A unified treatment of segments and clusters

Péter Rebrus and Péter Szigetvári
rebrus@nytud.hu / szigetvari@elte.hu
Hungarian Research Centre for Linguistics / Eötvös Loránd University
ICNGL12: Phoneme Inventories Workshop, Oslo 2021-06-18

Issues

1. Mismatches between **typological** markedness (implications) and **representational** markedness (information content): ə, ɨ, ʔ, h, or geminates are typologically marked, but representationally unmarked.
2. Problems of phonological **segmentation**: in certain cases, there is no clear-cut difference between **segments** and **sequences** of segments (e.g. prenasalization, pre-/post-aspiration, affricates, diphthongs). There might be valid phonological generalizations in a language supporting both analyses.
3. Representational problem of **gradual phonotactics**: attested language types exceed the number that can be modelled by means of an economically designed (universal) representation.

Phonological constructions

- are formal units of phonological expressions
- several kinds: segments or sequences (clusters, stress/harmonic domains etc.)
  but: similar formal and distributional properties
- the sets of **available constructions** are limited in a language
- a uniform treatment of **inventory** and **phonotactics**:
  - inventory: the available “segmental” constructions
  - phonotactics: the available “sequential” constructions
- systematic restrictions on available constructions:
  - universal: **implicational universals** between available constructions
  - language specific: the most “marked” available constructions in a language → **complexity**

Phonological constructions examined in this talk

Consonantal constructions:

1. sequences: **consonant+plosive** cluster subtype
2. segments: non-nasal stop (**plosive**) subtype

Vocalic constructions:

3. segments: short **non-low monophthong** subtype
4. sequences: **vowel+offglide** (falling diphthong) subtype

Complexity of constructions

- universal **complexity scale**: linear ordering between constructions of the same type:

  least complex cnstr. < … < moderately complex cnstr. < … < most complex cnstr.

- based on articulatory/perceptional information content (not discussed here)
- the available constructions in a language form a **contiguous** range in the scale, i.e. the available **minimal and maximal complexity** define all available constructions (emboldened) within the complexity scale:

  least ≤ ... ≤ **minimal** ≤ **...** ≤ **all others** ≤ **...** ≤ **maximal** ≤ ... ≤ most

- minimal/maximal complexities (= **complexity limits**) set up a **typology** of languages

Example 1: types of consonant+plosive (CT) clusters

- universal complexity scale for CTs:

TT <	**NT** <	**RT** <	**ST** <	**PT** <	**MT**
geminate	**hom. nasal**	**liquid**	**fricative**	**heterorg. plosive**	**heterorg. nasal**
tt kk pp ...	nt ŋk mp ...	rt lt rk lp ...	st sk ft fk ...	kt pt pk ...	mt nk ŋt ...

- cannot be derived from the sonority scale: TT < NT; NT < RT; PT < MT
- language types (examples):

*TT < **NT** < *RT < *ST < *PT (Manam)		**TT** < **NT** < *RT < *ST < *PT (Japanese)	
*TT < **NT** < **RT** < *ST < *PT (Diola Fogny)		**TT** < **NT** < **RT** < *ST < *PT (Sidamo)	
*TT < **NT** < **RT** < **ST** < *PT (Basque)		**TT** < **NT** < **RT** < **ST** < *PT (Italian)	
*TT < **NT** < **RT** < **ST** < **PT** (Russian)		**TT** < **NT** < **RT** < **ST** < **PT** (Hungarian)	

Example 2: types of non-nasal stops (T)

- universal complexity scale for Ts (simplified):

ʔ	<	t	<	k	<	p, q	<	c, ʈ, ...
glottal		**dental/alveol.**		**velar**		**labial, uvular**		**other**

- language types (examples):

*ʔ < **t** < **k** < *p,*q < *c,*ʈ	(Tamasheq)	**ʔ** < **t** < **k** < *p,*q < *c,*ʈ	(Tigre)
*ʔ < **t** < **k** < **p**,*q < *c,*ʈ	(many)	**ʔ** < **t** < **k** < **p**,*q < *c,*ʈ	(Tagalog)
*ʔ < **t** < **k** < *p, **q** < *c,*ʈ	(Aleut)	**ʔ** < **t** < **k** < *p, **q** < *c,*ʈ	(Itelmen)
*ʔ < **t** < **k** < **p**, **q** < *c,*ʈ	(Kazakh)	**ʔ** < **t** < **k** < **p**, **q** < *c,*ʈ	(Chukchi)
*ʔ < **t** < **k** < **p**,*q < **c**,*ʈ	(Hungarian)	**ʔ** < **t** < **k** < **p**,*q < **c**,*ʈ	(Nganasan)
*ʔ < **t** < **k** < **p**, **q** < **c**,*ʈ	(Koryak)	**ʔ** < **t** < **k** < **p**, **q** < **c**,*ʈ	(Nuxalk)
*ʔ < **t** < **k** < **p**,*q < *c, **ʈ**	(Bengali)	**ʔ** < **t** < **k** < **p**,*q < *c, **ʈ**	(Kurukh)
*ʔ < **t** < **k** < **p**, **q** < *c, **ʈ**	(Pashto)	**ʔ** < **t** < **k** < **p**, **q** < **c**, **ʈ**	(Hmong)

Context sensitivity

- complexity limits may **vary** depending on context: marked (≈ less sonorant) contexts constrain the available constructions, while observing contiguity:
  - minimal complexity can be higher in a “difficult” context
  - maximal complexity can be lower in a “difficult” context
- **graduality**:
  - complexity limits can differ in several ways depending on the “difficulty” of the context
  - complexity limits can change gradually by context

Context sensitivity: right-hand contexts

Occurrence of CT-clusters and plosives in different right contexts in Hungarian (monomorphemic, **voiced & voiceless,** **voiceless only**)

_ V	**TT** < **NT** < **RT** < **ST** < **PT** < **MT**	**t** < **k** < **p** < **c**	(all)
_ #	**TT** < **NT** < **RT** < **ST** < **PT**	**t** < **k** < **p** < **c**	(all)
_ r	**NT** < **RT** < **ST** < **PT**	**t** < **k** < **p**	
_ l	**NT** < **RT** < **ST**	**t** < **k** < **p** < **c**	(all)
_ n	**NT** < **RT**	**t** < **k** < **p**	
_ t/d	**NT** < **RT**	**t** < **k** < **p**	
_ k/g	**NT**	**t** < **k** < **p** < **c**	(all voiceless & gg)
_ j	**NT**	**k** < **p**	
_ s/z	**NT**	**k** < **p**	
_ ʦ/ʣ	**NT**	**k** < **p**	
_ ʃ/ʒ	—	**k** < **p**	
_ p/b	—	**p** < **c**	
_ c/ɟ	—	**c**	(only cc, ɟɟ)

Context sensitivity: left-hand contexts for CT-clusters

Occurrence of voiceless CT-clusters after **a**: and **e**: in Hungarian (monomorphemic, content words, not proper nouns, **word-finally** & **prevocalically**)

	0 TT	1 NT	2 RT	3 ST	4 PT	min … max compl.	
						_V	_#
T=coronal	*a:tt *e:tt	**a:nt(#)** **e:nt(#)**	**a:Rt(#)** **e:Rt(#)**	**a:St(#)** **e:St(#)**	**a:Pt**(*#) *e:Pt	1…4 1…3	1…3
T=velar	*a:kk *e:kk	**a:ŋk(#)** **e:ŋk**(*#)	**a:Rk**(*#) *e:Rk	**a:Sk**(*#) *e:Sk	**a:Pk**(*#) *e:Pk	1…4 1…1	1…1 —
T=labial	*a:pp *e:pp	**a:mp**(*#) *e:mp	**a:Rp**(*#) *e:Rp	**a:Sp**(*#) *e:Sp	*a:Pp *e:Pp	1…3 —	—
T=palatal	*a:cc *e:cc	*a:ɲc *e:ɲc	**a:Rc**(*#) *e:Rc	**a:Sc**(*#) *e:Sc	*a:Pc *e:Pc	2…3 —	—

Context sensitivity: left-hand contexts of plosives

Occurrence of plosives after fricatives in Hungarian (monomorphemic, content words, not proper nouns, **word-finally** & **prevocalically**)

	1 t/d	2 k/g	3 p/b	4 c/ɟ	min … max compl.	
					_V	_#
s _ z _	**st**(#) **zd**(#)	**sk**(#) **zg**(*#)	**sp**(*#) **zb**(*#)	**sc**(*#) *zɟ	1…4 1…3	1…2 1…1
ʃ _ ʒ _	**ʃt**(#) **ʒd**(#)	**ʃk**(*#) **ʒg**(*#)	**ʃp**(*#) **ʒb**(*#)	**ʃc**(*#) **ʒɟ**(*#)	1…4	1…1
f _ v _	**ft**(#) **vd**(*#)	**fk**(*#) **vg**(*#)	*fp *vb	*fc *vɟ	1…2	1…1 —
x _	**xt**(#) *xd	*xk *xg	*xp *xb	*xc *xɟ	1…1 —	1…1 —

Example 3: types of non-low vowels

- universal complexity scale for backness/roundness (simplified):

ɨ (ɯ), ə (ɤ/ʌ)	**<**	**u, o**	**~**	**i, e**	**<**	**ü, ö**
central/back unrounded		back rounded		front unrounded		front rounded

- different sets of salient properties and subsumption (partial ordering):

NONE ⊂ ROUND ~ FRONT ⊂ FRONT ∪ ROUND

- implicative structure:
  - ü typically implies u and i, but not always (e.g. Breton, Classical Attic Greek: i, ü, *u).
  - ɨ/ɯ can appear without u (e.g. Japanese, Alawa: i, ɯ *u; Arawak, Hopi: i, ɨ, *u)

Example 3: typology of short non-low vowels

min → max.							total number of vowels: languages
ɨ	<	*u	~	*i	<	*ü	3: Abelam;
ə	<	*o	~	*e	<	*ö	2: Arrernte (*ɨ), Adyghe (*ɨ)
*ɨ	<	**u**	~	**i**	<	*ü	5: Spanish, Czech; 4: Navajo (*u), Alawa (*o);
*ə	<	**o**	~	**e**	<	*ö	3: Cl. Arabic(*o,*e), Alabama(*u,*e), Amuesha(*u,*i)
ɨ	<	**u**	~	**i**	<	*ü	7: Romanian;
ə	<	**o**	~	**e**	<	*ö	6: Polish (*ə), Bulgarian (*ɨ); 4: Yupik (only i,u,ə)
*ɨ	<	**u**	~	**i**	<	**ü**	8: Finnish, non-reference Hung.; 7: Hungarian;
*ə	<	**o**	~	**e**	<	**ö**	5: Cl. Attic (*ö), Yukaghir (*ü)
ɨ	<	**u**	~	**i**	<	**ü**	10: Votic, Võro;
ə	<	**o**	~	**e**	<	**ö**	9: Estonian (*ɨ); 8: Turkish (*ə); 7: Albanian (*ɨ,*ö)

Example 4: types of vowel+offglide sequences

- universal complexity scale for “falling diphthongs”:

ej, ow	**<**	**aj, aw**	**<**	**oj, ew**	**<**	**iw, uj, uw, ij, …**
homorganic mid+high		low+high		heterorganic mid+high		other (high+V)

- **minimal & maximal complexity = ej/ow**: languages in which only homorganic diphthongs occur, usually as variants of long (mid) vowels
- **min. & max. complexity = aj/aw**: in many languages, only low-vowel initial diphthongs occur
- **min. complexity = ej/ow & max. complexity = aj/aw**: both types occur
- **max. complexity ≥ oj/ew**: non-low-vowel initial heterorganic diphthongs can occur with or without less complex types

Example 4: types of vowel+offglide sequences

*ej	<	*aj	<	*oj	(no diphthongs)	**ej**	<	*aj	<	*oj	non-ref. Hungarian (+öų)
*ow	<	*aw	<	*ew	several languages	**ow**	<	*aw	<	*ew	
						ej	<	*aj	<	*oj	Nambakaengo
						ow	<	*aw	<	*ew	Hopi
*ej	<	**aj**	<	*oj	Hausa, Arabela	**ej**	<	**aj**	<	*oj	Mandarin, Burmese
*ow	<	**aw**	<	*ew		**ow**	<	**aw**	<	*ew	
*ej	<	**aj**	<	*oj	Angas	**ej**	<	**aj**	<	*oj	Breton
						ow	<	*aw	<	*ew	
*ow	<	**aw**	<	*ew	Hungarian (limited)	**ow**	<	**aw**	<	*ew	Southern Kiwai
						ej	<	**aj**	<	*oj	Italian, Baining
						*ow	<	**aw**	<	*ew	
*ej	<	**aj**	<	**oj**	Koine Greek (+üj)	**ej**	<	**aj**	<	**oj**	Fijian, Kurdish, Old Latin,
*ow	<	**aw**	<	**ew**		**ow**	<	**aw**	<	**ew**	Cl. Attic Greek (+üj), Spanish
*ej	<	**aj**	<	**oj**	German	**ej**	<	**aj**	<	**oj**	English, Dani
*ow	<	**aw**	<	*ew		**ow**	<	**aw**	<	*ew	
*ej	<	**aj**	<	**oj**	Iranxe	*ej	<	**aj**	<	**oj**	S. Welsh
						ow	<	**aw**	<	**ew**	

Tentative parallelisms between different complexity scales 1

- zero-grade complexity: no inherent salient property (all other features come from the context)
  - (obstruent) consonants (**ʔ**, **h**): no buccal gesture involved
  - (non-low) vowels (**ɨ**, **ə**): no frontness and roundness involved
  - CT-clusters: **geminate** plosives (**TT**): no specific place and manner present in C_1
  - (vocalic sequences: **long V**: V_1 and V_2 identical, no specific place and manner present in V_2)
- 1st-grade complexity: least saliency (a minimal set of features present)
  - obstruents: unmarked dental/alveolar **t/d** or **s/z** which have the least salient properties
  - vowels: either (back) round **u/o** or front (unrounded) **i/e**
  - clusters: **partial geminates** (homorganic nasal+plosive **NT**): C_1 and C_2 minimally different (only in nasality)
  - diphthongs: **homorganic** mid V + glide (**ej**, **ow**): V_1 and V_2 minimally different (only in height)

Tentative parallelisms between different complexity scales 2

- 2nd-grade complexity: cooccurrence of polar salient properties
  - plosives: salient place features (labial: **p**, palatal: **c**, open: **q**) cooccur with occlusion
  - complex vowels: **front** and **rounded** (**ü ö**)
  - clusters: approximant+plosive (**RT**): both place and manner are specified in C_1, but unmarked: the greatest sonority difference: R > T
  - diphthongs: **low V+glide** (**aj, aw**): both height and frontness are specified for V_2, but unmarked: the greatest sonority difference: a>u,i
- higher-grade complexities: complex constructions with several different features
  - complex articulations: **secondary articulations** (**k^w, p^j, $t^ç$** etc)
  - clusters: heterorganic stop + plosive (**PT, MT**): place and/or manner are specified in C_1 and marked: smaller sonority difference
  - diphthongs: heterorganic **V+glide** (**oj, ew**): both frontness and roundness are specified,and marked: smaller sonority difference

Summary

- segmental and sequential phonological constraints can be uniformly captured by postulating phonological **constructions**
- constructions of the same type are ordered by **complexity**
- language specific restrictions can be expressed by postulating **available constructions** that constitute **contiguous** range in the complexity scale
- the typological markedness of representationally unmarked constructions can be explained by the limit on **minimal complexity**
- universal **representation** is not needed (and often not sufficient) to capture
  - (un)markedness effects
  - graduality
  - segment/sequence parallelisms

THANK YOU

& NKFI grant #119863