Tobias Scheer
Université de Nice
CNRS 6039

8th Manchester Phonology Meeting
18-20 May 2000

Why Moroccan Arabic tolerates anything word-initially, but Slavic does not

(1) Typology of word-initial clusters (T=Obstruent, R=Sonorant), e.g. Clements (1990)

		#CV	#TR	#RT	example
a.	#CV-only	yes	no	no	e.g. Ticuna (native indian, Colombia)
b.	#TR-only	yes	yes	no	English, French etc.
c.	#RT-only	yes	no	yes	does not exist
d.	#TR and #RT	yes	yes	yes	modern occidental Semitic, Berber, Slavic

(2) problems

a. syllabic theory was built in the 70's and 80's on the sole basis of #TR-only languages, i.e. the typical IE pattern.
==> Sonority Sequencing: "within a branching Onset, sonority increases (must increase)"

b. what about #TR-and-#RT languages? Two possible solutions:
   1. Sonority Sequencing does not operate in these languages, i.e. anything is a good branching Onset. ==> the properties of syllabic constituents are not universal, they are distributed accidentally over languages.
   2. there are no branching Onsets in these languages, they are underlyingly CVCV supported by semitic morphology

c. if Sonority Sequencing did not operate in #TR-and-#RT languages, there would be no Coda-Onset sequences because any CC is a good branching Onset ("if you want to know which cluster is a good branching Onset, look at initial clusters"). A language with branching Onsets but no Codas is contrary to typological implications: branching Onset ==> Coda, not the reverse. Hence, CVCV.

d. in any case, the theory is unable to say why #TR-only languages are possible, but not #RT-only languages, rather than the reverse.

(3) goals

a. build a theory that accounts for #TR-and-#RT languages without releasing any of the principles driving #TR-only languages.

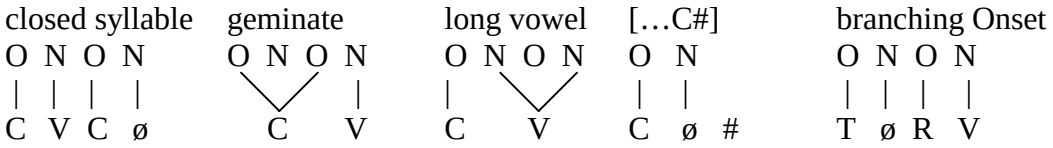
b. predict that #RT-only languages may not exist.

c. non-circularity: achieve b) without simply implementing what we observe word-initially.
   For #TR-only languages, why does the constraint say "within a branching Onset, sonority increases (must increase)" rather than the reverse? Because we observe that in these languages, it always does. This theory can do as well with a putative world where #RT-only languages do exist, but #TR-only languages do not. Build a theory that is unable to describe this kind of anti-world.

d. Slavic and Moroccan Arabic are both #TR-and-#RT languages, but still very different. Characterise this difference and understand it.

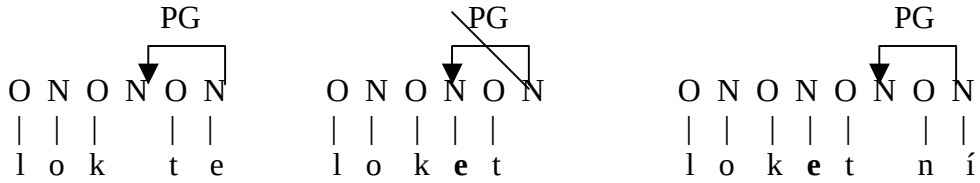
(4) tools (Government Phonology)
   a. CVCV
   b. Government (Proper and Infrasegmental)
   c. Licensing
   d. # = CV

(5) CVCV (Lowenstamm 1996, Scheer 1998,1999,2000)
syllable structure boils down to a strict sequence of non-branching Onsets and non-branching Nuclei. Hence, the following representations for basic phonological objects ensue:

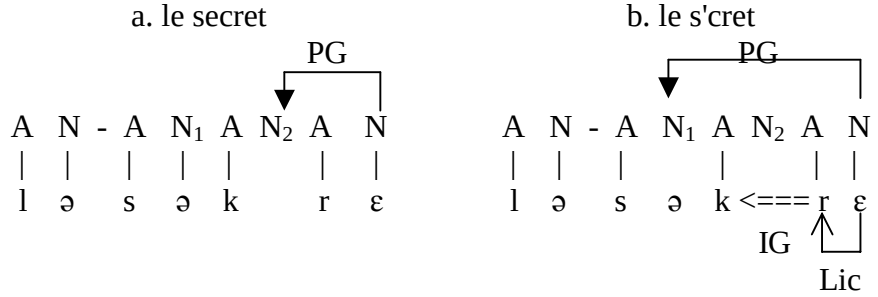


(6) Phonological Empty Category Principle (Kaye et al. 1990, Scheer 1998,1999)
an empty Nucleus may be phonetically unexpressed iff it is
   a. properly governed, or
   b. enclosed within a domain of Infrasegmental Government, or
   (c. final)

(7) Proper Government (PG) drives vowel-zero alternations (Kaye et al. 1990, Scheer 1997, 1998)
e.g. Czech *loket* "elbow"
a. *lokø**t**-e* GENsg b. *loke**t**-ø* NOMsg c. *loke**t**-ní* adj.

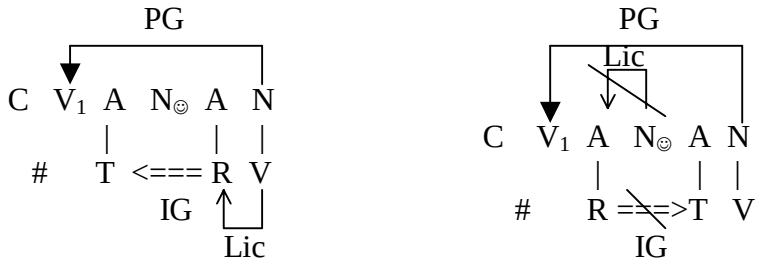


(8) Infrasegmental Government (IG) (Scheer 1996,1999)
a consonant A may govern infrasegmentally another consonant B iff
1. A is more complex than B (Harris 1990) and
2. A is licensed to do so (Charette 1990)
against the standard view, Sonorants are more complex than Obstruents.
Illustration: optional syncope of schwa in French *secret* "secret": PG reaches schwa under b) because [k] and [r] interact and thus satisfy the ECP of N_2.



(9) # = CV (Lowenstamm 1999, Ségéral&Scheer in press)
idea: morphological information enjoys a truly phonological identity in Phonology (Scheer 2000). Diacritics such as "#", "+" have no scientific status anyway, we could as well have eggs and apples in phonological theory. At best, they are variable whose identity is not understood.
Proposal: the phonological identity of the left margin of a word is an empty [CV] unit.

(10) Word-initial situation:
#RT is ruled out because, as may be seen under b), R is necessarily unlicensed because its Nucleus is empty. Therefore, it cannot interact with T, and $N_{☺}$ calls for PG from V, which is unable to govern both V_1 and $N_{☺}$.
a. well-formed structure: #TR b. ill-formed structure: #RT



(11) a. Hence, are #TR-and-#RT languages predicted not to exist?
No, the existence of #RT clusters in a language implies the absence of the initial [CV].
b. There is nothing wrong with the initial [CV] being sometimes present and sometimes absent: we know that the injection of morphological information into Phonology is an idiosyncratic matter of languages or even affixes: some suffixes build analytic phonological domains, others do not.
c. prediction if there is no initial [CV]: the distribution of word-initial clusters is free, since the empty Nucleus under d) is always governed by the first vowel of the word.
d. #TR-only: presence of the initial [CV] e. #TR-and-#RT: absence of the initial [CV]

(12) Typology predicted:

		phonological properties
a.	#CV-only	no empty Nuclei occur
b.	#TR-only	presence of the initial [CV]
c.	#RT-only	*
d.	#TR and #RT	absence of the initial [CV]

#RT-only languages are predicted not to exist because the presence of #RT-clusters implies the absence of the initial [CV]. This, in turn, releases any restriction on word-initial clusters. Hence, you can't get #RT without #TR (=absence of the initial [CV]), but you can get #TR without #RT (=presence of the initial [CV]).

(13) benefits
   a. typology is predicted by a simple paramter setting: presence vs. absence of the initial [CV]
   b. this parameter setting does not harm any of the theoretical generalisations that explain the absence of #RT clusters in #TR-only languages
   c. it is not circular: none of the tools implied have been build on the basis of word-initial data: Proper Government, Infrasegmental Government, Licensing, CVCV.

(14) a. Comparison of two #TR-and-#RT languages:
   modern occidental Semitic, e.g. Moroccan Arabic and Slavic
   b. The prediction made under (11)e is borne out in Moroccan Arabic: the distribution of #CC is free. Any sequence of two consonants that exist in this language may be observed word-initially, and the mirror sequence occurs as well.

#TR	#RT	
brɨd	rbɨT	cool down, bind
Drɨb	rDa	hit, accept
glɨʕ	lga	remove, find
bka	kbɨr	cry, grow larger
nzɨl	zna	descend, commit adultery
dna	ndɨm	come near, regret
bqa	qbɨl	stay, accept

(15) a. In Slavic, only a small subset of logically possible initial sequences do occur.
   b. the existing vs. unattested initial clusters do not appear to reduce to any regularity, nor do they constitue a natural class according to whatever criterion (sonority etc.). This is a classical problem of Slavic phonology, especially in the Polish tradition, cf. Kuryłowicz (1952), Cyran&Gussmann 1998,1999).
   c. I have tried to collect all words that bear an initial Sonorant-Obstruent cluster in all Slavic languages. The result is certainly to be amended, but it gives an idea of the situation: a total of **47** roots possess #RT-occurrences.
   The entire table showing the detail of 14 modern Slavic languages is available upon request.

(16)		Common Slavic	#RT	gloss CS	modern example
j	1	j-\-dọ	jd	walk 1sg	Cz jdu
	2	j\go	jh	yoke	Cz jho
	3	j\m	jm	seize	Cz jmout
	4	\n-	jm	name	Cz jméno
	5	j-es-m\	js	be 1sg	Cz jsem
r	6	štrXbX	rb	fragment	S-Cr rbina
	7	rXbadiga	rb	Herbaticum	Cr rbadiga
	8	r\k	rc	say, imper 2sg	Cz arch rci !
	9	uncertain	rč	hamster	S-Cr rčak
	10	rXd	rd	go red, flush	Cz rdít se
	11	str\ža	rd	core, essential	Pol rdzeń
	12	gXr(t)+dusiti	rd	strangle, choke	Cz rdousit
	13	rXdXky	rd	radish	S-Cr rdakva
	14	rufijanX	rf	procurer, pimp	Sle rfjan
	15	rusX	rs	yellow, blond	Sle rsa
	16	rXta	rt	ice-skate	Rus rta
	17	rXtXt\, rXtont\	rt	quicksilver	Cz rtut'
	18	rXt\	rt	peak, point	Cz rty (NOMpl)
	19	rXvati	rv	tear, rip, snatch	Cz rvát
	20	rXjọ	rv	dig	Cz rva (GENsg)
	21	rjuti	řv	roar, scream	Cz řvát
	22	rXž\	rž	rye	Cz rži
	23	rXzati	rž	neigh, whinny	Cz ržát
	24	drXg-	rž	tremble	U-Sor ržeć
	25	rěz-	rž	cut	Pol rżnąć

(16)		Common Slavic	#RT	gloss CS	modern example
l	26	lXb-	lb	skull	Cz lbi (GENsg)
	27	lXg-ati	lg	lie inf	Cz lhát
	28	l\g-	lg	light	Cz lhostejný
	29	lXk	lk	mourn	Cz lkát
	30	l\p-	lp	cling, stick	Cz lpět
	31	l\sk-	ls	shine, twinkle	Cz lštíti se
	32	l\st\	ls	cunning, ruse	Cz lsti (GENsg)
	33	l\v\	lv	lion GENsg	Cz lva (GENsg)
	34	sl\z	lz	tear	Pol łza
	35	lXž-	lž	spoon	Cz lžíce
m	36	mXd-lX	md	faint, weak	Cz mdlý
	37	mXchX	mch	moss	Cz dial mšina
	38	mXk	mk	sudden movement yielding an unforeseen result	Pol mknąć
	39	m\t-t\	ms	revenge	Cz msta
	40	mXstX	ms	must, fruit juice GENsg	Cz arch mstu
	41	mXtX	mt	gym swing GENsg	Cz arch mtu
	42	m\zda	mz	salary	Cz mzda
	43	mXzg-	mz	spoil	Rus mzgnut'
	44	m\ša < lat missa	mš	mass	Cz mše
	45	mXšica	mš	greenfly, aphid	Cz mšice
	46	m\chelX	mš	earnings, profit	Rus mšelX
	47	m\g-	mg	fog	Cz mhlavý

(17)

diachronic origin of modern Slavic #RT clusters				
#RT	nb roots < #RvT		uncertain origin	
	< #RyerT	< #RvT		
#jC	4	1 (5 j-es-m\)		
#rC	15	4 (14 rufijan\ 15 rusX 21 rjuti 25 rez)	1 (9 s-cr rčak)	
#lC	10	0		
#mC	12	0		
	41	5	1	Total 47

(18) diachronic generalisation

all Slavic #RT < #RvT

(19) the same holds true for Moroccan Arabic:

M. Arabic		Classical Arabic
$\#C_1 \text{ø} C_2 \text{ɨ} C_3$	<	$\#C_1VC_2VC_3$-a
k ø t ɨ b	<	k a t a b -a

diachronic correspondence: VV > V
V > ɨ

in sequences of successive schwas, every other is zero, counting from the left, cf. Amini (1997), Amimi&Bohas (1996)

(20) comparison Semitic – Slavic
- a. in both cases, **all** #RT < #RvT
- b. in the evolution towards Moroccan Arabic, ALL short vowels were centralised and further started to alternate with zero. In the pf active paradigm of Classical Arabic, the first vowel is always [a], i.e. the active morpheme. Hence, ALL first vowels fell out in this form, producing adjacency of C_1 and C_2.
- c. in Common Slavic, there were 11 vowels. Only two of them, the "soft" and the "hard" yer [\] et [X], have been centralised and further started to alternate with zero.
- d. hence, assuming a ponderated distribution of those 11 vowels in the environment #C__V, 2/11 of the Slavic lexicon was subject to $\#C_1 \text{yer} C_2 > \#C_1C_2$. Assuming a ponderated distribution of CS #TyerT, #TyerR, #RyerR and #RyerT, one fourth of those were #RyerT. Hence, 1/22 of the lexicon are the 47 roots mentioned, which supposes a total of 1034 roots in CS, a number which seems reasonable.

(21) results
- a. MA and Slavic are #RT-and-#TR languages that were #TR-only languages before (true for other #RT-and-#TR languages, or for all of them?)
- b. the diachronic change that has occurred between the former and the modern stage is the loss of the initial [CV]
  (where it is not clear whether the initial [CV] was lost because the plain vowels were centralised and started to fall out, or these vowels could start to alternate with zero because the initial [CV] was lost)
- c. the number of #RT clusters and the identity of R and T are accidental, they depend on the centralisation-process:
  all vowels are centralised ==> everything is possbile (Moroccan Arabic)
  only 2 vowels are centralised ==> C_1 and C_2 of an arbitrary subset of roots are put into contact (Slavic).
- d. hence, the distribution of initial #RT in modern Slavic is simply the reflex of the distribution of yers in CS. The key to initial #RT clusters in Slavic is a lexical property, and no phonological conditioning (natural classes etc.) is expected. This expectation meets the facts.
  Ex.: there is not a single word in Slavic where [n] is followed by an Obstruent: *nT vs. ok jT, lT, rT, mT. Reason: there was no IE root *niT,*nuT that has survived in Slavic (IE i,u > CS yers), cf. Pokorny (1959).
- e. Slavic grammar does not object to any possible initial sequence. Hence, prediction: new words that instantiate #RT clusters that are unknown may enter the language. Possible source for new #RT words: acronyms. Test in Czech:

(22)

a.	ČVUT	České vysoké učení technické
b.	LFUK	Lekařská Fakulta University Karlova
c.	JČU	Jihočeská Universita
d.	JSA	Jazyk symbolických adres
e.	LFOP	Lidová Fronta pro Osvobození Palestiny
f.	LSU	Liberální Sociální Unie
g.	LŠU	Lidová Škola Umnění

References

Amimi, Abdeljabber 1997. Contribution a l'étude de la phonologie et du lexique des parlers marocains. These de Doctorat, Université Paris 8.

Amimi, Abdeljabber & Georges Bohas 1996. Les formes nominales [CeCC] et [CCeC] en arabe marocain ou la persistance des schèmes. Revue canadienne de linguistique **41**, 1-28.

Charette, Monik 1990. Licence to govern. Phonology Yearbook **7.2**, 233-253.

Clements, George N. 1990. The role of the sonority cycle in core syllabification. Papers in Laboratory Phonology I, édité par J.Kingston & M.Beckmann, 283-333. Cambridge: Cambridge University Press.

Cyran, Eugeniusz & Edmund Gussmann 1998. Polish consonantal sequences: a phonological testing ground. Structure and Interpretation, édité par Eugeniusz Cyran, 127-138. Lublin: Pase.

Cyran, Eugeniusz & Edmund Gussmann 1999. Consonant clusters and governing relations: Polish initial consonant sequences. The syllable, Views and Facts, édité par Harry van der Hulst & Nancy Ritter, 219-248. Berlin, New York: de Gruyter.

Harris, John 1990. Segmental complexity and phonological government. Phonology Yearbook **7.2**, 255–300.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud 1990. Constituent structure and government in phonology. Phonology Yearbook **7.2**, 193-231.

Kuryłowicz, Jerzy 1952. Uwagi o polskich grupach spółgłoskowych [Remarks on Polish consonantal groups]. Biuletyn Polskiego Towarzystwa Językoznawczego **11**, 54-69.

Lowenstamm, Jean 1996. CV as the only syllable type. Current trends in Phonology. Models and Methods, édité par Jacques Durand & Bernard Laks, 419-441. Salford, Manchester: ESRI.

Lowenstamm, Jean 1999. The beginning of the word. Syllables?!, édité par John Rennison & Klaus Kühnhammer, 153-166. La Hague: Holland Academic Graphics.

Pokorny, Julius 1959. Indogermanisches Etymologisches Wörterbuch, 3e édition 1994. Tübingen, Basel: Francke.

Scheer, Tobias 1996. Une théorie de l'interaction directe entre consonnes. Thèse de doctorat, Université Paris 7.

Scheer, Tobias 1997. Vowel-zero alternations and their support for a theory of consonantal interaction. Certamen Phonologicum III, édité par P.M.Bertinetto, L.Gaeta, G.Jetchev et D.Michaels, 67-88. Torino: Rosenberg & Sellier.

Scheer, Tobias 1998. A unified model of Proper Government. The Linguistic Review **15**, 41-67.

Scheer, Tobias 1999. A theory of consonantal interaction. Folia Linguistica **32**, 201-237.

Scheer, Tobias 2000. De la localité, de la morphologie et de la phonologie en phonologie. Thèse d'Habilitation, Université de Nice.

Ségéral, Philippe & Tobias Scheer in press. The Coda Mirror. To appear in the Bulletin de la Société Linguistique de Paris.