

The syllable in domain generalization: Evidence from artificial language learning

Maya C. Wax Cavallaro

2024 Linguistics at Santa Cruz (LASC)

March 11, 2024

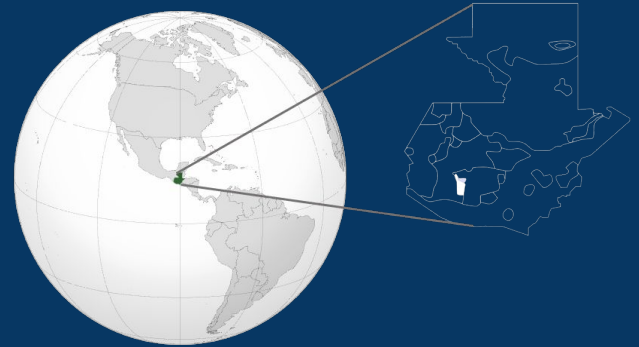


UNIVERSITY OF CALIFORNIA
SANTA CRUZ

Tz'utujil (Mayan) syllable-final approximant devoicing

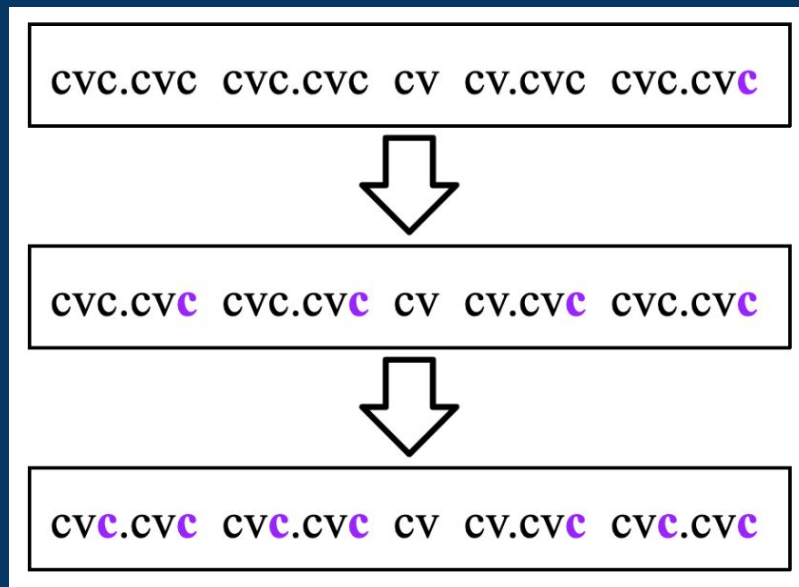
(1) Syllable-final approximant devoicing in Tz'utujil (Dayley 1985)

- | | | | |
|----|----------------|------------------------|------------------------|
| a. | <i>way</i> | [waj] | 'tortilla' |
| b. | <i>Moysees</i> | [mojse:s] | 'Moses' |
| c. | <i>kow</i> | [kɔϕ] | 'hard' |
| d. | <i>tewlaj</i> | [teϕlax] | 'very cold' |
| e. | <i>uleew</i> | [ule:ϕ] | 'land' |
| f. | <i>jul</i> | [xuɺ] | 'hole' |
| g. | <i>elnaq</i> | [eɺnaq ^h] | 'he has left' |
| h. | <i>q'or</i> | [q'ɔɺ] | 'lazy' |
| i. | <i>warnaq</i> | [waɺnaq ^h] | 'he has gone to sleep' |



Domain generalization

- Diachronic origins of certain domain-edge phenomena
 - Final devoicing, high tone retraction, vowel length neutralization...
 - (Phonetic motivation at phrase level)
- A shift in the domain of a distribution pattern from a longer, more inclusive phonological domain (e.g. the utterance) to a shorter, less inclusive one (e.g. the word).



Myers & Padgett (2014)

- 2 Artificial Language Learning experiments showing learning and generalization of final obs. devoicing from utterance-final to word-final position

Question: Is the word special?

- Do learners generalize to other (prosodic) domains?
- Is there a preference for word-level phonology?

Question: Is the word special?

- Do learners generalize to other (prosodic) domains?
- Is there a preference for word-level phonology?

Upshot: Domain generalization to the syllable is possible.

- Generalization can occur across *phonological* domains
- The syllable is accessible to the grammar/memory for necessary calculations

In this talk:

- 1) Generalization beyond the word - Why might we (not) expect it?
- 2) Artificial learning experiment
- 3) Discussion/Future directions

Generalization beyond the word

Generalization beyond the word

- General preference for word-level phonology?
 - Plenty of syllable-level phonology
 - Neutralizations, Spanish /s/-debuccalization, Tz'utujil approximant devoicing
- M&P - More words than utterances experienced/stored
 - Ah, but more syllables than words!
- Generalization beyond the word - Zec & Zsiga (2022)
 - High tone retraction in Štokavian varieties of Serbian
 - Generalization to the word and, in some varieties to the foot

The syllable as different from the word

- Primary difference
 - Words = morphosyntactic, semantic, and phonological units/domains
 - Syllables = phonological
- Two potential points of departure:
 - What structures are accessible in memory/grammar?
 - What type of domains do we generalize over?

What structures are accessible in memory / grammar?

- Are syllables even phonological domains?
 - Direct reference to syllable structure unnecessary (e.g., Ohala & Kawasaki-Fukumori 1997, Steriade 1999)
 - Some syllable structure in URs/lexicon (e.g., Vaux 2003)
- Do syllables get exemplar clouds? Is syllable structure encoded in stored words?

What types of domains do we generalize over?

- Bermudez-Otero & colleagues (e.g. Bermudez-Otero & Trousdale 2012)
 - Morphosyntactic-based stratal phonology
 - Phrase stratum → word stratum → stem stratum
 - Syllable is not a domain

Experiment

Experiment

- Based closely on Myers & Padgett (2014)
- Artificial language learning
 - Control language material participants are exposed to during a learning phase
 - Test acquisition of linguistic knowledge during a test phase
- Poverty of the stimulus
 - Classes of stimuli present in the test phase that were not included during the learning phase (but related)
 - To test generalization to novel classes

Participants

- 28 participants
- UCSC undergrad participant pool
- English-dominant

Experiment design

- Stimuli
 - Tri-syllabic words with codas in one of two positions:
 - *final coda* .CV.CV.CVC.
 - *medial coda* .CV.CVC.CV.
 - Consonants (C): [t,k,m,n,l,r] Vowels (V): [i,a,u]

Experiment design

- Stimuli
 - Tri-syllabic words with codas in one of two positions:
 - *final coda* .CV.CV.CVC.
 - *medial coda* .CV.CVC.CV.
 - Consonants (C): [t,k,m,n,l,r] Vowels (V): [i,a,u]
- Liquid neutralization pattern: /r, l/ → [l]

Experiment design

- Learning phase:
 - [l] but not [ɺ] occurs in word-final position (neutralization pattern)
 - No liquids ([l,ɺ]) in non-final codas (poverty of the stimulus)

Learning phase stimuli:

Block 1. CVVCVC.			Block 2. CVVCVC.	
nasal_coda ni ni man ki la kum lu ki kan ta ki nim ma nu tin tu la lan ri ka kin ra mi kin ri li kam su ku rin lu na lum la ra num	obst_coda ti nu lut nu ti rak ra ru lak ki ma rut mu ku lik na la nuk nu ti tat ra lu tak ti ku mit mi tu ruk ka ru tut mi ri mat	ll_coda ka na nil ta ma mul lu ta nal ra lu mil na ti nil ki ri rut ti mi kil ka ma tal ma ra lal li ni nul mu ta mil lu mu rul	nasal_coda ti mam ki na kin ka la lan ra ku nim li ti kum ka ka run ma ti mum li ma lam tu la tun mu ma run ti ki ram nu nu kam ra	obst_coda ra muk na ru kik mu li nat li mu tat nu li nut ku na nak ta na rik ta mi tat ni ki lok mu ru mit li tu lit ri ru tik ru

- Testing phase:
 - All consonants in all positions - Will participants accept coda [l]/[ɺ]?

Testing phase stimuli:

Block 1. CVVCVC.				Block 2. CVVCVC.			
nasal_coda mu lu tim la nu run nu ta lam ri ki tun ti ti tam li ki nam ma mi mun li li kum ma ra rim li ki rin nu ra nin mu nu lam mi la nin tu nu tum na ru num ku ni ram la li nam ri tu run	obst_coda ta ri kat tu ru lak ji ma tit lu ka muk mi ni nit ri ta mit ma ta kak na ru nak ra mu lut ki na mik su ri tut tu ma mat la ma mit ra ri kik ru la nak ra na lut na ru muk na ka rak	ll_coda ma nu mul ku la nul lu mi ril ka la kul ta ka lil mu nu kul ra tu ril ka lu kal ji nu kal ki mu kal li ru ral mi mi lil tu ka nul ma mi hul ku li nal ti lu lil ki ku mil nu na rul	rl_coda mi ki mir ki li tar li ra kar ri ma tir lu la kir na na tur ku ra lar lu ta mur tu mi rir ru ta tar ki tu mur ri mi kas mu ka tir ra ru lir mi ku nar lu ki tar na tu mur ku ti kir	nasal_coda ka tum na ni tan ru ra kim ka li rin ra ru mim ta mu lim tu ni mun tu ru rim lu mu man ti la mun ka ra ram lu la lin ma ta kim la ru rin ri ru mam ru na mam la ti kum lu la kun la	obst_coda ti ruk tu ni lak ma ta tik ri ri lut ra ri mik ri ma nat mu na mit ki ra luk ru ti nut ku lu kit ku li nak li lu rak ma nu rik li ru rik ki ka muk ta ki kut li ta tit nu ma kak ra	ll_coda li nul mi ta tik ri ka nal ma lu tul ti ru nil ma na tal ni ta lil ta ma nil ku li lul nu mu lak ri li mool ku ni mal na ti ril ma mu lul ri ku nul mu tu kir na ku tul ri	rl_coda tu tar ni mi mar na mi rur ku ka nar nu nu rik ki ru kir tu mu nur ki mi lur nu na mur ku ki nar la tu mur la na kas ni ka tur la ta nar ma ka kir ti ki kur ma tu kir na mi lir ma

Hypotheses

1. **Learning** - Participants will show a preference for [ɪ] over [ʊ] in *final* codas in the testing phase.

CV.CV.CV[ɪ] > CV.CV.CV[ʊ]

2. **Generalization** - Participants will show a preference for [ɪ] over [ʊ] in *medial* codas.

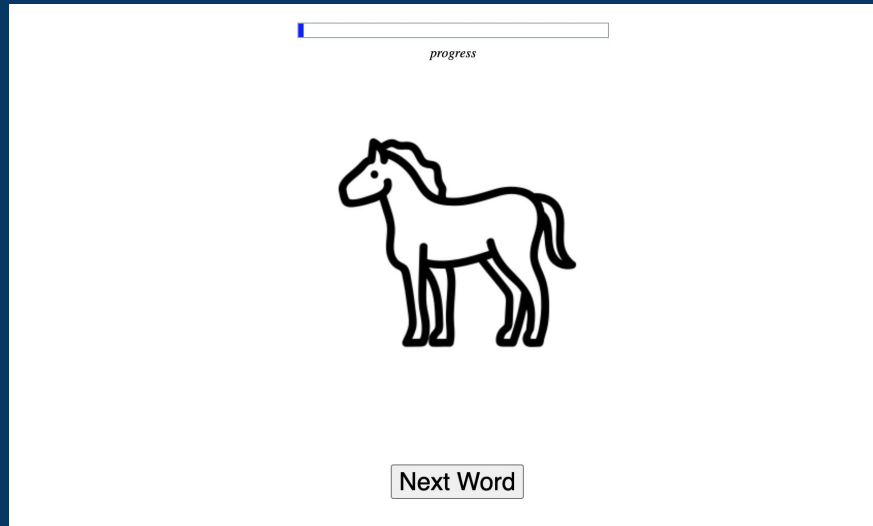
CV.CV[ɪ].CV > CV.CV[ʊ].CV

Procedure

- Online - PCIBex
 - Participants' own computers and headphones
- Self-paced

Learning phase

- Listening to trisyllabic words in a made-up language. Listen and repeat aloud.
 - No orthographic representation
 - Images of simple nouns - Each utterance is a single word (and more realistic learning task)



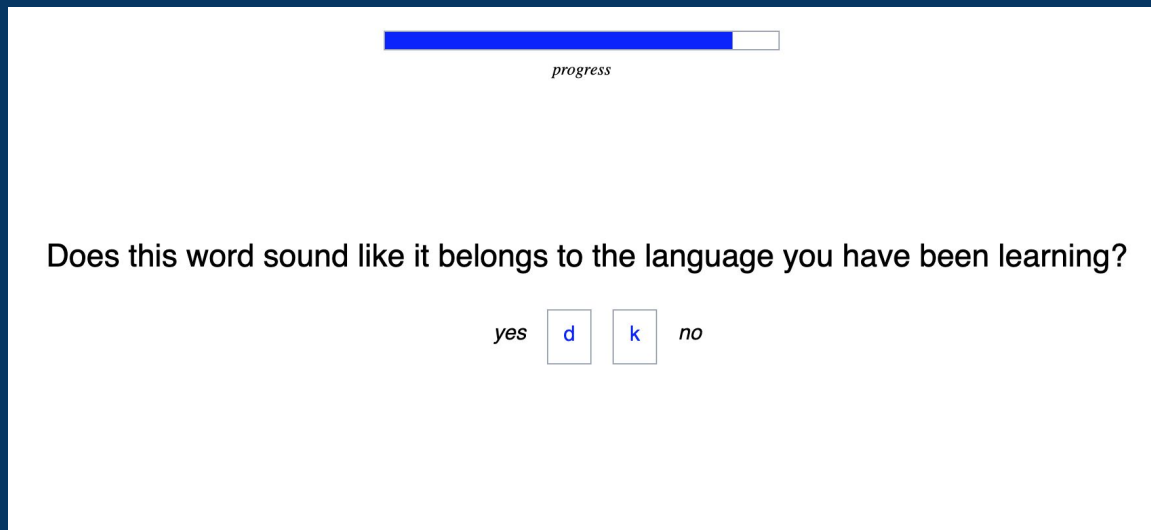
Learning phase

- Listening to trisyllabic words in a made-up language. Listen and repeat aloud.
 - Stimuli presented in blocks by coda position (final/medial)
 - Each block presented 3 times - random order within block



Test phase

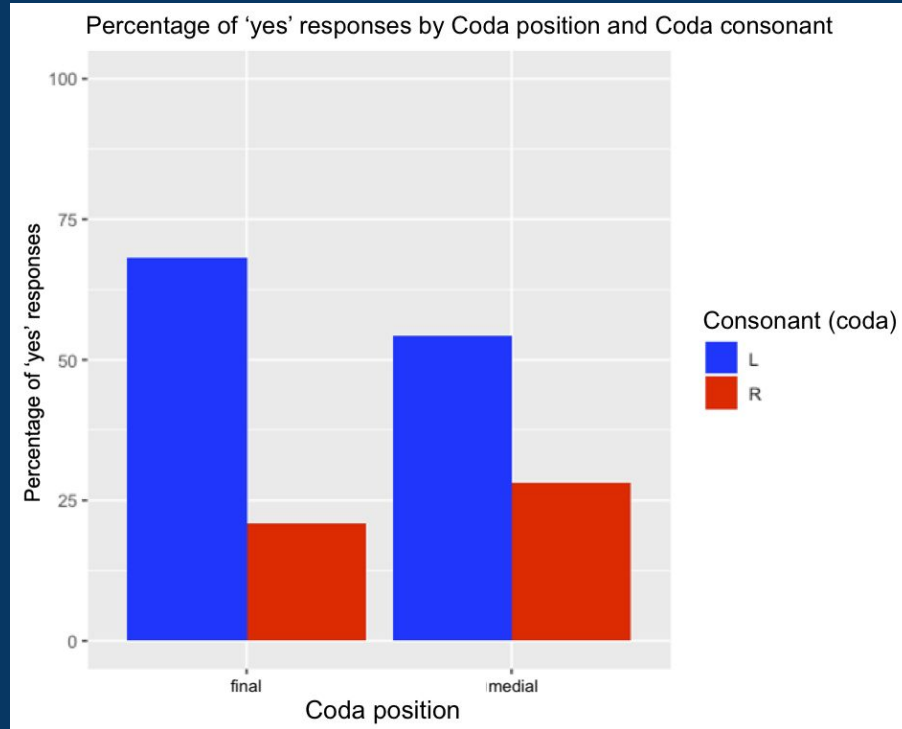
- Listen to new words and answer (yes/no)



A screenshot of a test phase interface. At the top, there is a blue progress bar followed by a small white box. Below the bar, the word "progress" is written in a small, italicized font. The main question is "Does this word sound like it belongs to the language you have been learning?". Below the question, there are two buttons: one labeled "d" and one labeled "k". The word "yes" is positioned to the left of the "d" button, and the word "no" is positioned to the right of the "k" button.

- 18 observations per condition per participant
 - 18 final [ɹ], 18 medial [ɹ], 18 final [l], 18 medial [l]

Results



Results

- Mixed effects logistic regression analysis
 - Dependent variable: response ('no'/reject or 'yes'/accept)
 - Fixed effects: liquid coda consonant ([l]/[ɹ]), coda position (final/medial)
- Significant main effects of coda consonant and coda position, as well as the interaction between the two
- Evidence of both *learning* and *generalization*

Discussion

Discussion

- Evidence of learning and generalization
 - Generalization to a smaller-than-word domain

Discussion

- Stimuli = single-word utterances, therefore word-final = utterance-final
 - Generalization from word → syllable or utterance → syllable?
 - Is generalization (necessarily) stepwise?
 - Future work could tease apart word → syllable vs. utterance → syllable generalization

Discussion

- Syllable structure?
 - Utterance/word → non-prevocalic?
 - Still a type of generalization
 - May be difficult to test with artificial learning

Thank you!

Special thank you to Jaye Padgett for sharing his research and insight and guiding me through this project, Ryan Bennett, Grant McGuire, and Rachel Walker for their ongoing support and mentorship, and Eric Baković and Scott Myers for taking time to provide insight to this banana slug. Also to Amanda Rysling, 2022 LING 290 participants, all my classmates and colleagues who have heard me discuss this research at many stages and provided fantastic questions and feedback, and to my Tz'utujil friends and collaborators who have shared their language and so much more with me and inspired much of my work. Maltyoox!

References

- Bermúdez-Otero, Rivcardo, and Graeme Trousdale. 2012. Cycles and continua: On unidirectionality and gradualness in language change. In T. Nevalainen, & E. Closs Traugott (Eds.), *The Oxford handbook on the history of English* (pp. 691-720). [55] (Oxford Handbooks Online). Oxford University Press.
- Myers, Scott, and Jaye Padgett. 2014. Domain generalisation in artificial language learning. *Phonology* 31.399-433.
- Ohala, John, and Haruko Kawasaki-Fukumori. 1997. Alternatives to the sonority hierarchy for explaining the shape of morphemes. In *Language and its ecology: Essays in memory of Einar Haugen*, ed. by Stig Eliasson and Ernst H. Jahr, 343-365. Berlin: Mouton de Gruyter.
- Pierrehumbert, Janet. 2000. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. & P. Hopper (eds.), *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- Steriade, Donca. 1999. Alternatives to the syllabic interpretation of consonantal phonotactics. In *Proceedings of the 1998 Linguistics and Phonetics Conference*, ed. by Osamu Fujimura, Brian Joseph, and Bohumil Palek, 205-242. Prague: The Karolinum Press.
- Vaux, Bert. 2003. Syllabification in Armenian, Universal Grammar, and the Lexicon. *Linguistic Inquiry*, 34(1).91–125.
- Zec, Draga, and Elizabeth Zsiga. 2022. Tone and stress as agents of cross-dialectal variation: The case of Serbian. In Kubozono, H., Ito, J & A. Mester (eds), *Prosody and Prosodic Interfaces*.

Additional materials

Syllable-level phenomena: Spanish [s] debuccalization

(1) Spanish /s/-debuccalization

	<u>orthography</u>	<u>non-debuccalizing</u>	<u>debuccalizing varieties</u>	
a.	<i>español</i>	[es'paɲol]	[eh'paɲol]	‘Spanish’
b.	<i>sostener</i>	[sos'tener]	[soh'tener]	‘sustain’
c.	<i>feliz cumpleaños</i>	[fe'lis kumple'aɲos]	[fe'lih kumple'aɲoh]	‘happy birthday’

Stimuli - Learning phase

Block 1 .CV.CV.CVC.			Block 2 .CV.CVC.CV.	
<u>nasal-coda</u>	<u>obs-coda</u>	<u>[[]-coda</u>	<u>nasal-coda</u>	<u>obs-coda</u>
.ni.ni.man	.ti.nu.lut.	.ka.na.nil.	.ti.mam.ki.	.ra.muk.na.
.ki.la.kum.	.nu.ti.rak.	.ta.mu.mul.	.na.kin.ka.	.ru.kik.mu.
.lu.ki.kan.	.ra.ru.lak.	.lu.ta.nal.	.la.lan.ra.	.li.nat.li.
.ta.ki.nim.	.ki.ma.rut.	.ra.lu.mil.	.ku.nim.li.	.mu.tat.nu.
.nu.nu.tin.	.mu.ku.lik.	.na.ti.nil.	.ti.kum.ka.	.li.nut.ku.
.tu.la.lan.	.na.la.tuk.	.ki.ri.rul.	.ka.run.ma.	.na.nak.ta.
.ri.ka.kim.	.mu.ti.tat.	.ti.mi.kil.	.ti.mun.li.	.na.rik.ta.
.ra.mi.kin.	.ra.lu.tak.	.ka.ma.tal.	.ma.lum.tu.	.mi.tat.ni.
.ri.li.kam.	.ti.ku.mit.	.mu.ra.lal.	.la.tun.mu.	.ki.lik.mu.
.nu.ku.rin.	.mi.tu.ruk.	.li.ni.nul.	.mu.run.ti.	.ru.mit.li.
.lu.na.lum.	.ka.ru.tut.	.mu.ta.mil.	.ki.ram.nu.	.tu.lit.ri.
.la.ra.num.	.mi.ri.mat.	.lu.mu.rul.	.nu.kam.ra.	.ru.tik.ru.

Stimuli - Testing phase

Block 1 .CV.CV.CVC.			
<u>nasal-coda</u>	<u>obst-coda</u>	<u>[l]-coda</u>	<u>[r]-coda</u>
.mu.lu.tim.	.ta.ti.kat.	.ma.nu.mul.	.ni.ki.mir.
.la.nu.run.	.tu.ri.lak.	.ku.la.nul.	.ki.li.tar.
.nu.ta.lan.	.ti.nu.tit.	.lu.mi.ril.	.li.ra.kar.
.ri.ki.tun.	.lu.ka.muk.	.ka.la.kul.	.ri.ma.tir.
.ti.ti.tam.	.mi.ni.nit.	.ta.ka.lil.	.lu.la.kir.
.li.ki.nam.	.ri.ta.mit.	.mu.mu.kul.	.na.na.tur.
.ma.mi.mun.	.ma.ta.kak.	.ra.tu.ril.	.ku.ra.lar.
.li.li.kum.	.na.ru.nak.	.ka.lu.kal.	.lu.ta.mur.
.ma.ra.rim.	.ra.mu.lut.	.ti.nu.kal.	.tu.mi.rir.
.li.ki.rin.	.ki.na.mik.	.ki.mu.kal.	.ru.ta.tar.
.na.ra.nin.	.nu.ri.tut.	.ti.ru.ral.	.ki.tu.mur.
.mu.nu.lam.	.tu.mi.mat.	.mi.mi.lil.	.ri.mi.lar.
.mi.la.nin.	.la.mu.mit.	.tu.ka.rul.	.nu.ka.tir.
.tu.nu.tum.	.ra.ri.kik.	.ma.mi.lul.	.ra.ru.lir.
.na.ru.nun.	.ru.la.nak.	.ku.li.nul.	.mi.ku.nar.
.ku.ni.ram.	.ra.na.lut.	.ti.lu.lil.	.lu.ki.tar.
.la.li.nam.	.na.ru.nuk.	.ki.ku.mil.	.na.tu.mur.
.ri.tu.run.	.na.ka.rak.	.nu.na.rul.	.ku.ti.kir.

Block 2 .CV.CVC.CV.			
<u>nasal-coda</u>	<u>obs-coda</u>	<u>[l]-coda</u>	<u>[r]-coda</u>
.ka.tum.na.	.ti.ruk.tu.	.ri.tal.ti.	.tu.tar.ni.
.ni.tan.ru.	.ni.lak.mu.	.li.tul.mi.	.mi.mar.na.
.ra.kim.ku.	.ta.tik.ri.	.ki.tal.ni.	.mi.rur.ku.
.li.rin.ri.	.li.lut.ra.	.mu.ral.ma.	.ka.nar.nu.
.ru.mim.ta.	.ri.mik.ri.	.lu.tul.ti.	.nu.rir.ki.
.mu.lim.tu.	.ma.nat.mu.	.ru.nil.ma.	.ru.kir.tu.
.ni.mun.tu.	.na.mit.ki.	.na.tal.ni.	.mu.nur.ki.
.ru.rim.lu.	.ra.luk.ru.	.ta.lil.ta.	.mi.lur.nu.
.mu.man.ti.	.ti.nut.ku.	.ma.nil.ku.	.na.nur.ku.
.la.mun.ka.	.lu.kit.ku.	.li.lul.nu.	.ki.nar.la.
.ra.ram.lu.	.li.nak.li.	.nu.lal.ti.	.tu.mar.la.
.la.lin.ma.	.lu.rat.ma.	.li.mul.ku.	.na.kar.ni.
.ta.kim.la.	.nu.rik.li.	.ni.nal.na.	.ka.tur.la.
.ru.rin.ri.	.ru.rit.ki.	.ti.ril.ma.	.ta.nar.ma.
.ru.mam.ru.	.ka.muk.ta.	.mu.lul.ri.	.ka.kir.ti.
.na.man.la.	.ki.kut.li.	.ku.nul.mu.	.ki.kur.ma.
.ti.kum.lu.	.ta.tit.nu.	.ku.lal.ri.	.tu.kir.na.
.la.kun.la.	.ma.kak.ra.	.ku.tul.ri.	.mi.lir.ma.

Fixed effects in a logistic regression model of responses

Significant values ($p < 0.05$) are indicated by *

	<i>b</i>	<i>z</i>	<i>p</i>
Intercept	0.8902	4.464	8.03e-06 *
Consonante coda ([ɾ])	-2.5419	-8.441	< 2e-16 *
Posición de coda (medial)	-0.7019	-2.946	0.003215 *
Coda consonant × Coda position	1.2211	3.852	0.000117 *

Effects of Coda consonant in data subsets defined by Coda position and Learning group

Significant values ($p < 0.05$) are indicated by *

Coda position	<i>b</i>	<i>z</i>	<i>p</i>
final	-2.7108	-7.19	6.5e-13 *
medial	-1.2621	-5.177	2.26e-07*