

Wordlikeness

Phonotactic theory (EGG 2025, Zagreb)

Much of the work on phonotactics uses statistical analysis of digital lexicons (see Gorman 2013:39f. for a long list of such studies up to that point), which is of unclear relevance to speakers' actual phonotactic knowledge. Today we'll review:

- the design of **wordlikeness tasks**, formal experiments gathering speakers' judgments about the wellformedness of nonce words, and
- computational models of phonotactic knowledge, and their correlation with experimental results from wordlikeness tasks.

1 Experimental design

- Table 1 summarizes the properties of various English wordlikeness tasks.

	<i>n</i>	<i>s</i>	pres. mod.	resp. mod.	notes
Scholes 1966, exp. 5	33	63	auditory	binary	tested as group
Bailey and Hahn 2001, exp. 1	22	341	orthogra.	Likert (9)	RW distractors
Bailey and Hahn 2001, exp. 2	12	328	auditory	Likert (9)	RW distractors
Albright and Hayes 2003, norm. exp.	24	87	auditory	Likert (7)	
Albright 2007	68	40	auditory	Likert (7)	unpublished

Table 1: Design properties of English wordlikeness tasks; *n* = number of participants; *s* = number of stimuli; pres. mod.: presentation modality; resp. mod.: response modality.

- Zimmer (1969) uses a different design:
 - Participants are presented with a pair of nonce words differing only in whether they obey (e.g., *törüz*) or violate (e.g., *töriz*) harmony generalizations.
 - They are then asked to pick which of the two “sounds more like a word that might actually occur in Turkish”.
- In ongoing work with K. Durvasula and J. Kahng, I am testing the effects of:
 - auditory vs. orthographic presentation,
 - veridicality of perception (by having participants transcribe auditorily-presented stimuli before responding),
 - binary vs. Likert response modality, and

- laboratory vs. online administration.

in English and Korean. We will release the de-identified experimental data for use by the research community.

2 Computational models of wordlikeness

- There is by now a cottage industry:
 - developing computational models of phonotactic knowledge, and
 - correlating their predictions with the results of wordlikeness judgment tasks.
- Bailey and Hahn (2001), for example, claim wordlikeness judgments derive from similarity to known words, a reflection of the lexical access procedure.
 - Let the **phoneme edit distance** (PED) between two (nonce or real) words x, y be the minimal number of phonemes that need to be substituted, inserted, or deleted to convert x onto y .
 - Let x, y be **neighbors** if $\text{PED}(x, y) = 1$.
 - Let the **neighborhood density** of x be the number of (real) words which are neighbors of x .
- More sensitive neighborhood density measures have since been developed (e.g., PLD20; Suárez et al. 2011), but these have not yet used to model wordlikeness judgments.
- Albright (2009) proposes a bigram phonotactic model trained on the lexicon. For example, the bigram probability of the sequence ijk is given by $\hat{p}(ijk) = p(i | \text{BOS}) \cdot p(j | i) \cdot p(k | j)$.
- Albright also proposes a variant defined over feature (rather than segment) features, but this variant is not described in sufficient detail to re-implement.
- Hayes and Wilson (2008), henceforth H&W, propose their **maximum entropy** (or **Max-Ent**) model of phonotactics.¹
- H&W conceptualize phonotactic knowledge as a grammar of weighted markedness constraints (à la Harmonic Grammar) projected from the lexicon.
- Training of their model proceeds in two steps:
 - First, they use a complex heuristic to induce candidate constraints over (sequences of) natural classes (or their complement classes).
 - Then, they use a generic gradient descent algorithm to weigh the constraints to best fit the observed lexicon.
- The model then can provide a “score” to each sequence based on the weights of the constraints that sequence violates.

¹Minor refinements of this model are described in Wilson and Gallagher 2018.

- Nearly all this work takes it for granted that wordlikeness (i.e., the construct itself) is inherently gradient:

...patterns of gradient well-formedness often seem to be driven by the very same principles that govern absolute well-formedness...I conclude that the proposed attribution gradient well-formedness judgments to performance mechanisms would be un insightful. Whatever “performance” mechanisms we adopted would look startlingly like the grammatical mechanisms that account for non-gradient judgments. (Hayes 2000:99)

A defect of current grammatical accounts of phonotactics is that they render simple up-or-down decisions concerning well-formedness and cannot account for gradient judgements. But when judgements are elicited in a controlled fashion from speakers, they always emerge as gradient, including all intermediate values. (Shademan 2006:371)

In the particular domain of phonotactics gradient intuitions are pervasive: they have been found in every experiment that allowed participants to rate forms on a scale. (Hayes and Wilson 2008:382)

When native speakers are asked to judge made-up (nonce) words, their intuitions are rarely all-or-nothing. In the usual case, novel items fall along a gradient cline of acceptability. (Albright 2009:9)

- However, Armstrong et al. (1983) argue that rating tasks induce intermediate ratings as an effect of the task *regardless of the gradience of the underlying construct*.

No integer seems to sit on the fence, undecided as to whether it is quite even, or perhaps a bit odd. No odd number seems odder than any other odd number. (Armstrong et al. 1983:274)

- However, for instance, when asked to rate “how odd” odd numbers are on a 7-point Likert scale participants freely use intermediate ratings!

Some have responded to these findings very consistently, by asserting that the experimental findings are to be interpreted as before: that, psychologically speaking, odd numbers as well as birds and vegetables are graded concepts... We reject this conclusion just because we could not explain how a person could compute with integers who believed that 7 was odder than 23. We assert confidently that the facts about subjects being able to compute and about their being able to give the definition of odd number, etc., are the more important, highly entrenched, facts we want to preserve and explain... (Armstrong et al. 1983:284)

- Durvasula (2020), Gorman (2013: ch. 2), and Kostyszyn and Heinz (2022) present binary-valued wordlikeness models of wordlikeness and find they they have a good fit even to Likert-scale wordlikeness data.

References

- Albright, Adam. 2007. Natural classes are not enough: Biased generalization in novel onset clusters. Ms., MIT.
- Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 9–41.
- Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90:119–161.
- Armstrong, Sharon L., Lila R. Gleitman, and Henry Gleitman. 1983. What some concepts might not be. *Cognition* 13:263–308.
- Bailey, Todd M., and Ulrike Hahn. 2001. Determinants of wordlikeness: phonotactics or lexical neighborhoods. *Journal of Memory and Language* 44:586–591.
- Durvasula, Karthik. 2020. O gradience, whence do you come? Plenary lecture at the Annual Meeting on Phonology, Santa Cruz.
- Gorman, Kyle. 2013. Generative phonotactics. Doctoral dissertation, University of Pennsylvania.
- Hayes, Bruce. 2000. Gradient well-formedness in optimality theory. In *Optimality Theory: Phonology, Syntax, and Acquisition*, ed. Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer, 88–120. Oxford University Press.
- Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.
- Kostyszyn, Kalina, and Jeffrey Heinz. 2022. Categorical account of gradient acceptability of word-initial polish onsets. In *Proceedings of the Annual Meeting on Phonology*.
- Scholes, Robert J. 1966. *Phonotactic Grammaticality*. Mouton.
- Shademan, Shabnam. 2006. Is phonotactic knowledge grammatical knowledge? In *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 371–379.
- Suárez, Lidia, Seok Hui Tan, Melvin J. Yap, and Winston D. Goh. 2011. Observing neighborhood effects without neighbors. *Psychonomic Bulletin & Review* 18:605–611.
- Wilson, Colin, and Gillian Gallagher. 2018. Accidental gaps and surface-based phonotactic learning: A case study of South Bolivian Quechua. *Linguistic Inquiry* 49:610–623.
- Zimmer, Karl E. 1969. Psychological correlates of some Turkish morpheme structure conditions. *Language* 45:309–321.