Bruce Hayes[1]
University of California, Los Angeles

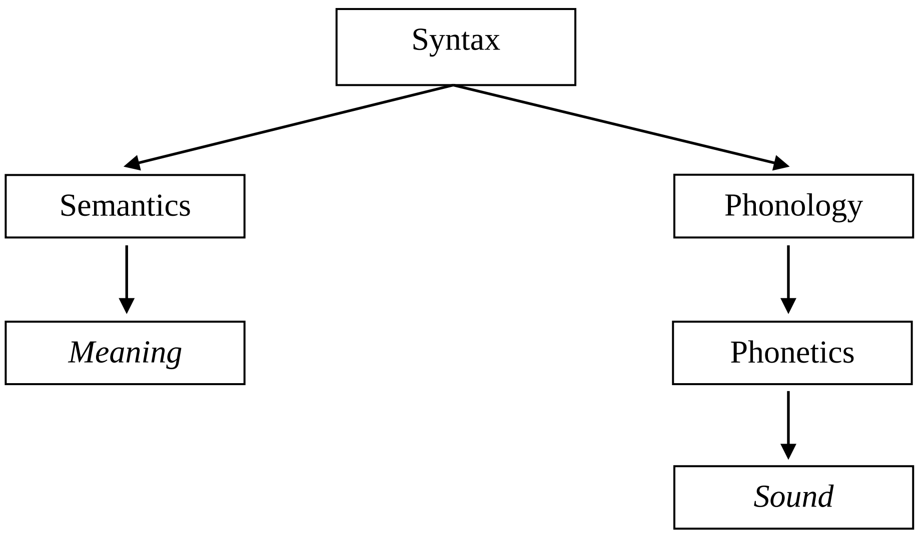
Arizona Linguistics Circle Conference
October 13, 2018

Phonological markedness effects in sentence formation: an investigation using MaxEnt

1. Research question

- How do the domains of linguistic knowledge (“components”) interact in the creation of sentences?

2. The classical feed-forward model (e.g. Chomsky 1965)



- Key prediction: the construction of sentences is blind to phonological consequences of word-concatenation.

3. There have been many challenges to the feed-forward model

- Speakers choose from **variant syntactic constructions** in the direction of avoiding phonological markedness violations.
    - ➢ Shih and Zuraw (2017): Tagalog speakers use the twin syntactic constructions [Adj. *na* Noun]$_{NP}$ vs. [Noun *na* Adj.]$_{NP}$ in ways that statistically avoid violating:

        — *[+nasal][+nasal]
        — *VV
        — others

[1] This talk gives results from a collaboration with **Canaan Breiss**, a graduate student in the UCLA Linguistics Department.

Forms like *túnay na ámaŋ* 'real elder' preferred to *ámaŋ na túnay*.

- Many other syntactic choices are shown to respond to Markedness, particularly for avoiding clashing stresses: Inkelas and Zec (1990), Schlüter (2005), Shih (2012, 2017), Zuraw (2015), Shih et al. (2015), Anttila (2016), Ryan (2018)

- Speakers choose among **synonyms** (e.g. *worse* vs. *worser*) to avoid Markedness (specifically, stress clash) violations — Schlüter (2005), Schlüter and Knappe (2018).

4. Our project

- Instead of being construction- or word- specific, be general and across-the-board.
  - I.e. look at whole sentences — *all* word concatenation.
- Why do this?
  - We can look at a **wider variety** of phonological constraints — anything that can get violated across a word boundary.
  - We hope to get a clearer sense of the **mechanisms** that underlie the patterns are being observed.

CONSTRAINTS WE WILL STUDY

5. In the case of English, what constraints would we like to test?

- Criteria:
  - Powerful: most of our constraints are close to inviolable within words
  - They mostly have good typological support.

6. The nine constraints we tested

- *CLASH (= stresses on adjacent syllables).
  - There are exceptions (*ìndèntátion*), but they are typically morphologically motivated (inherit stress of *indént*).
- *IAMBIC CLASH (= [stressless + stressed] followed by stress)
  - *Very* few exceptions in words (e.g. *ĕlèctrónic*), all morphologically motivated and often repaired by particular speakers (*èlĕctrónic*)
  - This has been noticed for over a century as phrasally relevant (Fijn van Draat 1910-12, Bolinger 1965); and documented as active in multiple corpora by Hammond (2016).
- *TRIPLE OBSTRUENT CLUSTER
  - Only about 200 violating words in Hayes's searchable lexical database of 18,000 words,[2] mostly with prefixes like *ex-* (*extract* [ɛkstrækt]).

[2] http://linguistics.ucla.edu/people/hayes/EnglishPhonologySearch/

- *TRIPLE CONSONANT CLUSTER
    - Of course, more common, but only 1300 examples in the database.

- *SIBILANT CLASH (two sibilants in a row)
    - Not a single stem violates this constraint (sample violation: *[ˈmɪsʃən]).
    - Just a handful of prefixed instances (e.g. *misshapen*).

- *GEMINATE (= identical consonant cluster)
    - Exceptions occur only with productive affixes, like *unknown* [ʌnnoʊn].
    - No monomorphemic cases at all; e.g. "*Hannah*" *[ˈhænnə].

- *HIATUS (= VV)
    - This is actually violated a lot (2300/18000), yet its typological validity tempts us to try it.

- *BAD SONORITY (= low sonority coda C + high sonority onset C)
    - Based on Hooper's (1976) Syllable Contact Law. Untestable in words,[3] but good typologically.

- *NÇ̥ (= no voiceless consonants after nasals)
    - Actually, *not* especially valid in English words, but good typological pedigree (Pater 2004).

WORD BIGRAMS AS A METHOD FOR DETECTING PHRASAL MARKEDNESS

7. Defining *word bigrams*

- = two consecutive words of a text

8. Restating the research question

- When English speakers form utterances, do they tend to create word bigrams that involve fewer violations of the constraints in (6) than might be expected on general grounds?

9. Example: the word bigrams of *Emma*

- Jane Austen's novel *Emma* begins:

> Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy disposition, seemed to unite some of the best blessings of existence.

[3] … because what would be violations get syllabified as branching onsets (*apron* [ˈeɪ.pɹən]).

- Its bigrams are:

[Emma Woodhouse]	[ɛmə wʊdhaʊs]
[Woodhouse, handsome]	[wʊdhaʊs, hændsəm]
[handsome, clever]	[hænsəm, klɛvɚ]
[clever, and]	[klɛvɚ, ənd]
(160,000 more) …	

10. Stating our problem probabilistically: Jane Austen as a stochastic device 

- Austen had an inventory of words — her mental lexicon — that she could use to create sentences.
- The **baseline probability** for Austen emitting the bigram **Word1 + Word2** is a product:
  - = (probability of emitting Word1) * (probability of emitting Word2)
- Our hypothesis can be stated thus: Austen *deviates from baseline probability for phonological reasons*:
  - If the word bigram Word1 + Word2 violates one of the constraints in (6), she is less likely to create it.
- Hammond (2016) calls this sort of underrepresentation **input optimization**; we won't use the term for the phenomenon because it doesn't match our analysis of it (below).

11. Excursus on dealing with probabilities

- It pays to use sound methods …
- We previously tested our hypothesis using the "shuffling" method of Martin (2011).
- But:
  - This method is just a stochastic version of the well-known **Observed/Expected statistic** of Pierrehumbert (1993 et seq.).
  - Example for C_1C_2 clusters:
    - — *Expected* is computed by multiplying the probabilities of C_1 and C_2, then multiply result by total cluster count.
    - — Divide *Observed* C_1C_2 count by this number.
    - — You have "underrepresentation" when O/E is well below 1.
  - O/E is shown by Wilson and Obdeyn (2009) to be **very unreliable —** it yields false, artifactual results when there are multiple overlapping constraints present.
  - As we have found, Wilson and Obdeyn's critique of O/E carries over completely to our own bigram application.
  - E.g. an artificial language designed to underrepresent *SIBILANT CLASH violations will also show low O/E values for *GEMINATE violations, since [ss], [ʃʃ], etc. are underrepresented.
- We need a better method …

12. Desiderata for a probabilistic system

- Our system should:
  - incorporate constraints

- generate probabilities
- be mathematically well-founded in probability theory
- be well-supported in software

13. MaxEnt (Smolensky 1986, Goldwater and Johnson 2003)

- Satisfies all four criteria just given.
- Is completely *un*-fooled, we find, by the cases that delude Observed/Expected.
  - We find that in contrived data such as in (11), where we know the pattern in advance, MaxEnt perfectly sorts out which constraints are responsible for the data, even when the effects of constraints overlap.

14. Quick summary of MaxEnt grammars

- A species of Optimality Theory (Prince and Smolensky 1993), with constraints selecting winner(s) from a set of candidates (GEN).
- Within OT, a species of Harmonic Grammar (Smolensky et al. 1990, Potts et al. 2010); the strength of constraints is expressed not by ranking them but by assigning them numerical **weights**.
- MaxEnt is a **stochastic** theory; not one single winner but a probability assigned to every member of GEN.
- Data-fitting: there are powerful, accessible algorithms, backed by mathematical proof, that reliably find the best weights to match the data you are working with.

15. The MaxEnt formula (Della Pietra et al. 1997, 1)

- This is the key mathematical formula.
- It takes in candidates, constraint violations, and constraint weights, and computes a probability for each candidate.

$$p(\omega) = \frac{1}{Z} e^{-\Sigma_i \lambda_i \chi_i(\omega)}, \text{ where } Z = \Sigma_j e^{-\Sigma_i \lambda_i \chi_i(\omega j)}$$

  - $p(\omega)$ denotes the predicted probability of a candidate ω
  - e is the base of natural logarithms
  - Σ_i denotes summation across all constraints
  - λ_i denotes the weight of the ith constraint
  - $\chi_i(\omega)$ denotes the number of times ω violates the ith constraint
  - Σ_j denotes summation across all possible candidates.

- Hayes (in progress) argues that every single element of this formula, suitably explicated, reflects our commonsense intuitions about how to draw conclusions from evidence.

APPLYING MAXENT TO THE PROBLEM AT HAND

16. Modeling Jane Austen's phonological preferences with a large MaxEnt grammar

- We have construed Austen as a stochastic device that emits hundreds of thousands of word bigrams.
- We will try **two** different MaxEnt grammars whose purpose is to predict the frequency with which she emits any given word bigram.
- The simpler of these two grammars — a baseline grammar — will simply describe in phonological terms the **population of word types** that Austen uses, including their frequency of use.
- The other, larger grammar will be a superset of the smaller one: it will *also* rely on the nine phonological constraints of (6), as applied to word bigrams.
- Do we do a better job of predicting the set of actually-emitted bigrams if we include these phonological constraints?
- If so, we can legitimately claim that these constraints have an influence on Austen's productions.

17. The simplest possible version of this scheme

- Let the baseline constraints regulate the words directly.
  - Thus, constraints like USE "SHRUBBERY"
- But Austen used about 14,000 distinct words in her writing, so this method is not very feasible, at least for us.

18. Achieving the same end by aggregating the data

- Specify the individual words according to their phonological type — these types are the ones that are relevant to the phonological word-bigram constraints we are testing ((6)).
- Example: In the bigram Word1 + Word2,
  - Does Word1 end in a vowel?
  - Does Word2 begin in a vowel?
  - Together, this would suffice to provide a baseline probability for how often Austen would emit a bigram violating *HIATUS, assuming there is no phonological effect of this constraint at the phrasal level (baseline model).
- We call FINAL VOWEL IN WORD1 a **constraint of convenience**; used merely to simplify the calculations.
  - A sufficient number of constraints-of-convenience can substitute for the infeasibly-numerous lexical constraints like USE "SHRUBBERY".
- We find (in small datasets) that constraints of convenience yield the same phonological constraint weights and statistical significance values as lexical constraints.

19. Our constraints-of-convenience

- There are 53 of them.

- For each, we give the phrasal word-bigram constraints for which it serves as part of the baseline.
- For Word 1:
  - FINAL VOWEL — needed to assess *HIATUS
  - FINAL STRESS — needed to assess *CLASH
  - IAMBIC STRESS — needed to assess *IAMBIC CLASH
  - FINAL [−son][−son] — needed to assess *THREE OBSTRUENTS
  - FINAL CC — needed to assess *CCC
  - {FINAL C} (separate constraint for every consonant) needed to assess most of the phonological constraints
- For Word2
  - INITIAL VOWEL — needed to assess *HIATUS
  - INITIAL STRESS — needed to assess *CLASH
  - INITIAL C — separate constraint for every C needed to assess most of the phonological constraints
  - INITIAL [−son][−son] — needed to assess *THREE OBSTRUENTS
  - INITIAL CC — needed to assess *CCC

20. Negative weights for constraints-of-convenience

- We allow the constraints-of-convenience to take on negative weights, meaning it is good to violate them.
- This is fine, in that we aren't taking them seriously as a model of phonology; they are a way of capturing the baseline phonological properties of the lexicon.

21. Setting up a candidate set (GEN)

- As before, proceeding word-by-word makes the problem too big for us to handle:
  - This would set up a candidate for every possible Word1 + Word2 bigram, given Austen's vocabulary.
  - The candidate count would be about 200,000,000 — again, too big for us.
- So, as before, we cope by abstracting away from individual words to the phonological distinctions we care about — we aggregate candidates together.
- Method: Classify all possible word bigrams according to which constraints-of-convenience they violate — how many *distinct patterns of asterisks* in tableau rows are there?
- It turns out that with our constraints, there are only 38,016 patterns, making the project feasible.

22. Tableau dimensions

- Rows with candidates
  - 38,016, as just defined above
- Columns with constraints; two possibilities:

  - 53 constraints-of-convenience (from (19)), used alone we are setting the grammar to match Austen's usage of the various types of Word1 and Word2 — **baseline mode**.
  - 62 columns (53 + 9): the bigger grammar that attempts to improve performance by including the 9 phrasal-bigram constraints of (6) — **testing mode**.
- A frequency column:
  - We plug in the empirical data we are trying to model — observed bigram counts for each row, taken from a corpus of Jane Austen's writings.

23. The actual work we carried out

- Gathering and sorting the data:
  - Pick a text.
  - Reduce it to its bigrams, assessing their violations, and forming a tableau, per above.
- Modeling work:
  - Using software created by Tim Hunter, find the best weights for the constraints, to predict as closely as possible how often Austen uses the various kinds of bigrams.[4]
  - We do this twice; first in baseline mode (53 constraints-of-convenience), then in testing mode (53 + 9 bigram constraints).

24. How to read meaningful results off the grammar?

- *If* word-concatenation respects phonological markedness, then:
  - The weights of the phonological word-bigram constraints should be **positive —** meaning they actually penalize, rather than rewarding, violations.
  - A grammar including them should be **more accurate** (we use log-likelihood criterion) in predicting the observed frequencies.
- We can also do a **statistical test** (likelihood ratio test), assessing the probability that the improvement in predictions is not a statistical accident.
  - This can be done both for the 9 phrasal-bigram constraints as a group, and for the constraints individually.

25. We analyzed 14 data corpora

- These were about 700,000 words each.
- 8 were written corpora, combining works from the oeuvres of famous (and out-of-copyright) figures of English literature:
  - Austen, Darwin, Dickens, Hawthorne, London, Melville, Trollope, Twain
- 6 were spoken, from various corpora publicly available or owned by our department.

[4] There are many other forms of software that can calculate MaxEnt weights. We used the Excel Solver to verify the weights found by Hunter's faster system.

- Collected transcripts of live questions and answers from White House press briefings (1994-1997)
- Michigan Corpus of Academic English: class lectures
- Michigan Corpus of Academic English: dialog
- 2016 Presidential Primary Debates corpus
- National Public Radio "Fresh Air" corpus
- A mélange of six smaller corpora of spoken English: 2016 Primary Debates corpus, Buckeye Corpus, Beatles corpus from Stanton (2016), Michigan Corpus of Academic English, British Academic Spoken English corpus, Human Communications Research Center Map Task corpus.

26. We pre-edited our corpora in various ways

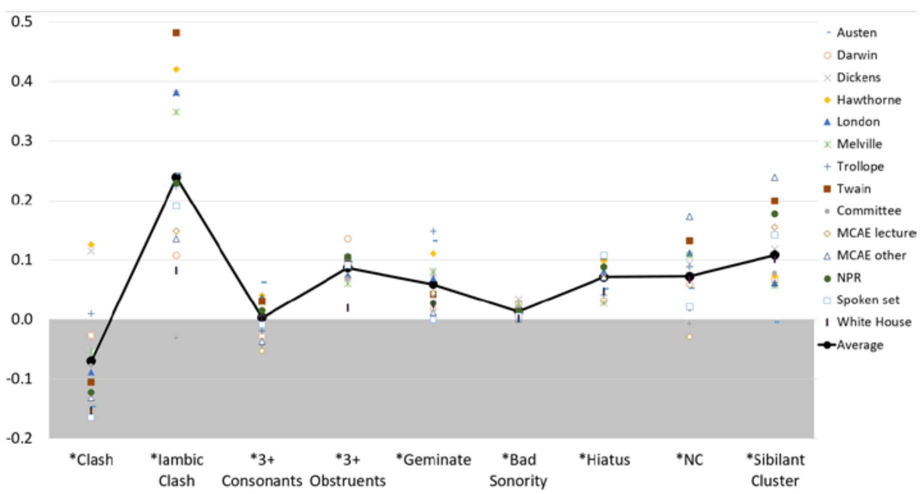
- This establishes "**modes of analysis**".
- We use these modes to diagnose the **mechanisms** by which markedness plays a role in sentence formation.

27. The "Core" mode of analysis

- We exclude bigrams that:
  - include a grammatical function word, or
  - occur more than once in the sample, or
  - occur separated by a major phonological break (assumed from the presence of a punctuation mark)
- The conditions of the Core mode are the most stringent; constraints that test significant in Core tend to test significant elsewhere.

RESULTS

28. Weights obtained for 9 constraints, 14 corpora (Core mode)



- Generally, the weights are positive (out of the gray) for all corpora.
- Taken as a group of 9, the constraint set comfortably passes significance testing for all 14 corpora.
- The "sore thumb" is *CLASH, which usually gets negative weights.

- We think that *CLASH is actually a strong constraint and that this is an artifact that arises only in the Core mode. See below.
- So, tentatively: yes, the phonological constraints are influencing the populations of bigrams emitted by writers and speakers.

CHECKING THE RESULT: TWO METHODS

29. Control method #1: Markedness constraints are often inapplicable at phrasal breaks

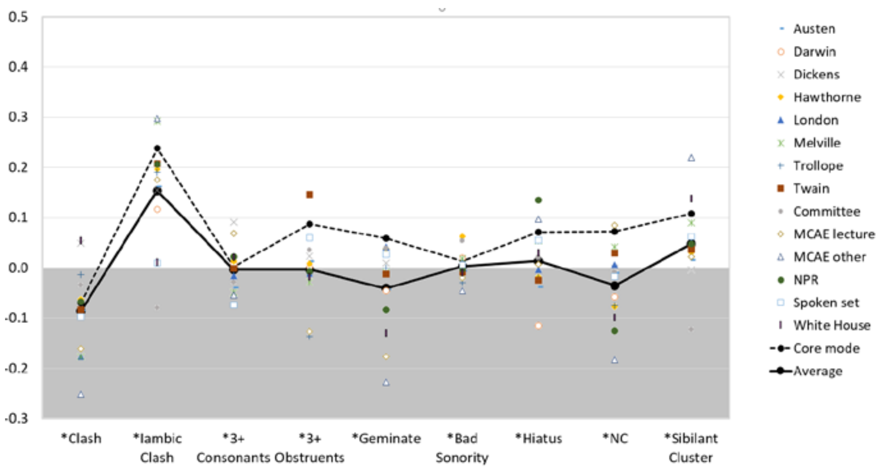
- One of many examples, from Korean:
  - The constraint against plain voiceless stops flanked by voiced sounds generally holds true, but *not* at an Accentual Phrase break (example from Jun 1993:81)

[kəmɨn kojaŋi-e]A	[palmok]A	phonemic form with Accentual Phrasing
[g]	[p]	phonetic output (voicing only within AP)
black cat-GEN	ankle	gloss: 'the ankle of the black cat'

30. "Phrase-break mode": Looking at the bigrams formed across punctuated breaks

- This is the complement set to Core mode.
- We expect phonological constraints to be diminished in force and get lower weights.
- This is true; see chart.

31. Results for 9 constraints, 14 corpora (Phrase-break mode)



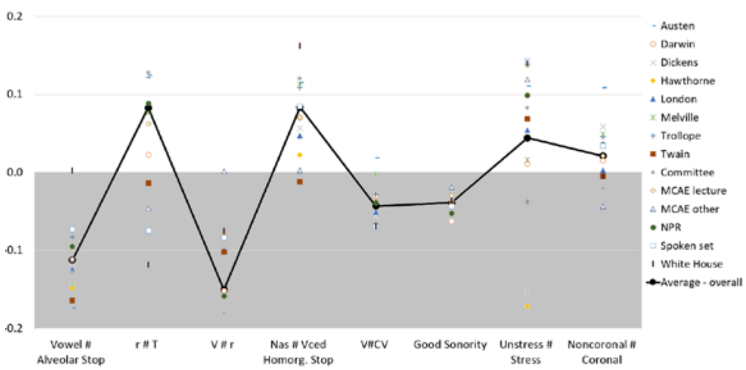
- Core mode result is given as a dotted line for comparison.
- Other than the powerful *IAMBIC CLASH, the Phrase-break constraint weights don't seem to be getting much out of the subzero gray zone.

32. Control method #2: Looking at "silly" constraints ("Pseudoconstraint mode")

- We made up a bunch of constraints out of our heads.
- For these, there is no reason we know of for languages to avoid violating them.

- Here they are:
    - Vowel # Alveolar Stop
    - r # Alveolar Stop
    - V # r
    - Nasal # Voiced Homorganic Stop
    - V # CV
    - In C1 # C2, C1 has higher sonority than C2
    - Unstressed # Stressed
    - Noncoronal C # Coronal C
- We repeated our Core mode, but with these silly constraints; calling it Pseudoconstraint Mode.

33. Weights obtained for 9 constraints, 14 corpora (Pseudoconstraint mode)



- Outcome: Weights are generally not high, and often negative (better to violate them).
- By comparison with (28), Core mode, this tells us that the effects of the phonologically-plausible constraints are probably authentic.
- Indeed, some of the pseudoconstraints that reflect phonological *goodness* are getting reliably-negative weights, as we might expect.

WHAT MIGHT BE THE CAUSES OF PHRASAL MARKEDNESS EFFECTS?

34. Our thoughts on this matter

- We think that more than one mechanism is responsible.
- By varying the conditions of the MaxEnt modeling, you can find evidence to bear on this.

35. Conjectured causal mechanism #1: listed phrases are phonologically optimized

- There is a research literature on listed phrases; researchers have asserted that the vocabulary of listed phrases in a language is extremely large, perhaps more than the number of words (e.g., Pauley and Syder 1983, Jackendoff 1997).

36. Examples of bigrams (from Austen) that my coauthor and I judge to be listed phrases

- *very well*, *great deal*, *young man*, *very good*, *few minutes*, *young ladies*, *next morning*, *soon afterwards*, *same time*, *young woman*, *next day*, *very pretty*, *soon after*, *very soon*, *only one*, *last night*

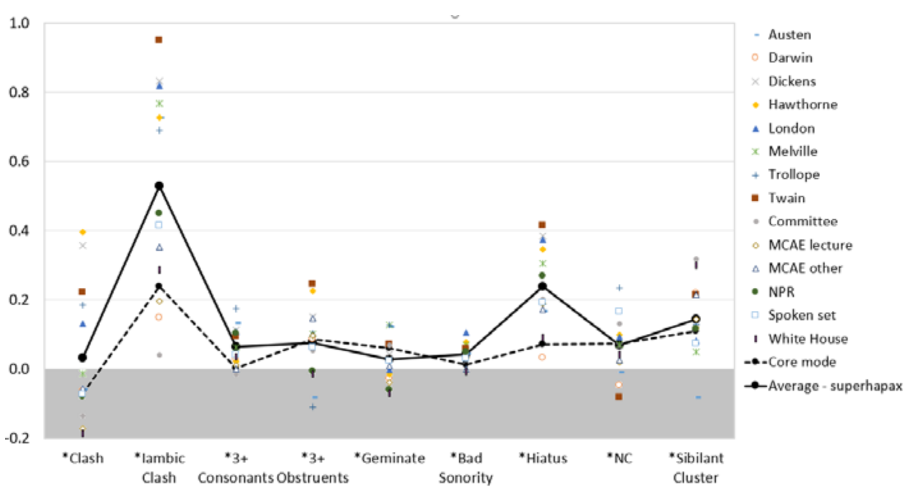
37. Earlier work on phonological markedness effects in listed phrases

- Martin (2011) was an inspiration for our study.
- He found statistical markedness effects in **compounds**.
- His interpretation:
    - Candidate compounds, created on line, get incorporated into the established lexicon preferentially if they are phonologically unmarked.
    - Thus the lexicon of listed compounds accretes well-formed members over time.
    - Then, because established compounds are used frequently, an observed corpus will tend to be phonologically unmarked.
    - The Martinian story extends straightforwardly to listed phrases.
- Some other work drawing similar conclusions: Benor and Levy (2006), Mollin (2012), Green (2016), Schlüter and Lappe (2018).

38. Evidence that listed phrases may play a major role in our own findings

- How to check: In the Core mode (see (28)), we reduced the data to nothing but *hapax* bigrams — occurring only once.
    - Suppose instead we look exclusively at "**superhapaxes**" — forms that occur *more* than once — these are more likely to listed.
    - This is our "Superhapax mode".

39. Constraint weights found for 9 constraints, 14 corpora (Superhapax mode)



- These are generally higher than under the Core mode (dotted line).
- So we think that preferential lexical listing, as applied mostly to phrases, is a major factor in the data patterning we have found.

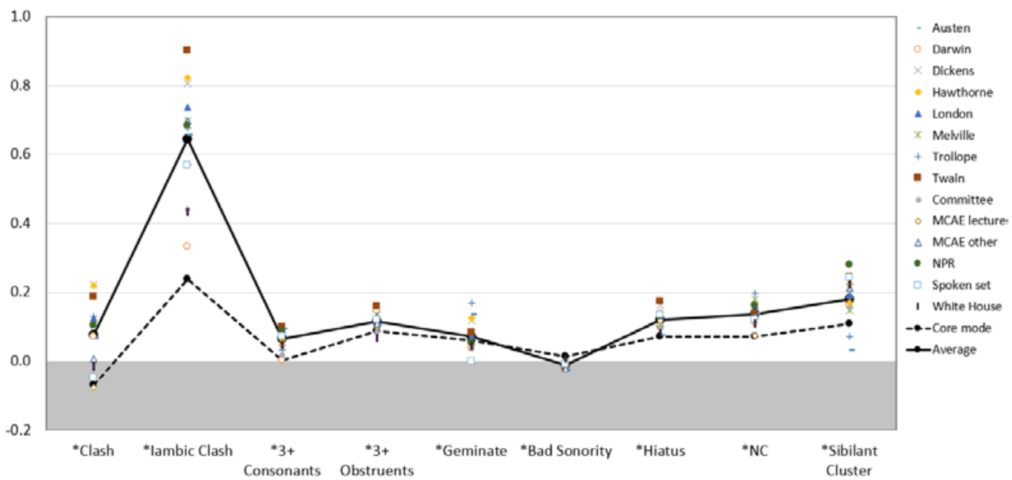
40. Conjectured causal mechanism #2: syntactic choices (and how to diagnose their effects)

- Particular syntactic constructions employ particular **function words**, shown in boldface below.
  - Genitive choice (Shih et al. 2015): *the whéel **of** the cár* >> *the cár's whéel* , in order to avoid violating *CLASH.
  - Dative choice (Anttila et al. 2010): *give hínts **to** Súe* >> *give Súe hínts* by *CLASH
  - Shih and Zuraw's Tagalog case (*túnay **na** ámaŋ* >> *ámaŋ **na** túnay* by *[+nasal][+nasal])

41. Our "Function mode" of bigram analysis

- Like Core mode: hapaxes only, avoid bigrams across phrase breaks.
- But this time *include* bigrams that have function words; e.g. [*the whéel*], [*whéel of*], [*of the*], [*the cár*], etc.

42. Constraint weights obtained for 9 constraints, 14 corpora (Function mode)



- Core weights shown dotted for comparison.
- Clearly there is a substantial increase.

43. Why does Function mode yield higher weights than Core mode?

- Key idea: Core mode, by omitting the function words, has skewed the constraint weights.
- The omission of function words can distort the constraint weights in three ways. We need a taxonomy.

44. I. Both syntactic choices include a function word

- E.g., Tagalog *na*, in (40) above.

- This is a domain where we expect statistically strong phonological Markedness effects, since the syntax give speakers the opportunity to opt for a phonologically unmarked pattern.
- Core mode throws these bigrams away, so the effect of Markedness weakens (weights go down), relative to Function mode.

45. II. Function words *mitigate* constraint violations

- Our canonical case is *CLASH: function words are stressless, so they are clash-mitigating.
- Core mode throws out bigrams in which the speaker avoided a *CLASH violation by using a better word order, e.g. every bigram of *the whéel of the cár* in (40).
- This leaves overrepresented the clash-violating bigrams, like *cár's whéel*.
- Thus, our Core condition actually *encourages overrepresentation* of *CLASH violations.
- This is our explanation for why *CLASH sometimes gets a negative weight in the Core mode — it's an (informative) artifact.

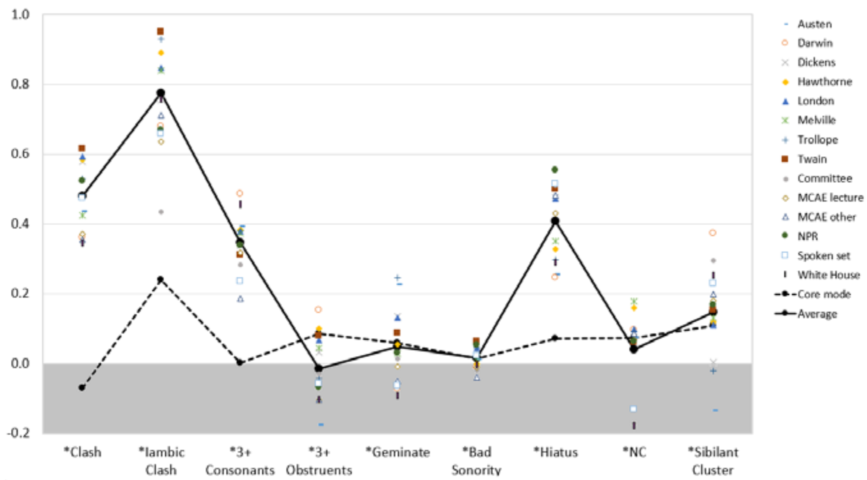
46. III. Function words *aggravate* constraint violations

- We have not found any such cases.
- But in principle, this should yield a *stronger* effect in the Core condition.

47. Putting the factors together

- We conclude the data presentation with a **Mass-comparison** mode: include *all* phrase-internal bigrams, and counting them by tokens, not types.

48. Constraint weights obtained for 9 constraints, 14 corpora (Mass-comparison mode)



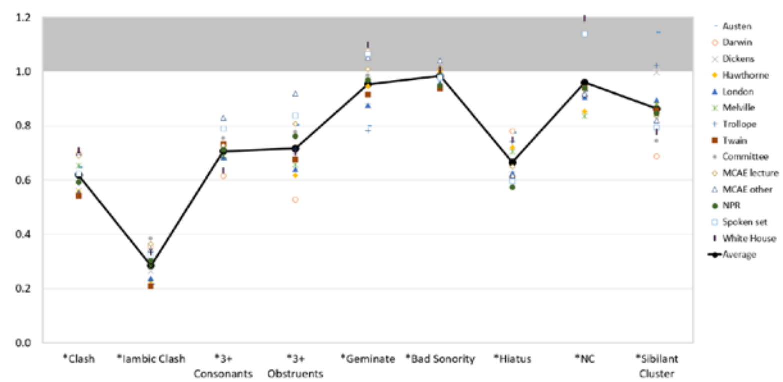
- This is the strongest outcome, since we are *combining* the effects of including superhapaxes and including function words.
- A additional comparison not reported here suggested that using token frequency is giving the weights an additional boost.

- An odd switch is the rise of weight in *3+ CONSONANTS and the fall in *3+ OBSTRUENTS.
  - ➢ However, these constraints are "ganged" (whatever violates *3+ OBSTRUENTS also violates *3+ CONSONANTS).
  - ➢ The weights mean that triple obstruent clusters are avoided, but not any more than any triple consonant clusters are.

49. What do the constraint weights mean in terms of actual probability of use?

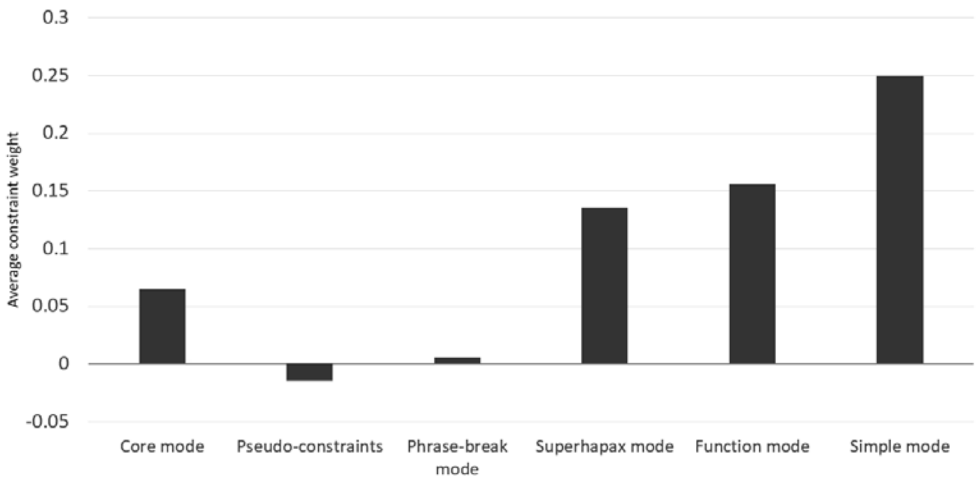
- From the MaxEnt math of (15):
  - ➢ Imagine two candidates, identical except that one violates a constraint with weight *w* and the other doesn't.
  - ➢ They will receive a **probability ratio** (odds), such that the non-violator is e^w times as likely to be output by the grammar.
- For intuition's sake, we can redo chart (48) to display these probability ratios (we also augmented the weights of *IAMBIC CLASH and *3+ OBSTRUENTS with the more general constraints that gang with them):

50. Results for 9 constraints, 14 corpora — Simple mode; ganging incorporated; plotted as probability ratios



- Average reduction in probability (one minus the value shown) ranges from the remarkable 71.5% for *IAMBIC CLASH to just 1.5% for *BAD SONORITY.

51. Comparing all the modes in one place: average constraint weight for each mode



- This averages the weights for all 9 constraints, giving values mode by mode
- We see the low values for our two control conditions (Pseudo-constraints and Phrase-break mode)
- We see the effects, synergistic, of the Superhapax mode (attributed to listed phrases) and the Function mode (attributed to syntactic choices) together in Simple mode.
- Even Core gets a significant overall effect, perhaps from lexical choices Schlüter (2005), Schlüter and Knappe (2018); or perhaps because it is an imperfect control for lexicalization and syntax.

GENERAL DISCUSSION

52. Summarizing

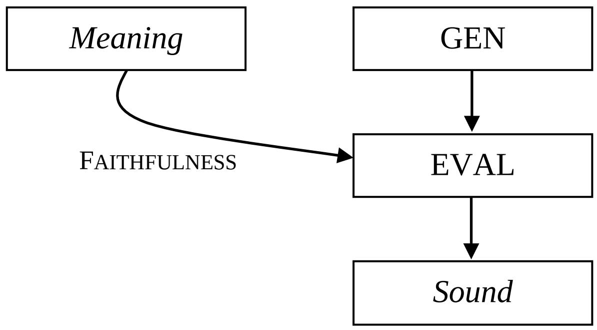
- Earlier work on phonological effects in sentence formation detected such effects in specific syntactic constructions and lexical items.
- Our scaled-up, across-the-board approach, implemented in MaxEnt, reinforces the view that such effects are pervasive, found for many constraints and many contexts.
- Our different modes of analysis illuminate the role of two factors in producing the observed Markedness effects:
  - preferential lexical listing of phonologically unmarked phrases
  - syntactic choices made on line

53. What sort of theory of the organization of grammar is compatible with our findings?

- Let us suppose that something like Optimality Theory is correct for all grammatical components.
  - Grammar is embodied by a great number of interacting constraints of all types (semantic, syntactic, morphological, phonological, phonetic).
  - Complexity in language arises from the prioritized application of many simple conflicting constraints.

- Let us suppose further that the *MaxEnt version* of OT is basically correct.
    - Then any constraint whatsoever — including phonological markedness constraints — can perturb output frequencies.

54. A pure-parallelist MaxEnt architecture compatible with this view (replacing (2))



- In this approach, candidates must be highly structured objects with semantic, syntactic, phonological, morphological, and phonetic structure.
- EVAL assigns a probability, in most cases vanishingly small, to every candidate.
- Faithfulness constraints encourage us to choose an output that does a good job of expressing our communicative intent.
- This is a MaxEnt-friendly conception of parallelist architecture; for established research programs see e.g. Jackendoff (2002, 2010), Bresnan et al. (2015).

55. The effect of constraints in grammar

- MaxEnt's key principle: *the only effect of a constraint is to lower the probability of candidates that violate it.*
- But different constraint weights can give radically different effects.
    - Smallish weights: the subtle effects of phonology on sentence construction that we and our colleagues are discovering.
    - Big weights: obligatory phonological patterns found in phonotactics or alternation (e.g., classical problem sets).

56. [if time] Consequences for rule-based phonology

- One unsympathetic view of Optimality Theory: the Markedness constraints are only an artificial portion of the phonological rules, which are the true objects.
    - The OT theorist is said to pointlessly extract from the rule A → B / C ___ D the markedness constraint *CAD.
- But the Markedness constraints of OT do other things as well.
    - Explain conspiracies (Kisseberth 1970)
    - Describe the detailed pattern of optionality in complex cases ("modulation" of process frequency; Anttila 1997 et seq.)
    - Predict frequency of speech errors (Goldrick and Daland 2009)

   - This project: they lower output probabilities in sentence formation.

- Rule-based phonology has nothing to say about our data — nothing is turning into anything else. Speakers are merely refraining — to some degree — from saying phonologically-marked things.

57. Where to go next?

- We have used MaxEnt here essentially as a method of statistical evaluation.
- But in the future we would like to start doing full-scale MaxEnt *grammars*, modeled on the architecture of (54).
- Working with such grammars would give us greater confidence that the results we obtained here are correct, by accurately modeling all syntactic influences on bigrams.
- Such grammars are also the future of linguistics, we think: large-scale, computationally-implemented, multicomponent grammars will

   - help us avoid myopia and anecdotalism
   - permit serious evaluation of *all* theoretical proposals (not just the ones tested here).

References

Anttila, Arto (2016) Phonological effects on syntactic variation. *Annual Review of Linguistics* 2:115-137.

Anttila, Arto, Matthew Adams, and Michael Speriosu (2010) The role of prosody in the English dative alternation. *Language and Cognitive Processes* 25:946-981.

Benor, Sarah Bunin and Roger Levy (2006) The Chicken or the egg? A probabilistic analysis of English binomials. *Language* 82:233–278.

Bolinger, Dwight (1965) Pitch accent and sentence rhythm. In his *Forms of English: Accent, Morpheme, Order*. Cambridge, MA: Harvard University Press.

Bresnan, Joan, Ash Asudeh, Ida Toivonen, and S. Wechsler. 2015. *Lexical-Functional Syntax, 2nd edition.* Wiley.

Chomsky, Noam (1965) *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Fijn van Draat, P. (1910) Rhythm in English prose. *Anglistische Forschungen.*

Goldwater, Sharon & Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120.

Goldrick, Matthew and Robert Daland (2009) Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. *Phonology* 26:147–185.

Hammond, Michael (2016) Input optimization: phonology and morphology. *Phonology* 33:459

Hayes, Bruce (in progress) MaxEnt grammars: rationale and spreadsheet tutorial. Ms., Department of Linguistics, UCLA.

Inkelas, Sharon and Draga Zec (1990) *The Phonology-Syntax Connection*. Chicago: U. Chicago Press.

Jackendoff, Ray. 2002. *Foundations of Language*. New York: Oxford University Press.

Jackendoff, Ray. 2010. *The Parallel Architecture and its place in cognitive science*. In B. Heine and H. Narrog (eds.), *The Oxford Handbook of Linguistic Analysis.* Oxford: OUP.

Kisseberth, Charles W. 1970. On the functional unity of phonological rules. Linguistic Inquiry 1:291–306.

Martin, Andrew (2011) Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language* 87:751-770

Mollin, Sandra (2012) Revisiting binomial order in English: ordering constraints and reversibility. *English Language and Linguistics* 16:81–103.

Pater, Joe. 2004. Austronesian nasal substitution and other NÇ̥ effects. In John McCarthy, ed., *Optimality Theory in Phonology: A Reader*. Oxford and Malden, MA: Blackwell.

Pauley, A. and F. H. Syder (1983) Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J.C. Richards & R.W. Schimdt, eds., *Language and Communication*. London: Longman.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt and Michael Becker. 2010. Harmonic Grammar with Linear Programming: From linear systems to linguistic typology. *Phonology* 27: 77-117.

Prince, Alan, and Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar*. Rutgers University Center for Cognitive Science Technical Report 2.

Ryan, Kevin (2018) Prosodic end-weight reflects phrasal stress. *NLLT* (online, print forthcoming)

Schlüter, Julia (2005) *Rhythmic grammar: The influence of rhythm on grammatical variation and change in English*. Berlin: Mouton de Gruyter.

Schlüter, Julia and Gabriele Knappe (2018) Synonym selection as a strategy of stress clash avoidance. in Sebastian Hoffmann, Andrea Sand, Sabine Arndt-Lappe, and Lisa Marie Dillman, eds., *Corpora and Lexis*, pp. 69-105. Leiden: Brill.

Shih, Stephanie S. (2012) Linguistic determinants of English personal name choice. Paper given at the Annual Meeting of the Linguistic Society of America, Portland, OR, 7 January 2012.

Shih, S. (2017) Phonological influences in syntactic choice. In Vera Gribanova and Stephanie S. Shih (eds). *The morphosyntax-phonology connection: locality and directionality at the interface.* OUP 223–252.

Shih, Stephanie S., Jason Grafmiller, Richard Futrell, and Joan Bresnan (2015) Rhythm's role in predicting genitive alternation choice in spoken English. In I. Vogel and R. van de Vijver (eds). *Rhythm in Cognition and Grammar: A Germanic Perspective*. Berlin: De Gruyter.

Shih, Stephanie S. and Kie Ross Zuraw (2017) Phonological conditions on variable adjective-noun word order in Tagalog. *Language: Phonological Analysis* 93:e317-e352.

Smolensky, Paul (1986) Information processing in dynamical systems: Foundations of Harmony Theory. In J. L. McClelland, D. E. Rumelhart and the PDP Research Group. *Parallel distributed processing, Vol. 2: Psychological and biological models.* Cambridge: MIT Press. 390-431.

Stanton, Juliet. 2016. Learnability shapes typology: the case of the midpoint pathology. *Language* 92: 753–791.

Wilson, Colin and Marieke Obdeyn (2009) Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay. Ms., Johns Hopkins University.

Zuraw, Kie (2015) Allomorphs of French *de* in coordination: a reproducible study. *Linguistics Vanguard* 1(1): 57–68.