HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

DFG Research Unit FOR 2373
Spoken Morphology: Phonetics and phonology of complex words

How morphological structure affects phonetic encoding

Modeling the duration of word-final S using Naive Discriminative Learning

Ingo Plag

presenting joint work with

Fabian Tomaschek*, Harald Baayen* & Mirjam Ernestus[++]

Universität Tübingen*, Radboud Universiteit Nijmegen[++]

Plenary talk, International Symposium of Morphology (ISMo) December, 13-15, 2017

Morpho-phonology as we know it from Ling 101

- The only formal level of representation of morphemes is phonological in nature.

Bauer, Lieber & Plag (2013) *The Oxford Reference Guide to English Morphology.* Oxford: OUP

"the allomorphs are /s/, /z/, and /ɪz/, where /ɪz/ occurs after sibilants, /s/ occurs after other voiceless consonants, and /z/ occurs elsewhere ... This allomorphy is easily understood in phonological terms (assimilation and epenthesis to break up illegal geminates), and is not controversial" (p. 15)

- Post-lexical phonology and phonetics have no access to lexical information.

Stratal OT (Bermúdez-Otero 2017: 9):

"stem-level, word-level, and phrase-level phonological constraints [in Stratal OT, IP] as defined here correspond roughly to the cyclic, postcyclic, and postlexical phonological rules of Booij & Rubach (1987)"

Lexical vs. post-lexical phonology

lexical rules

- Cyclic
- Have lexical exceptions
- Structure-preserving (output is a possible underlying representation)
- Not necessarily phonetically natural
- Never apply across words
- Apply only in derived environments
  (Trisyllabic shortening)

post-lexical rules

- Non-cyclic
- No lexical exceptions
- Not necessarily structure-preserving
- May apply across words
- May not refer to word-internal morphological information
  (Flapping in Am. English)

Speech production (Levelt et al. 1999)

Concept

Lemma ‘cap‘ Noun [+ concrete] [+count] ‘more than one‘ [plural]

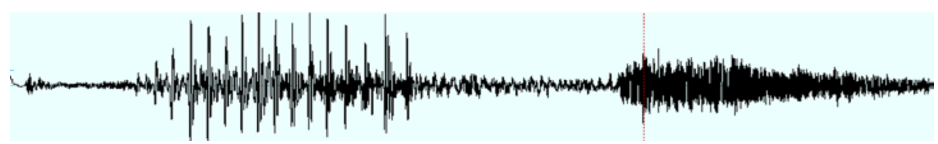
Phonological representation /kæp/ /-z/

(Morpho-)phonological rules [k^{h}æps]

Syllabary [$ kæps $]

Articulation <movement of articulators>

Acoustic signal



Crucial point

No morphological information available post-lexically

Problems

Role of phonetic detail

- Free and bound variants of a base differ acoustically (Kemps et al. 2005, Blazej & Cohen-Goldberg 2015)
- Duration of Dutch compound linking morphemes depends on paradigmatic probability (Kuperman et al. 2007)
- Vowel frontness of Russian verbal suffix depends on paradigmatic probability (Cohen 2014a)
- Duration of 3sg S depends on syntagmatic probability (Cohen 2004b)

Challenges

- for models that are strictly categorical in nature
- for models that build on the strict separation of lexical and post-lexical phonology

Research questions

- How does paradigmatic and syntagmatic morphological structure affect the articulatory, acoustic and phonological properties of complex words?

- What are the implications for the organization of the mental lexicon and for models of morpho-phonology, of lexical processing, of speech production and speech perception?

English final S

Traditional assumptions

- homophony of
  - plural
  - 3sg
  - genitive
  - genitive plural
  - clitics of *has, is, us*

⇨ no difference between different /s/ morphemes

- morphemic and non-morphemic sounds are the same in speech production

⇨ no difference between morphemic and non-morphemic /s/

Two previous studies

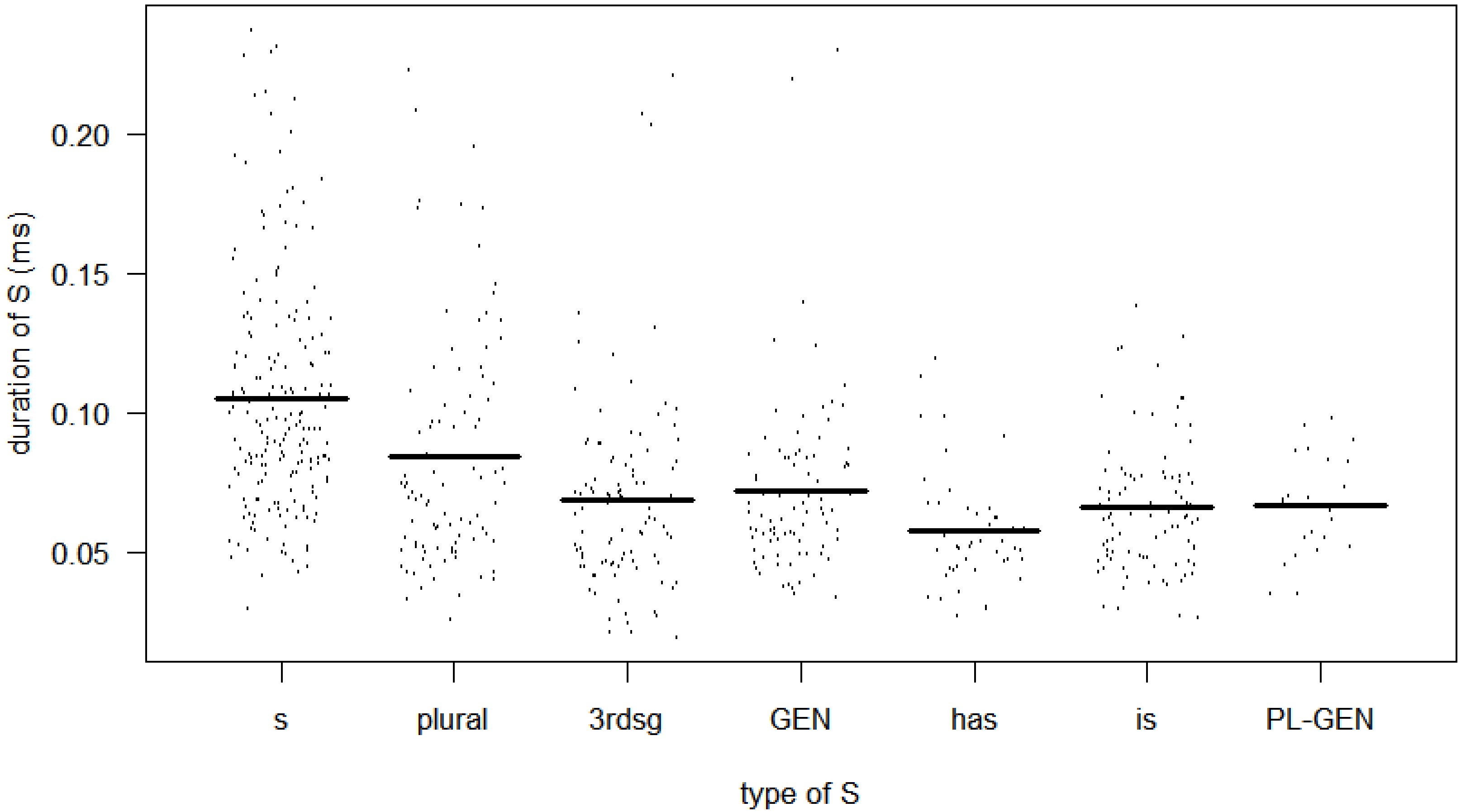
- final /z/ and /s/ (henceforth ‘S‘)
- non-morphemic, plural, 3sg, genitive, genitive plural, clitics of *has, is*
- Study 1: Buckeye Corpus, manual annotation, acoustic analysis, N=644, up to 100 per category, conversations, North American English (Plag et al. 2017)
- Study 2: Quakebox Corpus, N=7081, automatic segmentation, up to 245 per type, conversations, New Zealand English (Zimmermann 2016)
- Statistical analysis: duration by morpheme type, LMER, beta regression, absolute and relative duration, many noise variables

Covariates (Plag et al. 2017)

Table 2: Summary of the dependent variables and covariates used in the initial models.

Dependent variables	N	Mean	St. Dev.	Min	Max
Absolute duration of S: DURATIONOFS	644	0.081	0.039	0.019	0.237
Relative duration of S: PROPORTIONS	644	0.206	0.082	0.055	0.688
Numerical predictors	**N**	**Mean**	**St. Dev.**	**Min**	**Max**
Local speech rate: SYLSEC	644	5.601	1.202	1.984	10.179
Base duration: BASEDURATION	644	0.329	0.134	0.029	1.052
Base frequency: LOGBASEFREQ	644	8.672	2.399	0.000	14.146
Previous mention: BASEREP	644	0.317	0.772	0	6
Bigram frequency: LOGRBIGRAM	548	2.542	2.739	0.000	9.884
Neighborhood density: PND	601	14.21797	14.8551	0	60
Categorical predictors:	**N**	**Levels**			
No. of cons. before S: CONSONANTS	644	0: 325	1: 259	2: 58	3: 2
Voicing: ISVOICED	644	yes: 81	no: 563		
Following context: FOLLCONTEXT	644	pause: 97	V: 170	APP: 68	N: 33
		AFF: 10	F: 143	P: 123	
Syntactic position: BOUNDARY	644	yes: 226	no: 418		
Explanatory variable	**N**	**Levels**			
Type of S: TYPEOFS	644	S: 196	PL: 95	3rdsg: 100	GEN: 88
		has: 47	is: 95	PL-GEN: 23	

The raw data (Plag et al 2017)



Unvoiced tokens, absolute duration

	S	PL	3SG	GEN	PL-G	is	has
S	///	**	*	***	**	***	***
PL		///				*	*
3SG			///			*	*
GEN				///			
PL-G					///		
is						///	
has							///

American English

(Plag et al 2017)

	S	PL	3SG	GEN	PL-G	is	has
S	///	***	***	***	***	***	***
PL		///		**	.	***	***
3SG			///		*	***	***
GEN				///		*	*
PL-G					///		
is						///	
has							///

New Zealand English

(Zimmermann 2016)

Summary S

- Results for both varieties are very similar
- We find very many and robust differences between different types of S (largest difference 38 ms)
- We find differences in absolute and relative duration
- We find different differences for voiced and for unvoiced S
- Unvoiced realizations
  - Non-morphemic S is longer than all morphemic S‘s
  - Duration hierarchy:

Non-morphemic S > suffix S > clitic S

Discussion

- Traditional analyses of English S morphemes do not cover or predict the acoustic differences found.
- The acoustic differences cannot be accounted for by purely phonetic processes (covariates are controlled for).
- Conclusion: Phonetic detail reflects morphological structure.
- **How?**

Explanation 1: Morpho-phonetics

- Morphological boundary strength directly translates into phonetic strength, even if negatively:

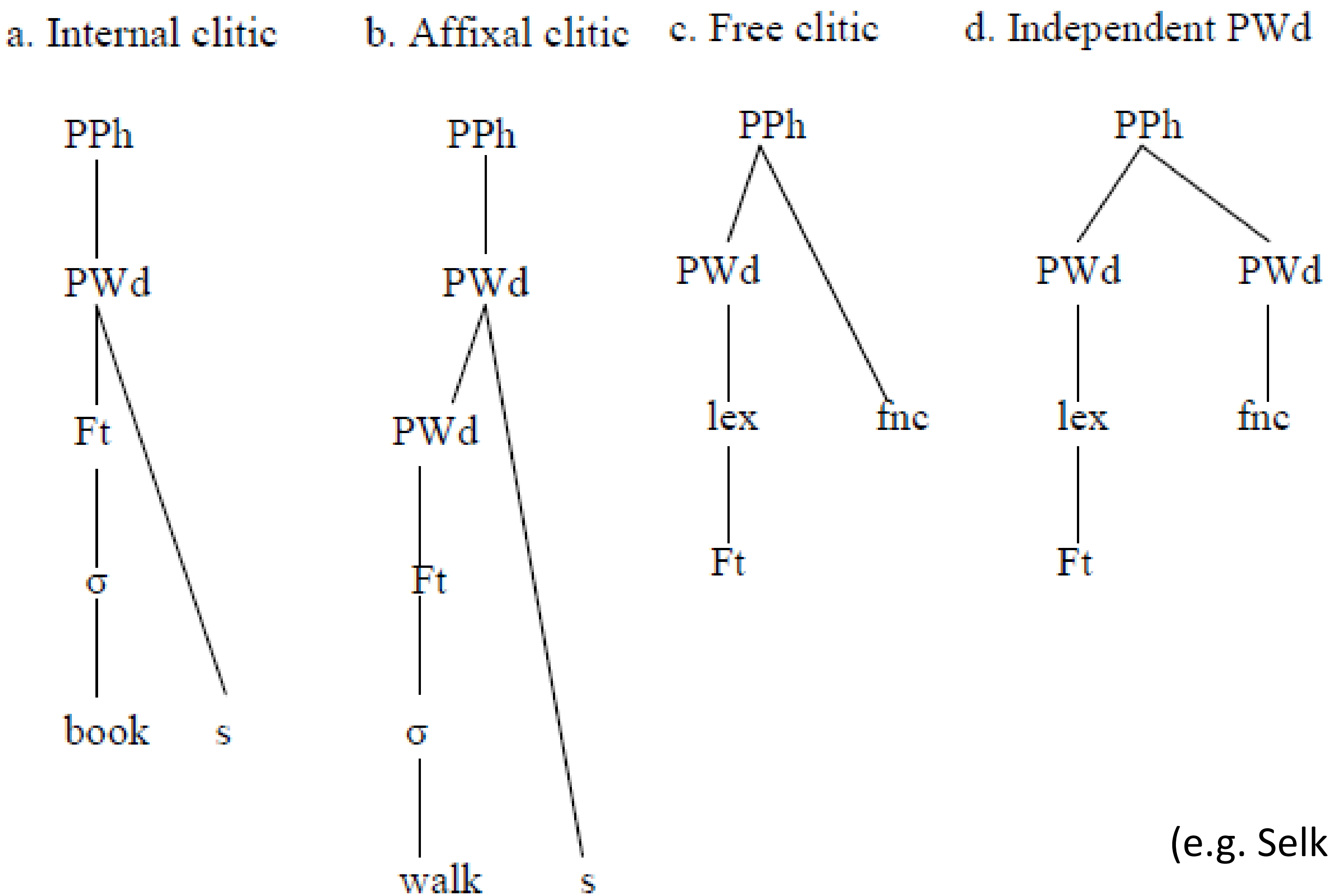
  No boundary > suffix boundary > clitic boundary

- Phonetic information is lexically represented

- *Pro exemplar-based models*
  Differential behavior w.r.t. voicing and duration
  Different distributions of properties across morphemes

- *Contra purely exemplar-based models*
  Effects of covariates

Explanation 2: Prosody



(e.g. Selkirk 1997)

Prosody: Problems

- Independent evidence for the proposed structures is weak
- Plural and 3rdsg do not differ in acoustic duration
- Interaction with voicing
- Negative correlation between boundary strength and duration

Interim summary

- The acoustic differences cannot be accounted for by purely phonetic processes (covariates are controlled for).
- Extant theories cannot explain the durational patterns
- How can phonetic detail reflect morphological structure?
- We hypothesize that we are seeing effects of linguistic experience.
- **Problem**
  The usual measures of experience cannot account for the S puzzle (e.g. lexical frequency, transitional phoneme probability, neighborhood density, bigram frequencies, etc.).

Problem: Which measures?

- Traditional measures do not take into account different linguistic levels (e.g. meaning and phonology) at the same time.
- Transitional phoneme probabilities may differ in complex ways by morphological function.
- Final /z/ after a vowel can be morphemic, a final /s/ cannot.
- Different phonotactics of verbs and nouns: 3sg S may have a different distribution of transitional probabilities than plural S.
- It is impossible to devise interpretable regression models with highly complex constellations of measures and specific properties of the target words (and their contexts).
- Other models are needed.

Naive Discriminative Learning

(Rescorla 1988 et seq.)

- Established learning theory, recently extended to language
  (Arnon & Ramscar 2012, Baayen et al. 2011, 2013, 2015, Baayen & Ramscar 2015, Blevins et al. 2015, Ramscar et al. 2010, 2013)

- Learning results from exposure to informative relations among events in the environment (co-occurrence of cues and outcomes)

- Association weights, adjusted according to new, informative experiences (‘Rescorla-Wagner equations‘)

- Association weights ≈ contextual and paradigmatic predictability

- General idea
  Association weights successfully predict durations of S in regression models

Rescorla-Wagner equations

cue	outcome
whiskers present	animal present
link to cat strengthened	cat
link to rabbit weakened	
link to rabbit strengthened	rabbit
link to cat weakened	

- Links are strengthend upon the **presence** of a particular cue and the **presence** of a particular outcome
- Links are weakened upon the **presence** of a particular cue and the **non-presence** of a particular outcome

A toy example: learning morphology

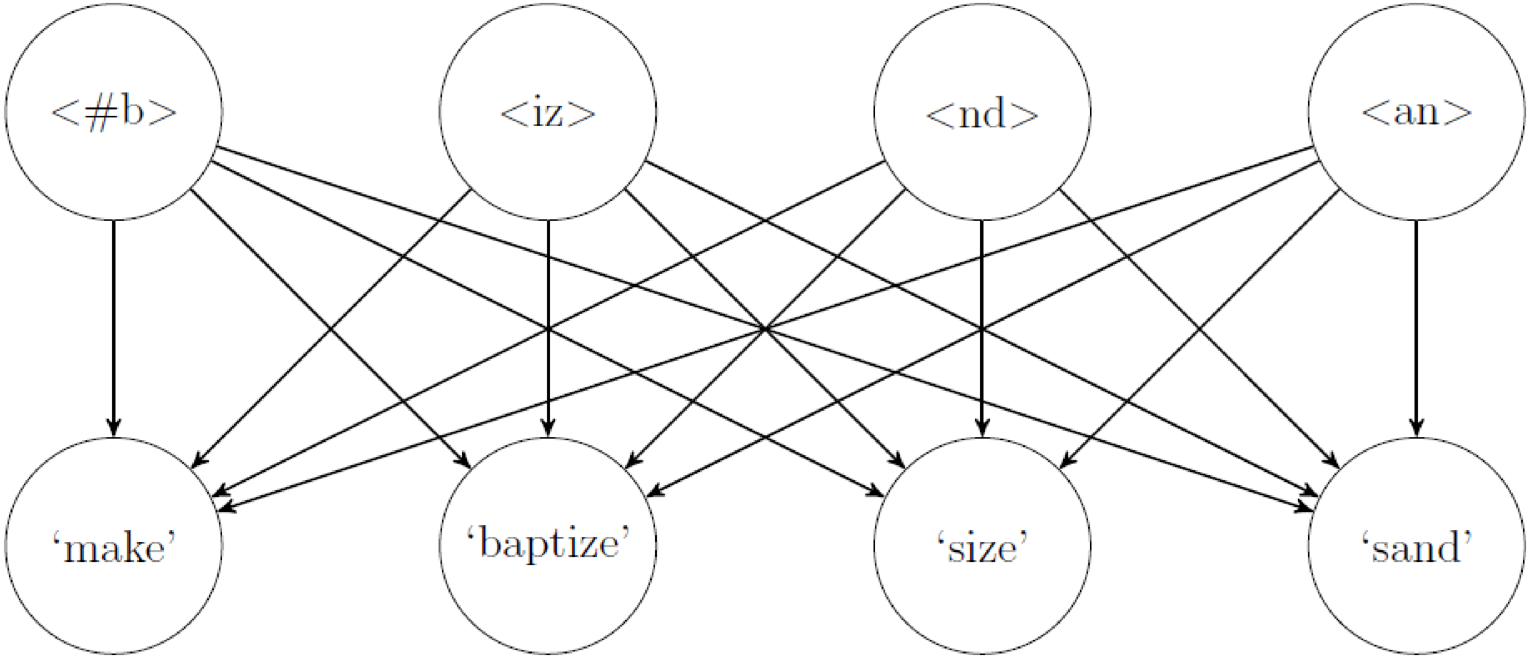
Plag & Balling (2017, with R code), Plag (2018)

Table 1: Toy lexicon (Frequencies are taken from the SUBTLEX-US corpus, Brysbaert & New 2009).

Words	Cues	Outcomes	Frequency
baptize	#b, ba, ap, pt, ti, iz, ze, e#	'baptize'	37
chance	#c, ch, ha, an, nc, ce, e#	'chance'	12303
extreme	#e, ex, xt, tr, re, em, me, e#	'extreme'	558
modernize	#m, mo, od, de, er, rn, ni, iz, ze, e#	'modern', 'make'	6
optimal	#o, op, pt, ti, im, ma, al, l#	'optimal'	15
optimize	#o, op, pt, ti, im, mi, iz, ze, e#	'optimal', 'make'	5
sand	#s, sa, an, nd, d#	'sand'	1035
size	#s, si, iz, ze, e#	'size'	2353

Cues and outcomes

40 cues, 8 outcomes, 320 connections, 320 association weights



Association weights: The weight matrix

Table 2: Weight matrix for four bigrams.

	'baptize'	'chance'	'extreme'	'make'	'modern'	'optimal'	'sand'	'size'
#b	0.213	-0.007	-0.004	-0.110	-0.016	-0.089	0.014	-0.062
iz	0.041	-0.032	-0.030	0.106	0.034	0.017	-0.019	0.128
nd	0.014	-0.030	0.006	0.036	0.012	0.006	0.220	-0.069
an	0.007	0.124	-0.009	0.017	0.006	0.003	0.190	-0.076

Table 3: Activation weights of bigrams for the meaning 'baptize'.

#b	ba	ap	pt	ti	iz	ze	e#	sum of weights
0.213	0.213	0.213	0.125	0.125	0.041	0.041	0.029	1

From association weights to lexical activation

Table 4: Activations of meanings by bigram strings.

	baptize	chance	make	modern	optimal
#b_ba_ap_pt_ti_iz_ze_e#	**1.00**	2.57e-16	0.00	9.71e-17	6.56e-16
#c_ch_ha_an_nc_ce_e#	-3.82e-17	**1.00**	1.18e-16	1.70e-16	3.56e-17
#m_mo_od_de_er_rn_ni_iz_ze_e#	4.09e-16	2.91e-16	**1.00**	**1.00**	-5.38e-16
#o_op_pt_ti_im_mi_iz_ze_e#	-3.82e-16	1.32-16	**1.00**	-4.30e-16	**1.00**

	‘baptize’	‘chance’	‘extreme’	‘make’	‘modern’	‘optimal’	‘sand’	‘size’
iz	0.041	-0.032	-0.03	0.106	0.034	0.017	-0.019	0.128
ze	0.041	-0.032	-0.03	0.106	0.034	0.017	-0.019	0.128
e#	0.029	0.106	0.085	0.076	0.024	0.012	-0.043	0.107
iz ze e#	0.111	0.042	0.025	0.288	0.092	0.046	-0.081	0.363

Back to S

- Predict the duration of S based on association weight measures derived from the weight matrix (association weights ≈ linguistic experience, i.e. ‘predictability‘)

- Make use of the whole Buckeye corpus (automatic annotation, N=28928)

Strategy

- Confirm Plag et al.‘s (2017) results with the whole corpus
- Then take out the predictor **TYPE OF S**, and add NDL weights as predictors, see what happens

Replicating Plag et al. (2017) with the full Buckeye corpus

	S	PL	3RDSG	GEN	HAS	IS	PL-GEN
S	n.a.	**	*	***	***	***	**
PL		n.a.			*	*	
3RDSG			n.a.		*	*	
GEN				n.a.			
HAS					n.a.		
IS						n.a.	
PL-GEN							n.a.

Table 1: Significant contrasts in duration between different types of unvoiced S in Plag et al. 2017. Significance codes: '***' p<0.001 '**' p<0.01, '*' p<0.05

	S	PL	3RDSG	GEN	HAS/ IS	PL-GEN
S	n.a.	***	***	***	***	
PL		n.a.			*	
3RDSG			n.a.		*	
GEN				n.a.		
HAS/IS					n.a.	
PL-GEN						n.a.

Table 2: Significant contrasts in the present study. Significance codes: '***' p<0.001 '**' p<0.01, '*' p<0.05

Replicating Plag et al. (2017): Summary

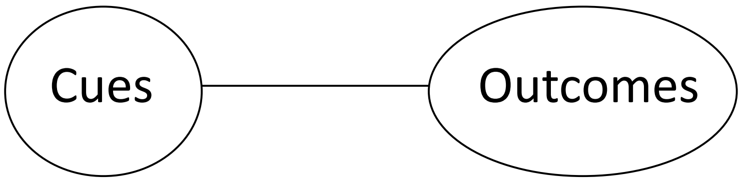
- Plag et al. 2017 can be replicated with the large data set
- Same predictors are significant
- Same types of effect of all predictors (morphological, lexical, phonetic). ☺

Method: NDL modeling

- Get the necessary information from Buckeye
- Rich cue-to-outcome structure (including semantics)
- Large integration windows (target word $\pm$ 2)
- Use the most influential and informative NDL-measures in Generalized Additive Mixed Effects Regression Models
- Most influential NDL-measures: random forest, other tests, and principled theoretical considerations

Cue-to-outcome structures illustrated

"... the small **dogs** bark at ..."



- word forms
- diphones (articulatory plans)

the small dogs bark at T@ @s sm m6 6l ld dO Og gz zb ba ar rk k@ @t	ld dO Og gz zb dogs DOG plural

- word forms
- lemmas
- morphologcical functions
- diphones (articulatory gestures)

Optimal cue-to-outcome structures

cues	***outcomes***
all cues	morphological function
all cues but last diphone	morphological function
only last diphone	morphological function

Measurements based on the weight matrix

- How well is a particular outcome (positively or negatively) connected in the network?

‘Prior’

= sum of absolute association weights of **ALL** cues in the network to a particular outcome

	$\mathcal{O}_{(1)}$	{plural} $_{(2)}$	…	$\mathcal{O}_{(n)}$	row no.
$\mathcal{C}_{(1)}$	$\mathcal{W}_{(1,1)}$	$\mathcal{W}_{(1,2)}$	…	$\mathcal{W}_{(1,n)}$	1
$\mathcal{C}_{(2)}$	$\mathcal{W}_{(1,1)}$	$\mathcal{W}_{(2,2)}$	…	$\mathcal{W}_{(2,n)}$	2
ld	$\mathcal{W}_{(3,1)}$	$\mathcal{W}_{(3,2)}$	…	$\mathcal{W}_{(3,n)}$	3
dO	$\mathcal{W}_{(4,1)}$	$\mathcal{W}_{(4,2)}$	…	$\mathcal{W}_{(4,n)}$	4
Og	$\mathcal{W}_{(5,1)}$	$\mathcal{W}_{(5,2)}$	…	$\mathcal{W}_{(5,n)}$	5
gz	$\mathcal{W}_{(6,1)}$	$\mathcal{W}_{(6,2)}$	…	$\mathcal{W}_{(6,n)}$	6
zb	$\mathcal{W}_{(7,1)}$	$\mathcal{W}_{(7,2)}$	…	$\mathcal{W}_{(7,n)}$	7
	$\Sigma\mathrm{W}_{(3-7,1)}$	$\Sigma\mathrm{W}_{(3-7,2)}$	…	$\Sigma\mathrm{W}_{(3-7,n)}$	8
…	…	…	…	…	9
$\mathcal{C}_{(k)}$	$\mathcal{W}_{(k,1)}$	$\mathcal{W}_{(k,2)}$	…	$\mathcal{W}_{(k,n)}$	10
	$\mathcal{P}_{(1)}$	$\mathcal{P}_{(2)}$	…	$\mathcal{P}_{(n)}$	11

Measurements based on the weight matrix

- How strong does a particular cue set CS_{Ω} in an event activate the associated outcome O_j?

'Activation'

= sum of association weights of a particular cue set to a particular outcome

	$\mathcal{O}_{(1)}$	**{plural}** $_{(2)}$	…	$\mathcal{O}_{(n)}$	row no.
$\mathcal{C}_{(1)}$	$\mathcal{W}_{(1,1)}$	$\mathcal{W}_{(1,2)}$	…	$\mathcal{W}_{(1,n)}$	1
$\mathcal{C}_{(2)}$	$\mathcal{W}_{(1,1)}$	$\mathcal{W}_{(2,2)}$	…	$\mathcal{W}_{(2,n)}$	2
ld	$\mathcal{W}_{(3,1)}$	$\mathcal{W}_{(3,2)}$	…	$\mathcal{W}_{(3,n)}$	3
dO	$\mathcal{W}_{(4,1)}$	$\mathcal{W}_{(4,2)}$	…	$\mathcal{W}_{(4,n)}$	4
Og	$\mathcal{W}_{(5,1)}$	$\mathcal{W}_{(5,2)}$	…	$\mathcal{W}_{(5,n)}$	5
gz	$\mathcal{W}_{(6,1)}$	$\mathcal{W}_{(6,2)}$	…	$\mathcal{W}_{(6,n)}$	6
zb	$\mathcal{W}_{(7,1)}$	$\mathcal{W}_{(7,2)}$	…	$\mathcal{W}_{(7,n)}$	7
	$\Sigma W_{(3-7,1)}$	$\Sigma W_{(3-7,2)}$	…	$\Sigma W_{(3-7,n)}$	8
…	…	…	…	…	9
$\mathcal{C}_{(k)}$	$\mathcal{W}_{(k,1)}$	$\mathcal{W}_{(k,2)}$	…	$\mathcal{W}_{(k,n)}$	10
	$\mathcal{P}_{(1)}$	$\mathcal{P}_{(2)}$	…	$\mathcal{P}_{(n)}$	11

Measurements based on the weight matrix

- To how many outcomes O is a cue set CS $_{\Omega}$ related?
  Or, how diverse is the activation by a given cue set CS_{Ω} across the different morphological categories?

'Diversity'

= sum of all absolute activations of a cue set

	$\mathcal{O}_{(1)}$	**{plural}** $_{(2)}$	...	$\mathcal{O}_{(n)}$	row no.
$\mathcal{C}_{(1)}$	$\mathcal{W}_{(1,1)}$	$\mathcal{W}_{(1,2)}$	...	$\mathcal{W}_{(1,n)}$	1
$\mathcal{C}_{(2)}$	$\mathcal{W}_{(1,1)}$	$\mathcal{W}_{(2,2)}$	...	$\mathcal{W}_{(2,n)}$	2
ld	$\mathcal{W}_{(3,1)}$	$\mathcal{W}_{(3,2)}$	...	$\mathcal{W}_{(3,n)}$	3
dO	$\mathcal{W}_{(4,1)}$	$\mathcal{W}_{(4,2)}$	...	$\mathcal{W}_{(4,n)}$	4
Og	$\mathcal{W}_{(5,1)}$	$\mathcal{W}_{(5,2)}$	...	$\mathcal{W}_{(5,n)}$	5
gz	$\mathcal{W}_{(6,1)}$	$\mathcal{W}_{(6,2)}$	...	$\mathcal{W}_{(6,n)}$	6
zb	$\mathcal{W}_{(7,1)}$	$\mathcal{W}_{(7,2)}$	...	$\mathcal{W}_{(7,n)}$	7
	$\Sigma W_{(3-7,1)}$	$\Sigma W_{(3-7,2)}$	...	$\Sigma W_{(3-7,n)}$	8
...	...	...	...	...	9
$\mathcal{C}_{(k)}$	$\mathcal{W}_{(k,1)}$	$\mathcal{W}_{(k,2)}$	...	$\mathcal{W}_{(k,n)}$	10
	$\mathcal{P}_{(1)}$	$\mathcal{P}_{(2)}$	...	$\mathcal{P}_{(n)}$	11

$\Sigma W_{(3\text{-}7,\ 1\text{-}n)}$

Modeling strategy: **Which** measure?

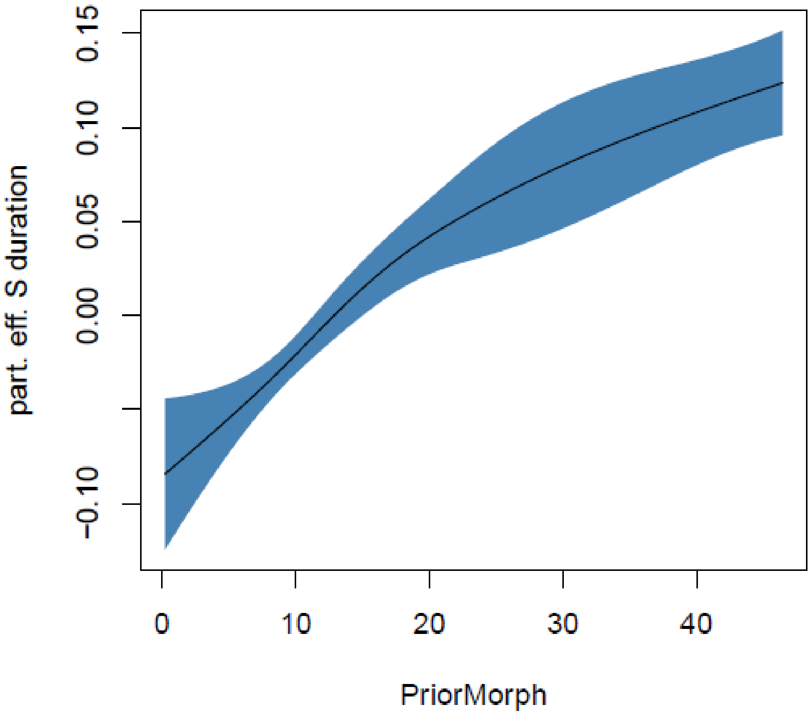
- Find the optimal cue-to-outcome structure (see above)
- Add NDL-measures in a bottom-up strategy to the Replication Model (minus **TYPE OF S**) maintaining its fixed and random effects structure.
- Exclusion of frequency of occurrence (collinearity with activations)
- When necessary, predictors were log-transformed to obtain normal distribution.
- We also excluded strong outliers (N= 1678, i.e. 5% of the data).

Results: NDL measures

- Only model for phrase-internal position of target word (N= 19588)
- Three NDL-measures remain in the final model
- In a GAMM model with the most influential NDL-measures, adding **TYPE OF S** does not improve the model.
- In the NDL model noise variables are still significant with the expected effects.

Predictive NDL measures: The Prior

- How well is a particular outcome (positively or negatively) connected in the network?



General trend

- Increasing support from the network goes together with longer durations.

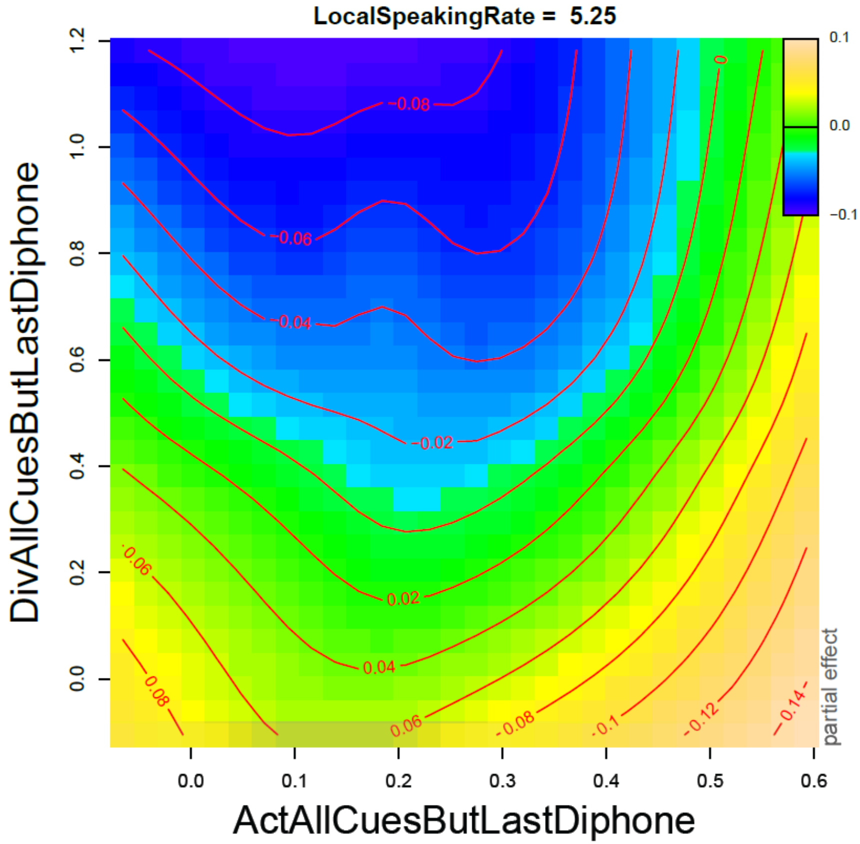
Predictive NDL measures: Activation * Diversity

all cues but the last diphone

Cue structure

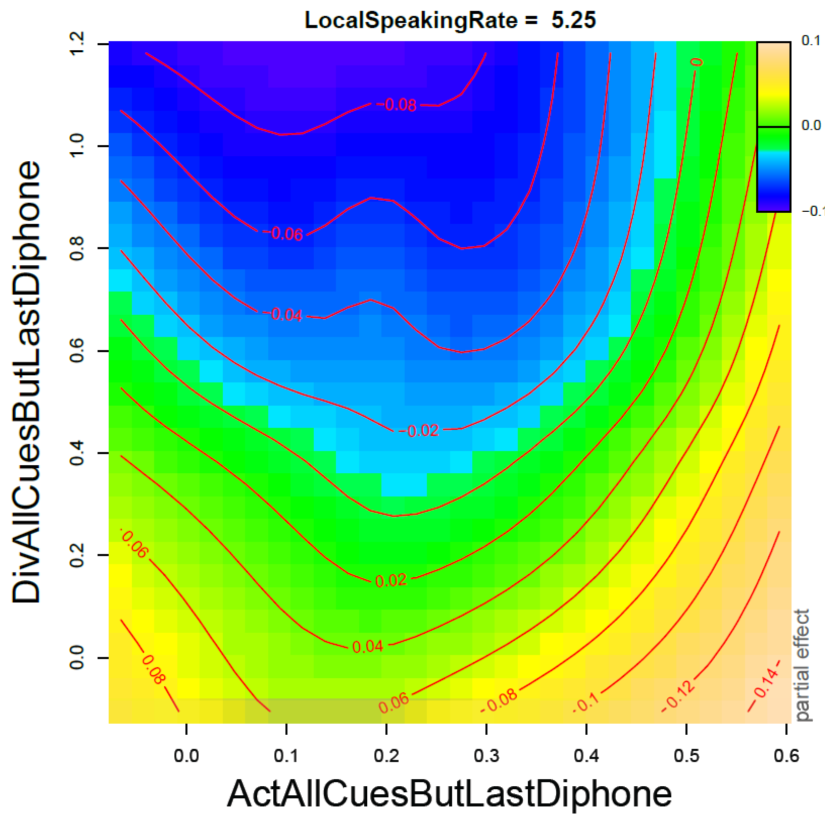
From all cues but last diphone to the morphological meaning expressed in the target word

How well is the given morphological meaning supported by everything in the context but the last diphone?



Activation * Diversity

all cues but the last diphone



General trends

- Increasing diversity goes together with shorter durations.
- Increasing activation goes together with longer durations.

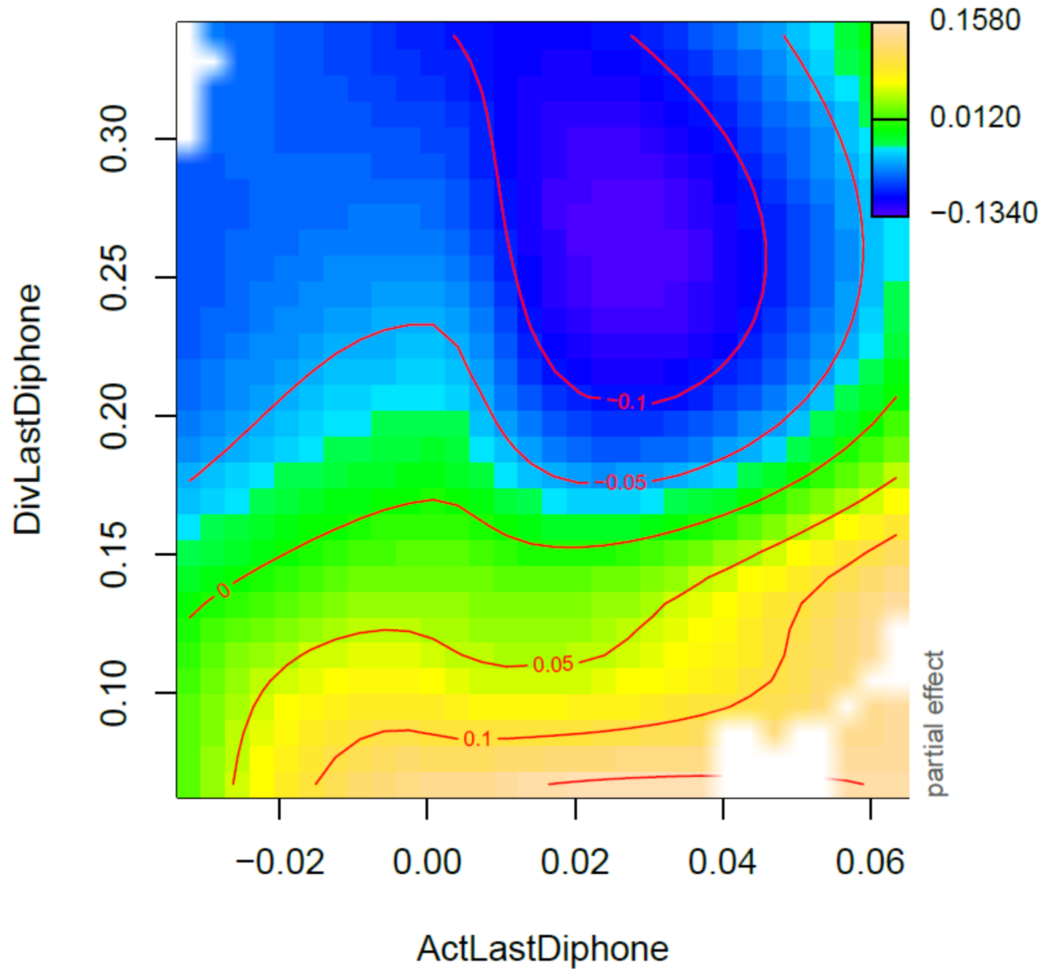
Activation * Diversity

only the last diphone as cue

Cue structure

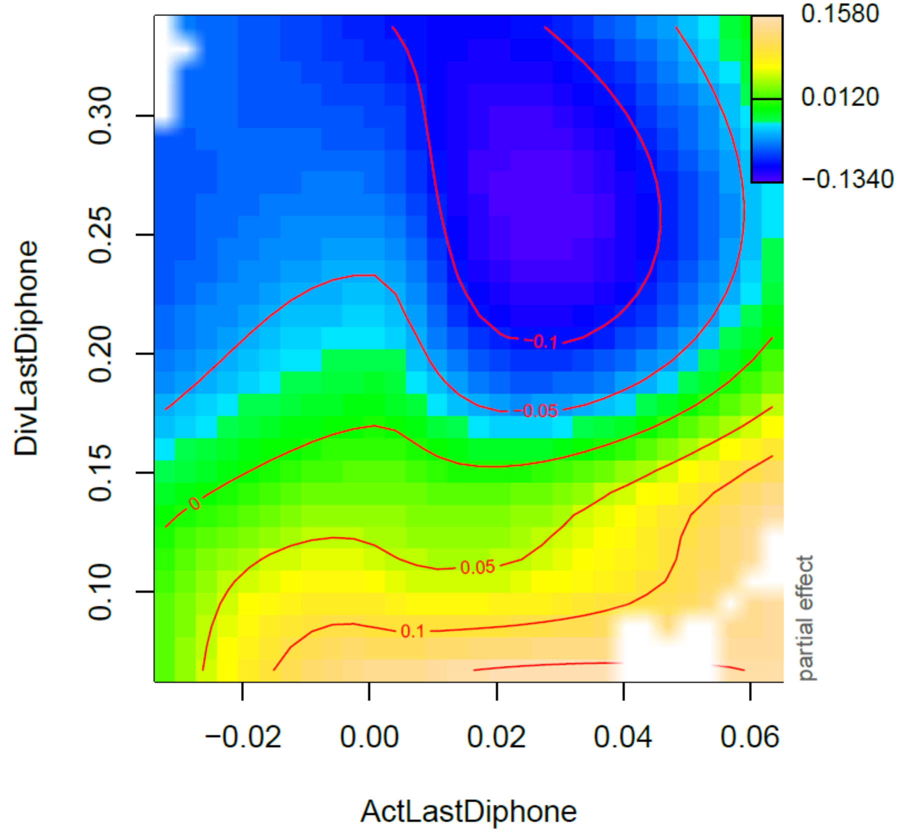
From the last diphone to the morphological meaning expressed in the target word

How well is the given morphological meaning supported by the last diphone?



Activation * Diversity

only the last diphone as cue



General trends

- Increasing diversity goes together with shorter durations.
- Only very small effect of Activation

Summary and interpretation

- NDL measurements involving phonological-phonetic, semantic and morphological information are predictive of S duration.
- General trend:
  More support for a particular morphological function increases durations

Well-known effect of ‘paradigmatic enhancement’
(e.g. Kuperman et al. 2007, Hay et al. 2012, Cohen 2014a, 2014b)

Understanding the effects

- The more support a given targeted outcome receives, the stronger the signal will be that is sent on to the articulators, and the longer the duration of the s will be.

- This longer S-duration is beneficial for the listener, as the acoustic signal is a reliable signal for the morphological function encoded.

- The more support is spread out over different outcomes (i.e. high entropy over the set of morphological functions., the weaker the signal to the articulators will be, resulting in a shorter s-duration.

- As a consequence, the S is not functional for the listener

The bottom line

- Linguistic experience shapes degrees of activation which then lead to structured variability in articulatory gestures, and to different durations of different S‘s.
- In this way morphology can leak into what used to be called post-lexical phonology and articulation.

Thank you very much for your attention!

Special thanks to…

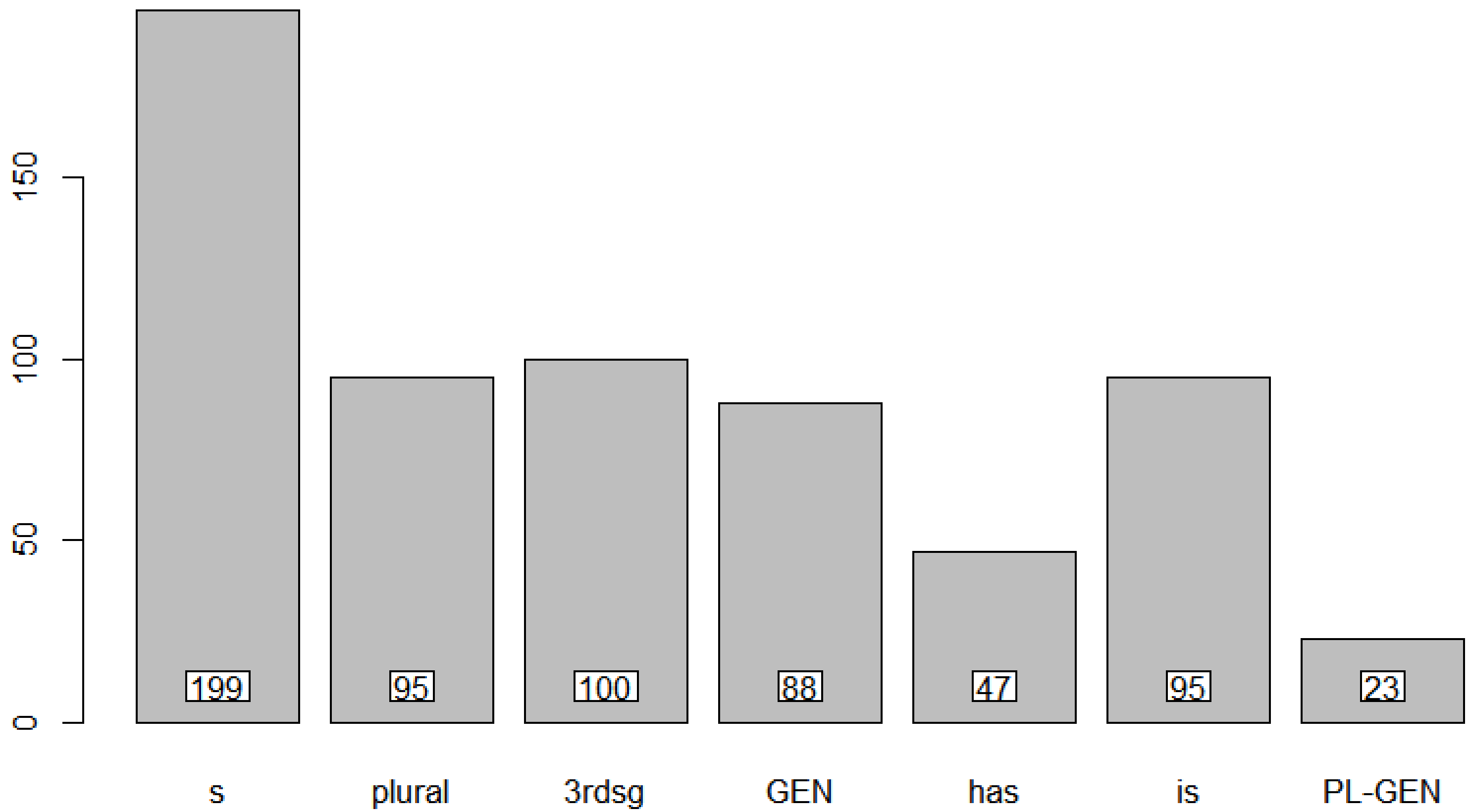
- Our student assistants for annotating the data

Funding

Deutsche Forschungsgemeinschaft: Research Unit FOR 2372

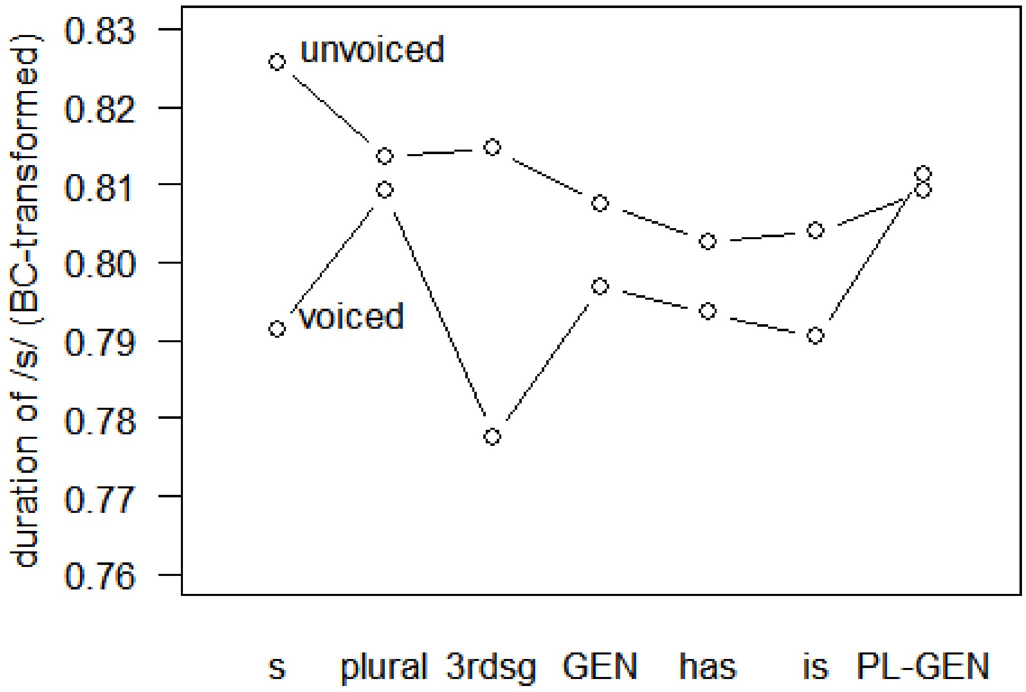
- Grant PL151/8-1 ‘Morpho-phonetic Variation in English‘
- Grant PL151/7-1 ‘FOR 2737 Spoken Morphology: Central Project’
- Grant BA 3080/3-1 ‘The articulation of morphologically complex words’

The data

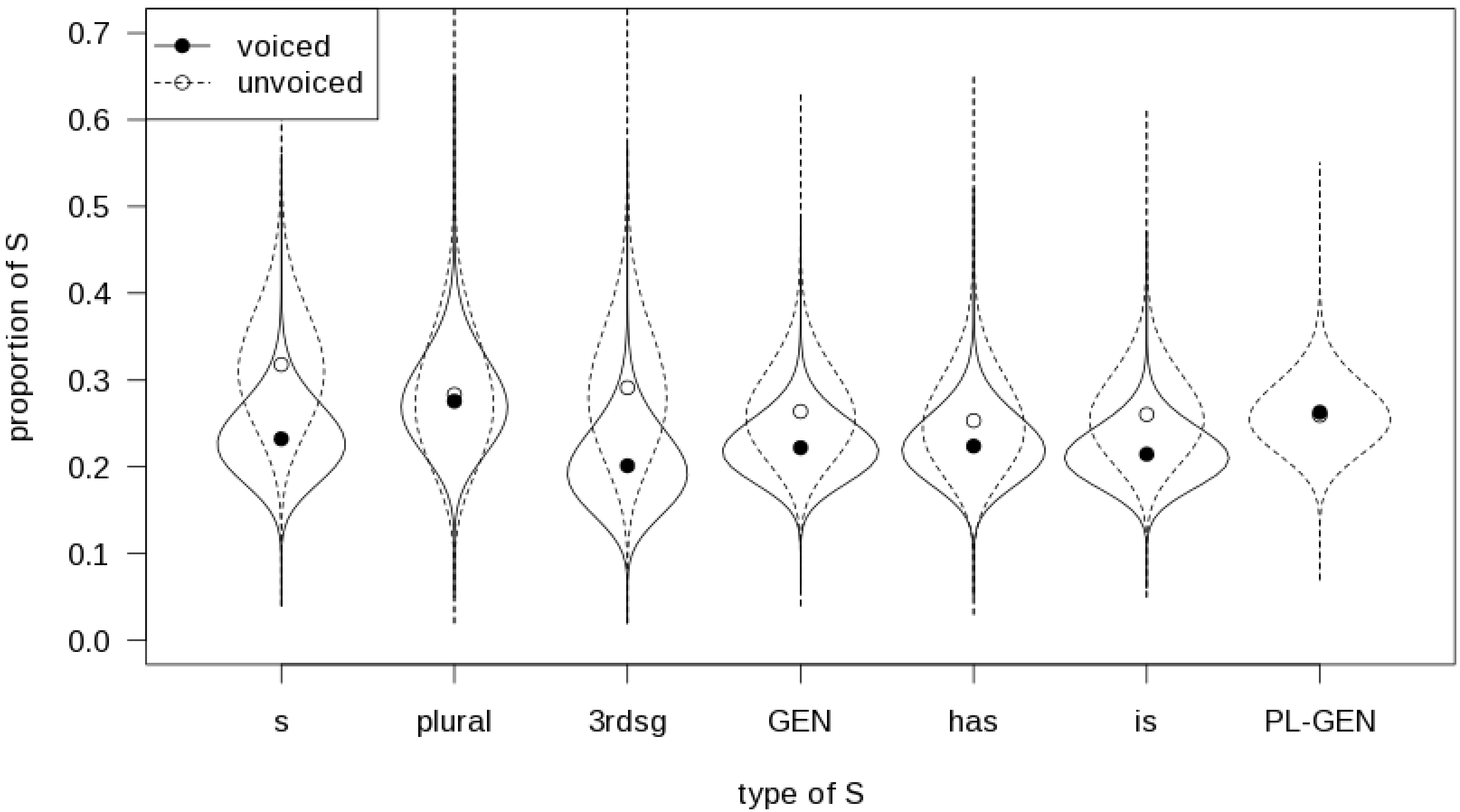


0
50
100
150
199
95
100
88
47
95
23
s
plural
3rdsg
GEN
has
is
PL-GEN

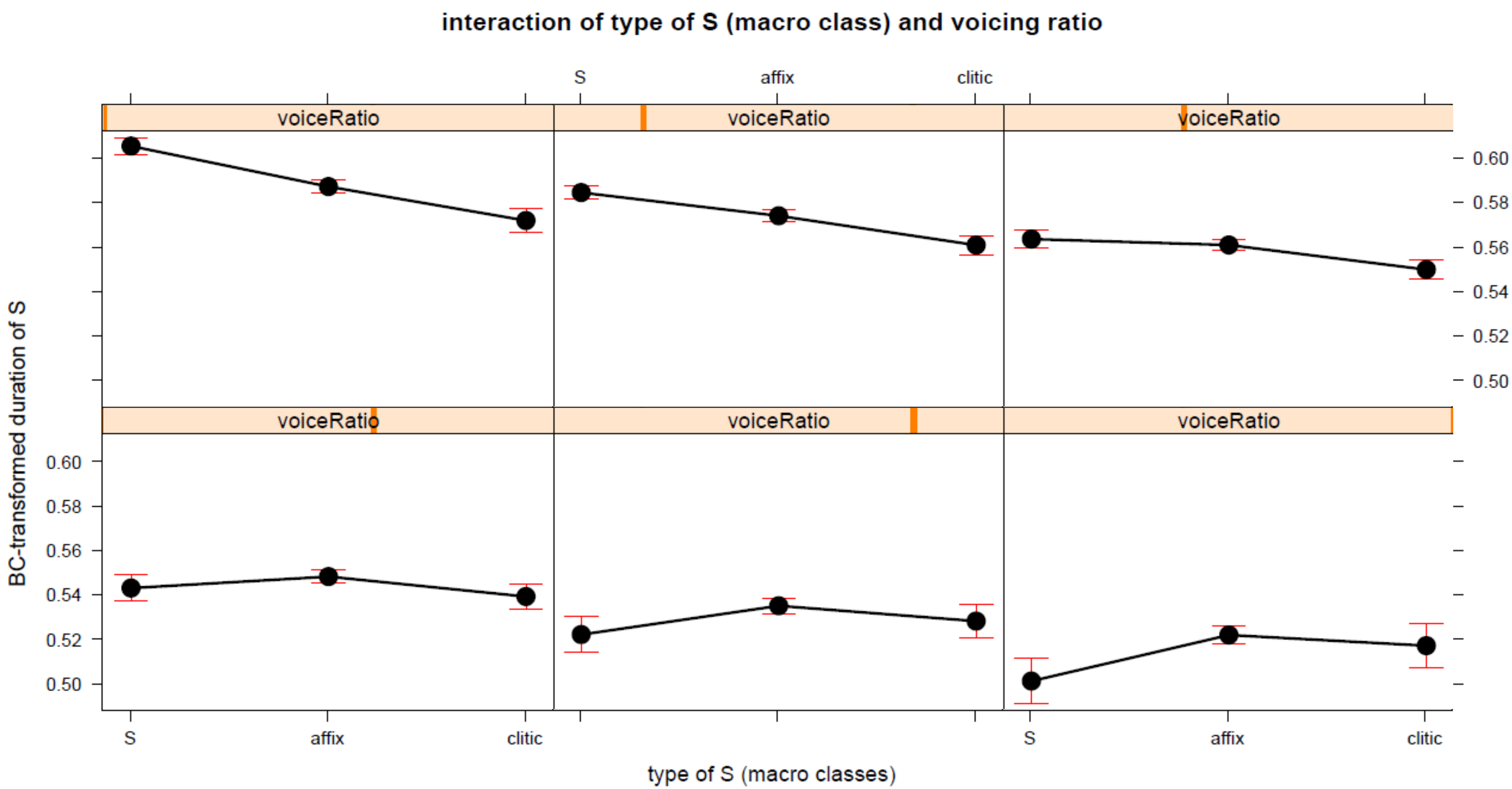
Absolute duration Buckeye: Type of S



Relative duration Buckeye: Type of S



New Zealand English



Rescorla-Wagner equations

$$w_{ij}^{t+1} = w_{ij}^{t} + \Delta w_{ij}^{t}$$

Weight of cue i to some outcome j at time point t+1 equals its weight at point t plus some change

$$\Delta w_{ij}^{t} = \begin{cases} 0 & \text{if } \textsc{absent}(C_i, t), \\ \alpha_i \beta_1 \left(\lambda - \sum_{\textsc{present}(C_k,\, t)} w_{kj}\right) & \text{if } \textsc{present}(C_i, t) \ \&\ \textsc{present}(O_j, t), \\ \alpha_i \beta_2 \left(0 - \sum_{\textsc{present}(C_k,\, t)} w_{kj}\right) & \text{if } \textsc{present}(C_i, t) \ \&\ \textsc{absent}(O_j, t). \end{cases}$$

C_i: the cue of interest

C_k: other cues present or absent

w_k: weights of other cues

O: outcome

t: a certain point in time

α and β: constants representing salience of cue and outcome (0.1)

λ: maximum amount of learning possible on a particular trial (1)

Replication Model

Table 2 shows the summary table of the partial linear effects in the final replication model. The R^2 of the present model was 0.48. All of our linguistic control variables show significant correlations with S durations, and do so in the expected directions.

	Estimate	Std. Error	df	t value	p(>|t|)
(Intercept)	-1.52	0.02	148.39	-69.93	0.00
TypeOfS3rdsg	-0.10	0.02	1372.72	-5.65	0.00
TypeOfSGEN	-0.15	0.03	5647.45	-5.46	0.00
TypeOfShasis	-0.15	0.02	1416.32	-7.33	0.00
TypeOfSPL-GEN	-0.12	0.11	5778.72	-1.08	0.28
TypeOfSplural	-0.10	0.01	1380.73	-8.98	0.00
CategVoicing.75unvoiced	0.23	0.01	28924.37	35.66	0.00
Endcluster.factor2	-0.19	0.01	5778.52	-26.03	0.00
Endcluster.factor3	-0.29	0.01	6103.94	-19.73	0.00
follContextAPP	-0.31	0.01	28822.04	-37.63	0.00
follContextFRI	-0.52	0.01	28900.28	-71.39	0.00
follContextNAS	-0.47	0.01	28872.42	-31.94	0.00
follContextPLO	-0.51	0.01	28906.19	-72.46	0.00
follContextVOW	-0.43	0.01	28909.55	-62.94	0.00
SylSec	-0.08	0.00	28837.16	-38.43	0.00
baseDuration	0.19	0.01	16193.21	32.88	0.00

Selection of cue-to-outcome structure

- three model comparisons based on principled considerations.
- three different cue-to-outcome networks.
- The first of these networks has a diphone-to-lexome structure (see network 1), which we used to test whether the additional inclusion of lexomes as cues is benificial for the durations o S durations.
- This network was less successful in predicting S duration than network 3
- Given the literature on conditional probabilities for upcoming information one might expect that this integration window would suffice for the prediction of S durations (e.g. Jurafsky et al. 2000, Bell et al. 2009). The second and third comparison, therefore, involved networks with the same cue-to-outcome structure as network 3, but with two different integration windows.
- In one, we only used the cues preceding and including the target, and in the other we used the target and the cues following it. Both networks were not as powerful as network 3