Bruce Hayes[1]
University of California, Los Angeles

Talk at Seoul National University
26 June 2018

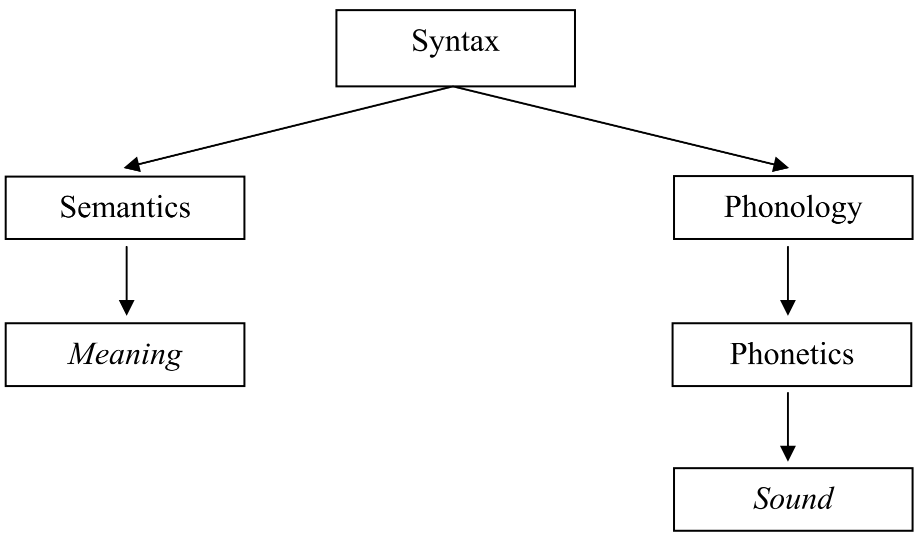
Phonological markedness effects in syntax: subtle but ubiquitous

CONTEXT

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 any phonological consequences of word-concatenation.

3. Challenges to the feed-forward model arrive from multiple directions

- Recent work:
    - Shih and Zuraw (2017): Tagalog speakers use the twin syntactic constructions Adj. $\begin{Bmatrix} \textit{ŋ} \\ \textit{na} \end{Bmatrix}$ Noun vs. Noun $\begin{Bmatrix} \textit{ŋ} \\ \textit{na} \end{Bmatrix}$ Adj. in ways that statistically avoid violating phonological constraints:

        — *[+nasal][+nasal]
        — *HIATUS
        — *NC̥

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

- Much other work, e.g. Inkelas and Zec (1990), Zuraw (2015), Shih et al. (2015), Shih (2012, 2017), Anttila (2016), Ryan (in press)

4. Goals

- Earlier work studies specific syntactic choices, such as Tagalog adjective-noun word order.
- We seek to learn how general these effects are from an **across-the-board, brute force** approach:
  - Look at whole sentences — *all* word concatenation.
- Why do this?
  - We can look at a wider variety of phonological constraints.
  - We may also get a clearer sense of *why* the patterns are being observed.

CONSTRAINTS WE WILL STUDY

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

- Criteria:
  - Powerful: close to inviolable within words
  - Have good typological support.

6. Examples of the constraints that we tested

- *CLASH (adjacent stresses).
  - 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. *ĕ̲l̲è̲c̲t̲r̲ó̲nic*), all morphologically motivated and often repaired by particular speakers (*èlĕctrónic*)
  - This has been noticed for over a century as phrasally relevant (van Draat 1910-12, Bolinger 1965).
- *TRIPLE OBSTRUENT CLUSTER
  - Only about 200 violating words in my searchable lexical database of 18,000 words,[2] mostly with prefixes like *ex-* (*extract* [ɛ<u>kst</u>rækt]).
- *TRIPLE CONSONANT CLUSTER
  - Of course, more common, but only 1300 examples in my database.

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

- *SIBILANT CLUSTER
  - *No exceptions at all* in words; and made-up cases sound strange: ??[ˈmɪsʃən].

- *GEMINATE
  - Exceptions occur only with productive affixes, like *unknown* [ʌnnoʊn].
  - No monomorphemic words at all with geminates; e.g. "*Hannah*" *[ˈhænnə].

- *HIATUS
  - This is actually violated a lot (2300/18000), yet its typological validity tempts us to try it.
  - In the future we will explore more specific types of hiatus that are strict in English (e.g. *[əV]).

WORD BIGRAMS AS A METHOD FOR DETECTING PHRASAL MARKEDNESS

7. Focusing the inquiry on *word bigrams*

- = two consecutive words of a text

8. Restating the question

- When English speakers form sentences, 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.

- *Emma* contains about 160,000 word bigrams, of which the first four, with our IPA translation, are:

[Emma Woodhouse]	[ɛmə wʊdhaʊs]
[Woodhouse, handsome]	[wʊdhaʊs, hændsəm]
[handsome, clever]	[hænsəm, klɛvɚ]
[clever, and]	[klɛvɚ, ənd]
…	…

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 (5), she is less likely to create it.

11. Dealing with probabilities

- We need a formal framework that
  - incorporates constraints
  - generate probabilities
  - is mathematically well-founded in probability theory
  - is well-supported in software
- We use **maxent grammars** (Smolensky 1986, Goldwater and Johnson 2003).
- Next section: a summary
- Then: back to the word-bigram problem

MAXENT GRAMMARS

12. What are maxent grammars?

- Intellectual ancestry:
  - The maxent framework is **constraint-based** — a version of Optimality Theory (Prince and Smolensky 1993). Full name could be “Maximum Entropy Optimality Theory”.
  - Maxent is **stochastic**, assigning a **probability** (sometimes substantial, sometimes vanishingly low) to every member of GEN (earlier frameworks, such as Boersma 1998, also do this).
  - In maxent, the strength of constraints is expressed not by ranking them but by assigning them numerical **weights** (hence, a type of Harmonic Grammar; Smolensky et al. 1990, Potts et al. 2010)
- What is unique to maxent?
  - There is a mathematical formula ((13)) that takes in candidates, constraint violations, and constraint weights, and **computes a probability** for each candidate.
  - The formula, when analyzed, can be shown to **behave intuitively**:
    —Higher weights give constraints a greater influence on the outcome of the grammar.
    —It takes a lot of evidence (aggregate weight) for a candidate’s probability to approach certainty (p = 0 or 1).
  - There are powerful, accessible **algorithms** (we use one in Excel) that reliably find the best weights to match the data you are working with.

13. The maxent formula (Della Pietra et al. 1997, 1)

- $p(\omega) = \frac{1}{Z} e^{-\Sigma_i \lambda_i \chi_i(\omega)}$, 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 scansions.

14. Classical phonology isn't going away

- Under feasible assumptions (Prince 1997) every classical OT grammar has a maxent translation, where strict ranking is rendered as large differences in weight.

APPLYING MAXENT TO THE PROBLEM AT HAND

15. Modeling Jane Austen's phonological preferences with a large maxent grammar

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

16. The simplest possible version of this scheme

- Regulate the words directly, with constraints like USE "KNIGHTLY".
- But Austen used about 10,000 distinct words in writing *Emma*, so this method is too hard for us.

17. Achieving the same end by aggregating the data

- Specify the 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 sibilant?
  - Does Word2 begin in a sibilant?

- Together, this would suffice to determine how often Austen would emit a bigram violating *SIBILANT CLASH — assuming there is no phonological effect of this constraint at the phrasal level (baseline model).
- So, a sufficient number of such **unigram baseline constraints** could substitute for the infeasibly-numerous lexical constraints like USE KNIGHTLY.

18. Our set of unigram baseline constraints

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

- For Word2
    - *INITIAL SIBILANT needed to assess *SIBILANT CLASH
    - *INITIAL STRESS needed to assess *CLASH
    - *INITIAL VOWEL needed to assess *HIATUS
    - *INITIAL C — separate constraint for every C needed to assess *GEMINATE, and others
    - *INITIAL [−son][−son] needed to assess *THREE OBSTRUENTS
    - *INITIAL CC needed to assess *CCC

19. Setting up a GEN function

- A GEN that is too big for us to handle:
    - Every possible Word1 + Word2 bigram, given Austen's vocabulary.
    - This would number about 100,000,000 — again, too big!
- Again, we can cope by abstracting away from individual words to the phonological distinctions we care about.
- Strategy: Classify all possible word bigrams according to what constraints they violate.
    - If any two word bigrams have different violation profiles, they must be assigned to a different category.
- It turns out that with our constraints, there are only 38,016 categories!
- Example: "0 1 ObsC T / P SonC 1"
    - This means: "any candidate with (a) a Word1 with a stressless penultimate syllable and stressed final syllable and an obstruent in penultimate position and a final [t]; (b) a Word2 beginning with [p] with a sonorant consonant in second position and initial stress".

   - This bigram violates a number of constraints, such as *IAMBIC CLASH, *THREE OBSTRUENTS, *FINAL T, *INITIAL P, etc.
   - Real-life example from Austen: *exact plan* [əgzǽkt plǽn][3]
- In sum: any type of word bigram that has distinct constraint violations is a separate member of GEN.

20. Tableau dimensions

- Rows with candidates
   - 38,016, as just defined above
- Columns with constraints:
   - 53 unigram baseline constraints (from (18)), used alone we are setting the grammar to match Austen's usage of the various types of Word1 and Word2 — **baseline mode**.
   - 60 columns (53 + 7): the bigger grammar that attempts to improve performance by including the 7 phrasal-bigram constraints of (5) — **testing mode**.
- One more column needed:
   - We still need to 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.

21. The actual work we carried out

- Gathering and sorting the data:
   - Pick a text.
   - Reduce it to its bigrams as in (7).
   - Sort the bigrams according to their type (in the 38,016-member taxonomy), obtaining the column of empirical frequencies to add to the tableau.
- Modeling work:
   - Using the Excel Solver utility, find the best weights for the constraints, to predict as closely as possible how often Austen uses the various kinds of bigrams.
   - We do this twice; first in baseline mode (53 unigram constraints), then in testing mode (53 + 7 bigram constraints).

22. 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.[4]
   - A grammar including them should be **more accurate** in predicting the observed frequencies.

---

[3] This is not from *Emma* but from *Sense and Sensibility*, which we also analyzed.

[4] Many of the baseline constraints get negative weights; this is fine, since they describe phonologically-good things.

- 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 phrasal-bigram constraints as a group, and for the constraints individually.

23. We did five different data corpora

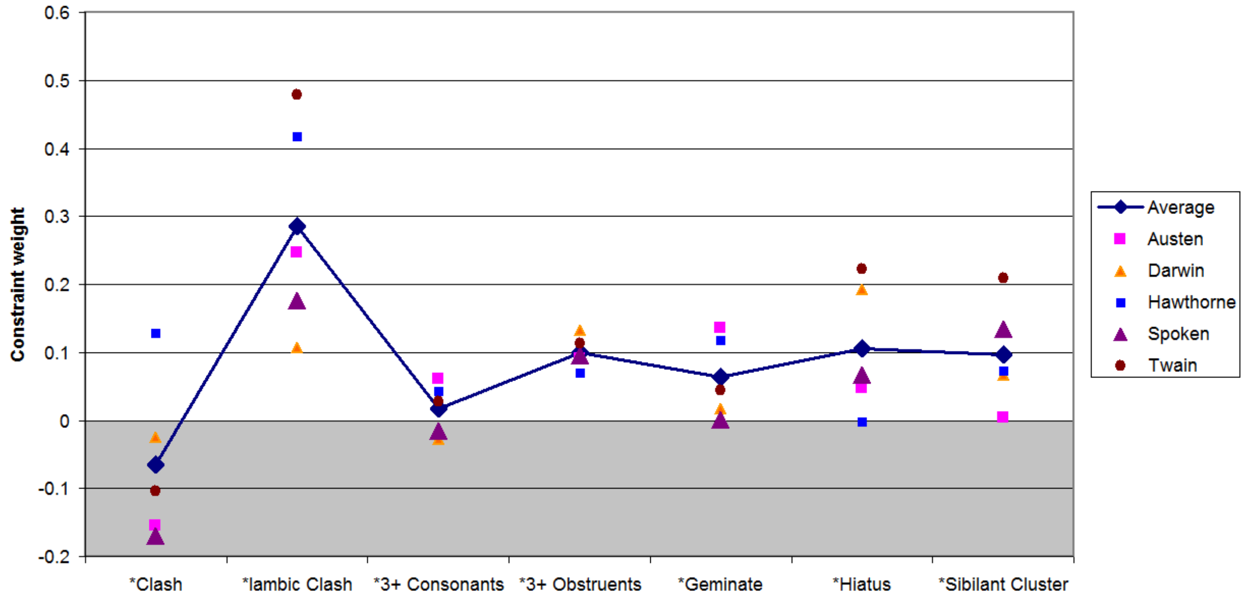
- List:
    - Six novels by Jane Austen, concatenated into one data file (722,000 words).
    - Six novels by Mark Twain (568,000 words)
    - Six novels by Nathaniel Hawthorne (592,000 words)
    - Six non-fiction works by Charles Darwin (935,000 words)
    - A mélange of conversations from six corpora of spoken English (~ 1,000,000 words; 2016 Primary Debates corpus, Buckeye Corpus, Beatles corpus (Stanton 2016), Michigan Corpus of Academic English, British Academic Spoken English corpus, Human Communications Research Center Map Task corpus).
- The spoken corpus is particularly important, since it is not the results of post-hoc, reflective prose editing.

24. Pre-editing of the bigram set (default mode of analysis)

- Before making a tableau, we throw out any bigrams that
    - include a grammatical function word, or
    - occur more than once in the sample
    - occur separated by a major phonological break (assumed from the presence of punctuation)
- Later, we'll try doing the maxent analyses with different assumptions — this will help us understand *why* we are seeing the patterns that occur.
- The conditions given here are probably the most stringent test of our overall hypothesis.

RESULTS

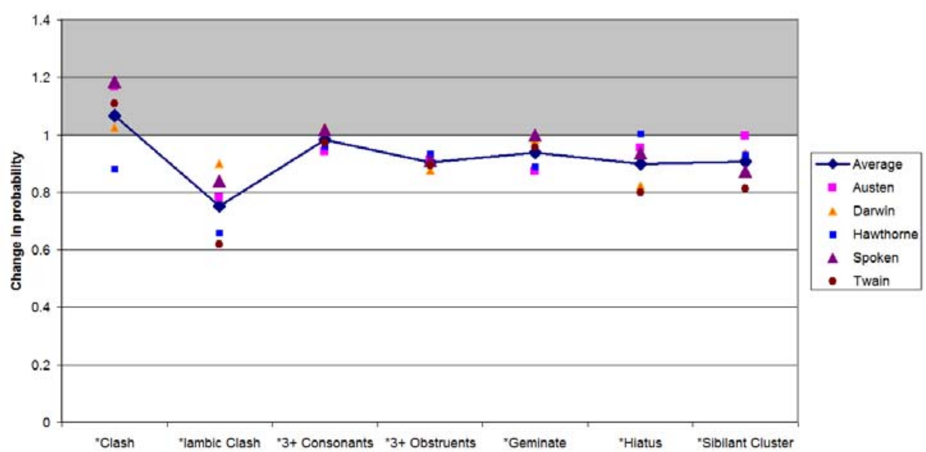
25. Results for the five corpora



- Generally, the weights are positive for all corpora.
- The exception is *CLASH, slightly negative. More on this later on.

26. What does the weights mean in intuitive numerical terms?

- From the maxent math:
  - 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 (25) to display these probability ratios:



- *CLASH is somewhat above one.

- For others, average reduction in probability ranges from very substantial 25% for *IAMBIC CLASH to just 1.7% for *3+ CONSONANTS.

27. Significance testing

- A likelihood-ratio test shows that in the aggregate, the improvement in the model from incorporating the extra bigram constraints is highly significant across corpora.
- For Austen, log likelihood of the data improves by 61.8, $p < .0001$.
- Other corpora similar.
- The individual constraints may also be tested, and generally emerge as significant.[5]

CHECKING THE RESULT IN VARIOUS WAYS

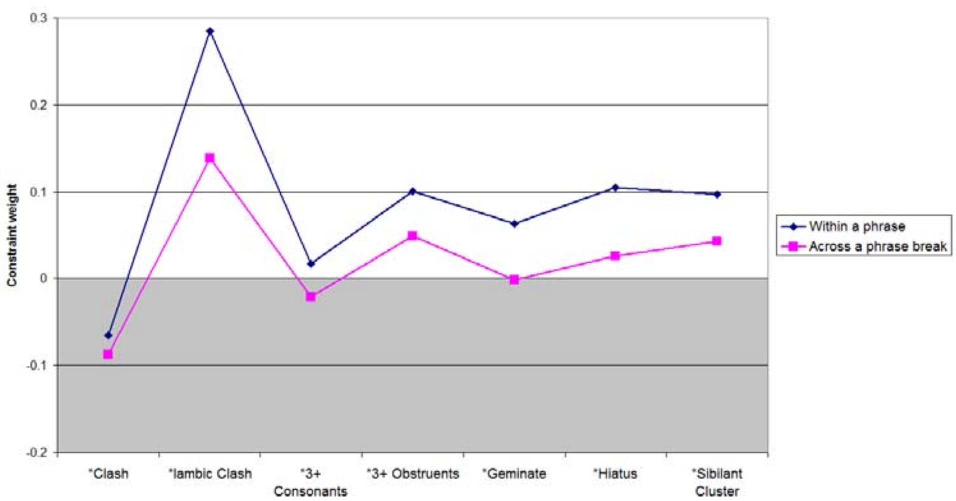
28. Markedness constraints are often inapplicable at phrasal breaks

- 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'

29. Looking at the English bigrams formed across punctuated breaks

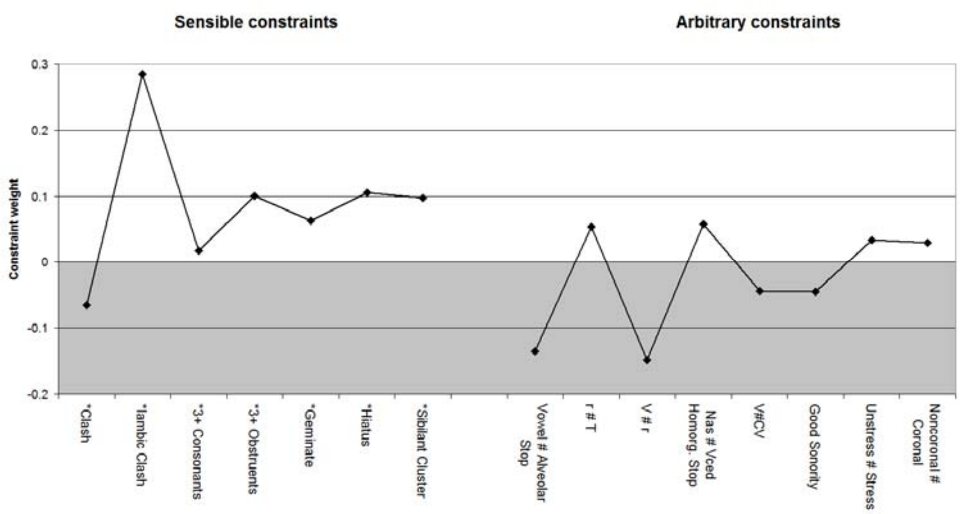
- We use punctuation as a rough approximation to Intonational Phrasing.
- We expect phonological constraints to be diminished in force and get lower weights.
- This is true: averages across the five corpora



[5] Caveat: this is still in progress for all the constraints in all the corpora.

30. Looking at "silly" constraints

- We just made them up out of our heads.
- 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 more sonority than C2
  - Unstressed # Stressed
  - Noncoronal C # Coronal C
- Weights are generally not high, and often *negative* (better to violate them).



METHODOLOGICAL EXCURSUS: WHY WAS IT NECESSARY TO USE MAXENT?

31. What we used to use in this project: the "Observed over Expected" statistic

- Here is a hypothetical example meant to illustrate this statistic.
  - Word1's of bigrams end with a sibilant **0.1** of the total population of words.
  - Word2's begin with sibilant **0.1** of the total.
  - Baseline expectation: bigrams should violate *SIBILANT CLASH 0.1 * 0.1 = **0.01** of the total.
- Suppose further that in reality, *SIBILANT CLASH is violated only **0.005** of the total.
- Then **Observed/Expected** for *SIBILANT CLASH = 0.005/0.01 = **0.5**.
  - This is less than one, and taken to mean, "Violations of *SIBILANT CLASH are underrepresented."

32. Uses of the Observed/Expected statistic

- It has been important to the study of **root cooccurrence constraints**, in Arabic and other languages (Frisch, Pierrehumbert, and Broe 2004, Coetzee and Pater 2008, Wilson and Obdeyn 2009).
- Martin (2011) offers a more sophisticated version of O/E, which we used in the early stages of our own project: expected values are created by **random scrambling** of bigrams, which helps with significance testing.

33. Why the Observed/Expected statistic leads to errors

- The probabilities in non-trivial cases result from **multiple overlapping constraints**.
- Experimentation with made-up examples shows that Observed/Expected cannot correctly attribute the data patterns to the constraints that underlie them.

34. Hypothetical example of the failure of Observed/Expected

- Premises:
  - Language X has 8 consonants: /p t b d s ʃ z ʒ/ (last four are sibilants).
  - All 8 consonants have equal frequency.
  - All words are of the shape CVC.
  - Speakers of Language X tend to avoid violating *SIBILANT CLASH when they make up sentences, such that they emit word bigrams like /tas ʃap/ only *half as often* as other bigrams.
- When we calculate the Observed/Expected statistic, it does indeed identify this effect:
  - For pairs violating *SIBILANT CLASH: O/E = **0.778**.
- Unanticipated consequence: the Observed/Expected for *GEMINATE turns out to be **0.840**, also indicating underrepresentation.
  - Surely this is wrong! There's nothing we did in the design of this language that says geminates should be bad.
  - Why did it happen? Because a substantial fraction of pairs violating *SIBILANT CLASH (/ss, ʃʃ, zz, ʒʒ/) also violate *GEMINATE.

35. Maxent is not fooled

- Running Language X through the maxent weighting procedure, with both *SIBILANT CLASH and *GEMINATE available, we find:
  - *GEMINATE receives a weight of *zero*, as seems appropriate.
  - *SIBILANT CLASH receives a positive weight (0.69) that exactly predicts the 0.5 underrepresentation.
- In general: when we need to disentangle the effects of multiple overlapping constraints, Maxent, based on sound principles of probability theory, is likely to home in on the truth and Observed/Expected, being ad hoc, is likely to obscure it.

36. Where we learned this

- The example just given is from us, but the lesson comes from **Wilson and Obdeyn (2009)**.
- They use a different example, from root structure, to show that maxent gives trustable results where Observed/Expected does not.

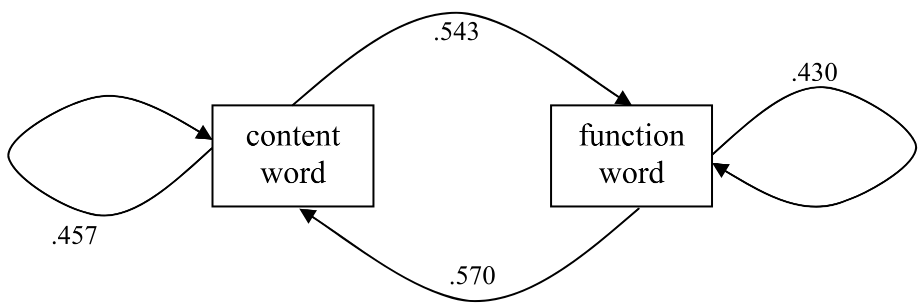
WHAT MIGHT BE THE CAUSES OF PHRASAL MARKEDNESS EFFECTS?

37. Our thoughts on this matter

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

38. I. The grammar of English is set up to be phonologically optimizing

- Most importantly, English syntax tends to **alternate content and function words**.
    - ➢ This illustrated by the following probabilistic transition diagram calculated for Austen's *Emma*:



- The function words of English are formed so that they tend to avoid violations when placed next to other words.
    - ➢ They seldom begin or end in consonant clusters.
    - ➢ They do not bear stress, so cannot trigger *CLASH.
    - ➢ They typically end in one consonant, averting *HIATUS violations.[6]
- In the primary modeling above, we controlled for function words, by examining only bigrams that consist of two content words (cf. (24))
- What happens when we leaving all the function words in? Constraint weights go up, by an average of **0.224**.[7]
    - ➢ This is *a lot* — review (25) for the original values.
- We suspect that English syntax/function words evolved diachronically to favor unmarked phonological sequences.

---

[6] It doesn't matter so much that they often *start* with vowels, because they are so often phrase-initial, and we are only counting phrase-internal violations.

[7] Austen corpus; others in progress.

39. II. Listed phrases are phonologically optimized

- Some examples of lexically-listed phrases: *a great deal*, *very good*, *a few minutes*
- Some researchers believe that the vocabulary of listed phrases in a language is *larger than the vocabulary of words* (Pauley and Syder 1983).
- Martin's (2011) study, which inspired ours, concludes:
  - Compounds get incorporated into the lexicon preferentially if they are phonologically unmarked.
  - We think the same may be true of listed phrases.
- Why do we think this?
  - In the main study (see (24)), we reduced the data to nothing but *hapax* bigrams — occurring only once.
  - These are much more likely to have been "formed on line," not listed.
  - But when we undo this effect, by *not* reducing to hapaxes, the constraint weights generally go up by an average of **0.216**, again a lot.[7]
- So we conjecture that much of our effect is the result of English containing a great number of listed phrases, and that per Martin phrases get listed more often if they are phonological unmarked.

40. III. Even so, on-line speech planning favors outputs that are phonologically unmarked

- Remember that when we control for both of the above factors, there are *still* significant markedness effects — as our primary modeling efforts show.
- So we are lead to the view that even in on-line speech production, speakers favor phonologically unmarked word bigrams.

41. An unstudied mechanism: word choice

- Consider a canonical case of synonymy in English, *sofa* vs. *couch*:
  - When speakers select between *sofa* and *couch*, are they guided by pressure to avoid constraint violations?
  - If so, they should mildly prefer:

    — *sofa selection* over *couch selection* (*SIBILANT CLASH)
    — *couch approval* over *sofa approval* (*HIATUS).

- We think we can study this but we haven't done it yet …

GENERAL ISSUES ARISING FROM OUR WORK

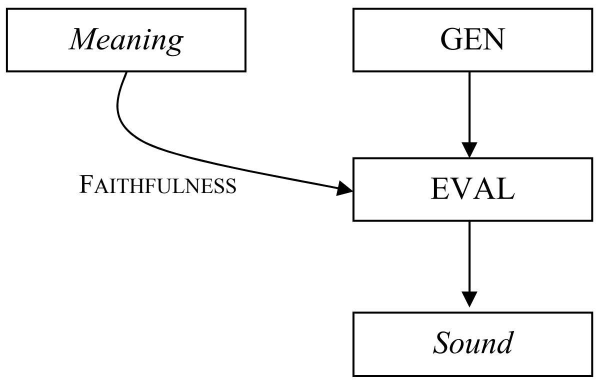
42. What sort of theory of the organization of grammar is compatible with our findings?

- 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 (the key lesson of OT).

- Let us suppose further that **stochastic** constraint based linguistics, as in the maxent version of OT, is basically correct.
  - Then **any constraints whatsoever** — including phonological markedness constraints — can perturb output frequencies in subtle ways.

43. A pure-parallelist grammatical architecture compatible with this view (replacing (2))



- In this approach, candidates must be highly structured objects with semantic, syntactic, phonological, morphological, and phonetic structure.
- This is just one conception of a parallelist architecture; for established research programs see e.g. Jackendoff (2002, 2010), Bresnan et al. (2015).

44. Markedness in phonology

- Markedness constraints have a very specific role in classical Optimality Theory: they serve to "drive" phonological processes.
  - Translating from traditional rule-based phonology, we could infer from the rule A → B / C ___ D the markedness constraint *CAD.
- In the maxent view, Markedness constraints could play a broader role.
  - They **lower the probability of candidates that violate them**.
  - … and, assuming (43), this can include whole sentences!
  - This yields our results.
  - It also implies that classical phonology is only one of consequences of phonology markedness.
  - N.B. Rule-based phonology has nothing to say about this.

45. Summarizing

- Earlier work on phonological effects in syntax detected such effects in specific constructions.

- Our scaled-up, "brute force" approach, implemented in maxent, suggests that such effects are pervasive, found for many constraints and many contexts.
- Our results offer encouragement to the pursuit of parallelist models of grammatical organization.

REFERENCES

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

Boersma, Paul. 1998. *Functional phonology*. The Hague: Holland Academic Graphics.

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.

Coetzee, A. W. and Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory* 26:289–337.

Frisch, S. A., Pierrehumbert, J. B., and Broe, M. (2004). Similarity avoidance and the OCP. *NLLT* 22:179–228.

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.

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.

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

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. 2002. Anything goes. In *A new century of phonology and phonological theory*, ed. Takeru Honma, Masao Okazaki, Toshiyuki Tabata, and Shin-ichi Tanaka, 66-90. Kaitakusha: Tokyo.

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

Ryan, Kevin (in press) Prosodic end-weight reflects phrasal stress. To appear in *NLLT*.

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. *Phonological Analysis.*

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.

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

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.