Gradual Phonotactics

A complexity-based typology of consonant clusters

Péter Rebrus & Péter Szigetvári

Research Institute for Linguistics, MTA & Eötvös Loránd University, Budapest

Lisbon, 2018-06-23

language typology by accessible CT constructions

cc	TT 0	NT 1	RT 2	ST 3	PT 4	MT 5	example
							Hawaiian (Maddieson 2013)
1		↔					Manam (Piggott 1999)
0–1	←	→					Japanese (Prince 1984), Pali (Zec 1998)
1–2		←	→				Diola Fogny (Piggott 1999)
0–2	←		→				Sidamo (Gouskova 2004)
1–3		←		→			Basque (Egurtzegi 2003)
0–3	←			→			Italian (Krämer 2009)
1–4		←			→		Spanish (Hualde 2014)
0–4	←				→		Hungarian (Siptár & Törkenczy)
1–5		←				→	Kashmiri (Wali & Koul 1997)
0–5	←					→	Hindi (Kachru 2006)

- MINIMUM REQUIREMENT: $\min cc \leq 1$
- MAXIMUM REQUIREMENT: $\max cc \geq 1$
- CONTIGUITY REQUIREMENT: $cc \geq \min_L cc \wedge cc \leq \max_L cc \Rightarrow cc \in L$

analogous implicational scales for segments

	0	1	2–
C+plosive cluster	tt kk pp	nt ŋk mp	rt rp rk lt lp lk...
oral stops (place)	ʔ	t k p	q c ʈ k͡p k^w...
vowels (place)	ə/ɨ	i a u	e o y ø ɯ...
approx's (manner)	w/j	r l	ʊ ɣ β...
fricatives (place)	h	s	f ʃ x θ...
diphthongs (?)	ej/ow	aj aw	oj ew uj iw...

plosives and the glottal stop

	ʔ	p t k	other	examples
				(no plosive: not attested)
0	↔			(only glottal stop: not attested)
1		↔		French, Karok, Ainu, Avar, Chuvash
0–1	←	→		Nama, Chamorro, Kanuri, Luo, Tagalog
1–2		←	→	Hungarian, Breton (c), Inuit, Uzbek (q), Diyari (c ʈ)
0–2	←		→	Bashkir (q), Wolof (c), Haida (c q), Hindi (q ʈ)

vowels

	ə	i a u	other	examples
				(no vowel: not attested)
0	↔			(only central vowel: not attested)
1		↔		Classical Arabic
0–1	←	→		Yupik
1–2		←	→	Czech (e o), Hungarian (e o y ø)
0–2	←		→	Bulgarian (e o), Albanian (e o y)

approximants

	w	l r	other	examples
				Pirahã (very rare)
0	↔			Fe?fe? (very rare)
1		↔		Nama (r), Vietnamese (l), Russian, Finnish (l r)
0–1	⟷	⟷		Japanese (ɾ), Navajo (l), Ainu (r), English (l r)
1–2		⟷	⟷	Hungarian (ʋ), Fijian, Ewe (ɣ), Koryak, Nahuatl (β)
0–2	⟷		⟷	Arrente, Lenakel (ɣ), Spanish (ɣ β)

fricatives

	h	s	other	examples
				Dyirbal (very rare)
0	↔			Hawaiian (very rare)
1		↔		Even, Pohnpeian, Akawaio, Kunimaipa
0–1	⟷	⟷		Ainu, A. Greek, Javanese, Kiowa, Khmer, Nepali, Pirahã
1–2		⟷	⟷	Maasai (ʃ), Songhai (f), French (f ʃ), Castilian (f θ x), Serbo-Croat (f ʃ x)
0–2	⟷		⟷	Chamorro (f), Yucatec (ʃ), Yoruba (f ʃ), Dutch (f x), Czech (f ʃ x), Eng (f ʃ θ)

markedness is multidimensional within a type

RT type: C_2: coronal < noncoronal; C_1: r < l

RT	+coronal	−coronal
−lateral	rt	rk rp
+lateral	lt	lk lp

ST type: C_2 and C_1: coronal < noncoronal

ST	+coronal	−coronal
+coronal	st	sk sp
−coronal	ft xt	fk xp

PT type: C_2 and C_1: coronal < noncoronal (TT excluded)

PT	+coronal	−coronal
+coronal	—	tk tp
−coronal	pt kt	pk kp

incomplete accessibility

- the accessibility of a CT-type can be INCOMPLETE
- the various CT subsets accessible are not random
- 5 (of 15) cases are predicted based on markedness:

- examples of ST subsets:

1: Lat __#

st	*sk
*ft	*fk

2a: Latin

st	sk
*ft	*fk

2b: Hun __#

st#	*sp#
ft#	*fp#

3: Eng, Finn

st	sk
ft/ht	*fk

4: Hun

st	sk
ft	fk

- examples of PT subsets:

1: Hun vd #

(dd)	*dg
*bd	*bg

2a: Hun affr #

(tsts#)	tsk#
*pts#	*pk#

2b: Lat, Eng

(X)	*tp
pt kt	*pk

3: Finnish

(tt)	tk
pt	*pk

4: Hun

(tt)	tk
pt kt	pk

gradual patterning of well-formed clusters

markedness differences between coronals (t ʦ ʧ), between noncoronals (k p/c), and in voicing in Hungarian ST clusters

ST	__t/d	__k/g	__p/b	__c/ɟ	__ʦ/ʣ	__ʧ/ʤ
s/z__	st/zd	sk/zg	sp/zb	sc/*zɟ	sʦ/*zʣ	*sʧ
ʃ/ʒ__	ʃt/ʒd	ʃk/ʒg	ʃp/ʒb	ʃc/ʒɟ	*ʃʦ	*ʃʧ
f/v__	ft/vd	fk/vg	*fp	*fc	*fʦ	*fʧ
x__	xt	*xk	*xp	*xc	*xʦ	*xʧ

accessibility statistics

ratio of accessible and potential clusters in CT types in Hungarian

	TT	NT	RT	ST	PT	MT	all
potential CTs	6	6	12	24/18*	30	15	95/87
voiceless	1	1	1	.50	.40	.07	.53
voiced	1	1	.83	.50	.13	0	.40
all	1	1	.92	.50	.27	.03	.46

* no voiced counterpart for x

context affects the accessibility of clusters

consequence: monotonically decreasing intervals of well-formed CTs

min. complexity will not be lower and max. complexity will not be higher

	TT	NT	RT	ST	PT	MT	
__V	←	→					Japanese: no CC#
__#							
__V		←	—	→			Spanish: no CC#
__#							
__V		←	—	—	—	→	Serbo-Croatian: limited CTs before #
__#		←	—	→			
__V		←	—	—	—	→	German: same CTs before V and #
__#		←	—	—	—	→	
__V	←	—	—	—	→		Estonian: final geminates
__#	←	—	—	→			
__V	←	—	—	—	→		Finnish: no final CC#
__#							

Hungarian CTs

ratios of intervocalic and word-final voiceless and voiced CTs

	TT	NT	RT	ST	PT	MT	all
all CTs	6	6	12	24	30	15	93
V__V	1	1	1	.50	.40	.07	.53
V__#	1	1	.92	.21	.10	0	.33

	DD	ND	RD	ZD	BD	MD	all
all CTs	6	6	12	18	30	15	87
V__V	1	1	.75	.50	.13	0	.39
V__#	1	.67	.50	.11	.03	0	.22

preconsonantally

like for CTV vs CT#, we find monotonically decreasing intervals in CTC

min. complexity will not be lower and max. complexity will not be higher

	TT	NT	RT	ST	PT	MT	
__V	←	→					Japanese: no CCC
__r							
__V	←			→			Italian: pre-r geminates
__r	←			→			
__V		←			→		Spanish: same CTs before V and r
__r		←			→		
__V	←				→		Hungarian: no pre-r geminates
__r		←			→		
__V	←				→		Hungarian: PTl limited (*ktl, *ptl)
__l		←		→			

CTC clusters in Hungarian

“sonority” and voicing hierarchies

	TT	NT	RT	ST	PT	MT	maximally complex example
__V	←				→	→	lɑbdɑ ‘ball’, tʃaːmtʃog ‘munch’
__r		←	→		→		ɛlɛktromoʃ ‘electric’, gɑrdroːb ‘wardrobe’
__l		←	→	→			ʃmirgli ‘sandpaper’, muskli ‘muscle’
__ʊ		←	→	→			hɑrdʊɛr ‘hardware’, uskʊɛ ‘about’
__n		←	→				—, pɑrtnɛr ‘partner’
__s		←	→				—, sfiŋks ‘sphynx’/mɑrksiʃtɑ ‘Marxist’
__t/ts		←	→				—, infɑrktuʃ ‘infarct’/apsorptsijoː ‘absorption’
__k		↔					—, pilintskaːzik ‘hesitate’
__p/c/f/ʃ							—, —

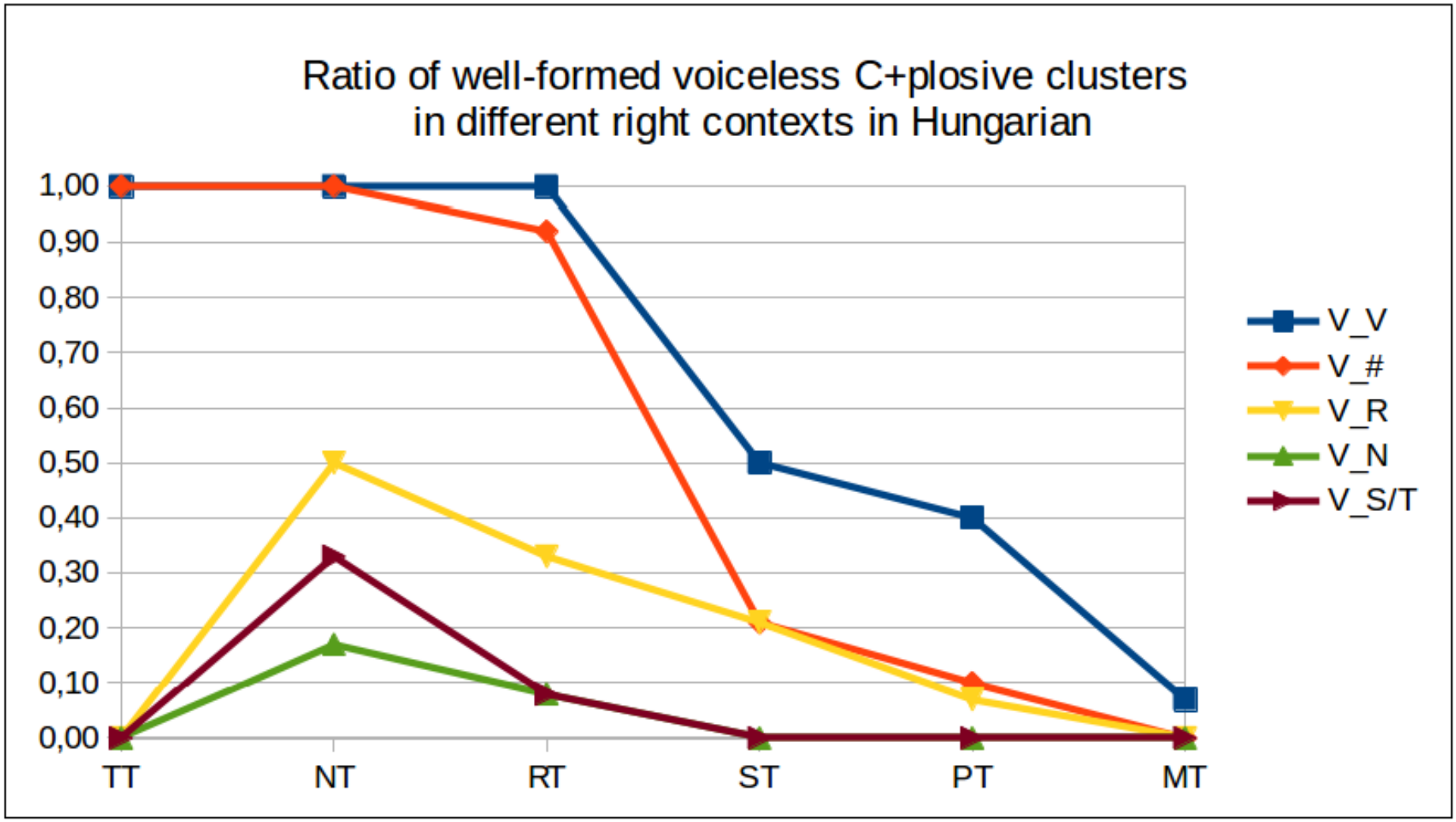
Hungarian CTs

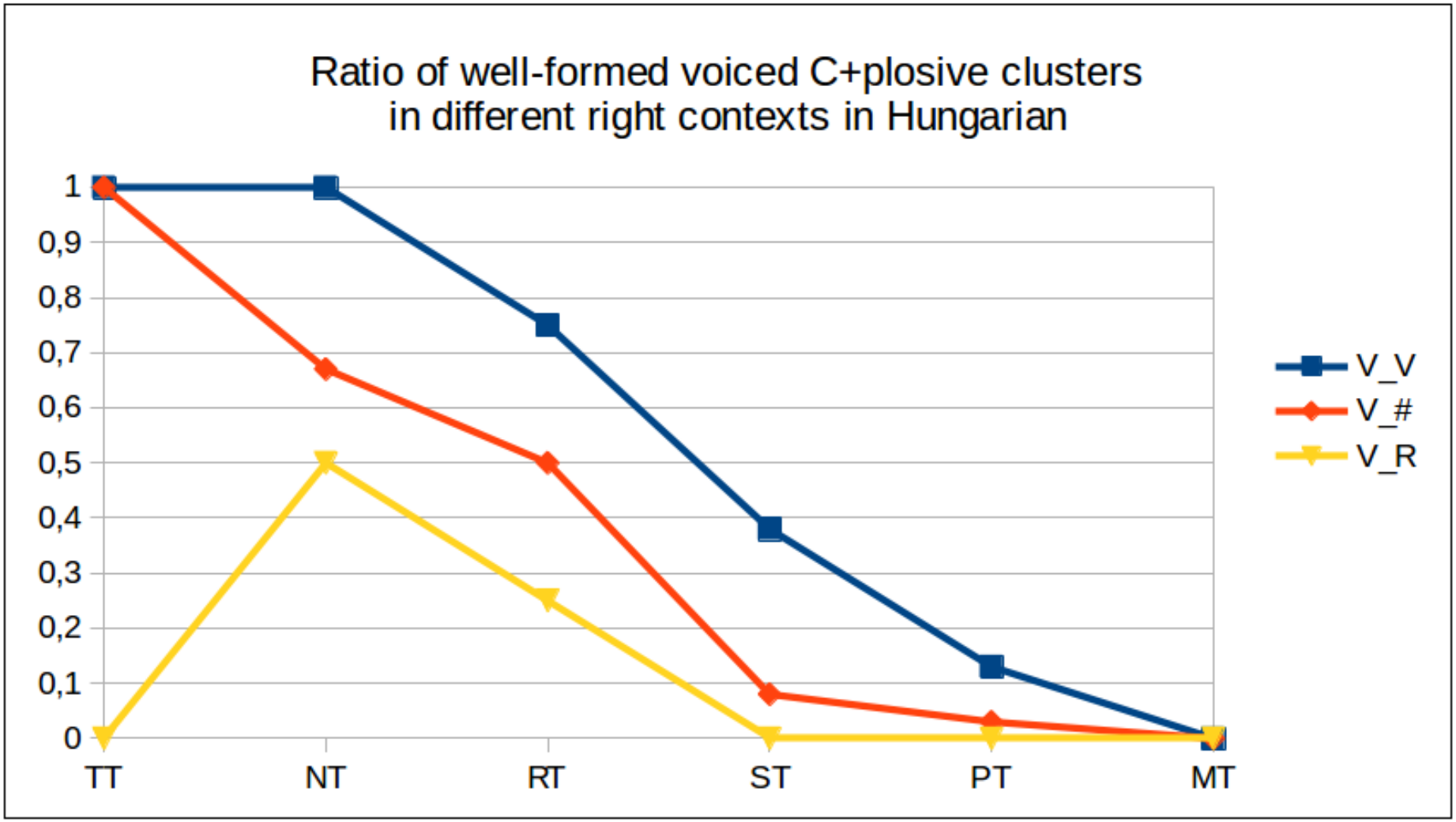
ratios of prevocalic and preconsonantal voiceless and voiced CTs

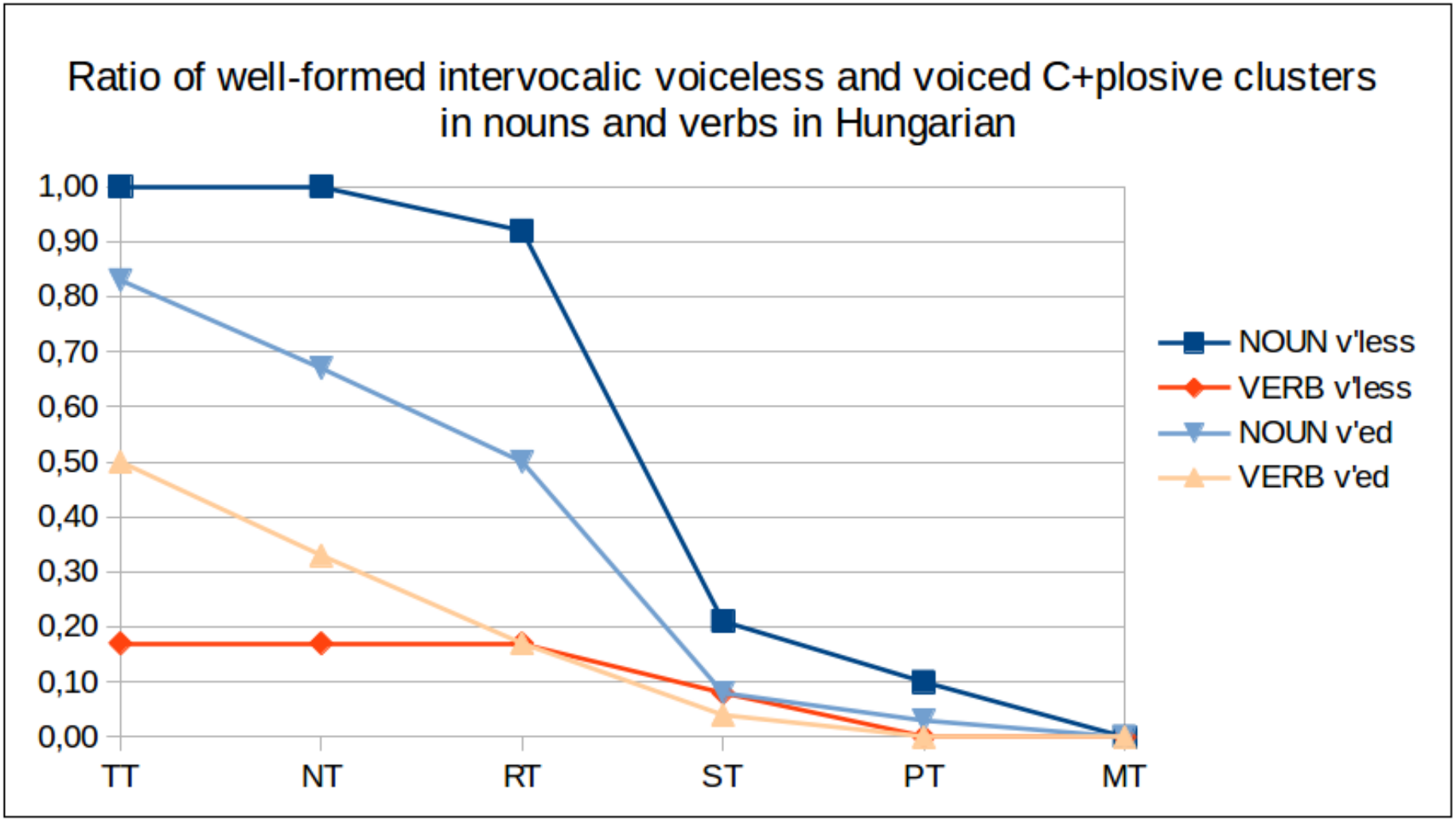
all CTs	TT 6	NT 6	RT 12	ST 24	PT 30	MT 15
V__V	1	1	1	.50	.40	.07
V__r	0	.50	.17	.21	.07	0
V__l	0	.50	.33	.08	0	0
V__ʊ	0	.33	.17	.08	0	0
V__n	0	.17	.08	0	0	0
V__s	0	.33	.08	0	0	0
V__t/ts	0	.33	.08	0	0	0
V__k	0	.17	0	0	0	0
V__p/c/f/ʃ0		0	0	0	0	0

all CTs	DD 6	ND 6	RD 12	ZD 18	BD 30	MD 15
V__V	1	1	.75	.50	.13	0
V__r	0	.33	.08	0	0	0
V__l	0	.50	.25	0	0	0
V__ʊ	0	.33	.08	0	0	0
V__n	0	0	0	0	0	0
V__s	0	0	0	0	0	0
V__t/ts	0	0	0	0	0	0
V__k	0	0	0	0	0	0
V__p/c/f/ʃ0		0	0	0	0	0

all	
v'less	v'ced
.53	.39
.13	.03
.10	.07
.06	.03
.02	0
.03	0
.03	0
.01	0
0	0


Ratio of well-formed voiceless C+plosive clusters
in different right contexts in Hungarian
1,00
0,90
0,80
0,70
0,60
0,50
0,40
0,30
0,20
0,10
0,00
TT
NT
RT
ST
PT
MT
V_V
V_#
V_R
V_N
V_S/T


Ratio of well-formed voiced C+plosive clusters
in different right contexts in Hungarian
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
TT
NT
RT
ST
PT
MT
V_V
V_#
V_R


Ratio of well-formed intervocalic voiceless and voiced C+plosive clusters in nouns and verbs in Hungarian
1,00
0,90
0,80
0,70
0,60
0,50
0,40
0,30
0,20
0,10
0,00
TT
NT
RT
ST
PT
MT
NOUN v'less
VERB v'less
NOUN v'ed
VERB v'ed

conclusions

- phonotactics is **too gradual** to be captured in a categorical manner (ie by syllable structure): the description of accessible clusters needs a very fine-grained scale
- the sets of CT clusters in a language can be profiled by contiguous intervals defined by **minimal** and maximal complexity
- a further refinement: the **edges** of the intervals are characterized by gradually descending ratios

thanks to

- you all
- the organizers
- NKFI #119863

slideshow available at

http://seas3.elte.hu/szigetva/papers.html#lisboa18