English compound prominence and compound semantics are mutually predictable

Dominic Schmitz, Melanie J. Bell & Ingo Plag
Heinrich-Heine-Universität Düsseldorf (DS, IP), Anglia Ruskin University Cambridge (MJB)

The problem

RECEIVED WISDOM (e.g., Bauer et al. 2013, Kunter 2011, Plag et al. 2008)

Left-stressed compounds: *campáign promise, chíld care, probátion officer*
Right-stressed compounds: *Boston hárbour, home phóne, silk tíe*

Phonetic correlates of prominence

- Pitch
- Intensity
- ...

Phonological interpretation

- left stress: one pitch accent, on left constituent
- right stress: two pitch accents, one on each constituent

Factors influencing prominence ratings

- *Semantic relation*	- *Semantic specificity*	- *Semantic class*
- Family size	- Analogy	- Lexicalization
- Length	- Region	- Individual speaker

MAJOR PROBLEMS

- Huge amount of variability
- Probabilistic effects of questionable categorical predictors
- How does that work, really?

HYPOTHESIS

- The observed effects emerge from a language system that originates in the speaker's experience, through a process of discriminative learning (Rescorla & Wagner 1972), as implemented in Linear Discriminative Learning (Baayen et al. 2019).

Discriminative learning

How it works

- Building association between representations ('cues' and 'outcomes').
- Association weight increases every time a given cue and a given outcome co-occur.
- Association weight decreases whenever the cue occurs without that outcome.

$$\mathbf{C} = \begin{array}{r} \\ \text{chief justice} \\ \text{retirement age} \end{array} \begin{array}{c} \begin{matrix} \text{C1} & \text{C2} & \text{C3} & \text{C4} & \text{C5} & \ldots \end{matrix} \\ \begin{pmatrix} 0.6960 & 0.6958 & 0.6953 & 0.6947 & 0.6941 & \ldots \\ 0.5047 & 0.5043 & 0.5030 & 0.5010 & 0.4989 & \ldots \end{pmatrix} \end{array}$$

Figure 1: $\mathbf{C}$ *matrix with form vectors (e.g., pitch values).*

$$\mathbf{S} = \begin{array}{r} \\ \text{chief justice} \\ \text{retirement age} \end{array} \begin{array}{c} \begin{matrix} \text{S1} & \text{S2} & \text{S3} & \text{S4} & \text{S5} & \ldots \end{matrix} \\ \begin{pmatrix} -0.0559 & -1.0561 & 0.8781 & -0.8009 & -0.8374 & \ldots \\ -1.0529 & -1.4735 & -1.5778 & -0.9707 & 1.7550 & \ldots \end{pmatrix} \end{array}$$

Figure 2: $\mathbf{S}$ *matrix with semantic vectors.*

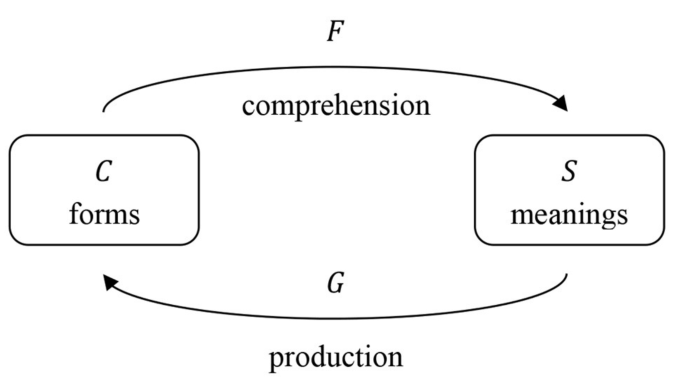

Figure 3: *Comprehension and production mapping in LDL. F and G are transformation matrices for comprehension and production.*

Methodology

Data and LDL modeling

- C matrix: Pitch and intensity contours of 760 compound tokens (193 types) from the Boston University Radio Speech Corpus (Ostendorf et al. 1996, Plag et al. 2008).
- S matrix: Context-sensitive semantic vectors for these compounds from BERT-base-uncased (Devlin et al. 2019).
- Fitting LDL models using the `JudiLing` package for Julia (Luo et al. 2021).

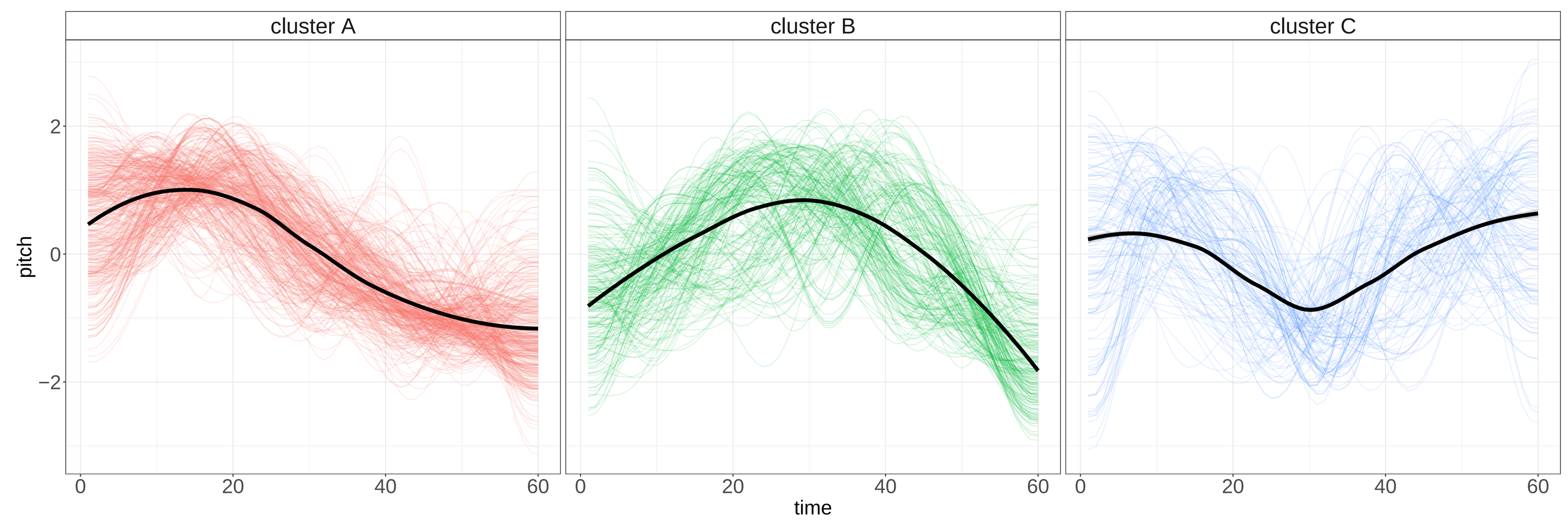

Figure 4: *Optimal clusters of pitch contours with non-linear average smooths.*

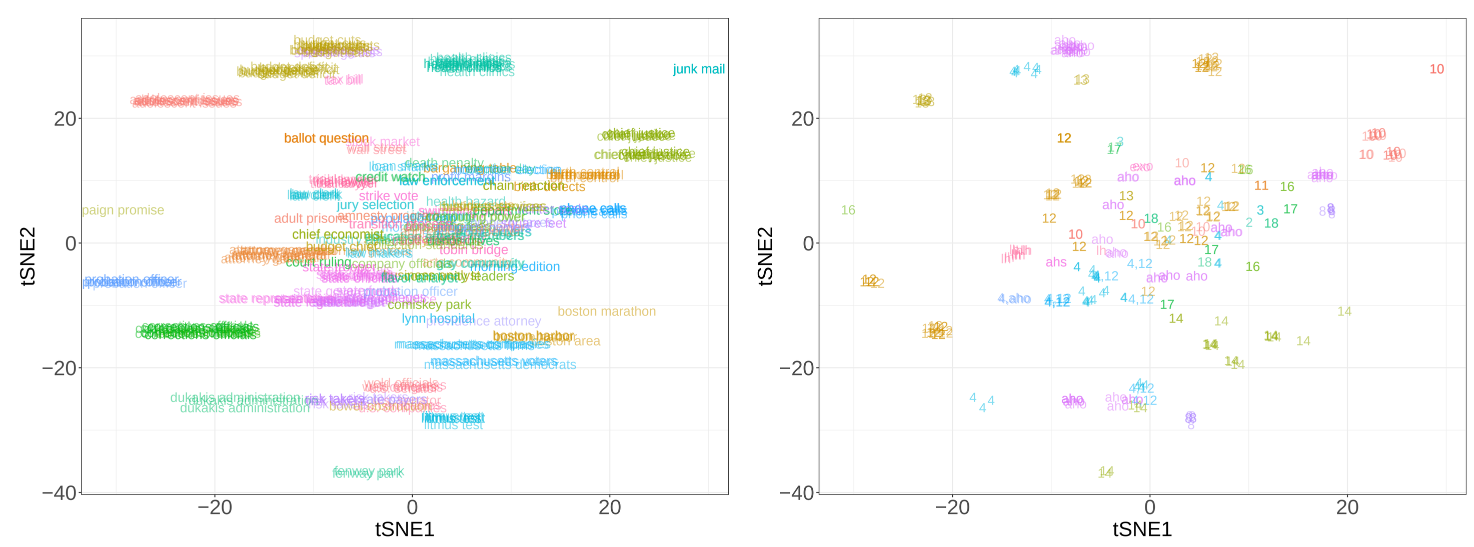

Figure 5: ***Left panel**: t-SNE plot of compound embeddings.* ***Right panel**: t-SNE plot of compound embeddings with their semantic relations (e.g., 4:* N1 HAS N2*, 14: N2 is located at N1, 12: N2 for N1,16: N2 during N1).*

Results

Accuracy

'correct': Nearest neighbor of predicted vector must be real vector of the same compound (type).

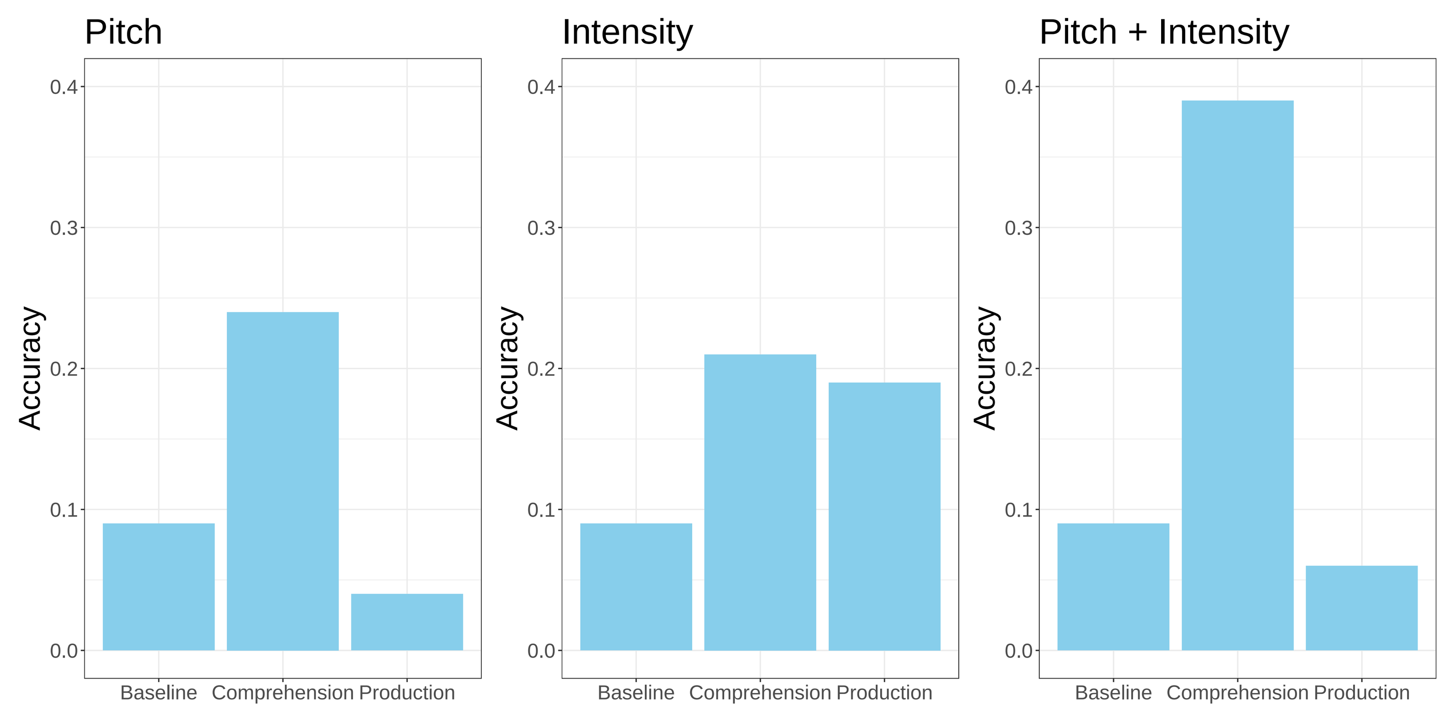

Figure 6: *Accuracies of different models.*

Conclusion

- Comprehension: It is possible to predict compound semantics on the basis of the acoustic signal significantly above chance level.
- Production: Predicting acoustics from semantics does not work as well.
- Pitch/intensity: Both acoustic cues are predictive in comprehension. In the production model, only intensity contours are predictable from the semantics.
- Bottom line: Compound prosody and compound semantics are not randomly mapped onto each other.

Acknowledgement: This research has partly been funded by the Deutsche Forschungsgemeinschaft (Grant PL 151/13-1 to Ingo Plag).

The references and the poster are available via the QR code: