Back
  – Close – Mid – Open
  – Rounded - Unrounded

• Typological issues
  – e.g. Front rounded vowels > back unrounded vowels (inventories like English are unusual)

• No correlation between vowel and consonant inventories (Justeson-Stephens, 1984)
Vowel Quality Inventories

Yimas (Papua New Guinea): 2
Vowel Quality Inventories

Yimas (Papua New Guinea): 2

French

German
Why such inventories and not others?

• Looking for cross-linguistic biases in the distribution of phonemes
• Most of research proposes explanations based on speech production and/or perception
  – Perceptual factors
  – Articulatory factors
    \[
    \{ \text{often compete (maximization of perceptual distinctness and minimization of articulatory effort)} \}
    \]
• Liljencrants and Lindblom (1972): Adaptive Dispersion Theory
  Phoneme inventories are preferable to the extent they possess contrasts that are maximally distinct in the perceptual domain.

• A number of other theories
  – Dispersion Focalization Theory (Schwartz et al. 1997)
  – Lindblom and Maddieson model (1988)
  – Stevens’s Quantal Theory (1972, 1989)

<table>
<thead>
<tr>
<th>3 vowels</th>
<th>Most common</th>
<th>Liljencrants and Lindblom predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>i u</td>
<td>i u</td>
<td>i u</td>
</tr>
<tr>
<td>a</td>
<td>19 of 24</td>
<td>a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 vowels</th>
<th>Most common</th>
<th>Liljencrants and Lindblom predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>i e o</td>
<td>i u</td>
<td>i u</td>
</tr>
<tr>
<td>a</td>
<td>9 of 34</td>
<td>a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3 vowels</th>
<th>Most common</th>
<th>Liljencrants and Lindblom predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>i u</td>
<td>i u</td>
<td>i u</td>
</tr>
<tr>
<td>a</td>
<td>19 of 24</td>
<td>a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 vowels</th>
<th>Most common</th>
<th>Liljencrants and Lindblom predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>i e o</td>
<td>i u</td>
<td>i u</td>
</tr>
<tr>
<td>a</td>
<td>9 of 34</td>
<td>a</td>
</tr>
</tbody>
</table>
Frequency Distributions Within Languages: Consonants

- There is a strong correlation between the typologically most common consonants and their frequency within languages
Frequency Distributions Within Languages: Vowels

• schwa occurs with greater frequency within languages than three of the cardinal vowels /e, o, u/, even though schwa is considerably less common across languages.
Phoneme Inventories: a Summary

• Although there are a large number of sounds attested in languages of the world, most languages only employ a relatively small subset of them to make contrasts.

• There are certain consonants and vowels that are much more common than others both cross-linguistically and within languages
  – sparked an extensive literature exploring the phonetic and phonological motivations shaping phoneme inventories

• Constantly evolving nature of the lexicon leads to changes in intralanguage frequency distribution of phonemes.
Phonological Rule Typology: Segmental Processes

• A more complex problem than segment inventory typology, requires more language-particular commentary and analysis

• All spoken languages have phonological rules, but not all rules are found in every language
  – may be in certain language families but not in the others, e.g. rounding harmony common in Turkic languages

• Most phenomena affecting segments may be explained by minimizing articulating effort and enhancing perceptual salience.
Segmental Processes

- Assimilation (bags [bægz])
  - Nom. sg: stol
  - Dim. (nom.sg.): stoljik
  - Loc.sg: stolje
  - ‘table’

- Long-distance assimilation (e.g. harmony)

- Dissimilation (pilgrim ← lat. peregrinus)

- Fortition, Lenition, Deletion and compensatory lengthening (p[ə]ˈtato, p[Ø]ˈtato)

- Epenthesis (e.g. oputimisuto in Japanese as syllable repair, etc.)

- Metathesis (more sporadic, more diachronic)

VC metathesis in Late Common Slavic (Townsend and Janda 1996: 60–1)

<table>
<thead>
<tr>
<th>Late Common Slavic</th>
<th>Gloss</th>
<th>Polish</th>
<th>Bulgarian</th>
</tr>
</thead>
<tbody>
<tr>
<td>gördů</td>
<td>‘enclosure’</td>
<td>grod</td>
<td>grad</td>
</tr>
<tr>
<td>golvá</td>
<td>‘head’</td>
<td>gwowa</td>
<td>glavá</td>
</tr>
<tr>
<td>sólma</td>
<td>‘straw’</td>
<td>wwoma</td>
<td>sláma</td>
</tr>
<tr>
<td>melkó</td>
<td>‘milk’</td>
<td>mleko</td>
<td>mlľako</td>
</tr>
</tbody>
</table>
Example of Assimilation: Vowel Harmony

- A type of long-distance assimilatory phonological process involving vowels
- A vowel or vowels in a word are changed to sound the same (thus "in harmony")
- In languages with vowel harmony, there are constraints on which vowels may be found near each other
- Many agglutinative languages have vowel harmony
Vowel Harmony

- **Vowel harmony**
  - Front-back (Turkish, Hungarian)
  - Height (N. Salentino)
  - Rounding (Turkish)

- **Variations in Rounding Harmony**
  - Kirghiz – all vowels assimilate in rounding to preceding vowels except that [a] does not assimilate to [u]
  - Turkish – only high vowels undergo, all round vowels trigger
  - Sakha (Yakut) – high vowels undergo, round vowels trigger; nonhigh vowels undergo if same height as trigger
  - Mongolian – only nonhigh vowels undergo, only nonhigh vowels trigger
  - Yawelmani – vowels undergo if same height as trigger
Vowel Harmony in Hungarian

háború  ‘war’
háborúról  ‘about war’
bűn  ‘guilt’
bűntelen  ‘guiltless’
bűnről  ‘about guilt’
út  ‘way’
útról  ‘about way’
keserű  ‘bitter’
keserűség  ‘bitterness’
keserűső  ‘bitter salt’
kör  ‘ring’
körút  ‘ring way’
körről  ‘about ring’
bátor  ‘brave’
bátorságról  ‘about braveness’
bátortalan  ‘not brave’
föld  ‘field’
földtelen  ‘fieldless’
burgonya  ‘potato’
burgonyaföld  ‘potato field’
sötét  ‘dark’
sötétség  ‘darkness’

1. Which words are compounds and why?
2. Which of the following words can be divided into parts?

földtan, földnek, háborúellenes, Budapest, burgonyalevés, óraütés, hőálló, bűnöző.

3. Translate into Hungarian:
guiltlessness, about field, about potato, wayless
A Syllable is:

- Well-recognized unit in linguistic analysis
- Explains the number of rhythmic units
- This number is usually equal to the number of vowels (but *little*, *eagle*)
- Easy concept: If listeners differ in syllabifying particular words, it is generally the case that both possible syllabifications can be permitted (pastry = past.ry or pas.try)
Canonical Syllable Patterns

• Which syllable types are permitted in a language, sequencing of segments within syllables

• May be different in different positions (in onset vs. in cora)
  – Italian: allows more than one C in onset position (pro.fon.do ‘deep’, tro.no ‘throne’, blat.ta ‘cockroach’) but only a single C in coda position (san.to ‘saint’, pal.ko ‘platform’, tor.ta ‘cake’)
Canonical Syllable Patterns

- CV – the most basic syllable, also within a language
- CV > CCV > CCCV and CV > V (markedness of the onset)
- CVC > CCV > CVCC > CVCCC
- CV > CVC > CCV
Syllable Structures in Languages

CV V CVC CCV CVCC ... CCCVCCCC

Simple Syllable Structure

Moderately complex syllable structure

Complex syllable structure
Simple Syllable Structures

CV    V    CVC    CCV    CVCC    ...    CCCVCCCCC

Simple Syllable Structure

Moderately complex syllable structure

Complex syllable structure

it is permitted not to have an initial consonant

Mba

Fijian

Yareba

Congo

Papua New Guinea
There are strict limits on what kinds of combinations are permitted: The second of two consonants is commonly limited to being one of a small set belonging to either “liquids” \((r, l)\) or “glides” ([w] in en. wet)
Complex Syllable Structures

CV V CVC CCV CVCC ... CCCVCCCCC

Simple Syllable Structure  Moderately complex syllable structure  Complex syllable structure

English

(C)(C)(C)V(C)(C)(C)(C)

strengths /stɛŋθs/
texts /tɛksts/
Distribution in WALS

http://wals.info/feature/12A#2/16.6/153.1
Correlations Between Syllable Complexity and Other Properties

- Maddieson (2007), based on data from WALS: finds a positive correlation between complexity of syllable structure and the number of consonants such that languages permitting more complex syllable types tend to have a greater number of consonants.

\[
\begin{array}{c}
\text{small consonant inventories} \leftrightarrow \text{simple syllable structure} \\
\text{large consonant inventories} \leftrightarrow \text{complex syllable structures}
\end{array}
\]

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Syllable structure</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple</td>
<td>Moderate</td>
</tr>
<tr>
<td>Small</td>
<td>20</td>
<td>42</td>
</tr>
<tr>
<td>Mod. small</td>
<td>13</td>
<td>70</td>
</tr>
<tr>
<td>Average</td>
<td>16</td>
<td>90</td>
</tr>
<tr>
<td>Mod. large</td>
<td>3</td>
<td>56</td>
</tr>
<tr>
<td>Large</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>273</td>
</tr>
</tbody>
</table>

Gordon (2016)
Syllable Structures: Japanese

*Given word in Japanese borrowed from English:*

redzonansu, oputimisuto, pen, endzin, medo in dz’apan, janki, noto-bukku, cupu, n’ujoku-tajmudzu, sekus’on, mota, dokuta, dzigudzagu, tikketto, indakus’on, s’okku, s’oppu, burokku, baransu, uisuki, majru, ojru, surogan, rajburari, ibuningu, bandaridzumu, intab’u, pasento, massadzi, ba, suta, atorakus’on, oba-koto, supido, dz’anaridzumu

1. *Find their English equivalents.*
2. *Translate to Japanese:*

   elevator, seal, yard, bolt, cook, trust, crane, knob, viktor, clerk, lucky, colour, supper, error

*Note: dz – a single phoneme [dz], like c [ts] in Czech*
Suprasegmentals

• Vowels and consonants: segments of which speech is composed.
• Segments are composed together to form syllables
• Suprasegmentals are superimposed on the syllables. These are other features that are known as:
  – Stress
  – Length
  – Tones
Suprasegmentals: Stress

• manifested by different acoustic properties
  – increased duration,
  – higher fundamental frequency
  – increased intensity

• Consonants and vowels in stressed syllables may undergo various fortition processes in stressed syllables ↔ Consonants and vowels in unstressed syllables may conversely display lenition effects
Databases on Stress Patterns

- **StressTyp**, **StressTyp2** (Goedemans et al. 2015) is a typological database containing information on stress and accent patterns in over 750 of the world's languages with nearly every language family represented.
  - [http://st2.ullet.net/](http://st2.ullet.net/)

- **WALS** (World Atlas of Language Structures), info on 502 languages
  - [https://wals.info/](https://wals.info/)
Suprasegmental: Stress

Stress is largely predictable based on phonological properties.

Stress is used to contrast lexical items or different morphological forms in a paradigm.

Fixed: Initial, Second, Third, Antepenultimate, Penultimate, Ultimate

Bounded: The placement of stress is sensitive to properties of syllables.

Unbounded: Stress can be anywhere.

Weight-Sensitive (variable) Stress
WALS: Fixed Stress Locations
WALS: Fixed Stress Locations

<table>
<thead>
<tr>
<th>Value</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No fixed stress (mostly weight-sensitive stress)</td>
<td>220</td>
</tr>
<tr>
<td>Initial: stress is on the first syllable</td>
<td>92</td>
</tr>
<tr>
<td>Second: stress is on the second syllable</td>
<td>16</td>
</tr>
<tr>
<td>Third: stress is on the third syllable</td>
<td>1</td>
</tr>
<tr>
<td>Antepenultimate: stress is on the antepenultimate (third from the right) syllable</td>
<td>12</td>
</tr>
<tr>
<td>Penultimate: stress is on the penultimate (second from the right) syllable</td>
<td>110</td>
</tr>
<tr>
<td>Ultimate: stress is on the ultimate (last) syllable</td>
<td>51</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>502</strong></td>
</tr>
</tbody>
</table>

Mapudungun/Araucanian

\( \text{\textit{t}i\'panto} \) ‘year’, \( \text{\textit{e\'lumu\,yu}} \) ‘give us’.

Winnebago, also known as **Ho-Chunk language** (**Hooc\=k, Hoc\=k**)

Siouan language family

\( \text{\textit{hochi\'chinik}} \) ‘boy’, \( \text{\textit{waghi\'ghi}} \) ‘ball’
Suprasegmentals: Stress

stress is largely predictable based on phonological properties

stress is used to contrast lexical items or different morphological forms in a paradigm

Fixed
Initial
Second
Third
Antepenultimate
Penultimate
Ultimate

Bounded: the placement of stress is sensitive to properties of syllables

Weight-Sensitive (variable) Stress

Unbounded (stress can be anywhere)
Weight-Sensitive Stress

https://wals.info/feature/15A#4/54.68/64.71
**Weight-Sensitive Stress: Unbounded**

**Russian**

- дорога
  - (1) dorOga /dərOgə/  
  - ‘a road’
  - (2) dorogA /dərʌgA/  
  - ‘dear’

**vowel reduction**

- a) to contrast lexical items:

  mOr’e – ‘a sea’ (Nom. Sg.)

- b) to contrast different morphological forms in a paradigm:

  mor’A – ‘seas’ (Nom. Pl.)
Weight-Sensitive Stress: Bounded

tátul — fox
nətyəłqin — hot
nuráqin — far
yəlyən — skin
néqəqin — quick
nəsəqəqin — cold
taplánətkən — he sews shoes
kəmyətək — roll up
?ítək — be
paqótkuk — run
nəλəqəqinat — white
púnəta — liver
qetúmyən — relative
píwtək — fall
nəmítəqin — skillful
túmyətəm — friend
tόtəka — walrus
kóttil — forehead
qalpúqal — rainbow
kəpərík — hold in arms
təvítətətkən — I work
píntəvəñək — throw at each other

Alyutor or Alutor is a language of Russia that belongs to the Chukotkan branch of the Chukotko-Kamchatkan languages.

Formulate the stress rules and put the stress for the following words:

Alyutor

sawat - lasso
pantawwi - boots
nəktəqin - solid
nəminəm - bouillon
Suprasegmentals: Vowel Length

Within languages that make length distinctions, short segments tend to vastly outnumber their long counterparts.

Estonian

saada /saːta/ – ‘to get’
saada /sa·ta/ – ‘send!’
sada /sata/ – ‘hundred’

Czech

šípk – ‘arrow’
šípk – ‘rosehip’ (Gen, SG)

The ratio of short-to-long vowels in 19 languages

Gordon 2016
Suprasegmentals: Tone

• The use of different pitch patterns to distinguish individual words or the grammatical forms of word
• Up to 60–70% of the world’s languages are tone languages (surprisingly sharp disagreements: different numbers in different classifications, cf. Maddieson’s (2013) in WALS)
• Tone languages are not distributed evenly throughout the world (widespread in Africa, Central America, and Southeast Asia)
WALS: Tones in languages
Suprasegmentals: Tone

No Tones

Simple tone systems

Complex tone systems

*English, Czech, German, Hebrew, Arabic, Finnish, French,* …

*Baltic*

\[ F \quad R \]

\[ F \quad R \quad L \quad F \quad B \quad F \quad \]

Lith. Latvian

\[ F \quad \text{falling (acute)} \]

\[ R \quad \text{rising (circumflex)} \]

\[ L \quad \text{level} \]

\[ B \quad \text{broken} \]

Stressed syllables containing a long vowel, diphthong, or a sonorant coda—may have one of two tones, falling (acute) tone or rising (circumflex) tone

*Serbian*

+ length, + variable stress

short falling ⟨ı⟩, short rising ⟨i⟩; long falling ⟨i⟩; long rising ⟨i⟩

*ne znam* = /nēznām/ - ‘I don’t know’

*Lithuanian, Latvian*

/khāá/ - ‘to trade’

/kh āā/ - ‘to get stuck’

/kàà/ - ‘galangal’

/kāá/ - ‘leg’

/káà/ - ‘leg’

*Thai*

Ideal tone language: Every syllable in a word is differentiated solely on the basis of tone

*Reality:* most tone languages have constraints on the distribution of tones (e.g. limited to roots and certain affixal domains)
# Serbian ↔ Russian Accents and Tones

<table>
<thead>
<tr>
<th>Serbian</th>
<th>Russian</th>
<th>Serbian</th>
<th>Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>мухоловка [mucholovka]</td>
<td>‘flytrap’</td>
<td>мрätz [moroz]</td>
<td>‘frost’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


2. Translate from Serbian into Russian:
   - нêбо [nêbo] вёдро [vedro] нôски [nôski]
References

References

• Thomas Stolz, Nicole Nau, Cornelia Stroh (Eds.) *Monosyllables. From Phonology to Typology*, Akademie Verlag, 2012.
• Charles Cairns and Eric Raimy (eds), 2011, *Handbook of the Syllable*.
• *Handbook of phonological data from a sample of the world's languages : a report of the Stanford Phonology Archive / compiled and edited by John H. Crothers ... [et al.]*
• Marilyn M. Vihman. *Phonology Archive Coding Manual*. Stanford University, 1974