# Phonotactics with[awt] rules: the learnability of a simple, unnatural pattern in English

The 24th Manchester Phonology Meeting

John Harris [1] & Nick Neasom[1] & Kevin Tang [2]

{john.harris, nicholas.neasom.10}@ucl.ac.uk & kevin.tang@yale.edu

[1]University College London

[2]Yale University

May 26th–28th, 2016

# Outline

# Introduction

# Main theme

- How much of the phonotactic patterning discovered by linguists is also discovered by speaker-hearers?
- Case study: /aw/ (MOUTH) in English

# Running order

- Phonotactic patterns: learnability factors
- /aw/ in English
- A nonword acceptability study

# Learnability

# Learnability of phonotactic patterns: factors

'Classic' factors

- Regularity: does the pattern have lexical exceptions?
- Productivity: is the pattern extendable to new words?
- Structural simplicity: how simple are the structural description and the derivation of the pattern? (Moreton & Pater 2012)
- Naturalness: is the pattern phonetically/substantively motivated? (Wilson 2006, Albright 2007, Hayes & White 2013, White 2014,...)

Lexical factors

- Type frequency/generality
- Token/usage frequency
- Lexicon size
- Lexical neighbourhood effects

# Learnability factors: interactions

Speakers can productively apply patterns that are...

- Regular, simple, natural
  - Classic wug-test studies of English -**(e)s**, -**(e)d** (Berko 1958 et passim)
- Irregular, structurally complex, not natural (at least synchronically)
  - English velar softening (e.g. **electric–electrician**), vowel shift (e.g. **vain—vanity**) (Ohala 1974, Pierrehumbert 2006)

Need for more case studies that allow us to test different permutations of potential learnability factors

- Example of English /aw/: regular, simple, unnatural – and general

# Rules versus analogy

Rule/constraint

- The pattern is stored off-line as an independent grammatical rule or constraint

Analogy

- The pattern is extracted on the fly from the lexicon
- Statistically inferred from the lexicon: phonotactic probability and neighbourhood density

# Rules vs analogy: nonword acceptability

- Predictions for nonword acceptability judgements
- Rule-based knowledge (strong version)
  - Structural simplicity: the pattern will generalise evenly across all the specified phonological contexts, uninfluenced by lexical statistics
  - Cf. wug tests: -s pattern productively applied to nonwords in dense lexical networks (e.g. [wʌdz]) as well as in sparse (e.g. [bɹɪlɪgz])
- Analogical implementation
  - The pattern will be unevenly applied across the specified phonological contexts
  - It will be influenced by lexical and usage effects e.g. neighbourhood density, lexicon size, frequency of real-word neighbours

# An /aw/ pattern in English

# The awT pattern

- /aw/ of MOUTH lexical set (shout, crowd, cow, round, etc.)
- The 'awT' pattern: a consonant following /aw/ must be coronal
  - shout, pout, crowd, loud
  - mouse, house, browse, carouse, mouth (n.), south, mouth (v.)
  - couch, slouch, gouge
  - town, brown
  - mount, fount, mound, ground, lounge, scrounge, pounce, flounce, joust
  - mountain, founder, council, frowsty
  - */lawp/, */lawk/, */lawf/, */lawm/, */lawŋk/
- awT is regular
  - No obvious counterexamples in CELEX2, CuBE (Lindsey and Szigetvári, 2016)

uc-rev-cmyk

# awT is simple

- Standard syllable-based analysis: C → [coronal] / aw_ within rime (Selkirk 1982, Anderson & Ewen 1987, Spencer 1996, Hammond 1999, Kubozuno 2001,...)
- Another awT context: before an unstressed vowel
  - chowder, doughty, dowdy, powder, rowdy, blowzy, frowsy, thousand, tousle, trouser
  - */lawbi/, */lawkl/, */lawmpə/
- Foot-based analysis (Harris, in press)
  - C → [Coronal] / aw _ ...]Foot
  - Monosyllabic foot: loud, mount
  - Disyllabic foot: powder, bounty
- awT is even simpler and lexically more general than once thought

# awT is unnatural

- Accent variation
  - MOUTH: [aw,ɑw,æə,əw,əʉ]
  - No special relation between [aw] quality and coronal
- awT can be overturned in neologisms and proper names
  - Baum, Smaug, Bowker, Taub,...
- awT has not established itself across all dialects of English, cf. Northumbrian (including Scots)
  - **cowp** 'tip over', **bowk** 'vomit', **howf** 'haunt, pub', **gowk** 'cuckoo'
- Recent sound changes
  - British English /t/-glottalling: /aw/ before [ʔ], e.g. out, shout
  - Labio-dentalisation of dental fricatives: /aw/ before [f], e.g. **mouth**, **south**
  - Vocalisation of /l/: /al/ > [aw], e.g. **talc**

# awT: natural history, unnatural outcome

- awT is the accidental result of an accumulation of unrelated sound changes
- /aw/ < earlier uː via Great Vowel Shift
- Main changes
  - Lenition/deletion of g after long vowel, e.g. bow (v.), fowl
  - Shortening of earlier uː > u (later > ʌ)
    - Before velars, e.g. **suck**, **duck**
    - Before labials, e.g. **sup**, **plum**
  - Together, the changes have left large gaps in the English lexicon by syphoning off potential sources of modern /aw/ plus velar or labial
  - The awT pattern is not synchronically natural

uc-rev-cmyk

# Do speakers know the pattern?

# Do speakers know awT?

# Rating Study

- An acceptability experiment designed to test the extent to which native speakers of English have tacit knowledge of the awT pattern
- Listeners presented with nonword auditory stimuli containing the diphthongs /aw/, /ow/, /ij/, followed by a range of consonants
- Listeners asked to judge how English-like they sounded individually on a scale of Englishness
- NB. We also conducted a forced-choice study: listeners made choices between paired words distinguished solely by whether the vowel was followed by a coronal versus a non-coronal consonant. This is not reported in this talk due to time reasons

# Auditory stimuli

- English-like nonwords, e.g. /tawm, plawt, strawk, sɹijʃ, brɪjg, kowð, nowb/
- Monosyllabic template: $[C_{1-3}VC_1]$
- Vowels
  - V is one of /aw/ (MOUTH), /ɪj/ (FLEECE), /ow/(GOAT)
- Read from IPA transcriptions by phonetically trained speaker of modern southern standard British English (/ow/ in southern British English = [əw, əɨ])
- Participants listen to nonword stimuli through headphones/speakers (it has no effect on the rating)

# Motivating our choice of non-words

- Onset size:
  - To maximise the size of the potential /aw/ lexicon without introducing interfering phonological conditions (as would happen if we varied, say, coda size)
- Control vowels
  - /ıj/ (FLEECE), /ow/ (GOAT)
  - Not subject to a coronal restriction (seem, seek, roam, broke)
- Monosyllables
- Single coda consonants

# Rating study

# Rating study: design

- Participants (N = 83)
  - Native speakers of British English
  - Age range: 16-60 (mean 25.6; SD 9.4)

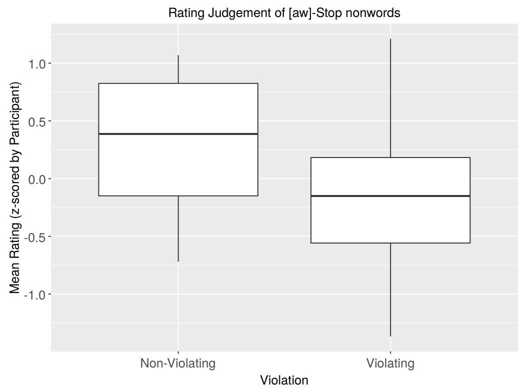
  Auditory stimuli
  - Total: ≈ 1200 nonwords
  - Each listener presented with random sample of 110
  - Total trials (after pre-processing): 8544
- Stimuli presented individually
- Listeners rated stimuli on a Likert scale
  - 1 = ‘completely UNNATURAL: not a good word of English at all’
  - 7 = ‘completely NATURAL: an absolutely fine word of English’

uc-rev-cmyk

# Final stops

- Focus here on stop-final nonwords
  - The only manner in English with all three places of articulation
  - **L**abial, **C**oronal, **D**orsal
- /aw/+stop nonwords
  - Total nonwords with this pattern: 156
  - Total ratings: 1104
  - Each item rated on average by five subjects out of the 83

# /aw/+stop: awT (non-)violations




| | $\beta$ | SE($\beta$) | $t$ | $p$-value |
|---|---|---|---|---|
| (Intercept) | 0.0313 | 0.0720 | 0.4346 | 0.6638 |
| Violation (Viol vs. Non-Viol$_{Ref}$) | -0.3817 | 0.1045 | -3.6518 | 2.6039e-04$^{***}$ |

uc-rev-cmyk

# awT-violation as a predictor

- awT on its own is a significant predictor
  - Non-words with non-coronal finals are less acceptable than those with coronals
- But maybe this effect is down to other factors
- Now we try a model that includes more predictors
  - Constraint: violation vs non-violation of awT
  - Lexical
    - Neighbourhood density
    - Phonotactic probability
  - Orthogonal phonological
    - Onset size
    - Voicing

# Lexical Statistics

- Neighbourhood density
  - Real-word neighbours – Number, Frequency and Phonological distance.
  - Generalised Neighbourhood Model (Bailey & Hahn 2001)
- Phonotactic probability:
  - Segment-based trigram model with Modified Kneser-Ney smoothing
- Reference lexicon
  - SUBTLEX-UK (van Heuven et al. 2014) – 201.7 million words and 160,022 word types
- Transcription: CUBE (Lindsey and Szigetvári, 2016)

# All Predictors: Results: Mixed Model (Maximal)

| | $\beta$ | SE($\beta$) | $t$ | $p$-value |
|---|---|---|---|---|
| (Intercept) | 0.1614 | 0.0726 | 2.2236 | 0.0261 |
| Violation (Viol vs. Non-Viol$_{Ref}$) | 0.1255 | 0.1771 | 0.7084 | 0.4786 (n.s.) |
| Voicing (Vl vs. Vd$_{Ref}$) | 0.0530 | 0.0661 | 0.8016 | 0.4228 (n.s.) |
| Neighbourhood | 0.4233 | 0.1652 | 2.5622 | 0.0104* |
| Phonotactic Prob. | 0.1295 | 0.05144 | 2.5171 | 0.0118* |
| Onset (CC vs. C$_{Ref}$) | 0.0985 | 0.0990 | 0.9960 | 0.3192 (n.s.) |
| Onset (CCC vs. C$_{Ref}$) | 0.9676 | 0.4063 | 2.3818 | 0.0172* |

# All Predictors: Summary

- The more neighbourhood support a nonword gets, the more acceptable it is
- The more probable the nonword is in terms of phonotactics, the more acceptable it is
- Nonwords with complex onsets are more acceptable than those with simplex.

# Is /aw/ special?

Does /aw/ show a stronger place preference than the control vowels?

- If yes, this supports awT
- If no, then it's probably just because coronal is special (Paradis & Prunet 1991, etc.)

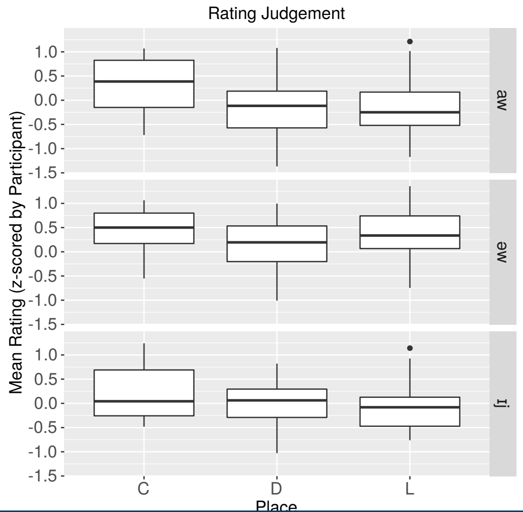
Mixed-model predictors

- Vowel: /ij, ow/ vs /aw/
- Place: Labial/Dorsal vs Coronal
- Interaction term: Vowel × Place

Totals

- 370 stop nonwords ([aw]:156, [ow]:111, [ij]:103)
- 3134 ratings
- Each item rated on average by 8 participants (of 83)

# /aw/ vs other vowels (stop-final nonwords)




uc-rev-cmyk

# Mixed model: Vowel and Place

| | $\beta$ | SE($\beta$) | $t$ | $p$-value |
|---|---|---|---|---|
| (Intercept) | 0.0936 | 0.0586 | 1.597 | 0.1102 |
| Place (LD vs. $C_{Ref}$) | -0.2864 | 0.0694 | -4.1230 | 3.7402e-05$^{***}$ |
| Vowel ([ow/ij] vs. $[aw]_{Ref}$) | 0.1348 | 0.0595 | 2.2652 | 0.0235$^{*}$ |
| Place × Vowel | 0.1736 | 0.1189 | 1.4609 | 0.1440 (n.s.) |

- Place is significant: there's a place preference for [coronal]
- Vowel is significant, /aw/ is dis-preferred compared to /ow/ or /ij/
- Crucially, the interaction term shows a mild tendency for more of a place preference with /aw/, but this is not significant and it can be dropped in a nested model comparison

# Rating Study: Conclusions

- The awT constraint by itself is a significant predictor of a nonword's acceptability
- But this effect disappears once we factor in lexical statistics (neighbourhood density and phonotactic probability) and other phonological variables (especially onset size)
- The Coronal vs Non-coronal pattern is not restricted to /aw/ but also shows up with /ij/ and /ow/
- The effect of place is not significantly stronger with /aw/ than with /ij/ and /ow/

# Conclusions

# Do speakers know awT?

- awT is a case where phonologists know more about a pattern than speakers know
- If speakers have any inkling of awT at all, it is buried so deep in their tacit phonological knowledge that it is much more difficult to get at than the kinds of patterns shown to be productive in previous work
- To the extent that speakers have any implicit knowledge of awT at all, it is probably not encapsulated in anything like a phonologist's rule or constraint
- Rather it's based on lexical statistics such as neighbourhood density and phonotactic probability

# awT versus other patterns

- What makes awT hard to learn or not worth learning?
- Compared to other patterns
  - awT is more regular than velar softening, vowel shift
  - awT is lexically more general than velar softening, vowel shift
  - awT is not natural, but neither are velar softening, vowel shift
  - awT is structurally simpler than -**(e)s**, -**(e)d**
- Static distribution
  - Unlike any of the above patterns, awT is wholly non-alternating (it's unwuggable)
  - If we think of alternations as reinforcing cues to morphology, then awT is simply not as salient

# Moral

- Not all phonotactic patterns that linguists discover in languages find their way into speakers' grammars
- We need to bear this in mind before building a synchronic phonological account of any given pattern

# Acknowledgements

- Audience from the London Phonology Seminar group
- Our participants
- My collaborators: John Harris and Nick Neasom

# References

References available on request.

# Appendices

# Data

- 156 [aw]-stop wugs
- 1104 ratings
- Each item rated on average by 5 people, divided amongst 83 people

# MixedModel *lmer* – Predictors

Constraint variable:

- Violation (Factor)

Lexical Variables:

- Neighbourhood density (GNM)
- Phonotactic probability

Control variables:

- Onset size (Factor)
- Voicing (Factor)

# Violation Alone: Mixed Model: Model structure (Maximal)

Testing if Violation is a good predictor for rating

- All continuous variables are log10-transformed and then z-transformed.
- All categorical variables are sum-coded.
- Random intercepts: Allow for participant and word level variability
- Random slopes: ‘Break’ the predictors
- Per-participant and per-word random intercepts
- Per-participant random slopes: Violation
- Per-word random slopes: Violation

# Violation Alone: Mixed Model: Maximal model

$$Rating \sim Violation + (Violation \mid Participant\ ) + (Violation \mid Word)$$

# Violation Alone: Results: Mixed Model (Maximal)

| | $\beta$ | SE($\beta$) | $t$ | $p$-value |
|---|---|---|---|---|
| (Intercept) | 0.0313 | 0.0720 | 0.4346 | 0.6638 |
| Violation (Viol vs. Non-Viol$_{Ref}$) | -0.3817 | 0.1045 | -3.6518 | 2.6039e-04$^{***}$ |

uc-rev-cmyk

# Violation Alone: Discussion: Mixed Model (Maximal)

- [awT] constraint is a significant predictor
- Non-coronal words are less acceptable than coronal words
- However, can this be explained using lexical statistics?
- Let's put in all the predictors (Violation, Neighbourhood, Phonotactic Probability, Voicing, Onset Size)

# All Predictors: Mixed Model (Maximal)

- All continuous variables are log10-transformed and then z-transformed.
- All categorical variables are sum coded.
- Random intercepts: Allow for participant and word level variability
- Random slopes: ‘Break’ the predictors
- Per-participant and per-word random intercepts
- Per-participant random slopes: Violation, Neighbourhood density (GNM) and Phonotactic probability
- Per-word random slopes: Violation, Neighbourhood density (GNM) and Phonotactic probability

# All Predictors: Mixed Model: Maximal model

$$Rating \sim Violation + Voicing + Neighbourhood + Phonotactic\ Prob. + Onset\ Size + (Violation + Neighbourhood + Phonotactic\ Prob \mid Participant\ ) + (Violation + Neighbourhood + Phonotactic\ Prob \mid Word)$$

# All Predictors: Results: Mixed Model (Maximal)

| | $\beta$ | SE($\beta$) | $t$ | $p$-value |
|---:|---:|---:|---:|---:|
| (Intercept) | 0.1614 | 0.0726 | 2.2236 | 0.0261 |
| Violation (Viol vs. Non-Viol$_{Ref}$) | 0.1255 | 0.1771 | 0.7084 | 0.4786 (n.s.) |
| Voicing (Vl vs. Vd$_{Ref}$) | 0.0530 | 0.0661 | 0.8016 | 0.4228 (n.s.) |
| Neighbourhood | 0.4233 | 0.1652 | 2.5622 | 0.0104* |
| Phonotactic Prob. | 0.1295 | 0.05144 | 2.5171 | 0.0118* |
| Onset (CC vs. C$_{Ref}$) | 0.0985 | 0.0990 | 0.9960 | 0.3192 (n.s.) |
| Onset (CCC vs. C$_{Ref}$) | 0.9676 | 0.4063 | 2.3818 | 0.0172* |

# All Predictors: Discussion: Mixed Model (Maximal)

- [awT] constraint is no longer significant, compared to a model with only Violation
- Voicing is insignificant
- Neighbourhood density and Phonotactic Probability are both significant
- The more neighbourhood support a nonword gets, the more acceptable it is
- The more probable the nonword is in terms of phonotactics, the more acceptable it is
- [awT] constraint is nowhere to be found after taking into account of lexical factors

# All Predictors: Mixed Model: Model Selection

- Maybe our model has been overfitted, therefore the results cannot be trusted
- Against model overfitting: backward model selection to remove predictors that do not significantly improve the model with chi-square test, alpha = 0.1.
- Start with random effects, then fixed effects

# All Predictors: Mixed Model (Best)

*Rating ∼ Neighbourhood + Phonotactic Prob. + Onset Size + (Neighbourhood | Participant ) + (1 | Word)*

- Dropped Voicing and Violation as fixed effect
- Dropped all slopes per word; and Violation and Phonotactic Prob. slopes per participant.

# All Predictors: Results: Mixed Model (Best)

| | $\beta$ | SE($\beta$) | $t$ | $p$-value |
|---|---|---|---|---|
| (Intercept) | 0.1457 | 0.0768 | 1.8974 | 0.0578 |
| Neighbourhood | 0.3176 | 0.1399 | 2.2697 | 0.02322* |
| Phonotactic Prob. | 0.1278 | 0.0398 | 3.2088 | 1.3324e-03** |
| Onset (CC vs. $C_{Ref}$) | 0.2082 | 0.1028 | 2.0238 | 0.04299* |
| Onset (CCC vs. $C_{Ref}$) | 0.6617 | 0.3344 | 1.9789 | 0.0478* |

uc-rev-cmyk

# All Predictors: Discussion: Mixed Model (Best)

- [awT] constraint is still nowhere to be found after taking into account of lexical factors
- The previous result is not due to over-fitting

# Vowel and Place: Mixed Model: Maximal model

- We test if the interaction term with vowel and place is significant or not
- Predictors: Vowel ([aw] or not [aw]), Place (Coronal or not Coronal) and their interaction term.
- Per-word and per-participant random intercepts
- Per-participant random slopes: Vowel, Place and Vowel:Place
- Per-word random slopes: None (Did not converge)

# Vowel and Place: Mixed Model (Maximal)

$$Rating \sim Place + Vowel + Place{:}Vowel + (Place + Vowel + Place{:}Vowel \mid Participant\ ) + (1 \mid Word)$$