U. Mass. Colloquium
March 28, 2025

Bruce Hayes
UCLA

A wug-test study of Catalan consonant alternations

1. Rationale for experiments in phonology

- Classical research: examine the phonological patterning of a language and characterize it precisely with a formal grammar, letting the result bear on phonological theory.
- But does the native speaker actually *know* the pattern?
  - The data usually consist of paradigms, and paradigms can, to a fair degree, be memorized (Baayen et al. 2002; Mayer to appear)
- So, since the 1970s phonologists have recognized that it is useful to do testing to see if speakers can *productively extend* the pattern.

2. The usual experiment for this purpose is a wug-test

- Berko (1958) was apparently the first to make up new stems and test how people (both adults and children) would inflect them.
- When her consultants gave [wʌgz] as the plural of [wʌg] *wug*, they showed productive knowledge of:
  - **Morphology**: The fact that English forms plurals with a particular suffix.
  - **Phonology**: The phonological requirement for voicing agreement in obstruent sequences, so the suffix appears as [z].
- Both types of knowledge will be important here: phonology by design, morphology by serendipitous accident.
- I have enjoyed doing Wug tests with my colleagues for some time now: Albright and Hayes (2003; English), Hayes and Londe (2008, Hungarian), Hayes, Zuraw, Siptar, and Londe (2009, Hungarian)

3. How I got involved in Catalan wug-testing

- My student Yang Wang and I designed a system to learn phonological underlying representations (*Linguistic Inquiry* in press; brucehayes.org/papers/WangHayesLearningUnderlyingRepresentations.pdf)
- We wanted to provide it with the challenge of *frequency-matching* the patterns of the Catalan lexicon (more on this below).
- But alas, there was no experimental work showing that Catalan speakers really do frequency-match.
- Why not give it a try?

4. My coauthors

- **Kevin Liang**, third-year at UCLA, is really capable and had an independent interest in Catalan and experimentation, so we joined forces.

- We then begged Prof. **Victoria Mateu** (UCLA Spanish), an outstanding experimentalist and native speaker, to join us, and are very glad she agreed.

5. This talk

- Methods and results of a wug-testing study of Catalan consonant alternations, assessing their implications for issues in phonological theory:
  - frequency matching
  - productivity of opacity and saltation
  - individual differences among speakers
  - … and something quite new to me, to be explained …

THE RELEVANT ASPECTS OF CATALAN PHONOLOGY

6. Catalan

- Romance, spoken in Northeastern Spain and neighbouring areas
- Excellent descriptive/analytical work, e.g. Mascaró (1976), Wheeler (2005).
- Focus: *consonant alternations at the right edge of stems*.

7. Our lexical database

- For our purpose we need *quantitative* characterization of the key lexical generalizations encountered by Catalan during the period of language acquistion.
- We sought paradigms of words with a masculine and a feminine form, e.g. [pəˈtit], [pəˈtit-ə] ‘small-m./f.’.
- Source: Catalan Wiktionary.
- Result was a lexical database of 5761 noun and adjective paradigms, inflected for gender.
  - The whole thing was hand-checked (thank you!) by Victoria, a native speaker

8. /n/ Deletion

$n \rightarrow \emptyset$ / V ___ $]_{word}$ ‘Delete postvocalic [n] in word-final position’

- I'll use rule-based phonology (Chomsky and Halle 1968) for now, switching to MaxEnt OT later on.
- The underlying /n/ is justified by the feminine.
- /plan/ 'flat' has: [plan-ə] (f.) ~ [pla] (m.)

'flat-masc.'	'flat-fem.'	
/plan/	/plan-ə/	underlying representation
Ø	—	/n/ Deletion
[pla]	[planə]	surface representation

9. Structured exceptionality in /n/ deletion

- As in many languages, there are exceptions, e.g. [ˈnɛn-ə] ~ [ˈnɛ**n**] 'child'
- As in other languages, exceptions are not random, but obey *patterns of frequency*:
  - Forms with the adjectival suffix [-in] ~ [-i]: unanimous deletion **(105/105)**
    - [əɾʒənˈt-in-ə] (f.) ~ [əɾʒənˈt-i] (m.) 'Argentine'
  - Polysyllabic /n/ stems with stem-final stress: application *near*-unanimous **(390/410)**
    - [kətəˈlan-ə] (f.) ~ [kətəˈla] (m.) 'Catalan'
  - Monosyllabic /n/ stems: only about half of the stems undergo. **(8/15)**
    - [ˈbɔnə] (f.) ~ [ˈbɔ] (m.) 'good', but [ˈnɛnə] (f.) ~ [ˈnɛn] (m.) 'child'
  - Polysyllabic /n/ stems with stem-penultimate stress: mostly do *not* undergo **(1/27)**
    - [əwˈtɔktunə] (f.) ~ [əwˈtɔktun] (m.) 'autochthonous'
- These patterns were discovered by Mascarò and Wheeler and confirmed by our survey.

10. Structured exceptionality in phonological theory

- Research on other languages shows that similar cases, speakers *know the statistical pattern*, and imitate it stochastically when they are given new stems to inflect.
- References: Zuraw (2000) on Tagalog, Ernestus and Baayen (2003) on Dutch, and much work since then.
- An intriguing research program is to study cases where frequency-matching is distorted or overridden by UG factors; Becker et al. (2011, 2012), Jarosz (2017).
- Here we focus just on the original question: if we do a wug-test on Catalan, will there be (at least some) frequency matching? This is **Research Question 1:**

11. /r/-deletion

$r \rightarrow \emptyset$ / ___ (s)]$_{word}$

‘clear-masc.’	‘clear-fem.’	
/klar/	/klar-ə/	underlying form
∅	—	/r/ Deletion
[kla]	[klarə]	phonetic form

- Not quite the same process as /n/ Deletion; applies before /s/ as well.
- But other than that, very similar.

12. Structured exceptionality in /r/ deletion

- Forms with the agentive suffix [-dor-ə] ~ [-do]: exceptionless **(205/205)**
  - [ədministrəˈdoɾə] (f.) ~ [ədministrəˈdo] (m.) ‘administrator’

- Polysyllabic /r/ stems with stem-final stress: very few exceptions (**250/256**)
  - [preˈmer-ə] (f.) ~ [priˈme] (m.) ‘first’ *common word type*
  - [eˈmir-ə] (f.) ~ [eˈmi**r**] ‘emir’: *unusual word type*

- Monosyllabic /r/ stems: about half of all stems undergo (**3/7**)
  - [ˈklarə] (f.) ~ [ˈkla] (m.) ‘clear’, but [ˈpurə] (f.) ~ [ˈpu**r**] (m.) ‘pure’

- Polysyllabic /r/ stems with stem-penultimate stress: very few undergo (**2/24**).
  - Normal case: [ˈprɔspərə] (f.) ~ [ˈprɔspə**r**] (m.) ‘prosperous’

- So, **Research Question 1** applies for /r/ stems as well as /n/ stems.

13. /nt/ Cluster Reduction

- Broadly: homorganic stop clusters are simplified word finally.
- Here, we focus on /nt/ stems, which simplify to [n].

/nt/ Cluster Reduction: $t \rightarrow \emptyset$ / n ___]$_{word}$

  - /sant/ ‘holy’: [sant-ə] (fem.) ~ [san] (masc.)

- In the Catalan lexicon, this process is exceptionless. (**40/40**)
  - Indeed, Catalan lacks any words that end in [nt] on the surface.

14. Opacity from /nt/ Cluster Reduction

- How does /nt/ Cluster Reduction interact with /n/ Deletion?
- Answer: the [n]’s left in final position by /nt/ Cluster Reduction are *never deleted*:
  - Example: /sant/: feminine [sant-ə], masculine [san] ‘holy’, never *[sa]

- This is a familiar pattern of rule-based phonology: /nt/ Cluster Reduction *counterfeeds* /n/ Deletion (applies too late to feed).

'holy-m.'	'holy-f.'	'good-m'	'good-f.'	
/sant/	/sant-ə/	/bɔn/	/bɔn-ə/	URs
—	—	bɔ∅	—	*/n/ Deletion*: n → ∅ / ___]$_{word}$
san	—	—	—	*/nt/ Cluster Reduction*: t → ∅ / n ___]$_{word}$
[san]	[santə]	[bɔ]	[bɔnə]	surface forms

- Such cases of counterfeeding are *opaque*, in the typology set up by Kiparsky (1973).
- The *psychological reality* of opaque phonology is sometimes called into question;
  - Sanders (2003), who argues that a case of opacity in Polish is not productive and is instead dealt with by memorization.
  - Mayer (to appear) suggests extensive memorization for an opaque Uighur example.
- So we have **Research Question 2**: *can opaque phonology be productive*?
- If it is not, we expect speakers to deal with /nt/ stems in ways that don't match the language. Perhaps:
  - Applying the rules transparently (/nt/ → [n] → ∅, yielding [sa])
  - Suppress /nt/ Cluster Reduction (/nt/ → [nt], yielding [sant])

15. The alternation [ʒ] ~ [tʃ]

- Background: Catalan has a garden-variety process of Final Obstruent Voicing.

  [−sonorant] → [−voice] / ___]$_{word}$

  - Example: /griz/: [gri**s**] (m. sg.) ~ [gri**z**-us] (m. pl.) 'grey'

- But there is a special case: stems with underlying final /ʒ/ (appearing as expected in the feminine).
- We expect these to devoice to /ʃ/ in the masculine, but they don't:

  /bɔʒ/ 'crazy' feminine [bɔ**ʒ**-ə], masculine [bɔ**tʃ**], *not* *[bɔ**ʃ**]

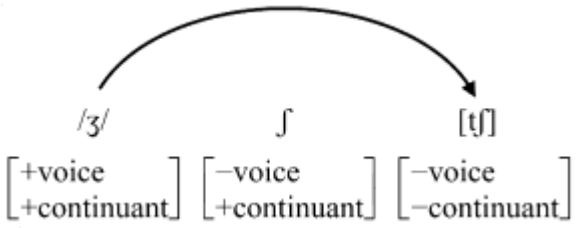
  All of the 15 or so /ʒ/ stems in Catalan behave this way.

- Note: there is nothing wrong with word-final [ʃ]; you get them from underlying /ʃ/:

  - Example: /kəlaʒ/: [kəˈlaʃ] (m. sg.) ~ [kəˈlaʃ-us] (m. pl.) 'drawer'

16. This is a case of "saltation"

- Latin for "jump".
- When a segment "leaps across" a phonetically intermediate segment which does not change:



Here, /ʒ/ leaps across phonetically intermediate [ʃ] to arrive at [tʃ].

17. Theoretical status of saltation

- An interesting challenge for Optimality Theory , which in its classical form (McCarthy and Prince 1995) cannot derive it (Łubowicz, 2002; Ito & Mester, 2003; Hayes & White, 2015).
- White's (2017) experiments suggest that people have trouble learning saltation,
- Hayes and White (2015) say that saltation is an unnatural process, with odd historical origins, vulnerable to mislearning (i.e. "repair") by subsequent generations.
    - The Catalan case is indeed the result of a complicated sequence of historical accidents.
- The natural "repair" Saltation in Catalan would be simply to devoice final /ʒ/ to [ʃ]
- **Research Question 3**: Do Catalan speakers tend to repair Saltation?

DESCRIPTION OF EXPERIMENT

18. A comment on how I will describe the experiment and its results

- Experimental work in phonology must satisfy many standards of quality.
- We have worked hard to achieve these standards, but I won't spend much time in this talk explaining what we have done, focusing instead on the phonology.
- You can email me for the paper (bhayes@humnet.ucla.edu), which is almost done, if you would like the details.

19. Making up good wug words

- They should be phonotactically normal in the language ("sound Catalan").
- They should cover a range of possible phonological shapes to cancel out unknown factors.
- Check that what we think is a wug word is not actually a real word.
    - ➢ … also that the likely participant *responses* are not real words.

20. Sample wug words, sorted by purpose

- Wug words to test /r/ Deletion: give the participants imaginary feminine **[ˈfɔrə]**, ask them what the masculine would be.
    - ➢ We expect either **[ˈfɔ]** or **[ˈfɔr]**.

- For /n/ Deletion: give the participants feminine **[məˈzinə]**, ask for the masculine.
  - We expect either **[məˈzi]** or **[məˈzin]**
- For /nt/ Cluster Reduction: give the participants feminine **[upsiˈkontə]**, ask for the masculine.
  - We expect **[upsiˈkon]**, or perhaps **[upsiˈkont]** or **[upsiˈko]**.

- Wug words to test Saltation: give participants feminine **[təlˈtiʒə]**, ask for the masculine.
  - We expect either **[təlˈtitʃ]** or **[təlˈtiʃ]** or perhaps **[təlˈtiʒ]**

- We made a total of 130 wug words, of which any one participant saw 26.[1]

21. A challenge to wug testing: often languages treat foreign words in special ways

- E.g. German uses the plural affix [-s] especially for words felt to be foreign, like *Auto* ~ *Autos* 'car', 'cars'
- We would like our participants to think that our wug words are "forgotten, but true" words of Catalan, not foreign words (the usual source of "wugs" in real life).
- Hence we present them in **frame paragraphs** designed to suggest the forgotten past, with topics like agriculture, the Middle Ages, and so on.

22. Sample frame paragraph

- Here is what the participant hears (they *see* nothing but control buttons; no text):

  [məˈzinə]

  Una obra [məˈzinə]$_1$ *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* ____$_2$*, feta de marbre, pedres precioses, i or. El primer quadre* ___$_3$ *no es va crear a Espanya fins al segle XVII.*

  'A ____ work was a piece of art where they had 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.'

23. How the experiment works: filling in the blanks

- Space #1: here, the wug word is actually read by the speaker (Victoria Mateu).
- Second blank space: this requires the participant to speak aloud — just repeating back what (s)he heard, i.e. [məˈzinə].

[1] Full disclosure: there were other stimuli testing for UR inference, as described in Dutch by Ernestus and Baayen (2003). The results were disappointing, for what we think is a boring reason, and I'll skip them here.

- Feminine is required because the adjective agrees in gender with feminine *escultura* ‘sculpture’.
- If the participant shows (s)he didn’t learn the word correctly, we throw out this trial.

- Third blank space: this requires the participant again to speak, providing the masculine.
  - Masculine is required by the frame, because the adjective agrees in gender with masculine *quadre* ‘painting’
  - They are likely to provide either **[məˈzi]** (with /n/ Deletion) or **[məˈzin]** (no /n/ Deletion)

24. Finding and testing participants

- We found native Barcelona Catalan speakers mostly by pursuing personal contacts.
- Testing took place online with LabVanced (Finger et al., 2017).
- After excluding participants for various reasons (e.g. wrong dialect) we ended up with 37.

25. Determining what the participants said

- Kevin Liang and I transcribed everything in IPA, with various controls/checks.

26. Experiment 2

- We took the most commonly volunteered forms from Experiment 1 and asked participants to rate them on a 1-7 scale.
- This gave broadly similar results, to be mentioned here only in passing.

RESULTS

27. An initial surprise: speakers sometimes contrive to not answer your question!

- We call this **avoidance**.
- Catalan offers two reasonable ways to “avoid the question”.
- A small number of masculines (41/5761) actually *have an ending*, namely [-u].
  - [buˈra.t͡ʃ-ə] (f.), [buˈra.t͡ʃ-**u**] (m.) ‘drunk’

- A small number of masculines (40/5761) are *identical to the feminine*, taking[-ə].
  - [əˈlegɾ-**ə**] (m.) ~ [əˈlegɾ-**ə**] (f.) ‘happy’).

- Participants used these unusual options in wug words (e.g. “[məzin-**u**]”) far more often than one would expect — and were rewarded by not having to make up their mind about the phonology.

28. A precedent in Korean children: Do (2018)

- Avoidance was, to my knowledge, first discovered in Do’s (2018) study.

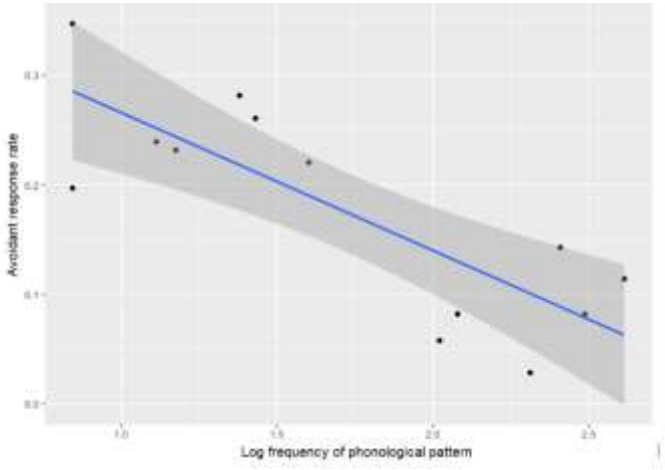
- This was an elicitation study (not wug) with Korean children, and found use of avoidant (in this case, *inappropriate*) affixes to get out of making decisions about stem-final consonant phonology.

29. Avoidance evidently reflects *difficulty*

- In our experiment avoidant responses took more time to provide:
  - 26.0 seconds vs. 19.7 seconds elsewhere; ($F(1, 915) = 24.00$, $p < 0.001$)
- The participants seem to be pondering!

30. Avoidance accompanies lexical sparsity

- Participants evidently opt for avoidance when the lexicon offers **few relevant data**.
- ARR E.g. the highest-avoidance item (37.3%) was /r/ monosyllables, with just 15 cases.
  - Log pattern frequency is correlated ($r = -0.82$, $p = 0.0007$) with avoidance rate.



31. We suggest the key factor is *uncertainty*

- In the Future Work section below, I will turn to the question of "uncertainty modeling" in phonology.
- In practical terms, we might learn more by running experiments that check the confidence of participants' responses.

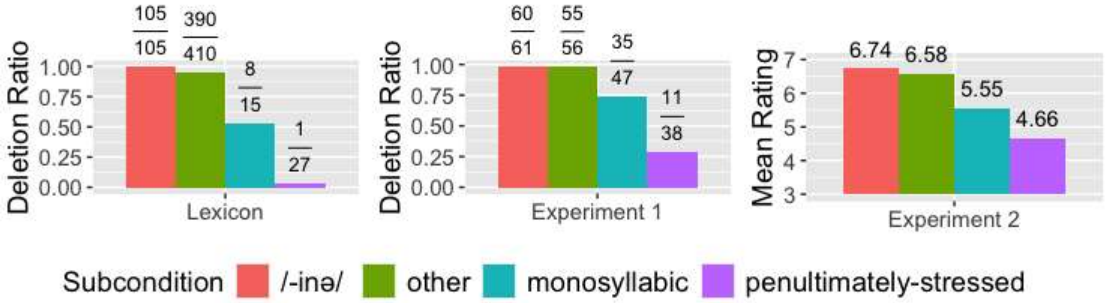
RESULTS FOR /N/ DELETION AND /R/ DELETION

32. /n/-deletion

- /n/ deletion is clearly productive, occurring quite frequently in the participants' responses.
- /n/ deletion is *frequency-matching*, with good agreement between the pattern of the lexicon and the frequency of application.

33. Graphing the /n/ deletion results

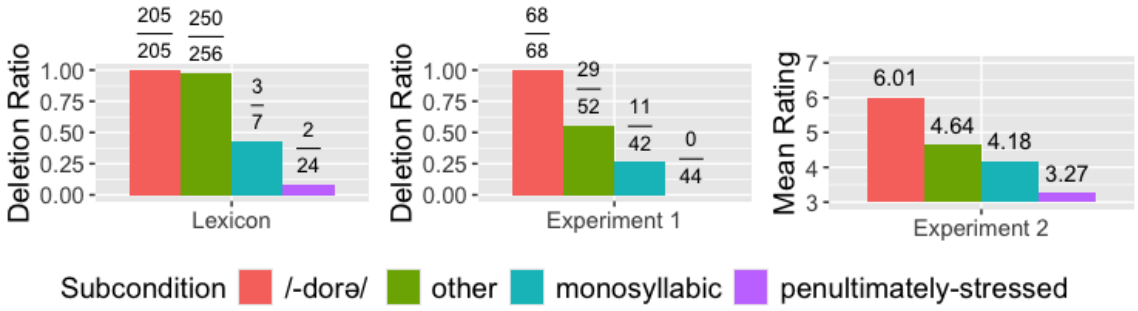
- These omit the avoidant and other aberrant responses; just phonology here.



- We are getting a pretty good matchup, in rough outline.
- Caveats to follow

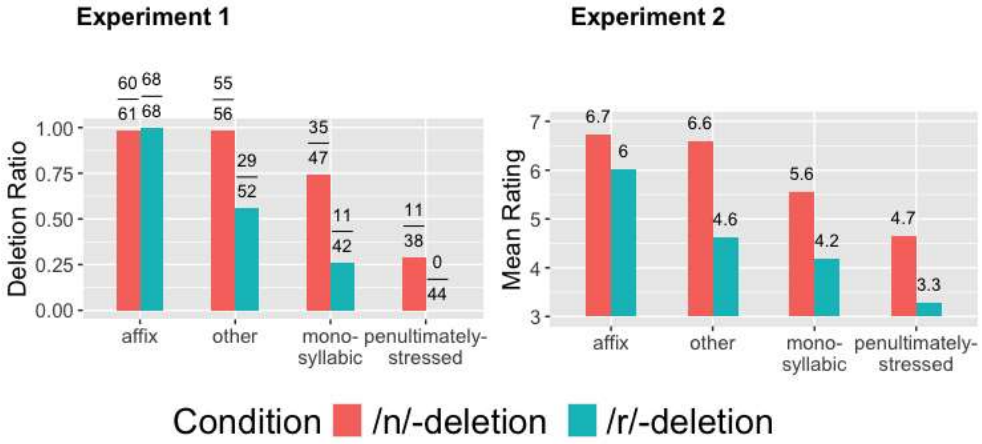
34. /r/-deletion results

- Again, the process is productive, applying frequently in novel forms.
- Again, there is frequency-matching, in the four categories examined.



35. Scrutinizing "Frequency matching" more closely: /r/-dispreference

- We see quite a *lower overall* deletion rate for /r/ than for /n/.



36. Why does /r/ deletion have a lower rate?

- Our conjecture, borrowed from Cena (1978), Kawahara (2018):
  - *Wug-test intuitions may be influenced by orthography.*
- The key difference:
  - /n/-deletion is "spelled out" in normal Catalan orthography.
    [kətəˈla] *Català* 'Catalan-m.' ~ [kətəˈlanə] *Catalana* 'Catalan-f.'
  - /r/-deletion is *not* "spelled out" in normal Catalan orthography.
    [ˈkla] *clar* 'clear-m.' ~ [ˈklarə] *clara* 'clear-f.'

37. What kind of mechanism could give rise to these patterns?

- At least some of the participants must be *making up spellings* for the words they hear.
- For /r/ Deletion
  - Hear [uliˈdar-ə], construct the orthography: *ulidara*
  - Form the orthographic masculine: *ulidar*
  - Then speakers tend to be faithful to the orthographic form, hence [uliˈdar].
  - Faithfulness to orthography is a widely observed linguistic behavior; cf. English **spelling pronunciations** like *often* [ˈɔften], *palm* [ˈpɑlm].
- For /n/ Deletion
  - Hear [səˈdanə], construct the orthography: *sedana*
  - Form the orthographic masculine: *sedan*
  - Apply "orthographic phonology" — a topic yet to be studied much by phonologists! Yields *seda*
  - Pronounce the orthographic form as [səˈda].
- The outcome of this orthographic process is perhaps blended in some way at the end with the outcome of purely phonological computation.

38. What explains the high response rates for monosyllabic and penultimately-stressed forms?

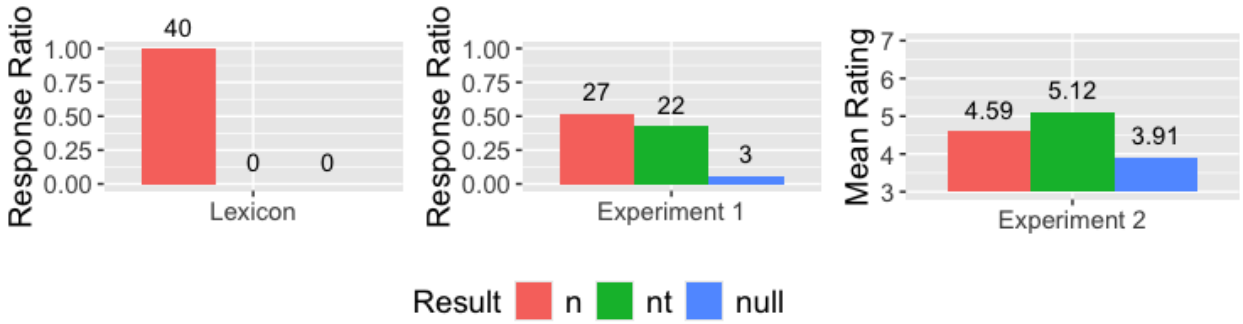
- A reasonable guess is: some combination of
  - low frequency of these cases (see (9) and (12) above
  - a bias for structural simplicity (Moreton and Pater (2012)
- These factors must be rather powerful for the case of monosyllables, in light of the evidence for a UG bias for initial-syllable (Becker et al. 2012).

RESULTS FOR /NT/ CLUSTER REDUCTION

39. Some a priori plausible outputs for the masculine of wug form [pəˈkunt-ə]

- [n] (apply the process) **[pəˈkun]** (attested in lexicon)
- [nt] (faithful) **[pəˈkunt]** (unattested in lexicon)
- Transparent **[pəˈku]** (unattested in lexicon) (pəkunt → pəkun → pəku)
- Avoidant, e.g. **[pəˈkunt-u]** (unattested in lexicon)

40. Results (just phonological outcomes, omit avoidance)



41. Discussion

- /nt/ Cluster Reduction is productive, in the sense that it does apply often to novel forms.
- But speakers often retained the final [nt].
  - Surprising: /nt/ Cluster Reduction is exceptionless in Catalan, which does not tolerate final [nt] in any word.
- One possible cause might be **widespread bilingualism**
- 43% of our participants self-report as fluent in English (which allows final [nt]).
- All speak Spanish, which permits word-final [nt] in loanwords (e.g., [esˈpɾĩn̪t̪] *esprint* 'sprint').
- L2 effects on L1 has been observed in the literature: native Russian speakers who also speak English (I remember my undergraduates!) often fail to apply final obstruent devoicing in Russian
  - Dmitrieva, Jongman, & Sereno (2010) for a phonetic study.

42. Can opaque phonology be productive?

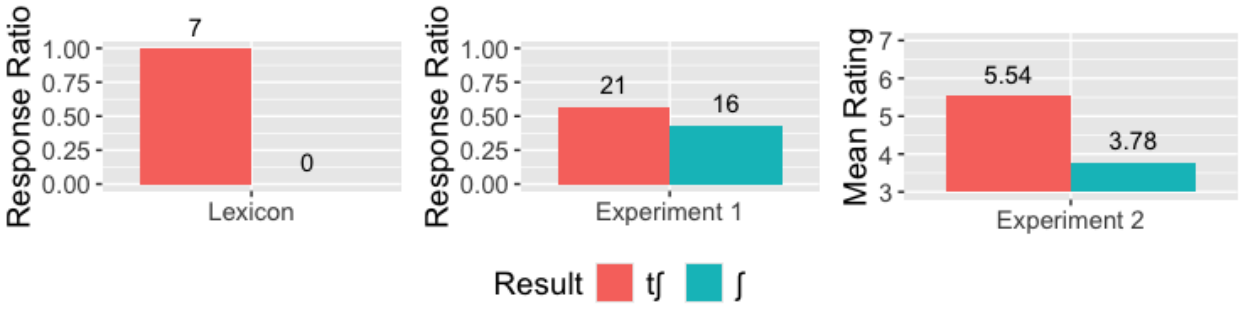
- Sure: 43% of Expt. 1 participants gave the opaque pattern, /nt/ → [n], [n] → ∅
- But a different 43% avoided opacity: /nt/ → [nt], /n/ → ∅
  - Are they suppressing /nt Cluster Reduction a way of avoiding opacity?
- Experiment 2 gives results that we are still trying to figure out …

RESULTS FOR [ʒ] → [tʃ] SALTATION

43. Sample outputs

- Sample outputs for the wug form **[ʎuˈdaʒə]**
  - **[ʎuˈdatʃ]** (attested in lexicon)
  - **[ʎuˈdaʃ]** (unattested in lexicon)

44. Experimental results



45. Discussion

- Experiment 1 yielded an impressive number of "saltation repairs", with phonetically sensible [ʃ] — these are never heard by speakers, yet they produce them.
- But still, the favorite response was the lexicon-faithful [tʃ].
- Like opacity, learning saltation seems to be difficult but not impossible.
- Further, the difficulty might be the result of low-frequency learning data.

DESCRIBING THE INDIVIDUAL SPEAKER

46. Moving beyond the "aggregate the data" approach

- The long-standing tradition in experimentation is to work with several dozen participants and pool the results.
- This is necessary, since only when you have a lot of data can you get statistically significant results.
- But do we know that all the speakers have the same internalized system?
- A new trend in phonology is to try to gain insight into individual speakers — see Yu et al. (2013) for a useful overview.

47. How we searched for speaker-specific traits

- Check responses against speaker-reported traits (education, sex, knowledge of other languages).
- This yielded only one effect — older speakers, educated in Spanish under Fascism, have a smaller difference between /n/ deletion and /r/ deletion — expected under our orthography-story.

48. Patterns that generalize across stimulus types

- **"Avoidant" speakers**: huge differences in how much speakers use the avoidance strategies described in (27)
- **Faithful speakers**: there are speakers who especially favor outputs that are obey the Faithfulness constraints of an OT analysis (so they tend to avoid /n/ and /r/ deletion, /nt/ Cluster reduction, and avoid saltating)

49. How can we detect avoidant or faithful speakers?

- Switch from rule-based phonology to **MaxEnt grammars** (Goldwater and Johnson 2003, Hayes and Wilson 2008)
- MaxEnt grammars are like Optimality Theory, only constraints are not ranked by weighted (given a number, expressing their strength).
- They output probabilities rather than yes/no outcomes, and so are ideal for data of the kind dealt with here.
- They also permit **statistical testing** of whether a constraint is meaningful/useful.

50. Here is a MaxEnt grammar for our data

Constraint	*Meaning*	*Lexicon Weight*	*Expt. Weight*
*U-THEMEVOWEL	Avoid final [-u] in masculine forms	6.61	24.29
*A-THEMEVOWEL	Avoid final [-ə] in masculine forms	6.96	24.02
*CODA-[n]	Avoid [n] in coda position	3.99	8.52
MAX(n)	Retain /n/	1.86	5.22
MAX(n)-MONO	Retain /n/ in monosyllables	2.08	2.23
MAX(n)-POSTATONIC	Retain /n/ in penultimately-stressed stems	5.46	4.20
[-i] for masc.	For /-in/$_{masc.}$, select allomorph [-i]	3.87	0.00
*CODA-[r]	Avoid [r] in coda position	5.40	13.93
MAX(r)	Retain /r/	2.53	13.70
MAX(r)-MONO	Retain /r/ in monosyllables	3.34	1.26
MAX(r)-POSTATONIC	Retain /r/ in penultimately-stressed stems	4.82	24.31
[-do] for masc.	For /-dor/$_{masc.}$, select allomorph [-do]	16.00	17.76
*nt]$_{word}$	Avoid word-final [nt]	18.68	16.78
MAX(t)	Retain /t/	2.15	8.06
MAX(CC)	Penalise complete deletion of two consonant sequences	11.24	5.50
*MAP(ʒ-ʃ)	Avoid correspondence between [ʒ] and [ʃ]	8.42	1.97
*MAP(ʒ-tʃ)	Avoid correspondence between [ʒ] and [tʃ]	0.00	1.70
*HIATUS	Avoid two adjacent vowels	0.30	6.39
[XV] is derived from /XVn/	Surface [XV] should have the UR /XVn/	6.86	1.87
[Xo] is derived from /Xor/	Surface [Xo] should have the UR /Xor/	8.82	0.46
[Xu] is derived from /Xur/	Surface [Xu] should have the UR /Xur/	7.28	0.00
[Xu] is derived from /Xu/	Surface [Xu] should have the UR /Xu/	1.33	0.00

51. Performance of the grammar

- If you just do the phonological constraints and candidates, it's perfect — there are enough constraints included to match lexical and Experiment 1 frequencies exactly.
- If you include the candidates and constraints for Avoidance, it's ok, but not perfect.
  - Conjectured reason: candidates like **[məˈzin-u]** should not be considered as part of the phonological competition, but are task-based responses to uncertainty (which is not in the tableau).

52. Back to the thread: our test for the presence of avoidant and faithful speakers

- These can be tested by adding additional, ad hoc MaxEnt constraints
- We added 74 new constraints — two for each participant.
  - *UNFAITHFUL(Speaker_i) i = 1 to 34
  - *AVOIDANT (Speaker_i) i = 1 to 34
- *UNFAITHFUL(Speaker_i) is violated every time a form is uttered by Speaker_i that violates any of the Faithfulness constraints in our analysis.
- *AVOIDANT(Speaker_i) is violated every time a form is uttered by Speaker_i that doesn't use the rare [-u] or [-ə] endings in the masculine form.
- As expected, the speakers we think of as especially avoidance had low personal weights for *AVOIDANT; ditto for *UNFAITHFUL.

53. Using the Likelihood Ratio Test to check our hypotheses

- We statistically test whether adding the two batches of 34 speaker-specific constraints improves the performance of the MaxEnt grammar.
- It does, and per Likelihood Ratio Test, the improvement is highly significant
  - Subject-specific *AVOIDANT: $p < 0.001$
  - Subject-specific *UNFAITHFUL: $p < 0.001$
  - The change in log likelihood is smaller for *UNFAITHFUL.

54. One other items we tested: consistency

- Add a constraint *INCONSISTENT, that is violated if you give a different response for the two wugs in the test that had the same UR configuration; e.g. different /n/ deletion outcomes for /ˈbatun/ and /ˈsodun/.
- This also tests significant, though it is a much smaller effect ($p = 0.007$)

55. Upshot

- We are essentially certain that individual Catalonians differ in how much they like to offer Avoidant responses — a testing strategy.
- We are pretty sure that there are individuals who are especially Faithful — does this mean different grammar, or a testing strategy?
- The Consistency outcome most clearly indicated speaker-specific grammars, but is the weakest effect of the three.

A DIRECTION FOR FUTURE WORK

56. The issue

- Modern-day stochastic linguistics can capably deal with *gradience* (conflicting preferences, intuitions of imperfect well-formedness).
- But what about *uncertainty*?

57. The uncertainty that arose in our wug-test

- Let us return to the avoidance phenomenon, from (30) above.
- Our avoidant responses carry two signatures of uncertainty:
  - long response time
  - occurrence in cases where lexical attestation is small

58. How big an iceberg is this the tip of?

- Catalan serendipitously provides the opportunity to "cheat" using morphological means, providing an outlet to speakers who are uncertain.
- But *how many are the cases*, cross-linguistically, where speakers feel uncertain and would cheat if they could? Standard wug tests don't check for uncertainty.
- Does this even carry over to *syntax*? Consider whether the following two descriptions correspond to your experience as consultant:
  - Gradience: "I think this sentence sounds somewhat bad but not horrible."
  - Uncertainty: "I don't even know what to think of this sentence."

59. To my knowledge the treatment of uncertainty is a lacuna in computational work

- MaxEnt grammars, my favorite, say nothing about uncertainty.
  - They will happily match an 80/20 frequency ratio (stronger constraint must be weighted higher by 1.39) — whether the lexicon has 8000/2000 or 8/2.

60. An idea that seems sensible for this purpose

- In statistics we can compute the "*95% confidence interval*" for the estimated probability when we hear *n*/*m* instances.
- For instance, the 95% lower confidence interval for 10/10 cases is not 1.00 but 0.69.

61. Scenario for Catalan: the experimental participant silently consults her grammar

- (We imagine that this is a grammar that can assign confidence intervals to its probability values.)

Native speaker: "Hello, grammar? Please tell me the probability of deriving **[**ˈ**fɔ]** from /ˈ**fɔr/**."

Speaker's grammar:	"Following currently accepted confidence intervals, I report that this is somewhere between 0.117 and 0.944"[2]
Native speaker:	"Well thanks a bunch, grammar! What do I do now?" [After pondering, responds with **[ʎuˈdaʒu]**.]

62. A challenge to computatationalists

- Is there a system for stochastic constraint-based grammars, perhaps incorporating confidence intervals, that handles both gradience and uncertainty?

SUMMARIZING THE TALK

63. Some (very tentative) bottom lines for phonological theory

- *All phonology examined is productive* (not massive memorization of paradigms)
- As found for other languages earlier, phonological responses tend to *frequency-match* the lexicon.
- *Opacity* is learnable — but maybe not all that well.
- *Saltation* is learnable — but maybe not all that well.

64. Other findings and speculations

- The frequent appearance of avoidant behavior.
- The suggested influence of orthography on /r/ Deletion
- Resistance by speakers to totally exceptionless phonology (/nt Cluster Reduction), perhaps due to foreign language influence.
- Differences between subjects, attributable to preference for avoidance or Faithfulness.

65. Should there be more wug-testing?

- So far, only a tiny fraction of the world's phonologies have been tested in this way.
- It seems there would be a payoff for doing quite a bit more — why?
- We often put forth multiple explanations for a particular deviation from frequency matching — only cross-linguistic study can show which of these are consistently valid.
- As seen here, whenever you do a wug test you are liable to encounter surprises, which may be useful in furthering the research program.

[2] These are the binomial confidence intervals at 99% for 4/7, the observed counts for monosyllabic /r/ deletion. Of course, the real grammar isn't just counts; for other reasons it has to work with constraints and weights.

References

Baayen, Harald, Robert Schreuder, Nivja de Jong, and Andrea Krott. 2002. Dutch inflection: the rules that prove the exception. In *Storage and computation in the language faculty*, ed. by Sieb Nooteboom, Fred Weerman, and Frank Wijnen, 61-92. Dordrecht: Springer.

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.

Cena, R.M., (1978). When is a phonological generalization psychologically real? *Indiana University Linguistics Club*.

Dmitrieva, O., Jongman, A., & Sereno, J.A. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics*, *38*(3), 483-492.

Do, Youngah (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.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: a unified JavaScript framework for online studies. In *International conference on computational social science*, 1-3. Cologne: University of Osnabrück.

Goldwater, S., & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory* (pp. 111-120).

Hayes, B., Siptár, P., Zuraw, K., & Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, 822-863.

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

Ito, J., & Mester, A. (2003). On the Sources of Opacity in OT: Coda Processes in German. In C. Féry & R. van de Vijver (Eds.), *The Syllable in Optimality Theory* (pp. 271–303). chapter, Cambridge: Cambridge University Press.

Jarosz, Gaja (2017). Defying the Stimulus: Acquisition of Complex Onsets in Polish. *Phonology* 34: 269-298.

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

Łubowicz, A. (2002). Derived environment effects in Optimality Theory. *Lingua*, *112*(4), 243-280

Mascaró, J. (1976). *Catalan phonology and the phonological cycle*. Doctoral dissertation, MIT.

McCarthy, John & Alan Prince (1995). Faithfulness and reduplicative identity. In J. Beckman, S. Urbanczyk & L. Dickey (eds.) *UMassoccasional papers in linguistics 18: Papers in Optimality Theory*. 249–384.

Mayer, Connor (to appear). A large-scale corpus study of phonological opacity in Uyghur. To appear in *Phonology*.

Moreton, Elliott and Joe Pater. (2012). Structure and substance in artificial-phonology learning. Part 1: Structure, Part II: Substance. *Language and Linguistics Compass* 6:686–701 and 702–718.

Prince, A., & Smolensky, P. (1993). Optimality Theory: Constraint interaction in generative grammar. *Optimality Theory in phonology: A reader*, 1-71.

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

Wasserman, Larry. 2004. All of statistics: A concise course in statistical inference. New York: Springer.

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.

Yu, Alan CL. (2010). Perceptual compensation is correlated with individuals'"autistic" traits: Implications for models of sound change. *PloS one* 5, no. 8 (2010): e11950.

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