Modeling derivative durations with linear discriminative learning

Internal Workshop, Mar 19, 2021

Simon David Stein
Ingo Plag

PROJECT VAR 1

DFG Deutsche Forschungsgemeinschaft
FOR2373

Previously on VAR 1

Derived English words from three corpora

AudioBNC | Quakebox | ONZE

Modeling durations with:

- morphological segmentability (relative frequency)
- other frequency measures (word frequency, base frequency)
- informativity (semantic information load, affix probability)
- prosodic structure (pword integration)
- a number of "traditional" covariates

  - In a nutshell: These variables produce very inconsistent results.
  - Both effects and null results are often not well explained at the (traditional) theoretical level.

LDL

We need to explore whether LDL is a fruitful alternative for predicting our data.

- How well can it account for the durational variation of derivatives?
- What do effects of LDL-derived measures tell us about speech production?
- What does LDL tell us about the role of morphological categories?

Dataset

		tokens	*types*	*derivational functions*
AudioBNC	audio data	4530	363	DIS, NESS, LESS, ATION, IZE

Dataset

		tokens	*types*	*derivational functions*
AudioBNC	audio data	4530	363	DIS, NESS, LESS, ATION, IZE
AudioBNC TASA Baayen et al. 2019	training data		363 + 4813	DIS, NESS, LESS, ATION, IZE, AGAIN, AGENT, EE, ENCE, FUL, IC, INSTRUMENT, ISH, IST, IVE, LY, MENT, MIS, NOT, ORDINAL, OUS, OUT, SUB, UNDO, Y, MONOMORPHEMIC

Matrices

Schematic examples

C matrix

	#k{	k{t	{t#	#h{	h{p
k{t	1	1	1	0	0
h{plnls	0	0	0	1	1
w$k	0	0	0	0	0
lEm@n	0	0	0	0	0

S matrix

	CAT	HAPPINESS	NESS	WALK
k{t	0.000000	-6.24E-05	-0.0003179	4.71E-05
h{plnls	-0.00056	0.0346008	0.032476	7.26E-05
w$k	0.000304	-0.0002334	-9.76E-06	0.00000
lEm@n	-7.28E-05	-2.41E-07	-0.0001247	-2.68E-05

Matrices

Schematic examples

NDL network in TASA

Baayen et al. 2019

lexome-to-lexome matrix

	CAT	HAPPINESS	NESS	WALK
CAT	0.000000	-6.24E-05	-0.0003179	4.71E-05
HAPPINESS	-0.000110	0.00000000	0.032476	0.000194
NESS	-0.000450	0.0346008	0.000000	-0.0001
WALK	0.000304	-0.0002335	-9.76E-06	0.000000
LEMON	-7.28E-05	-2.41E-07	-0.0001247	-2.68E-05

C matrix

	#k{	k{t	{t#	#h{	h{p
k{t	1	1	1	0	0
h{plnls	0	0	0	1	1
w$k	0	0	0	0	0
lEm@n	0	0	0	0	0

S matrix

	CAT	HAPPINESS	NESS	WALK
k{t	0.000000	-6.24E-05	-0.0003179	4.71E-05
h{plnls	-0.00056	0.0346008	0.032476	7.26E-05
w$k	0.000304	-0.0002335	-9.76E-06	0.00000
lEm@n	-7.28E-05	-2.41E-07	-0.0001247	-2.68E-05

Building the S matrices

Two networks

M-Network

$\overrightarrow{happiness} + \overrightarrow{NESS}$

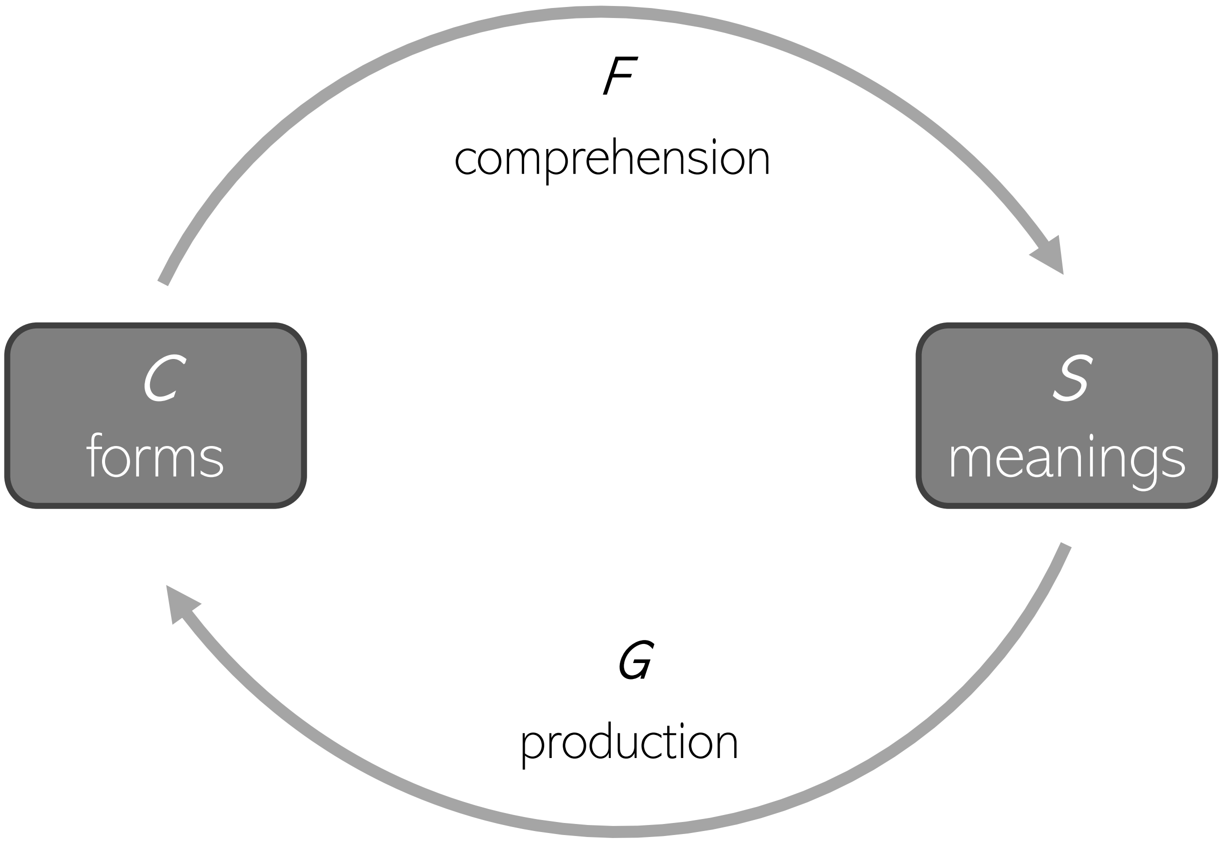
Vectors contain idiosyncratic information and information about morphological category.

I-Network

$\overrightarrow{happiness}$

Vectors contain only idiosyncratic information.

Comprehension and production mapping



Obtaining estimated vectors

predicting meanings

$$\hat{S} = CF$$

predicting forms

$$\hat{C} = SG$$

Predictor variables

predictor	defined as	represents

Predictor variables

predictor	defined as	represents
MEAN WORD SUPPORT	$\frac{sum\ of\ path\ supports}{number\ of\ path\ nodes}$	articulatory certainty

Predictor variables

predictor	defined as	represents
MEAN WORD SUPPORT	$\frac{sum\ of\ path\ supports}{number\ of\ path\ nodes}$	articulatory certainty
SEMANTIC VECTOR LENGTH	L1 distance of $\hat{s}$	activation diversity, polysemy

Predictor variables

predictor	defined as	represents
MEAN WORD SUPPORT	$\frac{sum\ of\ path\ supports}{number\ of\ path\ nodes}$	articulatory certainty
SEMANTIC VECTOR LENGTH	L1 distance of $\hat{s}$	activation diversity, polysemy
SEMANTIC DENSITY	mean correlation of $\hat{s}$ with top 8 neighbors	semantic transparency

Predictor variables

predictor	defined as	represents
MEAN WORD SUPPORT	$\frac{sum\ of\ path\ supports}{number\ of\ path\ nodes}$	articulatory certainty
SEMANTIC VECTOR LENGTH	L1 distance of $\hat{s}$	activation diversity, polysemy
SEMANTIC DENSITY	mean correlation of $\hat{s}$ with top 8 neighbors	semantic transparency
PATH ENTROPIES	Shannon entropy of path supports	articulatory uncertainty

Predictor variables

predictor	defined as	represents
MEAN WORD SUPPORT	$\frac{\boldsymbol{sum\ of\ path\ supports}}{\boldsymbol{number\ of\ path\ nodes}}$	articulatory certainty
SEMANTIC VECTOR LENGTH	L1 distance of $\hat{\boldsymbol{s}}$	activation diversity, polysemy
SEMANTIC DENSITY	mean correlation of $\hat{\boldsymbol{s}}$ with top 8 neighbors	semantic transparency
PATH ENTROPIES	Shannon entropy of path supports	articulatory uncertainty
TARGET CORRELATION	correlation between $\hat{\boldsymbol{s}}$ and $\boldsymbol{s}$	accuracy in predicting meaning from form

Predictor variables

predictor	defined as	represents
MEAN WORD SUPPORT	$\frac{sum\ of\ path\ supports}{number\ of\ path\ nodes}$	articulatory certainty
SEMANTIC VECTOR LENGTH	L1 distance of $\hat{s}$	activation diversity, polysemy
SEMANTIC DENSITY	mean correlation of $\hat{s}$ with top 8 neighbors	semantic transparency
PATH ENTROPIES	Shannon entropy of path supports	articulatory uncertainty
TARGET CORRELATION	correlation between $\hat{s}$ and s	accuracy in predicting meaning from form
SPEECH RATE	$\frac{number\ of\ syllables}{utterance\ duration}$	

Response variable

DURATION DIFFERENCE

residuals of a linear model absolute duration ~ baseline duration

absolute duration = actual acoustic duration

baseline duration = sum of mean segment durations in corpus

Performance

Network accuracy

M-Network

Comprehension	82 %
Production	99 %

I-Network

Comprehension	81 %
Production	99 %

traditional model with RELATIVE FREQUENCY, BIGRAM FREQUENCY, BIPHONE PROBABILITY, AFFIX, SPEECH RATE: 37 %

Explained variance of variables predicting duration

M-Network

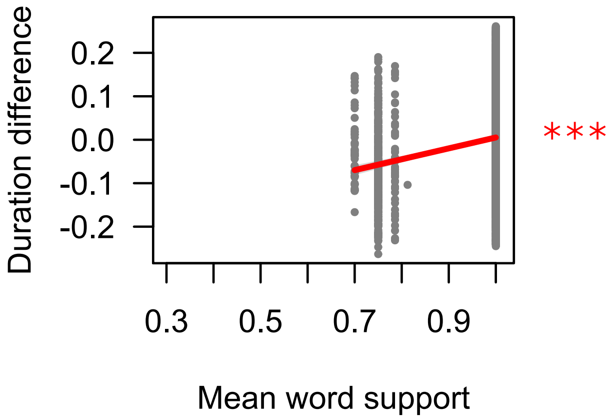
Adjusted R^2 37 %

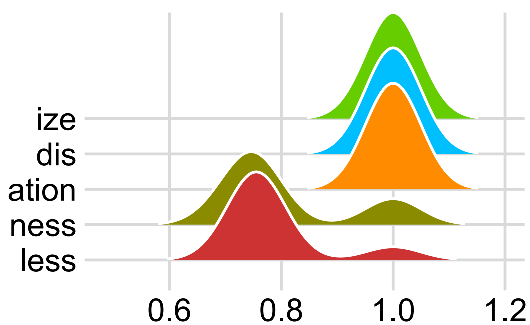
I-Network

Adjusted R^2 38 %

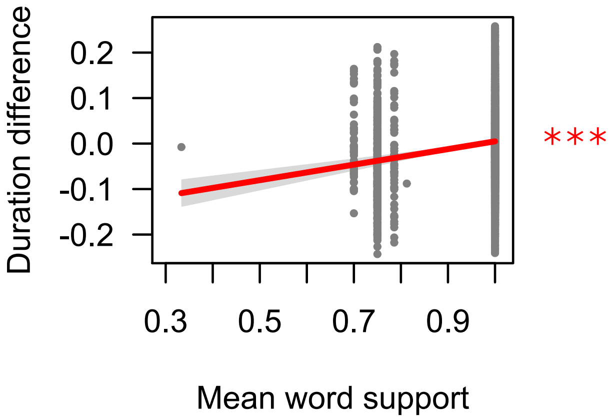
MEAN WORD SUPPORT

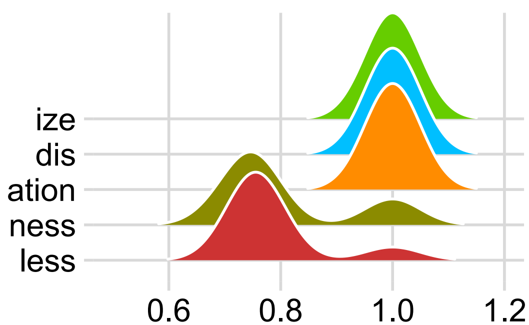
M-Network





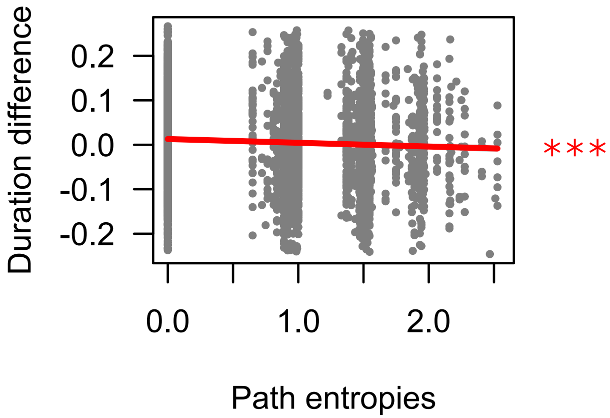
I-Network

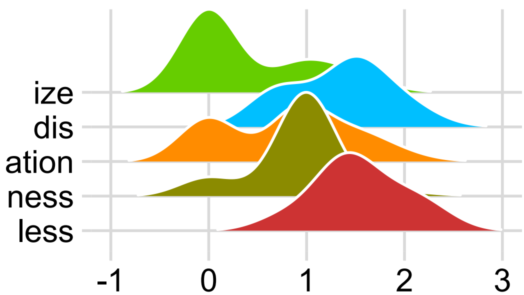




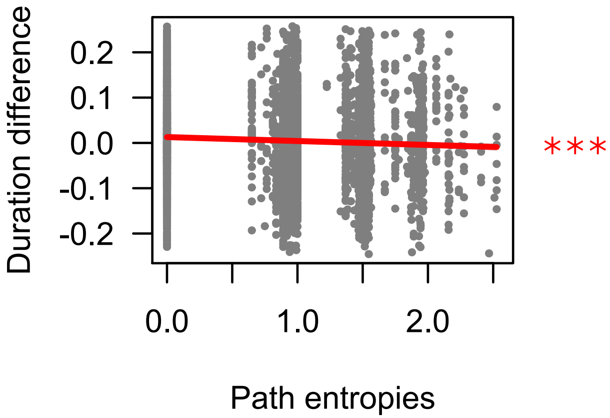
PATH ENTROPIES

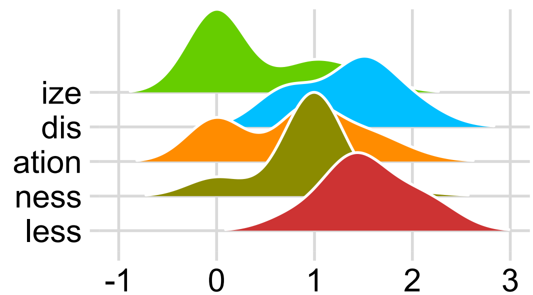
M-Network





I-Network

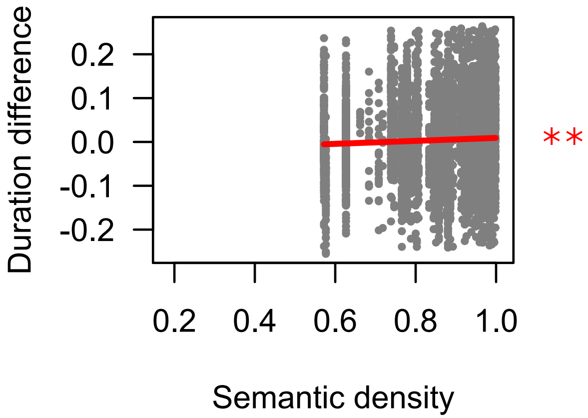


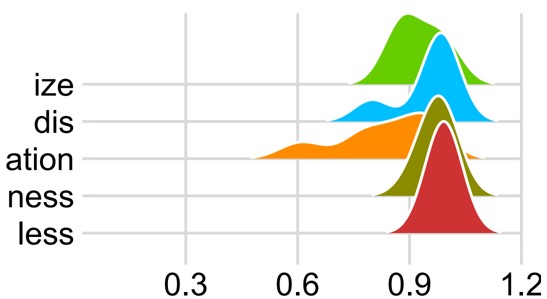


Results

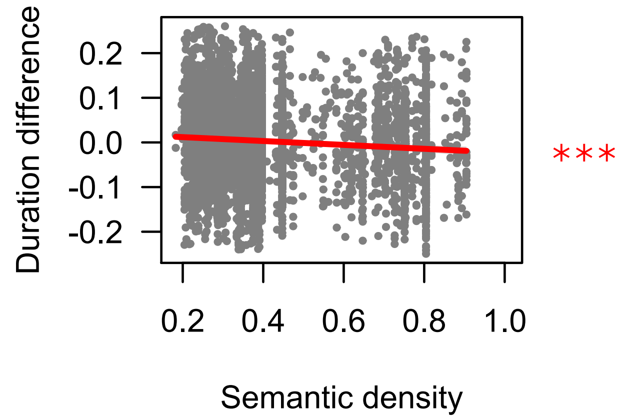
SEMANTIC DENSITY

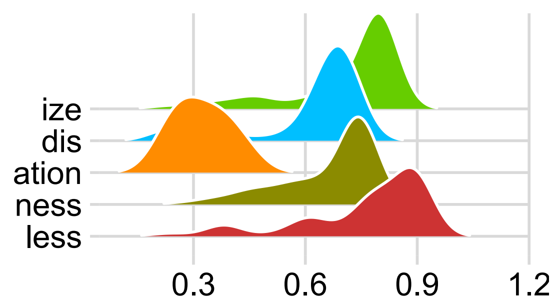
M-Network





I-Network

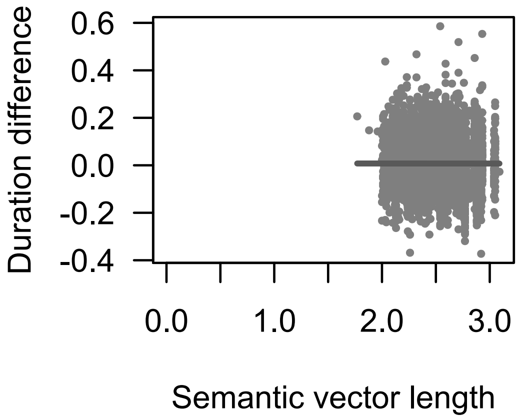


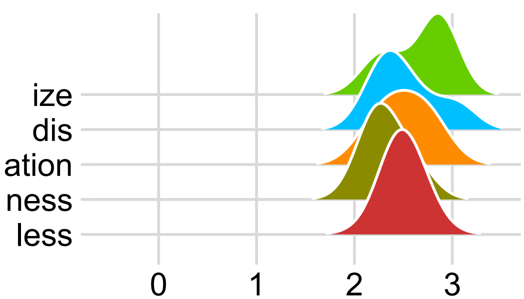


Results

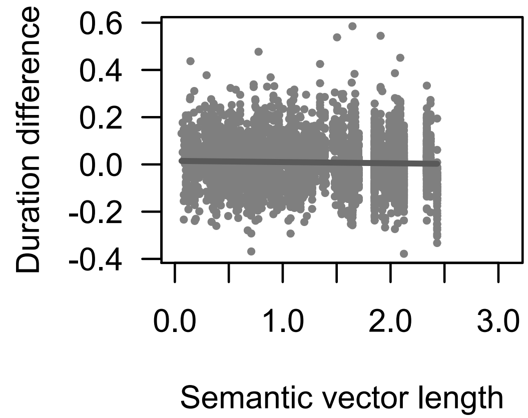
SEMANTIC VECTOR LENGTH

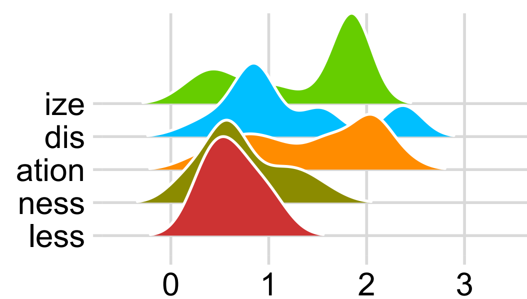
M-Network





I-Network

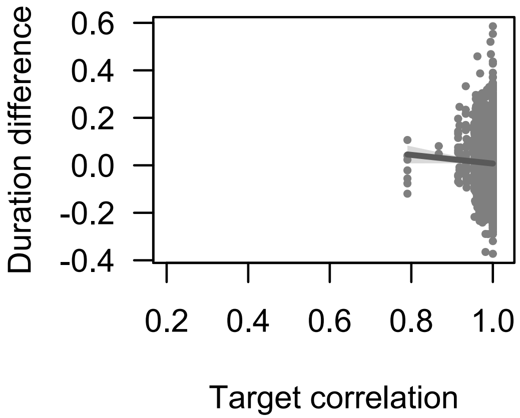


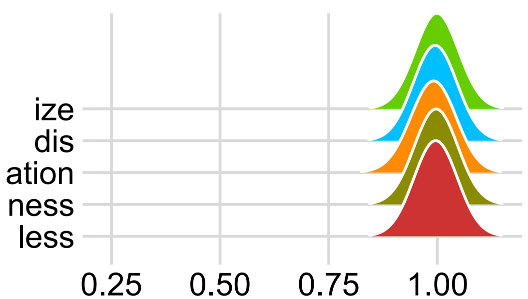


Results

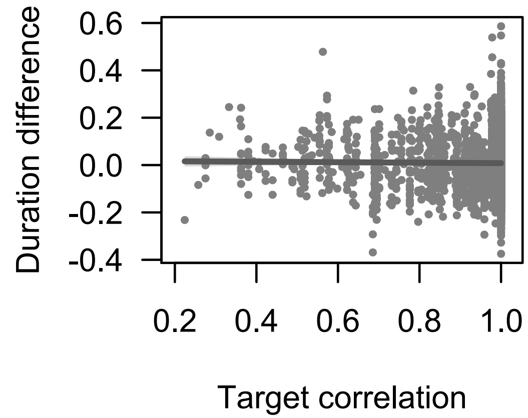
TARGET CORRELATION

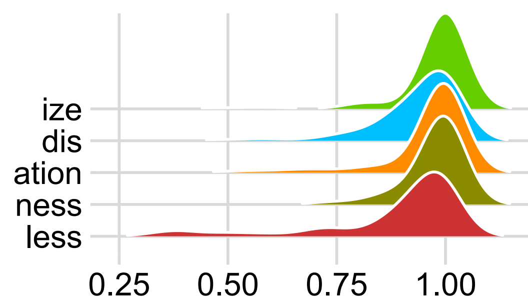
M-Network





I-Network





General implications

LDL-derived variables are successful in predicting derivative durations.

- This is further evidence that discriminative models are a promising approach to speech production where morpho-phonetic effects are not unexpected.
  cf., e.g., Baayen et al. 2019, Chuang et al. 2020, Tomaschek et al. 2019, Tucker et al. 2019

LDL can discriminate derivational functions from sublexical and contextual cues.

- This provides more support for the idea that morphology is possible without morphemes.

Effects on duration

Higher certainty is associated with lengthening.

- In the discussion of whether certainty has an effect of enhancement or reduction, much recent evidence points towards enhancement.
  cf. Tomaschek et al. 2019, Kuperman et al. 2007, Cohen 2014, Cohen 2015, Tucker et al. 2019, this study

Higher semantic transparency can be associated with lengthening and with shortening.

- Traditional lines of argumentation would expect lengthening.
  cf. Hay 2003, 2007, Plag and Ben Hedia 2018, Zuraw et al. 2020
- If interpreted with regards to activation diversity, we could also expect shortening.
  cf. Tucker et al. 2019

Differences between morphological functions

Differences between morphological categories can emerge even from the network without any information about derivational functions.

Some of these differences mirror traditional classifications from the literature.

- Semantic density is higher for words with NESS, LESS and DIS than for words with ATION.
  - *-ness*, *-less* and *dis-* are regarded as producing more transparent derivatives than *-ation* (exception: IZE vs. *-ize*).
    cf. Bauer et al. 2013; Plag 2018
- Semantic vector length was highest for IZE and ATION words.
  - *-ize* and *-ation* are traditionally described as having highly multifaceted semantics, while *-less*, *dis-*, and to a lesser extent *-ness* have clearer and narrower semantics.
    cf. Bauer et al. 2013; Plag 2018

Future directions

We think it could be worthwhile to…

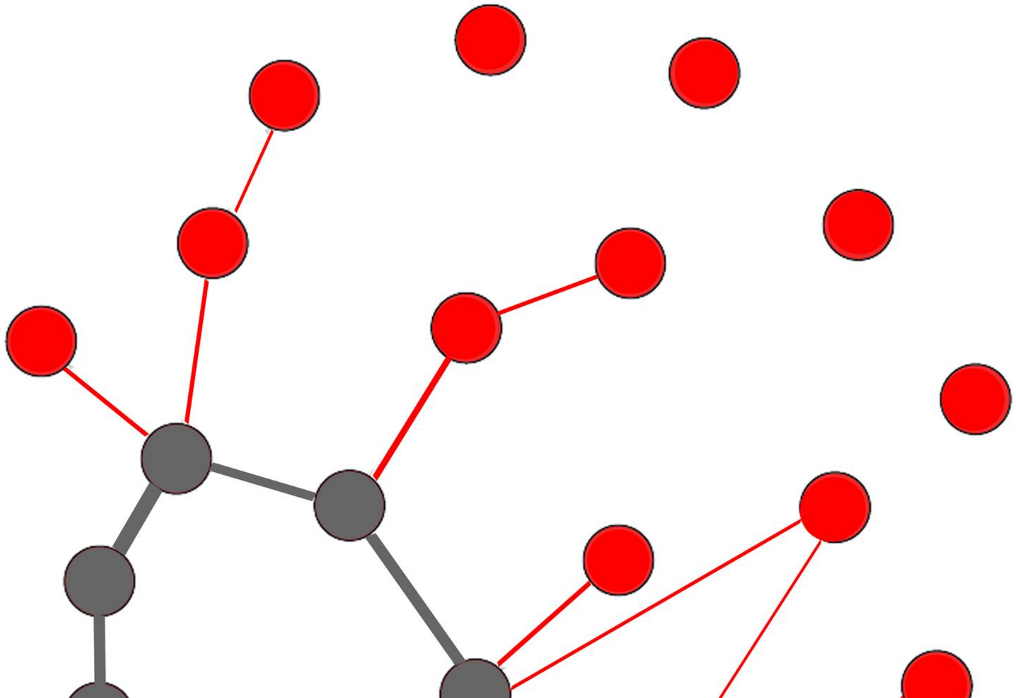
- analyze durations for a larger dataset with more derivational functions.
- somehow control the response variable for segmental makeup without referring to segments.
- explore how to build vectors for words with multiple derivational functions.

We also need to think about how to interpret semantic transparency effects:

- Why does articulation slow down both with high and with low semantic density, and is fastest for medium densities?
- Which behavior would be expected based on which theoretical perspective, and why?

Future directions

Thank you for listening.



1/5

- Baayen, R. H., Chuang, Y.-Y., and Heitmeier, M. (2019a). *WpmWithLdl: Implementation of Word and Paradigm Morphology with Linear Discriminative Learning*.
- Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019b). The discriminative lexicon. A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 2019, 1–39.
- Baayen, R. H., and Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research* 3, 12–28. doi: 10.21500/20112084.807.
- Baayen, R. H., Milin, P., **Đurđević**, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118, 438–481. doi: 10.1037/a0023851.
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database*. Philadelphia: Linguistic Data Consortium.
- Bauer, L., Lieber, R., and Plag, I. (2013). *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Ben Hedia, S. (2019). *Gemination and degemination in English affixation: Investigating the interplay between morphology, phonology and phonetics*. Berlin: Language Science Press.
- Ben Hedia, S., and Plag, I. (2017). Gemination and degemination in English prefixation. Phonetic evidence for morphological organization. *Journal of Phonetics* 62, 34–49.
- Blevins, J. P. (2016). "The minimal sign," in *The Cambridge handbook of morphology*, ed. A. Hippisley and G. Stump (Cambridge: Cambridge University Press), 50–69.
- Boersma, P., and Weenik, D. J. M. (2001). *Praat: Doing phonetics by computer*.
- Burnage, G. (1990). *CELEX: A guide for users*. Nijmegen: Centre for Lexical Information.
- Caselli, N. K., Caselli, M. K., and Cohen-Goldberg, A. M. (2016). Inflected words in production. Evidence for a morphologically rich lexicon. *The Quarterly Journal of Experimental Psychology* 69, 432–454.

2/5

- Chatterjee, S., and Hadi, A. S. (2006). *Regression analysis by example*. 4th ed. Hoboken: John Wiley & Sons.
- Chomsky, N., and Halle, M. (1968). *The sound pattern of English*. New York, Evanston, London: Harper and Row.
- Chuang, Y.-Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., and Baayen, R. H. (2020). The processing of pseudoword form and meaning in production and comprehension. A computational modeling approach using linear discriminative learning. *Behavior Research Methods*. doi: 10.3758/s13428-020-01356-w.
- Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement. Contextual and paradigmatic effects on morpheme pronunciation. *Morphology* 24, 291–323.
- Cohen, C. (2015). Context and paradigms. *The Mental Lexicon* 10, 313–338. doi: 10.1075/ml.10.3.01coh.
- Coleman, J., Baghai-Ravary, L., Pybus, J., and Grau, S. (2012). *Audio BNC: The audio edition of the Spoken British National Corpus*. University of Oxford.
- Davies, M. (2008). *The Corpus of Contemporary American English: 450 million words, 1990–present*.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93, 283–321. doi: 10.1037/0033-295X.93.3.283.
- Divjak, D. (2019). *Frequency in language: Memory, attention and learning*. Cambridge: Cambridge University Press.
- Firth, J. R. (1957). "A synopsis of linguistic theory, 1930–1955," in *Studies in linguistic analysis* (Oxford: Blackwell), 1–31.
- Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66, 789–806. doi: 10.1016/j.jml.2011.11.006.
- Grömping, U. (2006). Relative Importance for Linear Regression in R. The Package relaimpo. *Journal of Statistical Software* 17. doi: 10.18637/jss.v017.i01.
- Hay, J. (2001). Lexical frequency in morphology. Is everything relative? *Linguistics* 39, 1041–1070.

References

3/5

- Hay, J. (2003). *Causes and consequences of word structure*. New York, London: Routledge.
- Hay, J. (2007). "The phonetics of un," in *Lexical creativity, texts and contexts*, ed. J. Munat (Amsterdam, Philadelphia: John Benjamins), 39–57.
- Ivens, S. H., and Koslin, B. L. (1991). *Demands for reading literacy require new accountability measures*. Brewster: Touchstone Applied Science Associates.
- Kiparsky, P. (1982). "Lexical morphology and phonology," in *Linguistics in the morning calm: Selected papers from SICOL*, ed. I.-S. Yang (Seoul: Hanshin), 3–91.
- Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America* 121, 2261–2271. doi: 10.1121/1.2537393.
- Ladefoged, P., and Johnson, K. (2011). *A course in phonetics*. 6th ed. Boston: Wadsworth Cengage Learning.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1–38.
- Lindeman, R. H., Merenda, P. F., and Gold, R. Z. (1980). *Introduction to bivariate and multivariate analysis*. Glenview, London: Scott, Foresman and Company.
- Machač, P., and Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Prague: Epocha Publishing House.
- Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society* 26, 394–395.
- Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* 51, 406–413. doi: 10.1017/S0305004100030401.
- Plag, I. (2018). *Word-formation in English*. 2nd ed. Cambridge: Cambridge University Press.
- Plag, I., and Ben Hedia, S. (2018). "The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration," in *Expanding the lexicon: Linguistic innovation, morphological productivity, and ludicity*, ed. S. Arndt-Lappe, A. Braun, C. Moulin, and E. Winter-Froemel (Berlin, New York: Mouton de Gruyter), 93–116.

References

- Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology. The acoustics of word-final S in English. *Journal of Linguistics* 53, 181–216.
- Plag, I., Lohmann, A., Ben Hedia, S., and Zimmermann, J. (2020). "An <s> is an <s'>, or is it? Plural and genitive-plural are not homophonous," in *Complex words*, ed. L. Körtvélyessy and P. Stekauer (Cambridge: Cambridge University Press).
- Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America* 118, 2561–2569.
- R Core Team (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Ramscar, M., and Yarlett, D. (2007). Linguistic self-correction in the absence of feedback. A new approach to the logical problem of language acquisition. *Cognitive Science* 31, 927–960. doi: 10.1080/03640210701703576.
- Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34, 909–957. doi: 10.1111/j.1551-6709.2009.01092.x.
- Rescorla, R. A., and Wagner, A. (1972). "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement," in *Classical conditioning II: Current research and theory*, ed. A. H. Black and W. F. Prokasy (New York: Appleton-Century-Crofts), 64–99.
- Roelofs, A., and Ferreira, V. S. (2019). "The architecture of speaking," in *Human language: From genes and brains to behavior*, ed. P. Hagoort (Cambridge, Massachusetts: MIT Press), 35–50.
- Saussure, F. de (1916). *Cours de linguistique générale*. Paris: Payot.
- Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., and Malouf, R. (2017). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience* 33, 32–49.

References

5/5

- Sóskuthy, M., and Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166, 298–313. doi: 10.1016/j.cognition.2017.05.032.
- Tomaschek, F., Plag, I., Ernestus, M., and Baayen, R. H. (2019). Phonetic effects of morphology and context. Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics*, 1–39.
- Tucker, B., Sims, M., and Baayen, R. H. (2019). Opposing forces on acoustic duration. Preprint submitted to Elsevier. *PsyArXiv*, 1–38. doi: 10.31234/osf.io/jc97w.
- Tucker, B. V., and Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The Mental Lexicon* 11, 375–400. doi: 10.1075/ml.11.3.03tuc.
- Turk, A., and Shattuck-Hufnagel, S. (2020). *Speech timing: Implications for theories of phonology, speech production, and speech motor control*. New York: Oxford University Press.
- Vitevitch, M. S., and Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers* 36, 481–487.
- Walsh, L., Hay, J., Derek, B., Grant, L., King, J., Millar, P., Papp, V., and Watson, K. (2013). The UC QuakeBox Project. Creation of a community-focused research archive. *New Zealand English Journal* 27, 20–32.
- Widrow, B., and Hoff, M. E. (1960). Adaptive switching circuits. *WESCON Convention Record Part IV*, 96–104.
- Zuraw, K., Lin, I., Yang, M., and Peperkamp, S. (2020). Competition between whole-word and decomposed representations of English prefixed words. *Morphology*, 10.1007/s11525-020-09354-6. doi: 10.1007/s11525-020-09354-6.

Models

M-Network

	Estimate	*Std. Err.*	*t-value*	*Pr(>|t|)*	
Intercept	0.090708	0.025887	3.504	0.000463	***
MEAN WORD SUPPORT	0.250262	0.020700	12.090	< 2e-16	***
SEMANTIC DENSITY	0.033868	0.012372	2.737	0.006217	**
PATH ENTROPIES	-0.008442	0.002309	-3.656	0.000259	***
SPEECH RATE	-0.058602	0.001159	-50.579	< 2e-16	***

I-Network

	Estimate	*Std. Err.*	*t-value*	*Pr(>|t|)*	
Intercept	0.216901	0.026210	8.276	< 2e-16	***
MEAN WORD SUPPORT	0.170726	0.023507	7.263	4.45e-13	***
SEMANTIC DENSITY	-0.043545	0.008925	-4.879	1.10e-06	***
PATH ENTROPIES	-0.008688	0.002242	-3.875	0.000108	***
SPEECH RATE	-0.058757	0.001148	-51.186	< 2e-16	***

Models

Traditional model

	Estimate	Std. Err.	t-value	Pr(>|t|)	
Intercept	3.299e-01	1.086e-02	30.379	< 2e-16	***
RELATIVE FREQUENCY	-2.383e-05	4.167e-05	-0.572	0.567504	
BIGRAM FREQUENCY	-4.169e-07	6.135e-07	-0.680	0.496818	
MEAN BIPHONE PROBABILITY	-4.835e+00	8.661e-01	-5.583	2.51e-08	***
AFFIX less					
ness	2.921e-03	9.242e-03	0.316	0.751941	
ation	5.843e-02	8.201e-03	7.125	1.21e-12	***
dis	6.504e-02	1.016e-02	6.399	1.73e-10	***
ize	3.451e-02	9.222e-03	3.742	0.000185	***
SPEECH RATE	-5.885e-02	1.161e-03	-50.680	< 2e-16	***

Models

Traditional model

	Df	*Sum Sq*	*Mean Sq*	*F-value*	*Pr(>F)*	
RELATIVE FREQUENCY	1	0.018	0.0182	2.1070	0.14669	
MEAN BIPHONE PROBABILITY	1	0.043	0.0433	5.0118	0.02522	*
AFFIX	4	0.581	0.1452	16.8251	1.069e-13	***
SPEECH RATE	1	22.223	22.2229	2574.5115	< 2.2e-16	***
BIGRAM FREQUENCY	1	0.004	0.0040	0.4618	0.49682	

Comparing matrices

M-Network

$\overrightarrow{happiness} + \overrightarrow{NESS}$

Vectors contain idiosyncratic information and information about morphological category.

I-Network

$\overrightarrow{happiness}$

Vectors contain only idiosyncratic information.

B-Network

$\overrightarrow{happy} + \overrightarrow{NESS}$

Vectors contain information about morphological category and the base, but no idiosyncratic information.

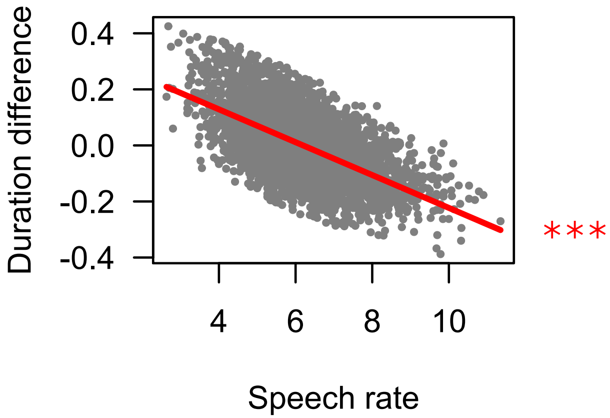
r = 0.08

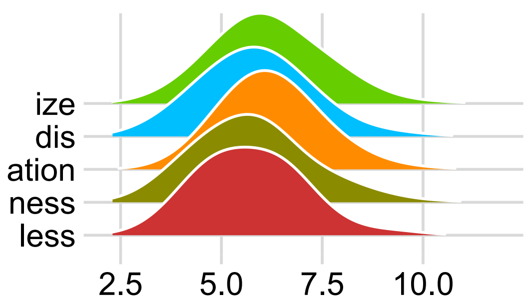
r = 0.1

r = 0.9

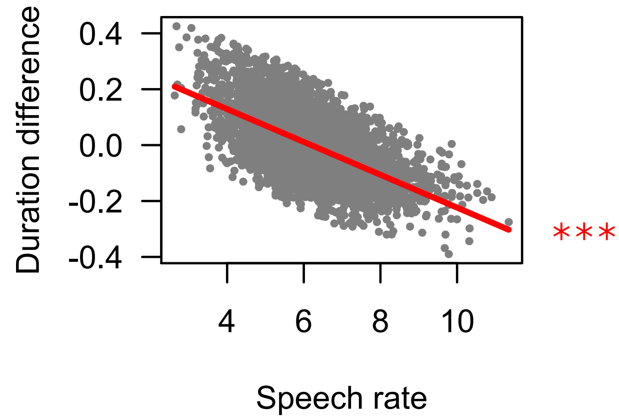
SPEECH RATE

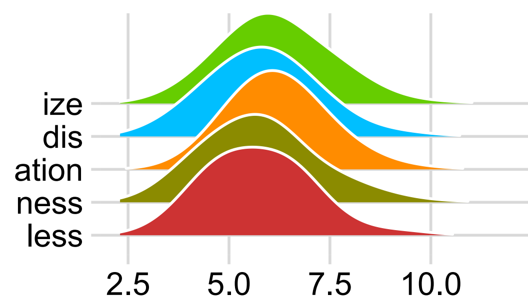
M-Network





I-Network





Path supports

Toy example