

# C1: The relationship between information rate and speech rate in several European languages

Zofia Malisz<sup>1</sup>, Yoon-Mi Oh<sup>2</sup>, Erika Schulz<sup>1</sup>, Bistra Andreeva<sup>1</sup>, Bernd Möbius<sup>1</sup>

<sup>1</sup>Universität des Saarlandes, Germany

<sup>2</sup>Université de Lyon, France

[sfb-c1@coli.uni-saarland.de](mailto:sfb-c1@coli.uni-saarland.de)

SFB 1102 Inaugural Colloquium



UNIVERSITÄT  
DES  
SAARLANDES

# C1: Information density and the predictability of phonetic structure

---

## Main goals

Investigate and model the relationship between information density and encoding density in phonetics and human speech processing

## Specific questions

- ▶ How can information density and encoding density be conceptualised in speech?
- ▶ Which acoustic phonetic features serve the modulation of encoding density?
- ▶ Which acoustic phonetic features are affected by predictability at different structural levels?
  - ▶ syllable structure (local effects)
  - ▶ collocations (non-local effects)

# Prosodic structure and information density

---

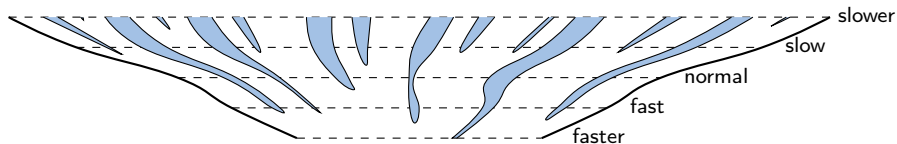
- ▶ Smooth Signal Redundancy Hypothesis (Aylett and Turk, 2004): SSR is achieved by a complementary relationship between language redundancy and acoustic redundancy.
  - ▶ E.g. if a word is **predictable**, phonemes or syllables are elided or acoustic-phonetic features are **reduced** to decrease acoustic signal redundancy.
  - ▶ If a word has **lower predictability**, phonemes or syllables are lengthened and **expansion** of acoustic-phonetic features is observed.

## SSR and prosody

**Prosodic structure** is used to control **acoustic salience** in order to signal **relative predictability** (Aylett and Turk, 2004)

# Speech rate, phrasing and information density

Malisz, O'Dell, Nieminen and Wagner (2015)



- ▶ In long phrases speech rate is higher than in shorter phrases (Quené, 2005, 2008; Yuan et al., 2006)
  - ▶ **Deleting** boundaries results in removing **boundary-associated lengthening**, leading to **higher** articulation rate
  - ▶ **Inserting** boundaries results in **more** boundary-associated lengthening, **slowing down** the articulation rate
- ▶ **Lower predictability** is associated with **insertions** of prosodic boundaries via pausing and boundary lengthening (Turk and Shattuck-Hufnagel, 2014).
- ▶ In conditions of lower predictability one must not be too fast.

# Speech rate and information rate universals

---

- ▶ Pellegrino et al. (2011): average information rate conveyed in different languages results from a trade-off between average information density and speech rate
  - ▶ Languages with an **inherently higher** information density due to linguistic structure (e.g. Mandarin Chinese) are **slower**.
  - ▶ Consistent average information rate at the global level.
  - ▶ Cross-linguistic strategies do not necessarily lead to a constant information rate at the local level.



The interface between prosody and information density offers fascinating insights into how speakers package meaning into the signal and structure it efficiently.

## Background

- 1 Prosodic structure and predictability are not independent
- 2 There are prosodic structure effects independent of information density (Aylett and Turk, 2004)
- 3 Duration is an exponent of both prosodic structure and information density
- 4 There is a possible trade-off between speech rate and information density, from a typological perspective

## Research hypothesis

Information encoding strategies are apparent in all tempos and show similar tendencies in all languages (Pellegrino et al., 2011).

# Material and methods

---

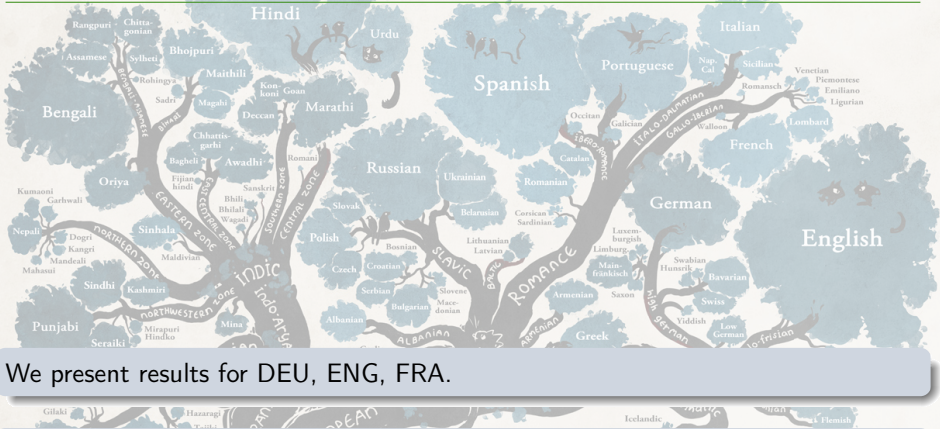
## Study methods

We **correlate** information density measures based on n-gram models with mean segmental duration. We also provide a **preliminary duration model** assessing the joint impact of information density and prosody: speech rate, prominence and boundaries, extracted from a multilingual speech corpus.

## BonnTempo corpus

- ▶ Several European languages in which speech rate variation was elicited from speakers in a reading task (Dellwo and Wagner, 2003; Dellwo et al., 2004).
- ▶ Self-paced in several steps, starting from the preferred 'normal' rate and extending towards the slowest and fastest extremes.
- ▶ The average duration of segments in 6 speakers (3 male and 3 female) for each language was computed

# Bonn Tempo corpus



We present results for DEU, ENG, FRA.

POL, FIN, CZE are in progress, resulting in a comparison of three different IE language subfamilies.

© Minna Sundberg

INDO-  
EUROPEAN

LANGUAGE FAMILIES  
SIZES OF THE BRANCHES REPRESENT THE RECORDED NATIVE SPEAKERS BEFORE YEAR 0.

Samoyedic  
Ezaya  
Komi  
Udmurt

BALTIC  
FINNO-UGRIC  
URALIC

## German original

Am nächsten Tag fuhr ich nach Husum. Es ist eine Fahrt ans Ende der Welt; hinter Gießen werden die Berge und Wälder eintönig, hinter Kassel die Städte ärmlich, und bei Salzgitter wird das Land flach und öde. Wenn bei uns Dissidenten verbannt würden, würden sie ans Steinhuder Meer verbannt.

Bernhard Schlink (1994) *Selbs Betrug*: Diogenes (p. 242)

## English

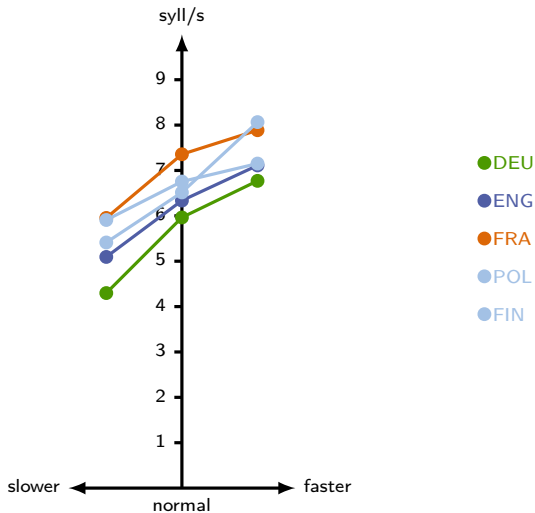
The next day I went to Falmouth. It is a voyage to the end of the world; after Lincoln the hills and woods become monotonous, after Bristol the towns get boring and near Saintsbury the countryside becomes flat and desolate. If dissidents were banned in our country, they would be banned to the Portishead Bay.

## French

Le jour suivant, je me suis rendue à Albi. C'est un voyage au bout du monde. Après Lisieux, les montagnes et la forêt deviennent monotones, après Châtel, les villes désolées, et près de Chartreville, la campagne devient plate et déserte. Si chez nous les dissidents étaient exilés, ils seraient alors exilés à Clermont-Ferrand.

# Speech rate estimation

adapted from (Malisz et al., 2015)



## Text corpora used for computing information density

- ▶ DEU

WebCelex,  $N_{tokens} = 4.6\text{mil.}$ ,  $N_{vow} = 33$ ,  $N_{cons} = 26$

- ▶ ENG

WebCelex,  $N_{tokens} = 17.2\text{mil.}$ ,  $N_{vow} = 24$ ,  $N_{cons} = 25$

- ▶ FRA

LEXIQUE 3.80,  $N_{tokens} = 9.1\text{mil.}$ ,  $N_{vow} = 15$ ,  $N_{cons} = 22$

# Measures of information density

---

## Surprisal definition

$$\text{Surprisal}(\text{Segment} = \vartheta) = -\log[P(\text{Segment} = \vartheta \mid \text{Segment} - 1 = \delta)]$$

# Surprisal models

---

The next day I went to Falmouth. ðə nekst 'deɪ aɪ went tə 'fælməθ

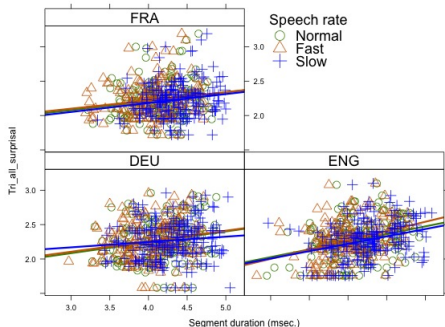
## Trigram models

$X_n | X_{n-1}, X_{n+1}$ : surprisal for [ðə] and [ðə] compared to  $\tilde{\theta}_{dur}$  and  $\theta_{dur}$

## Word class

- ▶ Open class words: related to content words
- ▶ All classes: open + closed (content and function words)

# Results: correlations



## Normal

- ▶ FRA  
 $r = 0.26, p < 0.01$
- ▶ DEU  
 $r = 0.23, p < 0.001$
- ▶ ENG  
 $r = 0.27, p < 0.001$

## Fast

- ▶ FRA  
 $r = 0.15, p < 0.01$
- ▶ DEU  
 $r = 0.21, p < 0.001$
- ▶ ENG  
 $r = 0.31, p < 0.001$

## Slow

- ▶ FRA  
 $r = 0.15, p < 0.01$
- ▶ DEU  
 $r = 0.15, p < 0.01$
- ▶ ENG  
 $r = 0.26, p < 0.001$

# Results: duration model (LMM)

---

## Dependent variable

Average (speaker normalised) segment duration in  $\log(\text{ms})$

## Prosodic predictors

- ▶ stress: segments that belong to a stressed syllable
- ▶ boundary: segments that belong to a syllable before a boundary
- ▶ intended speech rate: normal, slow, fast

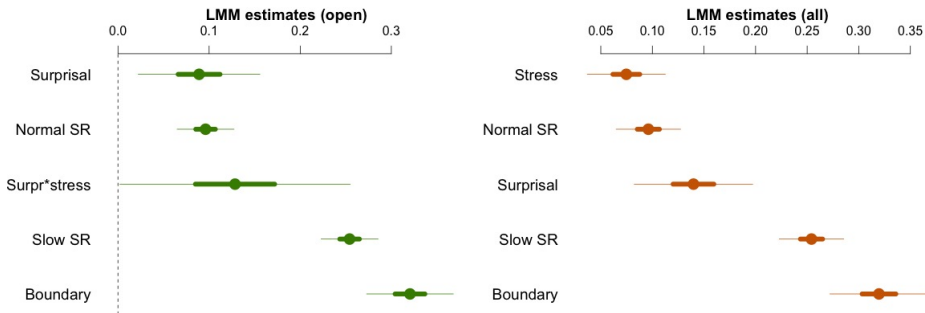
## Information density predictors

- ▶  $\log(\text{surprisal})$  based on a trigram, open word class model
- ▶  $\log(\text{surprisal})$  based on a trigram, all word classes model

## Random factors

- ▶ Segment identity (SAMPA labels)
- ▶ Language (DEU, ENG, FRA)

# Duration LMM: trigrams



- ▶ Only significant effects on  $\log(S_{dur})$  are presented.

# Conclusions

---

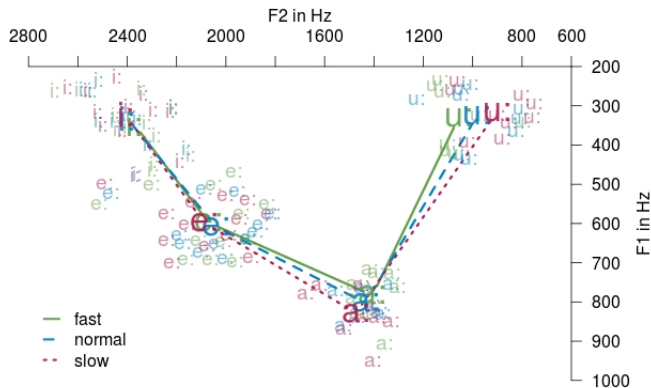
## Results summary

- ▶ Consistent positive correlation for surprisal in trigram, all word classes model.
  - ▶ The higher the surprisal, the longer the duration in all languages and speech rates
- ▶ Expected effects of prosodic factors on duration in all models
- ▶ In the open class, trigram model, we find a significant interaction between surprisal and lexical stress (Aylett, 2006)
- ▶ Surprisal also affects the mean duration of segments in the all classes, trigram model
- ▶ Multiple model comparisons showed that trigram models (both open class- and all classes-based surprisal) have a superior fit to bigram models.

## Vowel reduction and information density

Investigate the relationship between vowel reduction and information density by means of calculating the Euclidean distance of the vowel space centroid and formant values for each vowel

# Future work



Euclidean distance correlated with surprisal

- ▶ FRA  $r=0.26$ ,  $p<0.001$
- ▶ DEU  $r=0.21$ ,  $p<0.001$
- ▶ ENG  $r=0.13$ ,  $p<0.001$

# Acknowledgements

---

This research was funded by the German Research Foundation (DFG) as part of SFB 1102 'Information Density and Linguistic Encoding' at the Universität des Saarlandes



UNIVERSITÄT  
DES  
SAARLANDES

- Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1):31–56.
- Dellwo, V., Steiner, I., Aschenberner, B., Dankovicova, J., and Wagner, P. (2004). Bonntempo-corpus & bonntempo-tools: A database for the study of speech rhythm and rate. In *Proceedings of Interspeech 2004*.
- Dellwo, V. and Wagner, P. (2003). Relations between language rhythm and speech rate. In *15th International Congress of Phonetic Sciences*, pages 471–474, Barcelona.
- Malisz, Z., O'Dell, M. L., Nieminen, T., and Wagner, P. (2015). Assessing speech tempo and its influence on rhythm. (presented at the XXIX Fonetikan päivät, Espoo, Finland).
- Pellegrino, F., Coupé, C., and Marsico, E. (2011). Across-language perspective on speech information rate. *Language*, 87(3):539–558.
- Quené, H. (2005). Modeling of between-speaker and within-speaker variation in spontaneous speech tempo. In *Ninth European Conference on Speech Communication and Technology*.
- Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America*, 123(2):1104–1113.
- Turk, A. and Shattuck-Hufnagel, S. (2014). Timing in talking: what is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658).
- Yuan, J., Liberman, M., and Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *INTERSPEECH*.