Morphophonetics and Naïve discriminative learning

Commentary on Fabian Tomaschek et al. (2021). "Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning." In: *Journal of Linguistics* 57.1, pp. 123–161

Olivier Bonami
olivier.bonami@u-paris.fr

Université Paris Cité — Lexical Matters — November 2023

Morphophonetics

- Mainstream linguistic theories, and models of speech production, assume a modular design with no direct relationship between morphology and phonetics:

  MORPHOLOGY ⟷ PHONOLOGY ⟷ PHONETICS

- Yet various recent studies document subphonemic effects of morphology and/or the lexicon:
  - Free and bound stems differ acoustically (Kemps et al. 2005).
  - The duration of a suffix is influenced by its contextual and paradigmatic probability (Cohen 2014).
  - The duration of an affix is influenced by its segmentability, i.e., how salient the stem-affix boundary is (Hay 2007).
  - 'Homophonous' affixes are found to have measurably different realizations.
  - In particular, Plag, Homann, and Kunter (2017) document differences in duration of word-final [s] or [z] depending on whether it is morphemic, and, if morphemic, on the identity of the suffix (nominal plural, verbal PRS.3PL, genitive, GEN.PL, reduced *has*, reduced *his*)

The present study

- Two goals:
    1. Large scale replication of the Plag, Homann, and Kunter (2017) study
    2. Attempt to understand why the differences in duration are the way they are.
        - They do this using measures of predictability derived from a network trained using discriminative learning principles.

Replication study I

- Buckeye corpus (Pitt et al. 2007): 300 000 words of conversational speech by 40 speakers from Colomus, Ohio. The corpus is fully transcribed with automatic but hand-corrected alignment of words and phones.
- 28,928 tokens of word-final /s/ or /z/ in the corpus.

	Voiced	Unvoiced
s	1470	10141
3rdSg	832	2846
GEN	42	180
Has/is	622	5133
PL-GEN	0	12
Plural	1367	6095

Replication study II

- Linear mixed-effect model predicting log duration from:
  - ExponentFor: Morphological type of s (reference level: nonmorphemic)
  - Voicing
  - Cluster: number of consonants in the coda, including the S
  - MannerFollowing: manner of articulation of the next segment (reference level: non next segment)
  - LocalSpeechRate: syllables/second in a 20 second window
  - BaseDuration: duration of the rest of the word, with the S stripped.
  - Random intercepts for speaker and word.
- Importantly, frequency was not included, as it correlates strongly with base duration ($r = -0.69$).

Replication study III

- Results:

	Estimate	Std. error	df	t-Value
Intercept	−1.52	0.02	148.39	−69.93
ExponentFor = 3rdSg	−0.10	0.02	1372.72	−5.65
ExponentFor = GEN	−0.15	0.03	5647.45	−5.46
ExponentFor = has/is	−0.15	0.02	1416.32	−7.33
ExponentFor = PL-GEN	−0.12	0.11	5778.72	−1.08
ExponentFor = plural	−0.10	0.01	1380.73	−8.98
Voicing = unvoiced	0.23	0.01	28924.37	35.66
Cluster = 2	−0.19	0.01	5778.52	−26.03
Cluster = 3	−0.29	0.01	6103.94	−19.73
MannerFollowing = app	−0.31	0.01	28822.04	−37.63
MannerFollowing = fri	−0.52	0.01	28900.28	−71.39
MannerFollowing = nas	−0.47	0.01	28872.42	−31.94
MannerFollowing = plo	−0.51	0.01	28906.19	−72.46
MannerFollowing = vow	−0.43	0.01	28909.55	−62.94
LocalSpeechRate	−0.08	0.00	28837.16	−38.43
BaseDuration	0.19	0.01	16193.21	32.88

Replication study IV

- All predictors highly significant in the expected direction, except ExponentFor = PL-GEN.
- No interactions.
- In addition, significant contrasts in duration between pairs of exponents: nonmorphemic S is shortest, reduced auxiliaries are longest.

	PL	PRS.3G	GEN	AUX
S	×	×	×	×
PL				×
PRS.3SG				×
GEN				×

- This broadly replicates Plag, Homann, and Kunter's (2017) results, with some minute differences.

Naïve discriminative learning

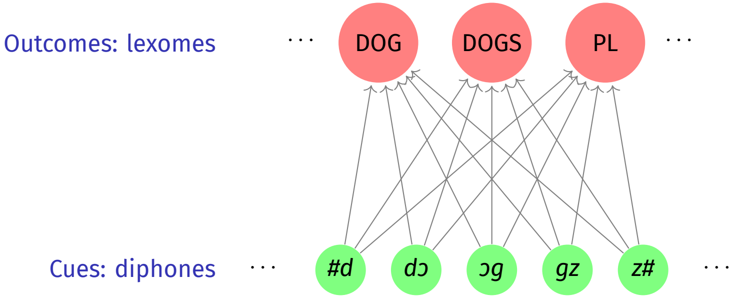
- Naïve discriminative learning is a direct implementation of the learning algorithm we discussed last week, based on the Rescorla-Wagner equations.

$$w_{ij}^{t+1} = w_{ij}^{t} + \begin{cases} 0 & \text{if} \quad \textsc{absent}(C_i, t) \\ \alpha \left(1 - \sum_{\textsc{present}(C_k,t)} w_{kj}\right) & \text{if} \quad \textsc{present}(C_i, t) \text{ and } \textsc{present}(O_j, t) \\ \alpha \left(0 - \sum_{\textsc{present}(C_k,t)} w_{kj}\right) & \text{if} \quad \textsc{present}(C_i, t) \text{ and } \textsc{absent}(O_j, t) \end{cases}$$

- It is ‘naïve’ by analogy to naive Bayes classifiers: the weights to outcomes are independent of one another.
- Although the implementation is generic, Baayen and colleagues have used this in a specific context:
  - Modeling phonological shapes as sets of *n*-phones (phoneme ngrams); in this study diphones are used.
  - Modeling content as “lexomes”. Lexomes are atoms representing the content of lexemes, words, and “morphological functions”
    - NB that in Linear Discriminative Learning, to be discussed in a later session, these are replaced by distributional vectors.

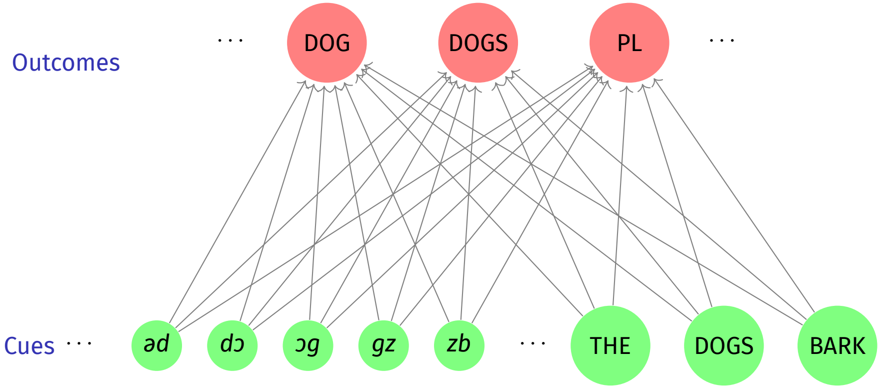
Cue to outcome structure I

- In most previous morphological work on NDL, the learning task was to learn word meanings from word forms:



Cue to outcome structure II

- Here (after testing various alternatives) they use a more elaborate learning task: learning from diphones and the collocational context coded as a set of lexomes.
- Training on the whole Buckeye corpus (286,982 tokens), with $\alpha = 0.001$, 5 word window.



Measures derived from the network I

- Activation: sum of all the weights from a given set of cues to a given outcome.
  - Sum of red weights in the example below.
  - This is akin to $P(\text{outcome} = o \mid \text{cues} = \{c_1, \ldots, c_n\})$
  - Tells us how well these cues discriminate this outcome.

	o_1	**{plural}** $_2$	$\ldots$	o_n
c_1	$w_{1,1}$	$w_{1,2}$	$\ldots$	$w_{1,n}$
c_2	$w_{2,1}$	$w_{2,2}$	$\ldots$	$w_{2,n}$
ld	$w_{3,1}$	$w_{3,2}$	$\ldots$	$w_{3,n}$
dO	$w_{4,1}$	$w_{4,2}$	$\ldots$	$w_{4,n}$
Og	$w_{5,1}$	$w_{5,2}$	$\ldots$	$w_{5,n}$
gz	$w_{6,1}$	$w_{6,2}$	$\ldots$	$w_{6,n}$
zb	$w_{7,1}$	$w_{7,2}$	$\ldots$	$w_{7,n}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
c_k	$w_{k,1}$	$w_{k,2}$	$\ldots$	$w_{k,n}$
	a_1	a_2	$\ldots$	a_n

Measures derived from the network II

- Prior: sum of the absolute values of the weights from all cues to a given outcome.
  - Sum of absolute values of light gray weights in the example below.
  - This is akin to P(outcome).
  - Tells us how much this outcome stands out among all outcomes.

	o_1	**{plural}** $_2$	$\dots$	o_n
c_1	$w_{1,1}$	$w_{1,2}$	$\dots$	$w_{1,n}$
c_2	$w_{2,1}$	$w_{2,2}$	$\dots$	$w_{2,n}$
ld	$w_{3,1}$	$w_{3,2}$	$\dots$	$w_{3,n}$
dO	$w_{4,1}$	$w_{4,2}$	$\dots$	$w_{4,n}$
Og	$w_{5,1}$	$w_{5,2}$	$\dots$	$w_{5,n}$
gz	$w_{6,1}$	$w_{6,2}$	$\dots$	$w_{6,n}$
zb	$w_{7,1}$	$w_{7,2}$	$\dots$	$w_{7,n}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
c_k	$w_{k,1}$	$w_{k,2}$	$\dots$	$w_{k,n}$
	a_1	a_2	$\dots$	a_n

Measures derived from the network II

- Activation diversity: sum of the absolute values of the weights from a given set of cues to all outcomes.
  - Sum of absolute values of $\boxed{\text{boxed weights}}$ in the example below.
  - This is akin to $H(\text{outcome} \mid \text{cues} = \{c_1, \ldots, c_n\})$.
  - Tells us how much these cues segregate outcomes overall.

	o_1	**{plural}** $_2$	…	o_n
c_1	$w_{1,1}$	$w_{1,2}$	…	$w_{1,n}$
c_2	$w_{2,1}$	$w_{2,2}$	…	$w_{2,n}$
ld	$w_{3,1}$	$w_{3,2}$	…	$w_{3,n}$
dO	$w_{4,1}$	$w_{4,2}$	…	$w_{4,n}$
Og	$w_{5,1}$	$w_{5,2}$	…	$w_{5,n}$
gz	$w_{6,1}$	$w_{6,2}$	…	$w_{6,n}$
zb	$w_{7,1}$	$w_{7,2}$	…	$w_{7,n}$
…	…	…	…	…
c_k	$w_{k,1}$	$w_{k,2}$	…	$w_{k,n}$
	a_1	a_2	…	a_n

Precise measures chosen for this study

1. PriorMorph: prior for the target lexome.
   - Because we have 9 lexomes, there are 9 discrete values to choose from.
2. ActFromBoundaryDiphone: activation of target lexome by final diphone of the word of interest.
   - 9 possible values for each boundary diphone.
3. ActFromRemainingCues: activation of target lexome by all other cues (diphones and lexomes) present in the 5 word window centered on the word of interest.
   - Very varied possible values
4. ActDivFromBoundaryDiphone: activation diversity of the boundary diphone.
   - 9 possible values for each boundary diphone.
5. ActDivFromRemainingCues.
   - Very varied possible values

The model I

- New model of basically the same data, but using NDL-derived measures instead of the nominal variable ExponentFor.
- This is a Generalized additive mixed model (Wood 2011), a class of models where the dependent variable is predicted from the linear combination of (unknown) smoothing functions applied to the predictor variable.
- Final model results from exploratory data analysis starting from the control variables and adding NDL-derived measures + interactions step by step.

The model II

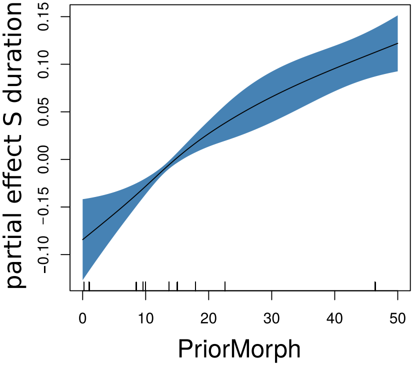
- Linear predictors in the final model:
  - As before: Manner of articulation of the segment Following S.
  - Manner of articulation of the segment Preceding S.
  - As before: Local speaking rate (20 second window).
  - Individual speaking rate of each speaker over the whole corpus.
- Smooth terms:
  - Interaction between ActFromBoundaryDiphone and ActDivFromBoundaryDiphone
  - Interaction between ActFromRemainingCues, ActDivFromRemainingCues, and LocalSpeakingRate.
  - PriorMorph
- Random intercepts for speaker and word.

Coefficients table

A. Parametric coefficients	Estimate	Std. error	t-Value	p-Value
Intercept	−2.9179	0.2294	−12.7173	<0.0001
Preceding = fricative	−0.0962	0.0299	−3.2151	0.0013
Preceding = nasal	−0.1335	0.0233	−5.7229	<0.0001
Preceding = plosive	−0.1869	0.0150	−12.4229	<0.0001
Preceding = vowel	0.0106	0.0144	0.7318	0.4643
Following = approximant	0.2839	0.1470	1.9315	0.0534
Following = fricative	0.1036	0.1470	0.7048	0.4809
Following = nasal	0.1089	0.1474	0.7390	0.4599
Following = plosive	0.0850	0.1469	0.5785	0.5629
Following = vowel	0.1310	0.1469	0.8919	0.3725
LocalSpeakingRate	−0.0463	0.0211	−2.1874	0.0287
IndividualSpeakingRate	2.3873	0.6633	3.5990	0.0003
B. Smooth terms	**edf**	**Ref.df**	**F-value**	**p-Value**
te(ActFromBoundaryDiphone, ActDivFromBoundaryDiphone)	14.4458	16.9557	548.4375	<0.0001
te(ActFromRemainingCues, ActDivFromRemainingCues, LocalSpeakingRate)	24.7081	32.1035	170.9787	<0.0001
s(PriorMorph)	2.0235	2.3027	84.2267	<0.0001
Random intercepts speaker	37.1278	38.0000	2118.9174	<0.0001
Random intercepts word	458.5028	2280.0000	2190.5616	<0.0001

Relevant partial effects I

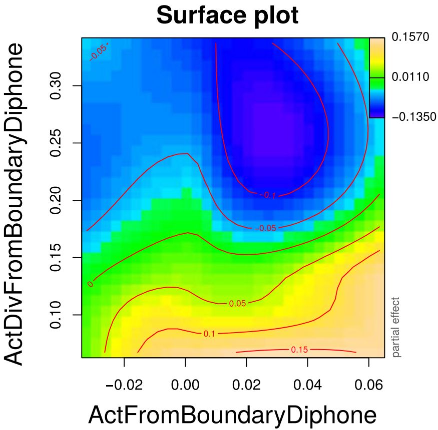
- Larger prior (i.e. overall salience of the lexome) lead to longer durations.



- Comparison with a model where the nominal variable ExponentFor from the previous study replaces Prior: model fit decreases while number of parameters increases.
- Hence the numerical variable Prior leads to better precision than the nominal variable.

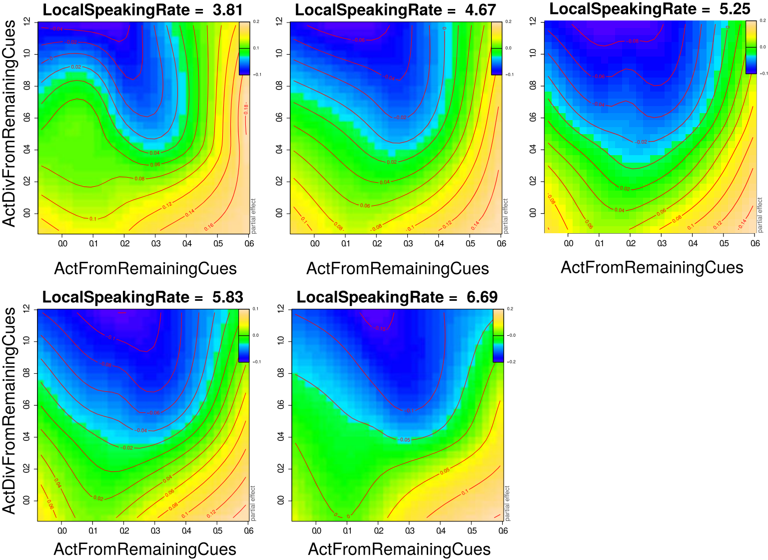
Relevant partial effects II

- Overall, larger activation leads to longer durations
- Overall, larger activation diversity leads to shorter durations
- Shortest durations are found for larger values of activation and largest values of activation diversity.
- Longest durations are found when lowest values of activation diversity combine with not too low values of activation.



Relevant partial effects III

- Similar looking effects of activation and activation diversity of remaining cues, but they are modulated by local speaking rate.



Discussion: pressures on duration

- Present results and previous literature suggest that opposing forces weigh on duration of S:
  - Enhance parts of the signal that support a meaning that is generally salient (prior).
  - Enhance parts of the signal that strongly support the intended meaning (activation).
  - Downplay parts of the signal that increase uncertainty (high activation diversity).
- The NDL-based model highlights the complex interaction between these forces.

Discussion: morphological theory

- The authors suggest that the results are more readily compatible with Word-and-Paradigm approaches to morphology than with Item and Arrangement approaches.
- The intuition seems to be that IA is inherently dependent on the postulation of discrete subword units, and hence cannot easily capture the patterns seen here that rely on a representation of form that ignores traditional morph boundaries.
- The authors concede that an IA approach is compatible with assigning probabilistic properties to morphemes and arrangements of morphemes, and hence could possibly capture the effects discussed here: they are just skeptical that this will lead to good results.

Discussion: Speech production

- The results clearly falsify modular models of speech production where the signal derives from a discrete phonological representation only (Dell 1986, Levelt et al. 1999).
- The results do not readily combine with the received view that less informative segments tend to be shorter (e.g. Jurafsky et al. 2001, Aylett and Turk 2004, Jaeger 2010).
    - Isn’t that a separate issue? The present model does not look at the specific support of the previous context for the use of a word.
- On the other hand, the results dovetail with the Paradigmatic Signal Enhancement Hypothesis (Kuperman et al. 2007): the more probable an exponent within a paradigm, the longer the articulation.

Evaluation

- Morphophonetic effects are beginning to make sense:
  - It is possible to read Plag, Homann, and Kunter (2017) as giving an argument the psychological reality of morphological segmentation: morphemes have individual phonetic properties.
  - Here we have a completely different picture: the data actually supports more a nondecompositional and fully gradient view of morphological knowledge.
- Interesting hypothesis on enhancement of discriminative signal.
- NDL as a practical, relatively tractable alternative to the use of either deep neural networks or explicit probabilistic modelling to capture the relation between form and meaning.
- The study raises at least as many questions as it answers:
  - Relationship between Prior, frequency, and duration?
  - Exact outcome structure and coding?
    - e.g. why {DOG DOGS PLURAL} rather than just {DOG PLURAL}?
  - Effect of cue structure and coding?
    - e.g. why diphones rather than triphones?
  - More generally, this is innovative in so many dimensions at once that it is hard to tell which are the usegul innovations.

References I

Plag, Ingo, Julia Homann, and Gero Kunter (2017). “Homophony and Morphology: The acoustics of word-final -S in English.” In: *Journal of Linguistics* 53, pp. 181–216 (cit. on pp. 2, 3, 7, 24).

Tomaschek, Fabian et al. (2021). “Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning.” In: *Journal of Linguistics* 57.1, pp. 123–161 (cit. on p. 1).