Phonetic Production Reflects Syntactic Probability: Evidence from Duration and Disfluency

Harry Tily, Neal Snider, Anubha Kothari,
Inbal Arnon and Joan Bresnan
Department of Linguistics
Stanford University

AMLaP, Turku
August 2007

Is linguistic knowledge probabilistic?

A growing body of evidence suggests that:

- speakers have knowledge about the probabilities of linguistic events
- those probabilities influence speakers' choices between constructions, and their phonetic realization

Resnik (1996); Jurafsky et al. (2001); Gahl and Garnsey (2004); Jaeger et al. (2005); Gahl et al. (2006); Pluymaekers et al. (2005); Bresnan (2006); Jaeger (2006); Jaeger et al. (2006); Levy and Jaeger (2006); Levy (2006); Bresnan et al. (2007); Wasow et al. (2007), etc

Probabilities influence phonetic realization

Words and syllables are **phonetically reduced** when:

- high frequency
  (Zipf, 1929; Bybee, 2000; Aylett and Turk, 2004)

- predictable given adjacent words and syllables
  (Gregory et al., 1999; Jurafsky et al., 2001; Bell et al., 2003; Aylett and Turk, 2004; Pluymaekers et al., 2005)

- repeated, or topical
  (Fowler and Housum, 1987; Aylett and Turk, 2004)

Probabilities influence fluency

Words are less likely to be disfluent when:

- they are part of a less complex NP
  (Clark and Wasow, 1998)

- they signify a previously mentioned referent
  (Arnold et al., 2003)

- they are more likely given the preceding words
  (Stolcke and Shriberg, 1996)

Different probability estimates for different effects

- Studies on construction choice have looked at “rich” probability measures
  (How likely is the construction given all/much of the information available to the speaker?)

- But most studies reporting phonetic effects look at **local** or **single-cue** probability measures
  (bigrams, frequency, etc)

- As do models of disfluency developed for speech recognition (Stolcke and Shriberg, 1996)

The current study

Phonetic realization and speech fluency reflect the probability of the linguistic structures being produced

Therefore:

- speakers use multiple, varied sources of information to estimate probabilities
- **syntactic** probabilities affect the **phonetic realization** and **fluency** of speech

The probability of a syntactic construction affects its phonetics

This is not a novel idea:

Gahl and Garnsey (2004): the probability of some argument structure given the verb **does** correlate with phonetic effects

Verb Bias: $P(Construction|Verb)$

The probability of a syntactic construction affects its phonetics

This is not a novel idea:

Gahl and Garnsey (2004): the probability of some argument structure given the verb **does** correlate with phonetic effects

Verb Bias: $P(Construction|Verb)$

Verb bias is an estimate of the probability of a construction — but not a very accurate one

The probability of a syntactic construction affects its phonetics

Do more accurate estimates of the probability of a construction that incorporate rich information sources correlate with phonetic effects?

The probability of a syntactic construction affects its phonetics

Do more accurate estimates of the probability of a construction that incorporate rich information sources correlate with phonetic effects?

We extend the Gahl and Garnsey finding:

- with a more accurate probability measure that incorporates semantic and contextual information
- with a novel construction
- using naturalistic data

A rich model of syntactic probability

Bresnan et al. (2007): a regression model predicts the construction choice in the dative alternation:

(1)a. Yeah, I haven't **given much thought to it**, I'm kind of busy raising my kids *(prepositional phrase)*

b. Yeah, I haven't **given it much thought**, I'm kind of busy raising my kids *(double object)*

A rich model of syntactic probability

Bresnan et al. (2007): a regression model predicts the construction choice in the dative alternation:

(1)a. Yeah, I haven't **given much thought to it**, I'm kind of busy raising my kids *(prepositional phrase)*

b. Yeah, I haven't **given it much thought**, I'm kind of busy raising my kids *(double object)*

$$Predictors = \left\{ \begin{array}{c} \text{verbal meaning, discourse accessibility,} \\ \text{relative argument length, structural parallelism,} \\ \text{definiteness, animacy, pronominality, ...} \end{array} \right\}$$

A rich model of syntactic probability

This model correctly predicts the choice on 94% of unseen data

Compare this to simple **verb bias** on a corpus of 2349 spoken dative sentences:

Model	Accuracy	Odds
Baseline (always choose DO)	79%	3.8
Verb bias	83%	4.9
Bresnan et al model	94%	15.7

Verb bias is 1.3 times improvement over the baseline, while the rich model is a 4.9 times improvement

A rich model of syntactic probability

The model outputs a number from 0.0 (meaning double object) to 1.0 (prepositional)

Model probability: $P(Construction = pp|Predictors)$

A rich model of syntactic probability

The model outputs a number from 0.0 (meaning double object) to 1.0 (prepositional)

Model probability: $P(Construction = pp|Predictors)$

Bresnan (2006):

- speakers' judgements agree with the model
- speakers are uncertain in the same cases that the model is uncertain

A rich model of syntactic probability

The model outputs a number from 0.0 (meaning double object) to 1.0 (prepositional)

Model probability: $P(Construction = pp|Predictors)$

Bresnan (2006):

- speakers' judgements agree with the model
- speakers are uncertain in the same cases that the model is uncertain

So we can use the output of the model as an estimate of speakers' knowledge of syntactic probability.

Study 1: Does syntactic probability affect word duration?

More predictable words tend to be phonetically reduced

(Gregory et al., 1999; Jurafsky et al., 2001; Bell et al., 2003)

Hypothesis:

- words within a construction will be **shorter** when it is assigned a **higher** probability

$$P(Cxt|predictors)$$

Study 1: Words of interest

Duration of an initial *"to"* in the PP outcome, and *"the"* in the NP outcome:

(2)a. ... they gave all that money **to** the people ...

b. ... if you're going to pay teachers **the** salary they're paid...

These words were chosen because

- reduction effects have been found on these words (Bell et al., 2001, 2003)
- they can be kept constant across all cases
- they are very rarely prosodically marked

Study 1: Data

The **Spoken Dative Database** (Recchia, 2006) : a heavily annotated dataset of 2349 spontaneous, spoken dative constructions extracted from Switchboard

We remove:

- utterances without time alignments (1.7%)
- outliers three s.d. from mean speech rate (5.5%)
- disfluent utterances (3%)

Study 1: Data

The **Spoken Dative Database** (Recchia, 2006) : a heavily annotated dataset of 2349 spontaneous, spoken dative constructions extracted from Switchboard

We remove:

- utterances without time alignments (1.7%)
- outliers three s.d. from mean speech rate (5.5%)
- disfluent utterances (3%)

total dataset:	2114
PPs (which all begin with *to*):	488
NPs beginning with *the*:	260

Study 1: Linear regression model

- **Dependent variable:** duration of critical word
- **Independent variable:** syntactic probability
- **Controls:**
  - **verb bias**
  - **forward** and **backwards bigram probabilities**
  - **speech rate** (syllables per second within the fluent region that contains the critical word)
  - **phonological context** (whether the previous and following segments are consonants or vowels, for the four levels *C_C, C_V, V_C, V_V*)

Study 1: Results

We model PP and DO outcomes separately, because they show prosodic differences

Study 1: Results

We model PP and DO outcomes separately, because they show prosodic differences

Syntactic probability is a significant predictor of the duration of *to* in the PP ($p < .05$):

- as $P(pp|predictors)$ increases, duration **decreases**

Study 1: Results

We model PP and DO outcomes separately, because they show prosodic differences

Syntactic probability is a significant predictor of the duration of *to* in the PP ($p < .05$):

- as $P(pp|predictors)$ increases, duration **decreases**

Syntactic probability is a marginally significant predictor of the duration of *the* in the NP ($p = .07$):

- as $P(pp|predictors)$ increases, duration **increases**

Study 1: Results

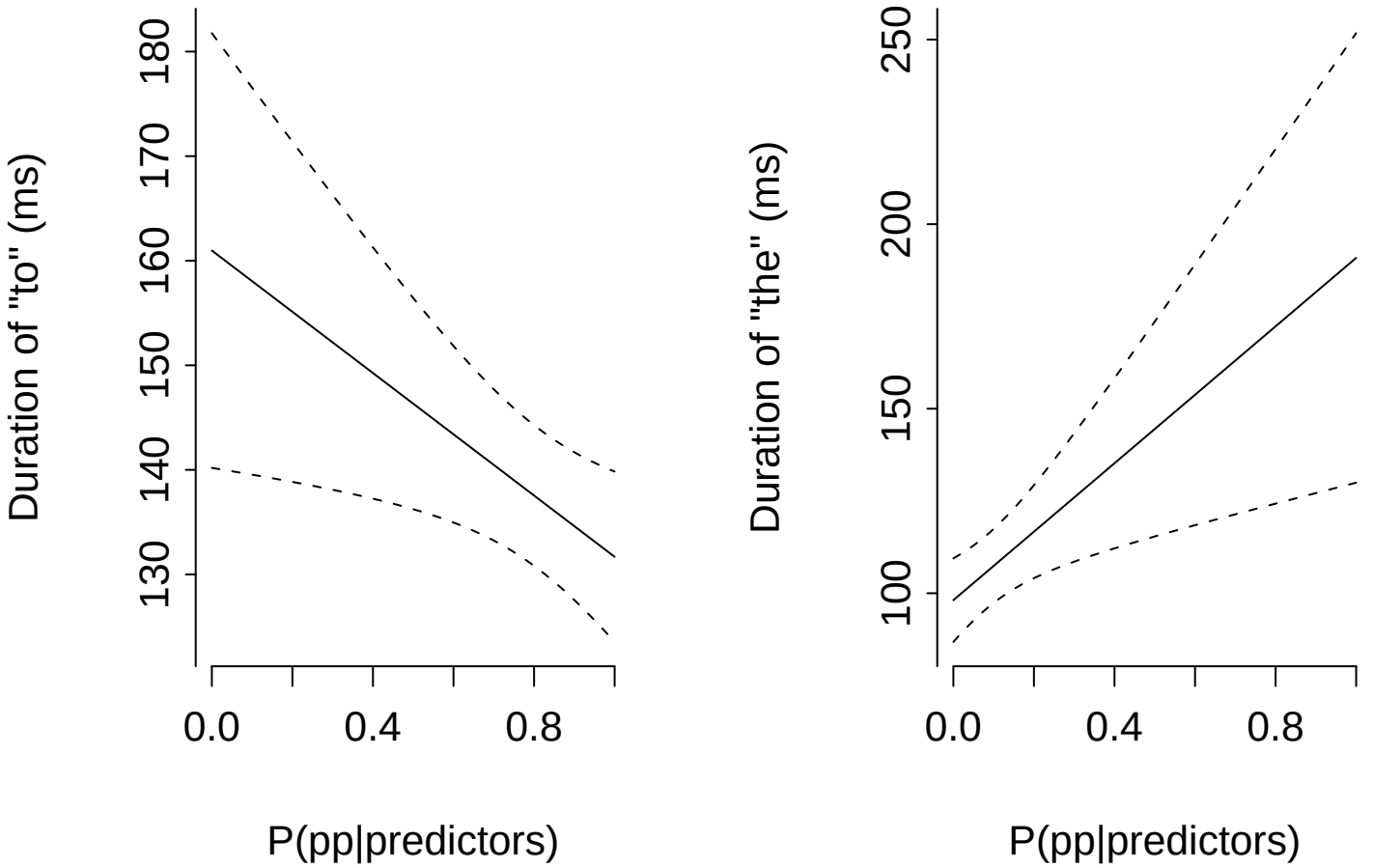

Figure 1: Model 1 predictions (fitted word durations)

Study 1: Results

- For both data sets, there are significant effects of bigram probability
- but no effect of verb bias

Study 1: Results

- For both data sets, there are significant effects of bigram probability
- but no effect of verb bias

There is no significant collinearity between any predictors:

- all vifs < 1.2 (variance inflation factors)

Study 1: Summary

- syntactic probabilities estimated from rich information affect articulation
- word durations are shorter in more predictable constructions
- this effect is significant beyond previously reported probability estimates
  - verb bias does not predict duration in this dataset

What is the relationship between fluency and probability?

- Words that are more likely given previous words are less likely to be disfluent (Stolcke and Shriberg, 1996)

What is the relationship between fluency and probability?

- Words that are more likely given previous words are less likely to be disfluent (Stolcke and Shriberg, 1996)

- Referents given in context are more probable
  - and given NPs are less disfluent (Arnold et al., 2003)

What is the relationship between fluency and probability?

- Words that are more likely given previous words are less likely to be disfluent (Stolcke and Shriberg, 1996)
- Referents given in context are more probable
  - and given NPs are less disfluent (Arnold et al., 2003)
- More complex phrases are less probable
  - and more complex phrases are more often disfluent (Clark and Wasow, 1998)

What is the relationship between fluency and probability?

- Words that are more likely given previous words are less likely to be disfluent (Stolcke and Shriberg, 1996)
- Referents given in context are more probable
  - and given NPs are less disfluent (Arnold et al., 2003)
- More complex phrases are less probable
  - and more complex phrases are more often disfluent (Clark and Wasow, 1998)

So several probability estimates have been (indirectly) shown to correlate with fluency

Study 2: Are less probable constructions less fluent?

Hypothesis:

- there will be **less disfluency** in a construction that is assigned a **higher** probability

$$P(Cxt|predictors)$$

Study 2: Operationalizing disfluency

- binary outcome: **fluent** or **disfluent**
- we code disfluency adjacent to the first word of the second argument (as in Study 1)
- disfluencies can be:
  - a pause of 500ms or more
  - a filled pause *(“uh”, “um”)*
  - repetition of a word
  - a “stumble” or restart
    (“give them ano- another trial”)

Study 2: Logistic regression model

- **Dependent variable:** disfluency at the critical word
- **Independent variable:** syntactic probability
- **Controls:**
  - **speech rate at that point** (this time, including the duration of the critical word)
  - **the length of the second argument in words** (to control for planning effects associated with the complexity of the phrase)

Again, we model PP and DO outcomes differently

Study 2: Results

- within the PP dataset, syntactic probability is a significant predictor of disfluency at the start of the second argument ($p < .01$)
  - as $P(pp|predictors)$ goes up, disfluency is less likely

Study 2: Results

- within the PP dataset, syntactic probability is a significant predictor of disfluency at the start of the second argument ($p < .01$)
  - as $P(pp|predictors)$ goes up, disfluency is less likely
- within the DO dataset, syntactic probability is a significant predictor of disfluency at the start of the second argument ($p < .05$)
  - as $P(pp|predictors)$ goes up, disfluency is more likely

Study 2: Results

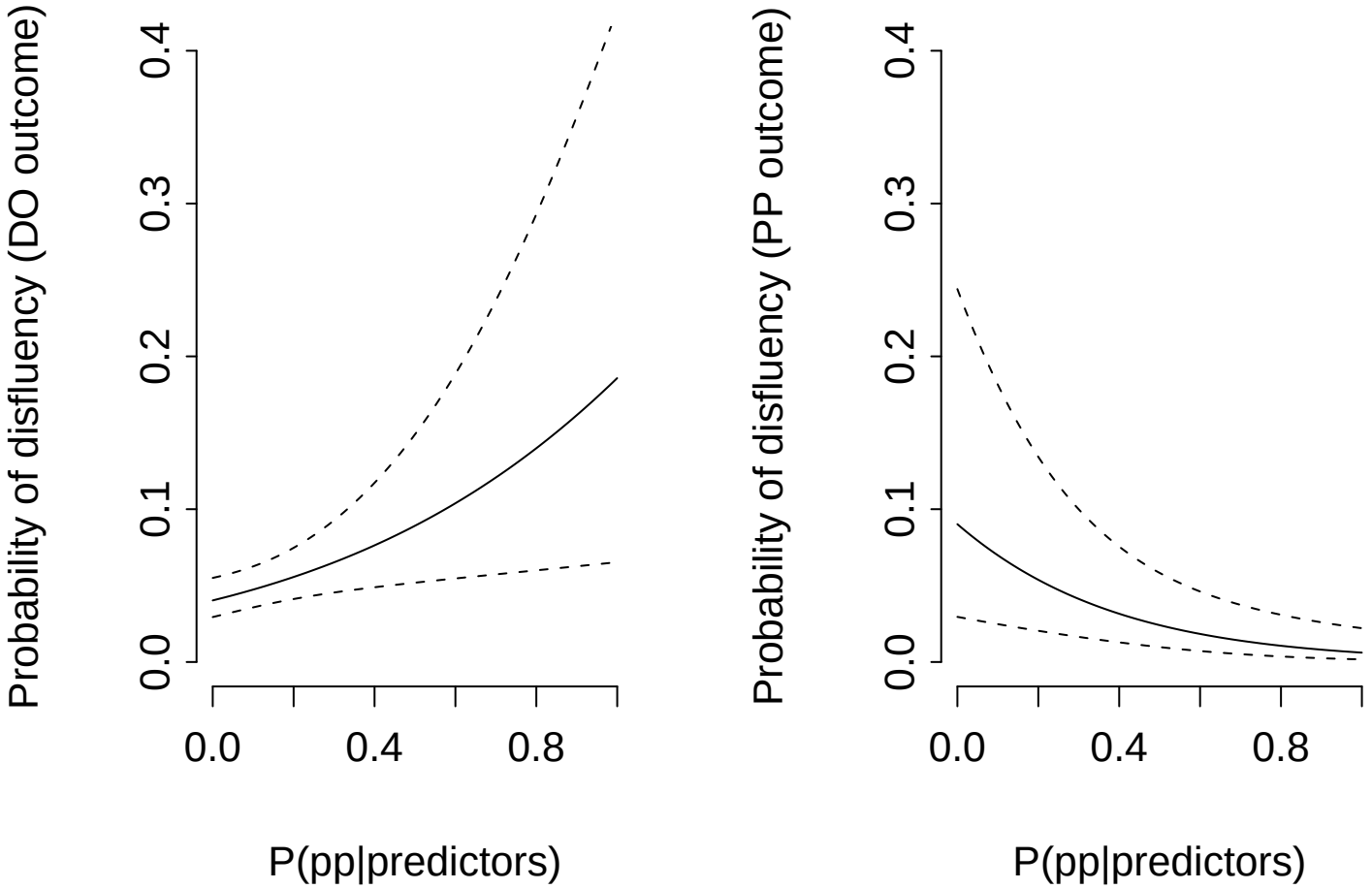

Figure 2: Estimated disfluency probabilities

Discussion

Previous work has shown that:

- speakers' choice between constructions is conditioned on many sources of semantic and contextual information
- temporal reduction is correlated with probability
- disfluency is correlated with probability

Discussion

Previous work has shown that:

- speakers' choice between constructions is conditioned on many sources of semantic and contextual information
- temporal reduction is correlated with probability
- disfluency is correlated with probability

We combine these findings:

- the probability of a construction conditioned on **semantic and contextual information** correlates with **temporal reduction** and **fluency**

Discussion

This strengthens the findings that:

- speakers estimate the probability of a construction based on rich information
  (Bresnan, 2006; Bresnan et al., 2007)

- syntactic probability affects phonetic realization
  (Gahl and Garnsey, 2004)

Implications for models of speech production

- The probability (or activation level) of a syntactic constuction and/or the information on which it is conditioned is available during the time-course of sentence production
  - and this influences articulatory planning

Further work

- Are the effects of construction probability in production mirrored by similar effects in comprehension?
  - Further work will test this possibility
- We are working on a controlled lab experiment to confirm these findings
  - Participants will produce more or less probable structures (à la Gahl and Garnsey (2004))

The End

Thanks for listening!

Thanks to Arto Antilla, Susanne Gahl, Florian Jaeger, Dan Jurafsky and Tom Wasow

This research was supported in part by NSF Award no. IS-0624345

References

Arnold, J. E., Fagnano, M., and Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, 32:25–36.

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and Speech*, 47:31–56.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2001). Form variation of english function words in conversation. ms University of Colorado.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in english conversation. *Journal of the Acoustical Society of America*, 113:1001–1024.

Bresnan, J. (2006). Is syntactic knowledge probabilistic? experiments with the english dative alternation. In Featherston, S. and Sternefeld, W., editors, *Roots: Linguistics in search of its evidential base*, Studies in Generative Grammar. Mouton de Gruyter, Berlin.

Bresnan, J., Cueni, A., Nikitina, T., and Baayen, R. H. (2007). Predicting the dative alternation. In Boume, G., Kraemer, I., and Zwarts, J., editors, *Cognitive Foundations of Interpretation*, pages 69–94. Royal Netherlands Academy of Science, Amsterdam.

Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In Barlow, M. and Kemmer, S., editors, *Usage-based Models of Language*, pages 65–85. CSLI Press, Stanford.

Clark, H. H. and Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37:201–242.

Fowler, C. and Housum, J. (1987). Talkers' signaling of new and old. words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26:489–504.

Gahl, S. and Garnsey, S. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunctiation variation. *Language*, 80:748–775.

Gahl, S., Garnsey, S. M., Fisher, C., and Matzen, L. (2006). Syntactic probabilities affect pronunciation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in

lexical production. In *CLS-99*, Chicago. University of Chicago.

Jaeger, T. F. (2006). *Probabilistic Syntactic Production: Expectedness and Syntactic Reduction in Spontaneous Speech*. PhD thesis, Stanford University.

Jaeger, T. F., Levy, R., Wasow, T., and Orr, D. M. (2005). The absence of "that" is predictable if a relative clause is predictable. Talk presented at AMLaP, September 2005.

Jaeger, T. F., Snider, N., Staum, L. W., and Jurafsky, D. (2006). (in)dependence of lexical and syntactic production: that-reduction and omission in spontaneous speech. Poster presented at 19th CUNY Human Sentence Processing conference, New York.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, *Frequency and the emergence of linguistic structure*, pages 229–254. John Benjamins, Amsterdam.

Levy, R. (2006). Expectation-based syntactic comprehension. To appear in Cognition.

Levy, R. and Jaeger, T. F. (2006). Speakers optimize information density through syntactic reduction. In *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62:146–159.

Recchia, G. (2006). The spoken syntax lab. Presented at the Stanford Linguistics Department Colloquium, October 20, 2006.

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, 61:127–159.

Stolcke, A. and Shriberg, E. (1996). Statistical language modeling for speech disfluencies. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 405–408, Atlanta.

Wasow, T., Jaeger, T. F., and Orr, D. (2007). Lexical variation in relativizer frequency. In *Workshop on Expecting the unexpected: Exceptions in Grammar, 27th Annual Meeting of the German Linguistic Association*, Germany. University of Cologne.

Zipf, G. K. (1929). *Relative frequency as a determinant of phonetic change*, volume Volume XL of *Harvard Studies in Classical Philology*. Harvard, Cambridge, MA.