

A guide to “Phonological markedness effects in sentence formation”



Canaan Breiss

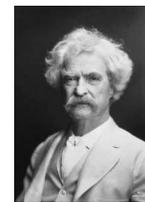
Bruce Hayes

UCLA



$$\Pr(x) = \frac{\exp(-\sum_i w_i f_i(x))}{Z}$$

Language (2020)
96:338-370



1. Welcome

- This video is meant to be a brief and informal guide to our paper “Phonological markedness effects in sentence formation”, which is about to come out in *Language*.
- We’re going to speak for about half an hour, giving a very general outline of our paper and putting it in context.

SETTING

2. Our hypothesis

- When people speak, they pervasively, but subtly, skew their outputs to reduce violations of phonological Markedness constraints.

3. Three background threads

- Markedness
- The Zwicky/Pullum Hypothesis and its demise
- Statistical evidence and probabilistic linguistics

4. Thread #1: The expanding role of Markedness

- The concept dates way back to Jakobson and Trubetzkoy.
- It took on much more importance and applicability with the advent of Optimality Theory, which gave markedness an explicit role in the formal theory.
 - Markedness constraints with a capital M serve as of the main elements of the grammar.
- This was a really fruitful, since the constraints thus formalized let us apply the concept to an ever-expanding set of domains.

5. Things you can do with Markedness constraints

- Rank them against Faithfulness to get **derivations**, UR → SR
- Deal with phonological **conspiracies** (Kisseberth 1970)
- Create all-markedness **phonotactic grammars**, using the MaxEnt probabilistic extension of OT, which can be trained with corpus data and match native-speaker intuition (Hayes and Wilson *LI* 2008)
- Use similar mechanisms to make Markedness constraints the basis of an explicit computational model of **speech errors** (Goldrick and Daland 2009)

6. Thread #2: The fall of the Zwicky/Pullum Hypothesis

- This is a claim about how syntax is related to phonology, that they made in a famous paper in 1986.
- It's actually a bit tricky to render what they said, but the *tentative* story that they give early in their paper is the claim that we think has been empirically falsified.

'the syntactic component determines the order in which words may be placed in sentences ... and the phonological component determines what pronunciations are associated with particular structured sequences of words that the syntax says are well-formed'

- We ignore the far more nuanced story they tell later in their paper, focusing only on the naïve generalization just given — which is very falsifiable.
- What subsequent work clearly showed is that sentences can be bad for phonological reasons.

7. Example: Reversed-sonority imperatives in Norwegian (Rice 2007)

- a. Sykl opp bakken
bike up the.hill
'Bike up the hill!'
- b. *Sykl ned bakken
bike down the.hill
'Bike down the hill!'

8. Examples become abundant when we over to probabilistic linguistics

- See work cited in our paper by (among others) Arto Anttila, Stephanie Shih, and Kie Zuraw.
- A frequently observed effect: when a language provides a choice between two syntactic constructions that express the same meaning, speakers tend to pick the variant that is phonologically less marked.

9. *CLASH influences the distribution of the two forms of the English dative construction

give books to Bill just one violation of *CLASH

statistically preferred to

give Bill books two violations of *CLASH

Effect disappears if *Bill* replaced by *Péter*, since this is one *CLASH violation in each case.

For details see §1 of our paper.

10. Thread #3: development of effective formal methods for dealing with gradient data

- Effective forms of **math and statistics**, notably MaxEnt OT grammars, which expand the power of Optimality Theory to cover gradient phenomena.
- Use of **data corpora**, so we can test hypotheses on a whole-language basis rather than with small bodies of selected data
- Use **significance testing**, to make sure our results are valid and not the results of wishful thinking

We each have been using these methods a lot in our own individual research, and it makes us excited to look for new applications.

11. In sum, three threads

- expanding role of markedness
- the intriguing body of work that has toppled the Zwicky Pullum Hypothesis
- the expanding inventory of formal means for exploring gradient effects in language

WHAT WE DID

12. Restating the original claim

- When people speak, they pervasively, but subtly, skew their outputs to reduce violations of phonological Markedness constraints.

13. Operationalizing the hypothesis to make it testable

- In probabilistic terms, we're saying that speakers tend to create sentences that respect Markedness to a great extent than would be expected at random.

14. A scenario: Jane Austen at her desk



- She is unconsciously trying to not violate the following four constraints:

<i>Constraint</i>	<i>Definition</i>	<i>Sample violation</i>
*CLASH	no adjacent stressed syllables	<i>Jáne Fáirfax</i>
*IAMBIC CLASH	no clash where the first word has iambic stress	<i>exáct plán</i>
*CCC:	no triple consonant clusters	<i>pleasánt mánners</i>
*HIATUS:	no consecutive vowels	<i>thrée or</i>

all of these are well-documented constraints, valid within English words

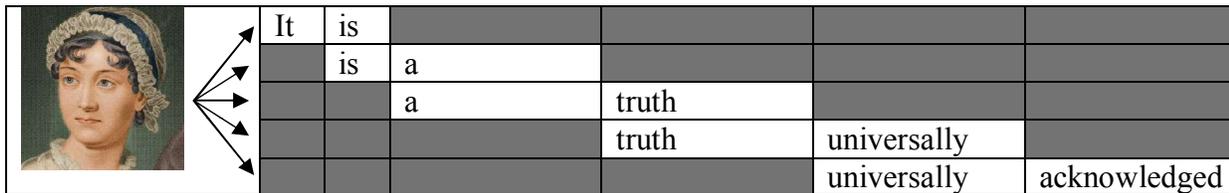
15. Side remark: writing vs. speech

- The Austen scenario may be a bit idealized, since she held a pen instead of speaking.
- But in fact we get the same results when we work with spoken corpora.

16. Idealizing further: word bigrams

- We need a model that is really simple, enough to be tested against raw corpora.
- So let's model Jane Austen as a probabilistic emitter of *word bigrams*
- Rationale: it's precisely at the join between two words where sentence formation can create phonological violations.
- This is the smallest domain that we could choose and is easiest to work with.

17. Jane Austen modeled as a probabilistic emitter of word bigrams



It	is				
	is	a			
		a	truth		
			truth	universally	
				universally	acknowledged

18. Detailed procedure

- To test our hypothesis, we want to see if the bigrams she emits are collectively more phonologically well-formed than we would expect if phonology were not involved.
- So let's compute a **baseline probability** for each bigram, that we can later use for assessing the phonological effects.

19. Computing the baseline probability

- In her works Jane Austen employs a vocabulary of about 14,000 distinct words
- Each of these occurs with a particular probability, which can be estimated empirically by dividing its frequency by the overall corpus size.
 - E.g. *Elizabeth*: $454/725,374 = 0.00063$
- Then the probability of a bigram is just the product of the probability of the words it contains.
- This defines a **baseline distribution** of bigrams, predicted on the crudest of terms.

20. How could we improve on the baseline distribution?

- In principle, lots of ways, like adding syntactic constraints, or a list of frequent idioms.
- But we will here go with a very simple model, that just adds in the phonological constraints

21. More technical details: the MaxEnt OT grammar

- 14,000 constraints:
 - Don't use *Elizabeth*
 - Don't use *Bennet*
 - Don't use *the*
 - ...
- 4 phonological constraints, as above
- This glosses over a lot of complications, which to learn about you will have to read the paper!

22. What the analysis outputs

- A **weight** for every constraint, which tells you how strong it is.
- We also do a **statistical test**.
 - This is the **Likelihood Ratio Test**, which yields a p -value for every constraint.
 - This which is informative in telling us whether the constraint is truly helping to explain the observed distribution of bigrams.

23. The results we got, in brief

- The four constraints (plus five others not mentioned here) play a statistically significant role in governing the bigrams emitted by Austen.
- And ditto for 13 other corpora —Twain, Dickens, Beatles interviews, and a whole bunch of other spoken corpora.

24. How big is the effect?

- In MaxEnt, you can convert the weights into **probability ratios**:
 - = estimated **percent reduction** in the number of violating bigrams caused by each constraint

25. The probability ratios for the four constraints discussed here

(Average across all 14 corpora.)

*CLASH:	38.2%
*IAMBIC CLASH:	71.5%
*CCC:	29.3%
*HIATUS:	33.6%

WHAT ABOUT MECHANISMS?

26. What are the *causes* of the patterns we are finding?

- We think there are basically two mechanisms:
 - idioms and fixed phrases
 - choices between syntactic constructions

and we'll tell you how we diagnosed them

27. Diagnosing the effect of the fixed phrases

- Throw away all bigrams of frequency great than 1 — these are likely to be fixed phrases
- Rerun our test on the remaining bigrams
- Result: a lot weaker.
- And — as we know from earlier research published in *Language* by Andy Martin — phonologically unmarked items are more likely to be come lexically listed.

28. Diagnosing the effect of syntax

- Throw away all bigrams that *include a function word* (preposition, aux, pronoun, etc.)
- Rerun our test on the remaining bigrams
- Result: a lot weaker.
- Why?
 - Work discussed above: speakers deploy syntactic choices for phonological reasons
 - Most syntactic constructions include at least one function word
 - So, removing bigrams with function words removes the cases in which speakers deployed syntax for phonologically-beneficial reasons, weakening the effect.

29. Summary of key results

- We think the effect is real — valid across all 14 corpora.
- The mechanisms behind the effect seem to be:
 - use of phonologically-optimizing fixed phrases
 - use of phonologically-optimizing syntactic choices

FUTURE RESEARCH DIRECTIONS

30. A conservative research direction

- We think that since our results are surprising, the first step ought to be simple some further **replication**.
- We were aware of danger while we worked on the paper: after five corpora, we fixed;
 - set of constraints to be studied
 - computational methods and software in place, and did nine more corpora.
- Since finishing the paper, we've done two more for other reasons: the Penn Treebank and a subset of the Childe corpus of child-directed speech; and it's still working.

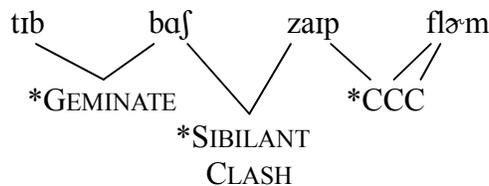
31. Replication of the basic effect in other languages

- We've looked at three classic Vowel Harmony languages — do they tend to obey vowel harmony across word boundaries?
- Turkish and Finnish worked out just fine.
- Hungarian bafflingly tends to *violate* Vowel Harmony across boundaries.
 - This result particularly makes us want to replicate further.

32. Could these effects be verified experimentally?

- E.g. some measurable effect of phonological difficulty?
- Canaan has done pilot work to *test if our constraints produce difficulty in production*.

33. Four nonsense words arranged in constraint-violating order



... and the same four words arranged in violation-free order:

fləʊm zɑɪp bɑʃ tɪb

- The violating sequence incurs a greater penalty in **planning time** (interestingly: not execution time).

34. Integration with syntax

- Our own research activities is in phonology, particular probabilistic, constraint-based phonology.
- But there is also some very intriguing work on constraint-based probabilistic syntax.
- Analytical:
 - Probabilistic syntax with Joan Bresnan and colleagues — which uses a MaxEnt model just like us.
- Experimental:
 - Probabilistic experimental syntax, also using a weighted constraint model, notably Sam Featherston and Frank Keller.

35. Aiming higher

- This leads us to wonder if these research programs could somehow be combined.
- The goal would be to produce a combined, multi-modular grammar that assigns probabilities to sentences, incorporating both phonological and syntactic factors.

36. Our current work

- We are collaborating with our syntactico-computational colleague **Tim Hunter**.
- Processing Penn Treebank data, we learn, all at once:
 - a constraint-based Probabilistic Left-Corner Grammar, which also includes
 - the four phonological constraints we've talked about here
- The phonological constraints get weights on the same order of what they got in Breiss and Hayes, which is encouraging.

37. Goals for the very long term

- We'd like to think this is a tiny little seed for future work
- Someday, linguists will work with grammars that:
 - are very large
 - include both syntax and phonology as elements influencing well-formedness.
 - are computationally implemented and learned
 - engage comprehensively with whole-language data.

We hope you've enjoyed our presentation and are feeling encouraged to take a look at our paper, which is available on the Language website.

Thanks for watching, and thanks to the LSA for giving us the opportunity to give this presentation.