Modeling the relationship between prominence and semantics in English compounds

Dominic Schmitz*, Ingo Plag* & Melanie Bell#

*Heinrich-Heine-Universität Düsseldorf, #Anglia Ruskin University Cambridge

Workshop ‘Morphological Variation‘, DGfS 2025, March 5-7, 2025, U Mainz

The problem

Variable stress in English compounds

ópera *glasses*
Óxford *Street*
campáign *promise*

Boston ***márathon***
Oxford ***Róad***
summer ***níght***

- Previous studies report that about one third of compound tokens are right-stressed (e.g. Bell & Plag 2012; Kunter 2011; Plag 2010)
- Many compound types show both patterns (Bell 2015)

Stylized pitch contours, adapted from Kunter (2011:95)

Factors influencing stress assignment

- semantic relation between constituents
- semantic specificity of constituents
- semantic class of constituents
- constituent family sizes
- analogy
- lexicalization
- length
- region
- individual speaker
- ....

Arndt-Lappe 2011; Bell 2015a,b; Bell & Plag 2012, 2013; Kunter 2011; Kunter & Plag 2007; Plag 2006; Plag et al. 2007; Plag et al. 2008; Plag & Kunter 2010; Plag 2010

Explanations?

Research questions

- How do speakers actually make use of these factors?
- How does the speaker learn to apply these factors?

Hypothesis

- The observed effects emerge from a language system that originates in the speaker's experience, through a process of discriminative learning

Aims

- Model the relation between compound prosody and semantics in a discriminative learning framework (NDL, LDL, Baayen et al. 2019)
- Investigate whether pitch and intensity contours are predictable from the compounds' context-specific semantics and vice versa

Data

- Boston University Radio Speech Corpus (BURSC) (Ostendorf et al. 1996)
- American English, news texts, professional speakers
- Number of NN compound tokens = 4327, number of types = 2476
- ‘Latin square’ set (at least 2 speakers in each context) tokens = 397, types = 79

type	context	1	2	…
	tokens	> 1	> 1	…

Sample from BURSC

The device is attached to a plastic wristband. It looks like a watch. It functions like an electronic probation officer. When a computerized call is made to a former prisoner's home phone, that person answers by plugging in the device. The wristband can be removed only by breaking its clasp, and if that's done the inmate is immediately returned to jail. The description conjures up images of big brother watching. But Jay Ash, deputy superintendent of the Hampton County jail in Springfield, says the surveillance system is not that sinister.

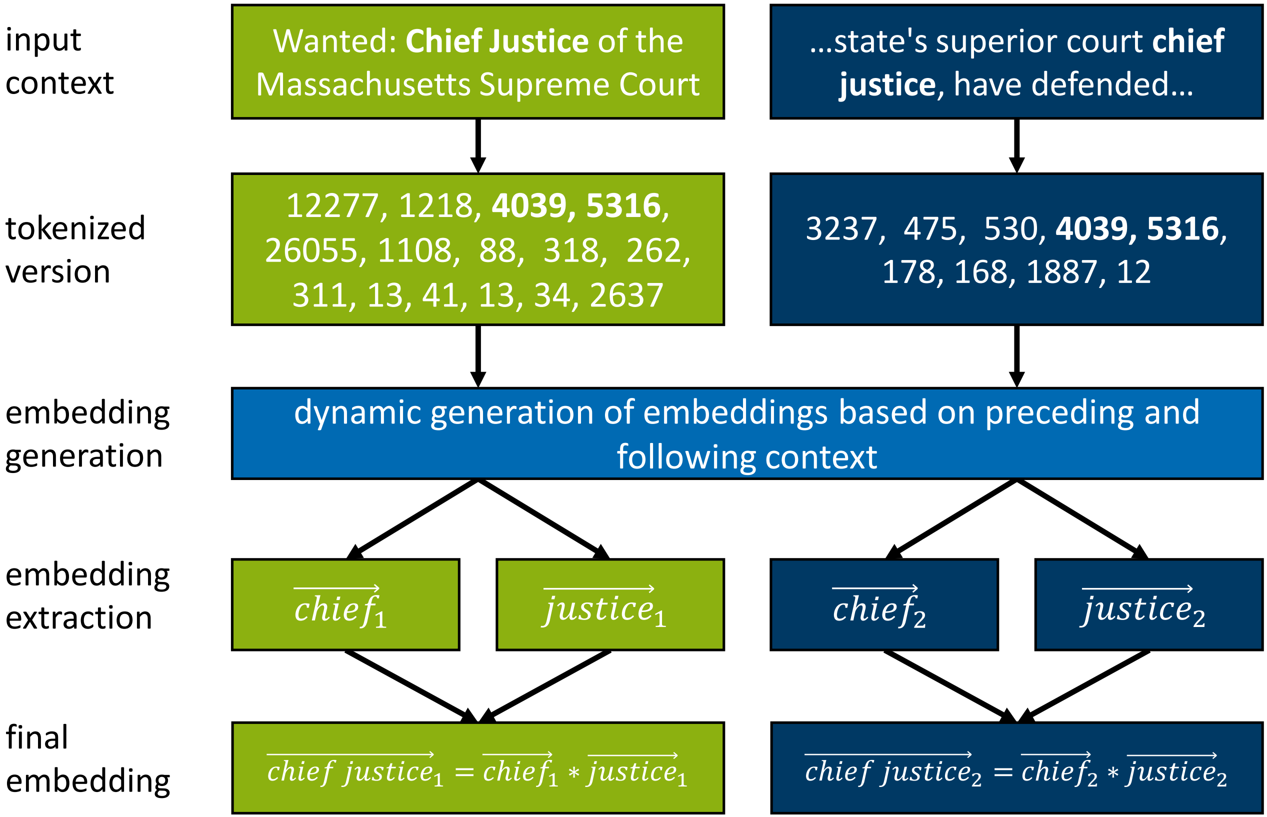
Overall procedure

1. Use BERT (a pretrained large language model) to get semantic vectors for the meaning-in-context of each compound token in the dataset
2. Use Praat to get pitch and intensity measurements for voiced sections of the audio tokens
3. Use PraatSmooth to create pitch and intensity curves for the whole of each token
4. Map **form** (pitch or intensity values) onto **semantics** (vectors), and vice versa, in LDL
5. Check whether form can be predicted from semantics, and vice versa

Semantics: Embeddings

- Context-dependent token-level semantic vectors (embeddings) were extracted from BERT (Devlin et al. 2018)
  - A transformers model pretrained on a very large corpus of English data in a self-supervised fashion
  - Trained to guess the next word in sentences
  - Uses both preceding and following context
  - 110M parameters
  - Texts are tokenized using a byte-level version of Byte Pair Encoding
  - Inputs are sequences of 512 consecutive tokens
  - For each compound, output is a vector with 768 dimensions

Embeddings



input context
Wanted: **Chief Justice** of the Massachusetts Supreme Court
...state's superior court **chief justice**, have defended...
tokenized version
12277, 1218, **4039, 5316**, 26055, 1108, 88, 318, 262, 311, 13, 41, 13, 34, 2637
3237, 475, 530, **4039, 5316**, 178, 168, 1887, 12
embedding generation
dynamic generation of embeddings based on preceding and following context
embedding extraction
$\overrightarrow{chief}_1$
$\overrightarrow{justice}_1$
$\overrightarrow{chief}_2$
$\overrightarrow{justice}_2$
final embedding
$\overrightarrow{chief\ justice}_1 = \overrightarrow{chief}_1 * \overrightarrow{justice}_1$
$\overrightarrow{chief\ justice}_2 = \overrightarrow{chief}_2 * \overrightarrow{justice}_2$

Embeddings: Semantic matrix S

		[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	
wristband	[1,]	90	710	803	515	331	284	221	588	672	95	...
probation officer	[2,]	556	214	969	577	39	193	198	350	667	863	...
home phone	[3,]	844	62	194	157	894	186	496	497	723	614	...
	[...]	...										

Form: Pitch and intensity

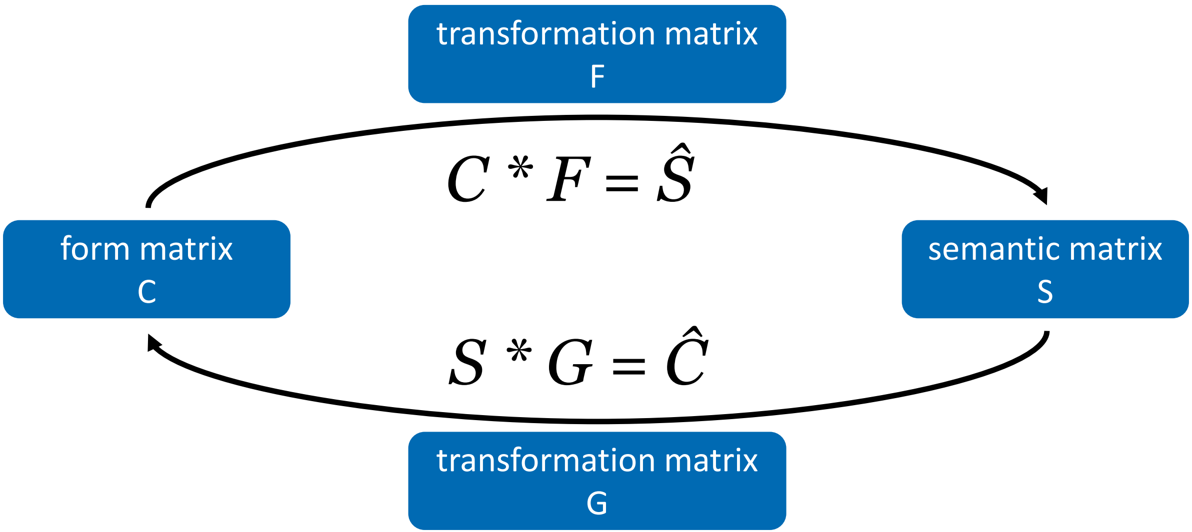
- Extracted using sound files and TextGrids with the rPraat package (Bořil & Skarnitzl 2016) in R
- From raw to final **pitch** data
  - Sound to pitch with speaker-specific parameters
  - Removal of octave jumps
  - Contour smoothing using the PraatSmooth algorithm
  - Transformation from Hertz to semitones, speaker-specific baseline
  - Centring and scaling
  - Sampling at 51 equally spaced points in time
- From raw to final **intensity** data
  - Contour smoothing using the PraatSmooth algorithm
  - Centring and scaling
  - Sampling at 51 equally spaced points in time

Form matrix C

		[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
wristband	[1,]	234	978	213	330	491	834	395	259	71	235
probation officer	[2,]	504	705	926	627	248	478	718	201	479	589
home phone	[3,]	72	68	562	472	41	816	931	422	182	249

Mapping form on semantics, semantics on form

- Linear discriminative learning (JudiLing: Luo et al. 2024)



Mapping form on semantics, semantics on form

$$C * \boldsymbol{F} = \hat{S}$$ transformation matrix *F*

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	
[1,]	90	710	803	515	331	284	221	588	672	95	...
[2,]	556	214	969	577	39	193	198	350	667	863	...
[3,]	844	62	194	157	894	186	496	497	723	614	...
[...]	...										

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	234	978	213	330	491	834	395	259	71	235
[2,]	504	705	926	627	248	478	718	201	479	589
[3,]	72	68	562	472	41	816	931	422	182	249

transformation matrix *G* $$\hat{C} = \boldsymbol{G} * S$$

Predicting forms and predicting meanings

- 397 compound tokens in a leave-1-out design
- Train the transformation matrix on 396 compounds then use the transformation matrix to predict the form or semantics of the 397th compound.
- 397 iterations, check accuracy of predictions
- ‘Correct’: the nearest neighbour vector is that of a token of the same type

$\overrightarrow{chief\ justice_2}$

$\overrightarrow{deputy\ superintenent_1}$

$\overrightarrow{house\ speaker_3}$

$\overrightarrow{chief\ justice_1}$

$\overrightarrow{governor\ Dukakis_2}$

$\ldots$

nearest neighbor

nearest neighbor

Results

form vector type	**chance** accuracy	**comprehension** form to semantics	**production** semantics to form
pitch	2 %		
intensity			

Results

form vector type	**chance** accuracy	**comprehension** form to semantics	**production** semantics to form
pitch	2 %		19.1 %
intensity			21.7 %

- Pitch and intensity contours are predictable from the semantics
- Accuracy is not significantly different for predicting pitch and intensity contours

Results

form vector type	**chance** accuracy	**comprehension** form to semantics	**production** semantics to form
pitch	2 %	11.6 %	19.1 %
intensity		5.5 %	21.7 %

- Pitch and intensity contours are predictable from the semantics
- Accuracy is not significantly different for predicting pitch and intensity contours

- Semantics is predictable from pitch and intensity
- Pitch performs significantly better than intensity
- Predicting acoustic contours is easier than predicting semantics

Interpretation

Why is comprehension worse?

- Mathematical reason: going from a lower number of dimensions to a higher number of dimensions is more complex
- Each compound is characterised by either 51 pieces of information (dimensions of form) or 768 pieces of information (dimensions of semantics)
- Estimating 768 on the basis of 51 is more difficult and error-prone than the reverse

Interpretation

Why is intensity worse than pitch in comprehension?

- Idea
  - Pitch provides richer information about a type
  - Pitch is more type-specific than intensity

Why is there a difference between pitch and intensity only in comprehension and not in production?

  - Production has the same input in both models

Conclusion

- It is possible to map between acoustic and semantic parameters
- Old model of compound stress
  - Only concerned with production
  - Computation of abstract stress on the basis of abstract semantic categories and other lexical properties
- New model
  - Concerned with production and comprehension
  - Direct mapping from speech signal to experience-based semantics and vice-versa
- The models suggest that
  - Human language could involve a direct mapping between the speech signal and semantics
  - Abstract categories may be emergent

Thank you for listening!

References

Arndt-Lappe, S. (2011). Towards an exemplar-based model of stress in English noun-noun compounds1. *Journal of Linguistics*, *47*(3), 549-585. https://doi.org/10.1017/S0022226711000028

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). 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*, 4895891. https://doi.org/10.1155/2019/4895891

Bell, M. J. (2015a). Basic Relations and Stereotype Relations in the Semantics of Compound Nouns. *Journal of Cognitive Science*, *16*(3), 224-260. https://doi.org/10.17791/JCS.2015.16.3.224

Bell, M. J. (2015b). Inter-speaker variation in compound prominence. *Lingue e Linguaggio*, *XIV*(1/2015), 61-78. https://doi.org/10.1418/80754

Bell, M. J., & Plag, I. (2012). Informativeness is a determinant of compound stress in English. *Journal of Linguistics*, *48*(3), 485-520. https://doi.org/10.1017/S0022226712000199

Bell, M. J., & Plag, I. (2013). Informativity and analogy in English compound stress. *Word Structure*, *6*(2), 129-155. https://doi.org/10.3366/word.2013.0042

Bořil, T., & Skarnitzl, R. (2016). *Tools rPraat and mPraat* (pp. 367-374). https://doi.org/10.1007/978-3-319-45510-5_42

Heitmeier, M. (2022). *JudiLingMeasures*. Eberhard Karls Universität, Seminar für Sprachwissenschaft.

Kunter, G. (2011). *Compound stress in English: The phonetics and phonology of prosodic prominence*. De Gruyter Mouton.

Kunter, G., & Plag, I. (2007). What is compound stress? In J. Trouvain & J. Barry (Eds.), *Proceedings of the 16th international congress of phonetic sciences* (pp. 1005-1008).

Luo, X., Heitmeier, M., Chuang, Y.-Y., & Baayen, R. H. (2024). *JudiLing: an implementation of the Discriminative Lexicon Model in Julia*. Eberhard Karls Universität, Seminar für Sprachwissenschaft.

Ostendorf, M., Price, P., & Shattuck-Hufnagel, S. (1996). *Boston University Radio Speech Corpus*. Linguistic Data Consortium.

Plag, I. (2006). The variability of compound stress in English: structural, semantic, and analogical factors. *English Language & Linguistics*, *10*(1), 143-172. https://doi.org/10.1017/S1360674306001821

Plag, I. (2010). Compound stress assignment by analogy: the constituent family bias. *Zeitschrift Für Sprachwissenschaft*, *29*(2), 243-282.

Plag, I., & Kunter, G. (2010). Constituent family size and compound stress assignment in English. *Linguistische Berichte*, *Sonderheft*(17), 349-382.

Plag, I., Kunter, G., & Lappe, S. (2007). Testing Hypotheses About Compoundstress Assignment in English: A Corpus-Based Investigation. *Corpus Linguistics and Linguistic Theory*, *3*(2), 199-232. https://doi.org/10.1515/CLLT.2007.012/MACHINEREADABLECITATION/RIS

Plag, I., Kunter, G., Lappe, S., & Braun, M. (2008). The Role of Semantics, Argument Structure, and Lexicalization in Compound Stress Assignment in English. *Language*, *84*(4), 760-794. https://doi.org/10.1353/LAN.0.0072

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners*.