AMP 2024
November 2, 2024
Rutgers University

An experimental study of Catalan consonant alternations

Kevin Liang, Victoria Mateu, and Bruce Hayes
University of California, Los Angeles

BACKGROUND AND GOALS

(1) The Catalan consonant alternations

- They occur at the right edge of stems.
- Often, the feminine form of a stem, with [-ə], preserves the UR intact, with phonological changes in (unsuffixed) masculine.
- We examine four alternations:

Deletion of /n/	[sa**n**-ə] ~ [sa]	‘healthy fem./masc.’
Deletion of /r/	[du**r**-ə] ~ [du]	‘hard fem./masc.’
Simplification of /nt/	[sa**nt**-ə] ~ [sa**n**]	‘holy fem./masc’
Affrication of /ʒ/	[bɔ**ʒ**-ə] ~ [bɔ**tʃ**]	‘crazy fem.masc.’

(2) Rationale for this work

- Catalan phonology has played an important role in phonological theory, in computational linguistics, and in pedagogy.
- But little work has been done assessing the *productivity* of the various processes.[1]
- By studying this productivity experimentally, we can shed light on a variety of current issues in theoretical phonology.

(3) Roadmap

- Description of the phenomena and their theoretical relevance
- Describe our experiments
- Results, with discussion
- Conclusions and directions for further work

[1] We know only of Jovanovich-Trakál (2021), who, in studying 7-and 8-year-olds, found modest productivity for /n/ deletion in alternations like [klaˈfons] ~ [klaˈfo(n)] ‘wug-pl./sg.’

PHENOMENA

(4) /n/-deletion and /r/-deletion

- /n/ and /r/ are deleted in final position, following very similar patterns.
  - (Not *quite* identical: /r/ deletion also applies before plural [-s] ([du-s] 'hard-masc.pl.')
- Both processes are *lexically specific*: applicability to individual items must be memorized, in some way.
- But the exceptionality is *patterned* (Zuraw 2000): various factors influence deletion rates across the lexicon.
- The patterning is essentially the same for both /n/-deletion and /r/-deletion.
- Here are the aspects of patterning we study:
  - **Penultimately-stressed stems** virtually never undergo deletion (e.g., [əwˈtɔk.tu.**n**ə] ~ [əwˈtɔk.tu**n**] 'autochthonous-fem./masc'; [ˈprɔs.pə.**r**ə] ~ [ˈprɔs.pə**r**] 'prosperous').
  - **Monosyllabic stems**: deletion is more frequent, about half the time ([ˈsa.**n**ə] ~ [ˈsa] 'healthy', but [ˈnɛ.**n**ə] ~ [ˈnɛ**n**] 'child'; [ˈkla.**r**ə] ~ [ˈkla] 'clear', but [ˈpu.**r**ə] ~ [ˈpu**r**] 'pure').
  - **Frequent suffixes**: deletion is exceptionless with *-i(**n**a)* 'related to' and *-do**r**(a)* 'agentive' ([ər.ʒənˈti.**n**ə] ~ [ər.ʒənˈti] 'Argentine', [əd.mi.nis.trəˈdo.**r**ə] ~ [əd.mi.nis.trəˈdo] 'administrator').
  - In all **other cases**, it is *near*-exceptionless ([kə.təˈla.**n**ə] ~ [kə.təˈla] 'Catalan', [sə.ˈgu.**r**ə] ~ [sə.ˈgu] 'safe').

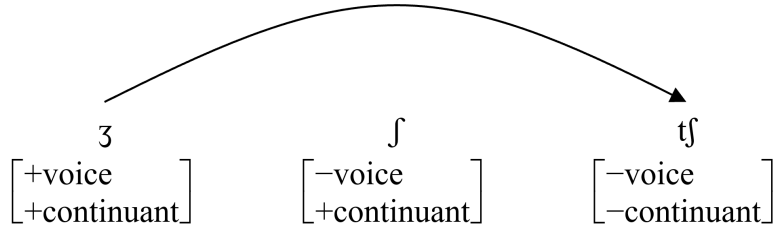
(5) /nt/ cluster simplification

- /t/ is deleted finally after /n/, as in [ˈsan.**t**ə] ~ [ˈsan] 'saint'
- Note that /nt/ cluster simplification and /n/-deletion show *counterfeeding opacity*:
  - Word-final [n] resulting from cluster simplification is never deleted (no cases like [sa**nt**-ə] ~ *[sa])
- Here is an illustration using (for brevity) classical rule-based phonology:

'holy-m.'	'holy-f.'	'healthy-m.'	'healthy-f.'	
/sant/	/sant-ə/	/san/	/san-ə/	URs
—	—	sa	—	*Final /n/ Deletion*: n → ∅ / ___]$_{word}$
san	—	—	—	*Cluster Simplification*: t → ∅ / n ___]$_{word}$
[san]	[santə]	[sa]	[sanə]	SRs

(6) **[ʒ] ~ [tʃ] alternation**

- This is a **saltatory** alternation, in the sense of Hayes and White (2015)
  - All voiced obstruents undergo devoicing in final position.
  - But [ʒ] devoices not to the expected [ʃ] but [tʃ], as in [ˈbɔ.ʒə] ~ [ˈbɔ**t͡ʃ**] 'crazy'.
  - Thus [ʒ] "saltates," jumping over intermediate [ʃ] in arriving at surface [tʃ]:



  - (We can't simply turn final [ʃ] into [tʃ]: final [ʃ] is well-formed, e.g. [ba**ʃ**] 'short').

(7) **The productivity of saltation**

- Hayes and White (2015) consider saltation to be marked, and document cases of diachronic breakdown.
- White's experimental and modeling work (artificial grammar learning) suggests a learning bias against saltation (White, 2014 in adult English-speakers; White & Sundara, 2014 in 12-month-old infants).

(8) **Three research questions**

a) What productive generalizations do learners make from exceptionful data?

- Current work suggest a two-part answer:
  - As a rough approximation (Zuraw 2000 et seq.): when using their grammar productively, language learners *frequency-match the lexicon.*
  - But they *deviate* from frequency-matching due to UG biases. (Becker et al., 2011 in Turkish; Becker et al., 2012 in English; Ernestus & Baayen, 2003 in Dutch; Hayes et al. 2009 in Hungarian)
- Catalan /n/-deletion and /r/-deletion are a good area to study, because they show clear patterns of structured exceptionality.
- How do Catalan speakers' responses differ from the lexical pattern, and why?

b) Can an opaque alternation be productive?

- As shown above in (5), /nt/ cluster simplification interacts opaquely with /n/-deletion.
- Sanders (2003) argues that Polish counterbleeding opacity is not productive and is instead dealt with by memorization.
- What of the opaque pattern in Catalan?

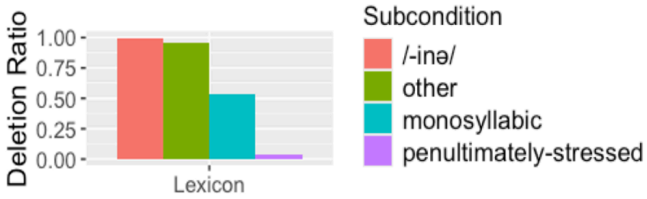
c) *Is the saltatory* [ʒ] ~ [t͡ʃ] *alternation of Catalan productive*?

- Do some speakers commit "saltation repair?"
  - Example: [səˈlɔʃ], instead of [səˈlɔt͡ʃ], for [səˈlɔʒ-ə]

(9) **Evaluating the lexical generalizations quantitatively: our database**

- The above generalizations are carefully covered in the analytical literature, notably Mascaró (1976) and Wheeler (2005).
- We reconfirmed the patterns and assessed them quantitatively by constructing and counting a corpus of 5,761 nominal and adjectival paradigms, compiled from Wiktionary.

Graph: rates of /n/ Deletion for the four environments given above in the lexicon.



- These data will appear in the graphs below as we compare the patterns seen in the wug test with the patterns of the lexicon.

OUR WUG-TEST EXPERIMENT

(10) Strategy

- A classical wug test (Berko, 1958)
- We gave the participants feminine forms, and designed the task to require them to construct the corresponding masculine, thus testing the productivity of the target phonological process.
  - **Experiment 1** (production task): given a feminine form, participants recorded themselves saying aloud the appropriate masculine form.
  - **Experiment 2** (rating task): participants rated the acceptability of two or three potential masculine forms on a scale from 1 to 7. Choices were as in table (13) below.
  - Example: asking for the masculine of [ˈfrun-ə] tests the productivity of /n/-deletion in monosyllables – will they respond with (Expt. 1) or prefer (Expt. 2) [ˈfru] or [ˈfrun]?

(11) Participants

- Adult, native speakers of Central Catalan who spoke Catalan at home and attended elementary school in Catalan.
- Participants who failed the pre-experiment training trials or control trials with real words were excluded.
- They participated remotely and they were compensated with a $15 electronic gift card.
- Participant count:
  - Experiment 1 (production task): **37** (+ 6 exclusions)
  - Experiment 2 (rating task): **37** (+14 exclusions)

(12) Materials

- We employed 100 different feminine wug forms, such as [səˈða.n-ə], for this study; any one participant saw a balanced selection of 20.
- In designing the wug forms we sought to achieve:
  - **Phonotactic acceptability** (wugs sound natural to a native speaker)
  - **Novelty** (wugs and their inflected forms were not real words of Catalan)
  - **Variegation:** they contained a wide range of distinct consonants and vowels.

(13) Sample wug forms

- There were 4 conditions and 10 subconditions as exemplified in the table below.
- The table entries correspond directly to the phonological phenomena and environments described earlier.
- Comment on possible outcomes for the masculine form:
  - First outcome: process applies
  - Second outcome: process does not apply

Condition	*Subcondition*	*Feminine form (presented to participants)*	*Anticipated masculine responses*
/n/-deletion	frequent affix /-inə/	[bəlunˈtrin-ə]	[bəlunˈtri], [bəlunˈtrin]
	monosyllabic	[ˈfrun-ə]	[ˈfru], [ˈfrun]
	penultimately-stressed	[ˈdɔstun-ə]	[ˈdɔstu], [ˈdɔstun]
	other	[gəˈmɛn-ə]	[gəˈmɛ], [gəˈmɛn]
/r/-deletion	frequent affix /-dorə/	[gruəˈdor-ə]	[gruəˈdo], [gruəˈdor]
	monosyllabic	[ˈlɛr-ə]	[ˈlɛ], [ˈlɛr]
	penultimately-stressed	[ˈsɔlir-ə]	[ˈsɔli], [ˈsɔlir]
	other	[kəˈnar-ə]	[kəˈna], [kəˈnar]
/nt/ final cluster reduction (opacity)	—	[mirbunt-ə]	[mirbun], [mirbunt], [mirbu] (feeding order)
/ʒ/ final obstruent devoicing (saltation)	—	[səˈlɔʒ-ə]	[səˈlɔt͡ʃ], [səˈlɔʃ] (final devoicing only)

(14) Frame paragraphs

- The feminine wug items were first presented once in isolation, and then embedded in frame paragraphs read by a female native speaker.
- Sample paragraph:

  WUG-fem.

  Una obra __WUG-fem__ era una peça d'art on s'havien aplicat tècniques mixtes amb ornaments de metalls i pedres precioses. Al segle XV, un artista català va crear la primera escultura ____, feta de marbre, pedres precioses, i or. El primer quadre ___ no es va crear a Espanya fins al segle XVII.

  'A ____ work was a work of art where they applied mixed media with precious metals and stone ornaments. In the 15th century, a Catalan artist created the first ___ sculpture, made of marble, precious stones and gold. The first ___ painting was not created in Spain until the 17th century.'

- The frame paragraphs were constructed with the goal of encouraging participants to interpret the stimuli as authentic Catalan words.
- The grammatical context was always one which would force the use of a masculine form of the wug word to fill the pause.

- The paragraphs were recorded such that there was a pause where a response was requested.
- The frames were presented both in spoken form and as text. However, the wugs never appeared in written form.

KEY RESULTS AND THEORETICAL INFERENCES

(15) How we report the findings

- Expt. 1 and Expt 2 yielded very similar results, so we report them together.
- We are not reporting statistical testing in this talk; generally, differences we report here test as significant; please ask us for the written paper to see full details.

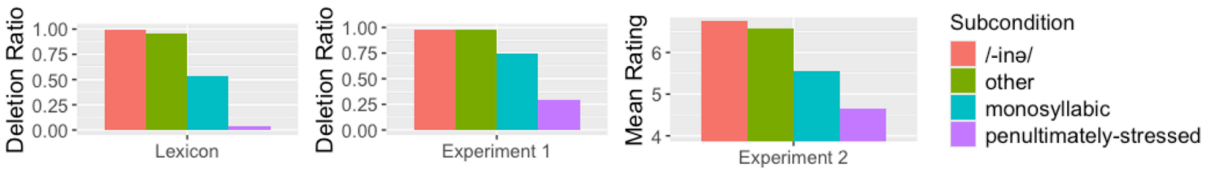
(16) General findings

- All of the processes we investigated were productive at least to some degree.
- In detail, the findings shed light on various theoretical questions.

(17) Frequency-matching in /n/-deletion

- We obtained clear evidence of frequency-matching (Zuraw 2000, Ernestus & Baayen 2003) for both experiments:

Graph: /n/-deletion in the lexicon, Experiment 1 (production), and Experiment 2 (ratings)

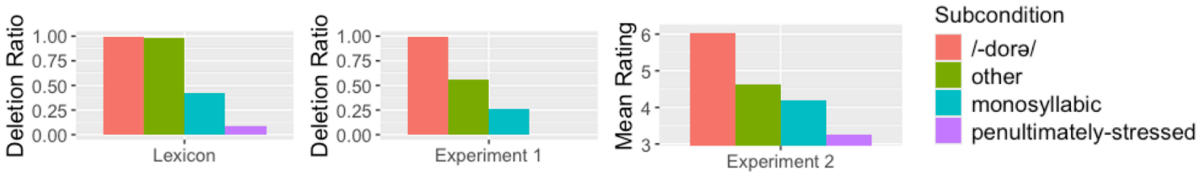


  - Experiment 1: Across four environments, the environments where /n/-deletion applies most often in the lexicon match the environments where speakers most often applied /n/-deletion: *frequent affix* > *other* > *monosyllabic stems* > *penultimately-stressed stems*
  - Experiment 2 (ratings): Same pattern (most to least acceptable)
  - *Not* "dialect mix": although there were participants who consistently deleted and others who consistently produced /n/ or /r/, most participants provided both types of answers (ditto for all other phenomena).

(18) Participants also frequency-matched for /r/-deletion

- The four contexts for /r/-deletion have similar *relative* frequencies in the lexicon and experiments.

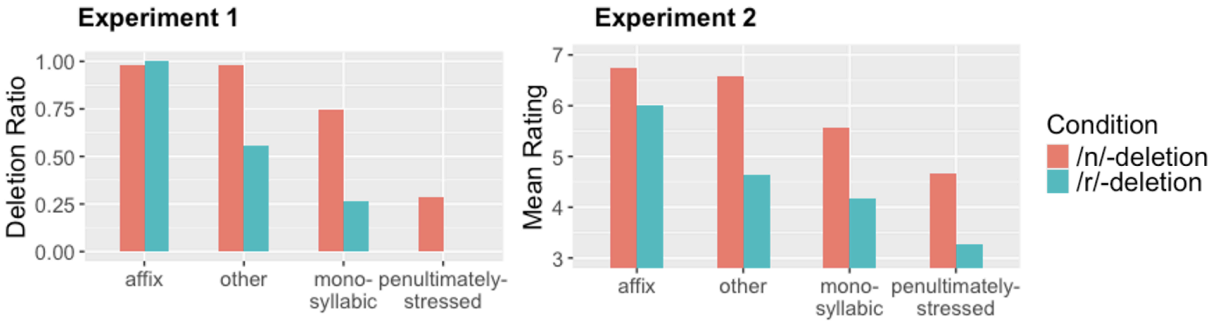
Graph: /r/-deletion in the lexicon, Experiment 1 (production), and Experiment 2 (ratings)



(19) A frequency-matching puzzle: why does /n/ delete far more often than /r/?

- [n]-deletion closely matched the lexical frequencies.
- But [r]-deletion matched only in *relative* terms:
  - Speakers consistently disfavored [r] deletion, relative to the lexical pattern

Graph: /n/- and /r/-deletion in Experiment 1 (production) and Experiment 2 (ratings)



(20) Why the /n/ - /r/ difference? Hypothesis I: dialect variation

- Speakers of Central Catalan encounter speakers of another major dialect, Valencian, which lacks /r/-deletion.
- /n/-deletion is pan-dialectal.
- For dialectology, see Wheeler (2005).

(21) Why the /n/ - /r/ difference? Hypothesis II: orthographic influence

- Previous work argues that phonological intuitions are often influenced by orthography (see Kawahara, 2018; Daland, Oh & Kim, 2015).
- In Catalan orthography, /n/-deletion is spelled out:
  - [ˈsan-ə] ~ [ˈsa] is spelled *sana* ~ *sa*
- /r/-deletion is not spelled out:

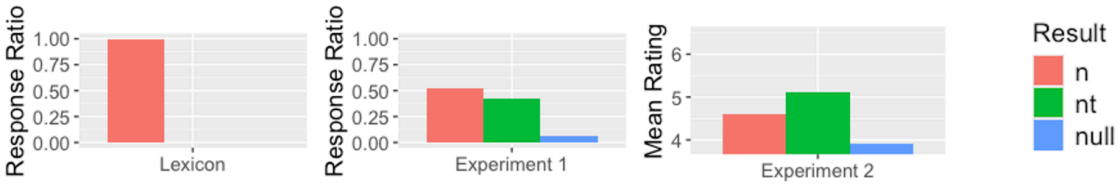
- [ˈklar-ə] ~ [ˈkla] is spelled *clara* ~ *clar*

- Rough idea: Our participants may have been constructing appropriate orthographic representations for what they heard, preferring to pronounce these representations faithfully.
- We are exploring models that could express and incorporate this influence on participant responses.

(22) /nt/ Cluster Simplification

- Despite this process being exceptionless in the lexicon, 42% of the responses in Experiment 1 had final [nt].
- Such forms were also rated higher than expected, better than forms undergoing cluster simplification.

Graph: /nt/ cluster simplification in the lexicon, Expt. 1 (production), and Expt. 2 (ratings)



- We conjecture three possibilities:
  - Exposure to other languages or other dialects of Catalan that allow final [nt] (e.g., Wheeler, 2005:221) weakens the native-language phonotactic constraint banning final [nt].
  - Orthographic influence, as above: /nt/ cluster simplification is *not* spelled out, e.g. [san] 'saint-masc.' is spelled *sant*.
  - Opacity repair: see immediately below.

(23) Results for opacity (/nt/ cluster simplification and /n/ deletion)

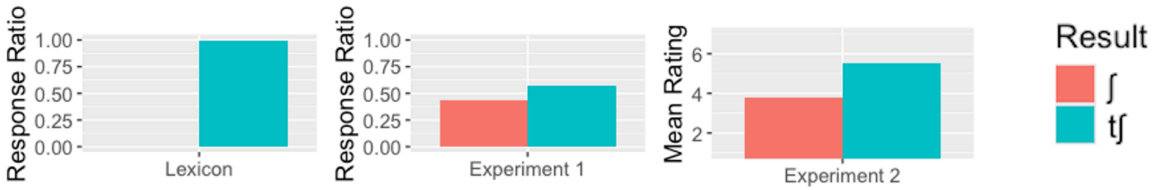
- We examined a subset of the data, in which usable responses were given for both /nt/ cluster simplification and /n/-deletion.
- This examination suggests that *counterfeeding opacity can be quite productive*: two thirds of the response patterns[2] are like [gəˈmɛn-ə] → [gəˈmɛ], [mirbuntə] → [mirbun].
- Of the remaining third, virtually all were of the type [gəˈmɛn-ə] → [gəˈmɛ], [mirbuntə] → [mirbunt]

[2] Participants received two wug words for each subcondition; where their responses treated both pairs identically, we counted the data as two response patterns.

- Opacity-related? – if you don't apply /nt/ deletion, the resulting output keeps /n/-deletion transparent.

(24) The saltatory [ʒ] ~ [tʃ] alternation

Graph: [ʒ] ~ [tʃ] *alternation in the lexicon, Expt. 1 (production), and Expt. 2 (ratings)*



- Many speakers produced forms that repaired saltation (e.g., [ʎudaʒə] → [ʎudaʃ]) and rated such forms highly.
- These saltation repairs have [ʃ], not [ʒ], because Final Devoicing remains a powerful phonotactic principle.
- Lexical basis: Forms with [ʒ] ~ [ʃ] are *not attested* in the lexicon, nor in any other dialect of Catalan.
- White (2014) and Hayes & White (2015) argue that saltation is a form of "unnatural phonology," liable to repair – we may be seeing such a case here.

SUMMARY OF FINDINGS

(25) Tentative answers to our research questions (8a-c)

- Catalan speakers generally:
  - **frequency-match the lexicon**, with deviations resulting perhaps from orthography or dialect difference
  - **can manage opacity**: Many participants gave the /nt/ → [n], /n/ → ∅ pattern.
  - **tend to repair saltation**

Moltes gràcies!

(26) Thanks to …

- Marta Camps, Roger Castells-Graells, Anna Gavarró, Mireia Marimón, Joan Mascaró, Jaume Mateu, Benet Oriol Sabat, Francesc Reda Coll, Gemma Repiso-Puigdelliura, Aina Soley Mateu, Mireia Toda Cosi
- CU Política Linguística (Generalitat de Catalunya), Institut d'Estudis Catalans, Societat Catalana de Llengua i Literatura
- Our experimental participants
- UCLA Dean of Humanities for research funding
- Members of the UCLA Phonology Seminar

References

Becker, M., Ketrez, N., & Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, 84-125.

Becker, M., Nevins, A., & Levine, J. (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language, 88*(2), 231–268.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*(2-3), 150-177.

Daland, R., Oh, M., & Kim, S. (2015). When in doubt, read the instructions: Orthographic effects in loanword adaptation. *Lingua, 159*, 70-92.

Do, Y. (2018). Paradigm uniformity bias in the learning of Korean verbal inflections. *Phonology, 35*(4), 547–575.

Ernestus, M., & Baayen, R. H. (2003). Predicting the Unpredictable: Interpreting Neutralized Segments in Dutch. *Language, 79*(1), 5–38.

Hayes, Bruce, Kie Zuraw, Peter Siptar, and Zsuzsa Londe (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85: 822-863.

Hayes, B., & White, J. (2015). Saltation and the P-map. *Phonology, 32*(2), 267–302.

Kawahara, S. (2018). Phonology and orthography: The orthographic characterization of rendaku and Lyman's Law. *Glossa, 3*(1).

Kiparsky, Paul (1971). Historical linguistics. In Dingwall, William O. (ed.), *A Survey of Linguistic Science*. College Park: University of Maryland, 576-649.

Mascaró, J. (1976). *Catalan phonology and the phonological cycle*. [Doctoral dissertation, Massachusetts Institute of Technology].

Sanders, N. (2003). *Opacity and sound change in the Polish lexicon*. [Doctoral dissertation, University of California, Santa Cruz].

Wheeler, M. W. (2005). *The phonology of Catalan*. Oxford University Press.

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition, 130*(1), 96–115.

White, J., & Sundara, M. (2014). Biased generalization of newly learned phonological alternations by 12-month-old infants. *Cognition, 133*(1), 85–90.

Zuraw, K. R. (2000). *Patterned exceptions in phonology*. [Doctoral dissertation, University of California, Los Angeles].