MORE THAN WORDS

A DISCRIMINATIVE LEARNING MODEL WITH LEXICAL BUNDLES

March 8th, 2017

Saskia E. Lensink, R. Harald Baayen
s.e.lensink@hum.leidenuniv.nl

Contents

- Multi-word units and their cognitive reality
- Experimental methods
- Computational model of multi-word units
- Eye-tracking study
- Production study
- Results and implications

A typology of multi-word units

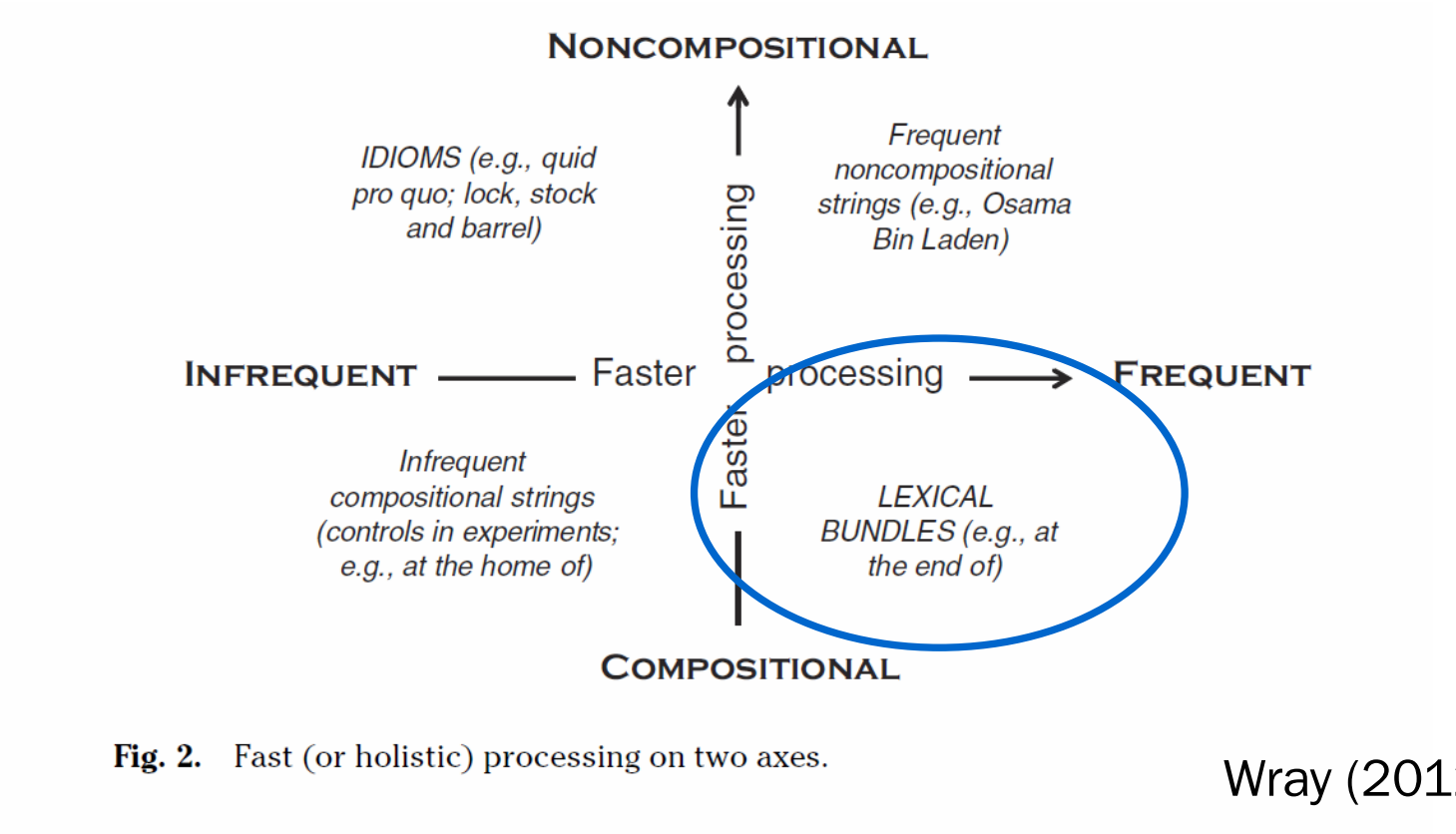


Fig. 2. Fast (or holistic) processing on two axes.

Wray (2012)

Multi-word units

- Indicator of nativeness
- Thought to be represented as a whole
- How can we experimentally test for the cognitive reality of these multi-word units?

Multi-word frequencies

Previous studies have found an effect of frequencies of regular multi-word units

suggests storage of wholes

Previous studies

- self-paced reading — Tremblay, Derwing, Libben, & Westbury, 2011
- phrasal decision tasks — Arnon & Snider, 2010; Ellis & Simpson-Vlach, 2009
- priming of the last word of the ngram — Ellis & Simpson-Vlach, 2009
- word reading tasks — Arnon & Priva, 2013; Ellis & Simpson-Vlach, 2009; Han, 2015; Tremblay & Tucker, 2011
- picture naming — Janssen & Barber, 2012
- sentence recall — Tremblay et al., 2011
- immediate free recall — Tremblay & Baayen, 2010
- eye-tracking — Siyanova-Chanturia, Conklin, & Van Heuven, 2011
- ERPs — Tremblay & Baayen 2010
- L1 language acquisition — Bannard & Matthews, 2008
- L2 speakers — Conklin & Schmitt, 2012; Han, 2015; Jiang & Nekrasova, 2007; Siyanova-Chanturia et al, 2011

Frequency is an impoverished measure

- Collapses counts of homophones
- Collapses counts of different senses
- Language always occurs in context – prediction also plays a large role in processing
- Salience and recency also play a role

Mind the neighbors!

- When studying words, we pay attention to
  - Frequency effects
  - Length
  - Neighborhood density effects
- When studying multi-word units, we pay attention to
  - Frequency effects
  - Length
  - But not to neighborhood density effects!

Motivation for our study

- We know that the framework of discriminative learning has given us some new insights into language
- A computational model implementing discriminative learning, NDL, provides us with a measure reflecting neighborhood density effects
- When adding features of discriminative learning to our models of the processing of multi-word units, we might gain new insights into the processing of multi-word units
- We conducted both an eye-tracking and a production study to study comprehension and production

NDL

Baayen et al., 2011

- Naïve Discriminative Learning
- Implements Rescorla-Wagner equations that specify how experience alters the strength of association of a **cue** to a given **outcome**
- Distributional properties of corpus data used, using basic principles of **error-driven learning**
- Weight from cues to outcomes **adjusted** depending on **correct/incorrect prediction** of an outcome given a certain cue

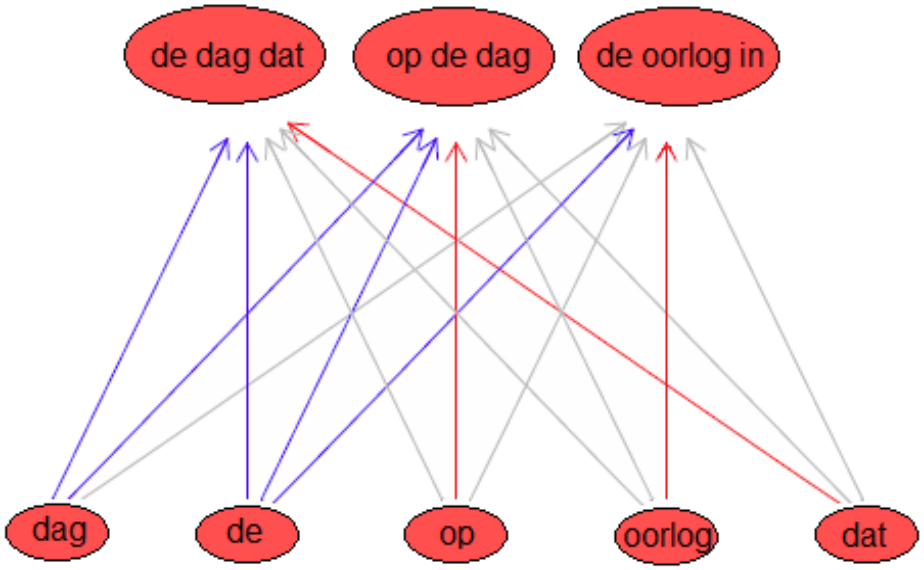
  This approach successfully predicted word frequency effects, morphological family size effects, inflectional entropy effects, and phrasal frequency effects

NDL

Baayen et al., 2011

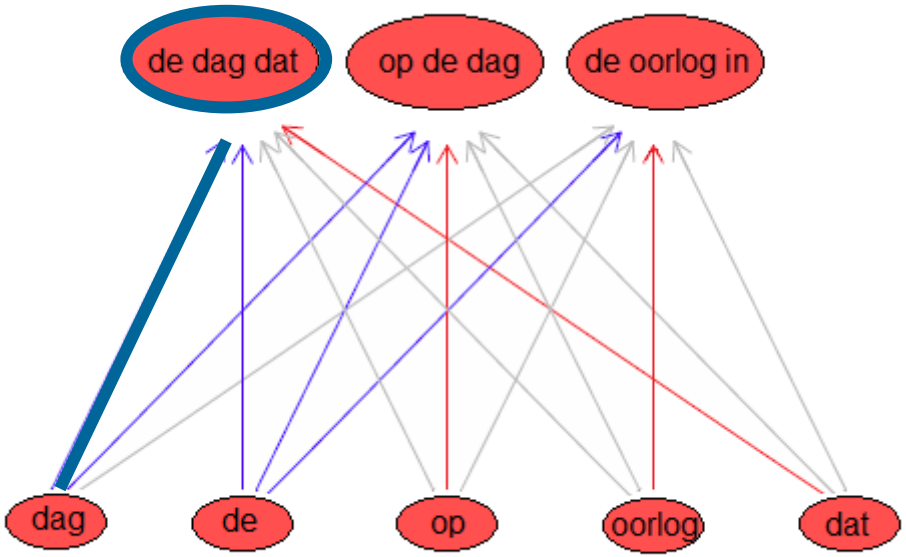
- Outcomes are thought of as pointers to locations in a multi-dimensional semantic space
- These locations are constantly updated by the experiences a language user has

NDL with lexical bundles



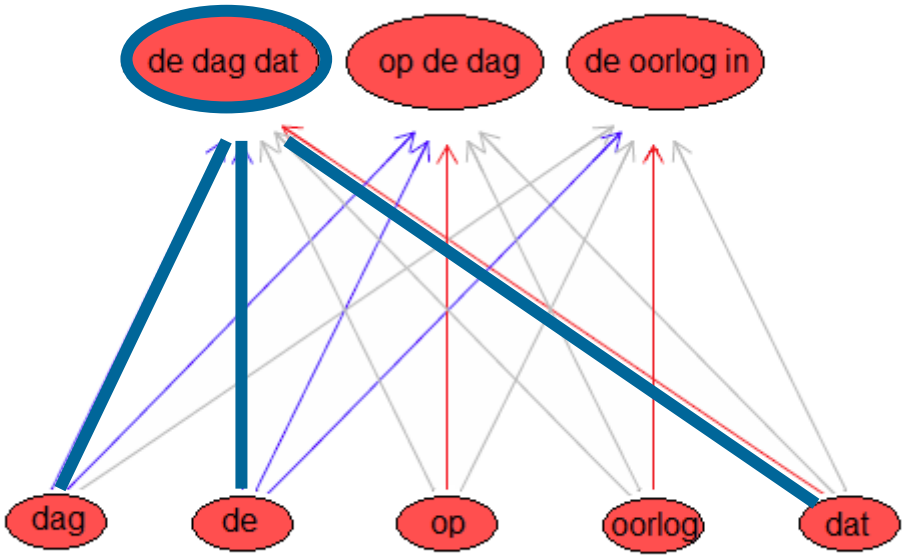
Weight word X

Bottom-up information



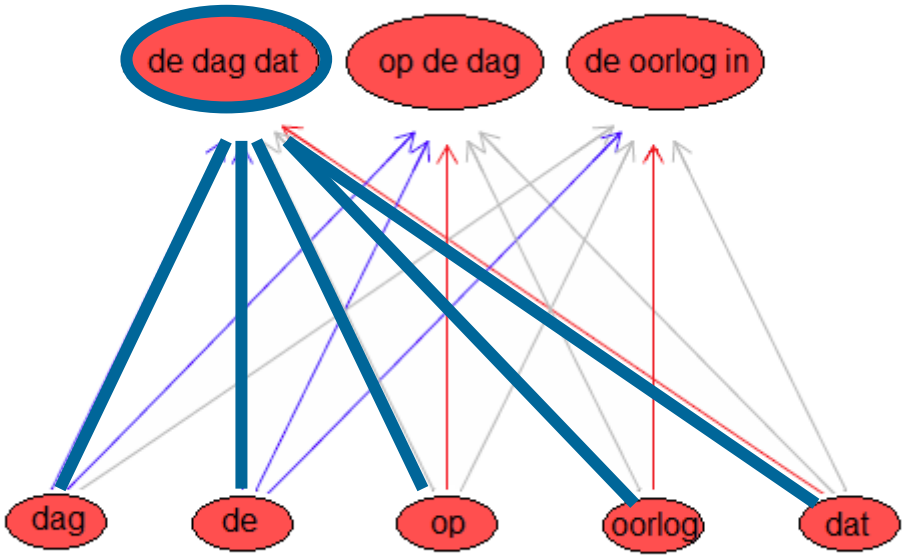
Total activation trigram (act)

Bottom-up information



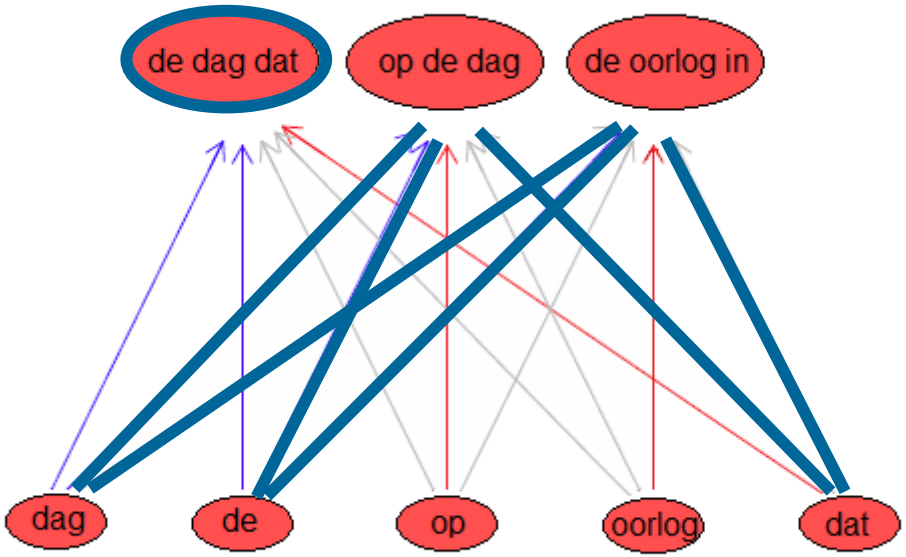
Prior activation trigram

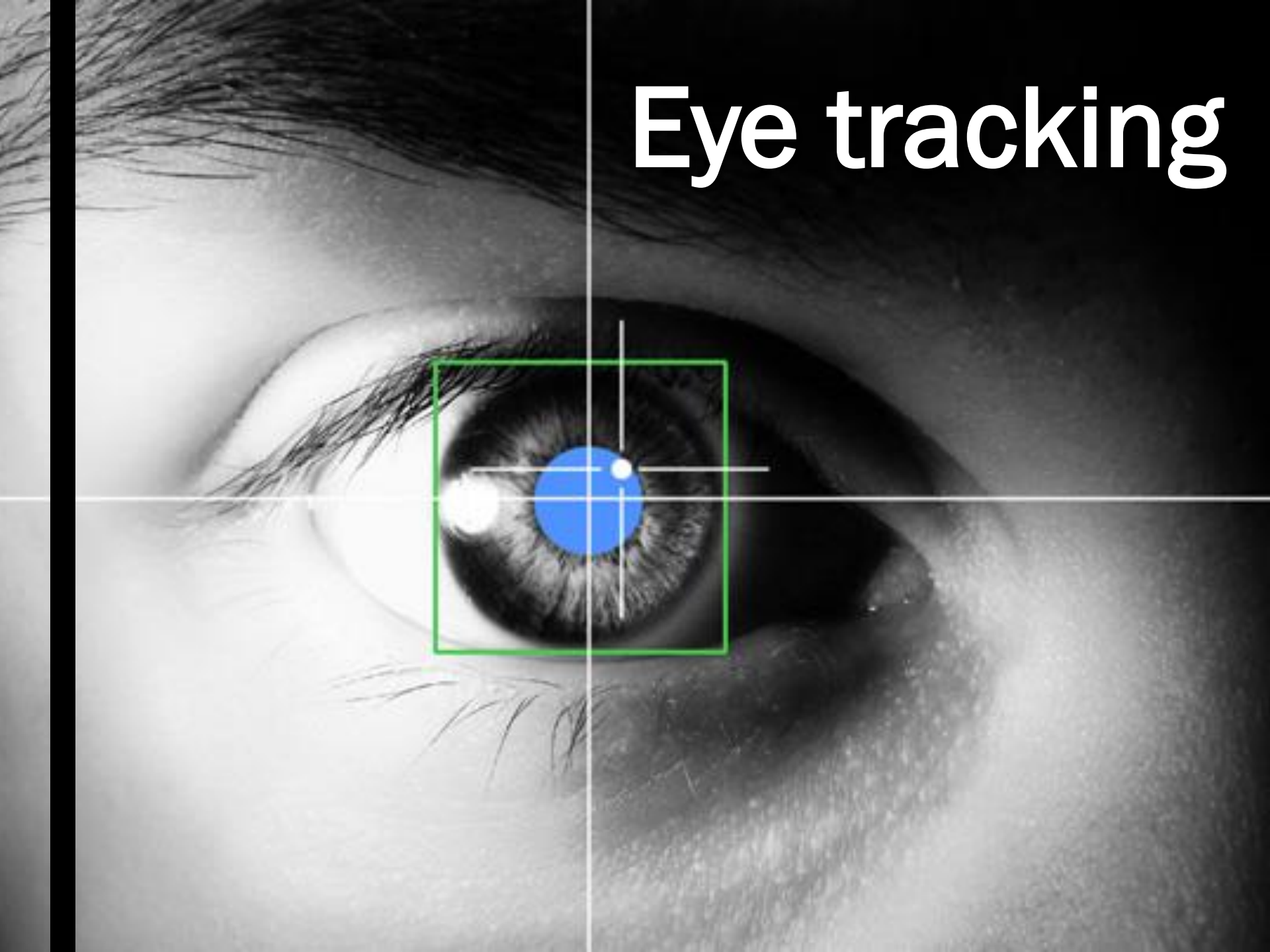
Top-down information



Activation diversity

Competing trigrams – neighborhood density



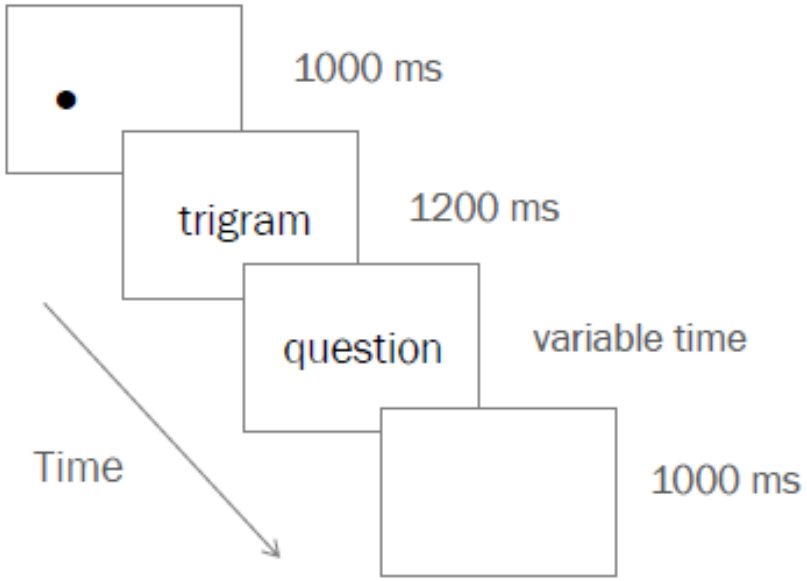

Eye tracking

Stimuli

- most common n-grams (trigrams) from corpus
- OpenSoNaR corpus
- Use frequencies extracted from a corpus of Dutch subtitles (N = 109,807,716)

Procedure

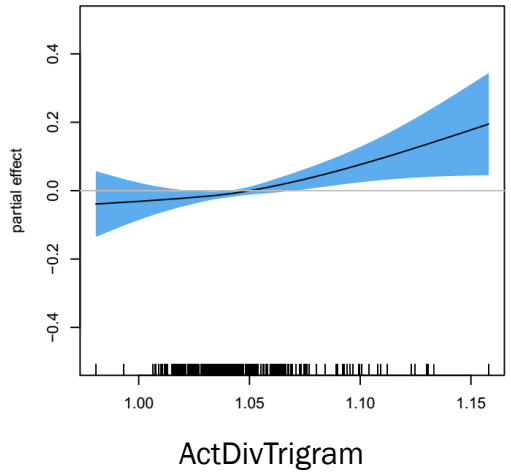
- Silent reading
- Comprehension questions to ascertain attentive reading
- 30 participants (10 male)
- Analyzed using generalized additive mixed-effects models (GAMMS)

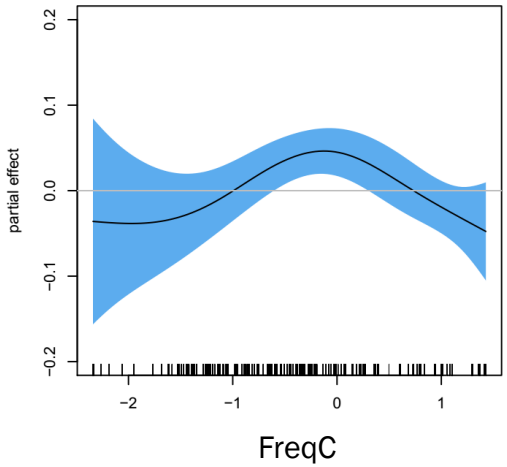


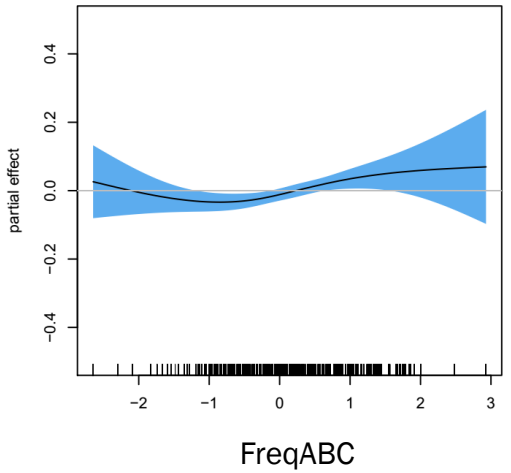
Modeling data

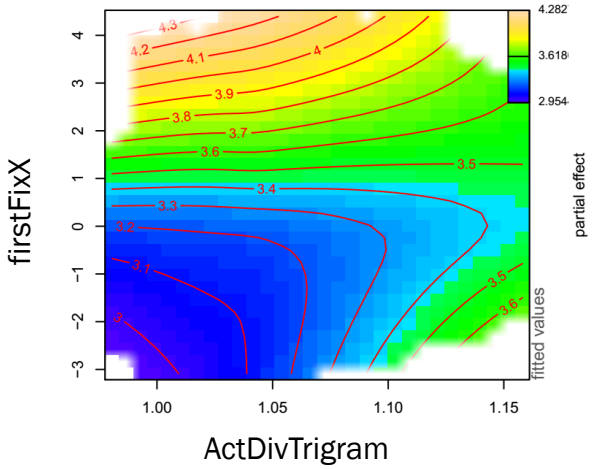
- See if and to what extent NDL measures gives us more insights over and above more traditional frequency measures
- Some frequency and NDL measures show high amount of collinearity – e.g. ‘freqABC’ and ‘prior’
- Models with just frequencies performed worse than models with both frequencies and NDL measures
- Neighborhood density effects are best reflected by the Activation Diversity measure, which was a significant predictor in several models

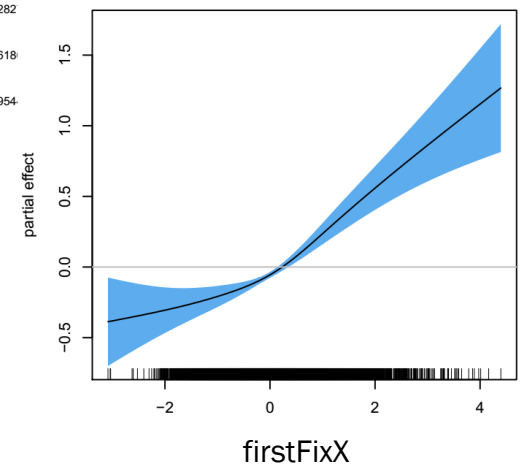
First fixation durations

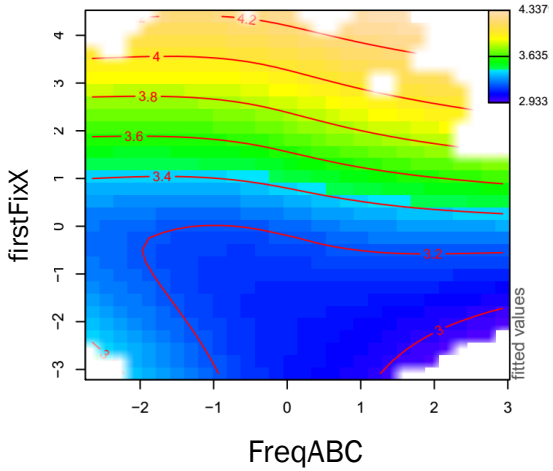




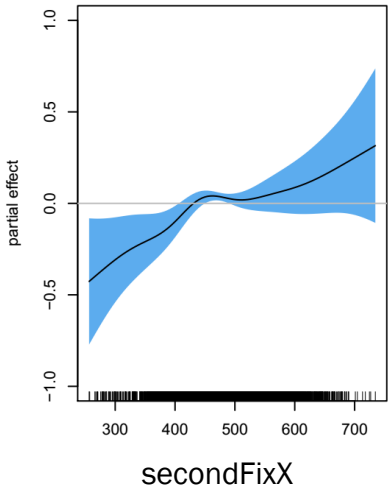


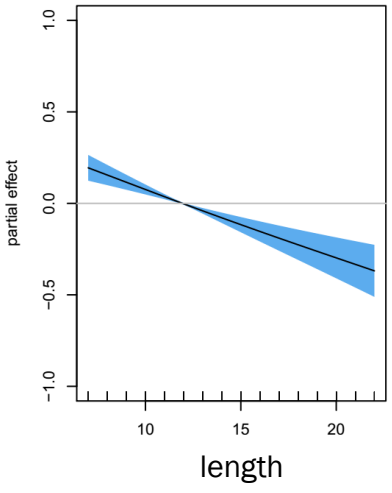


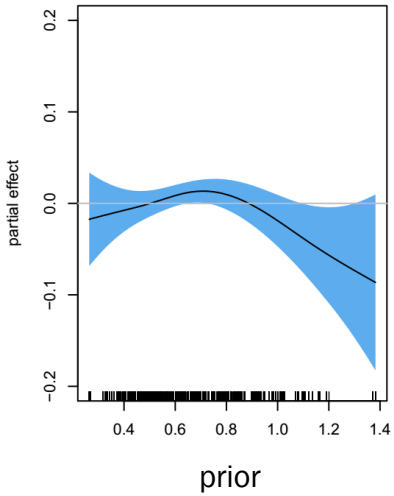


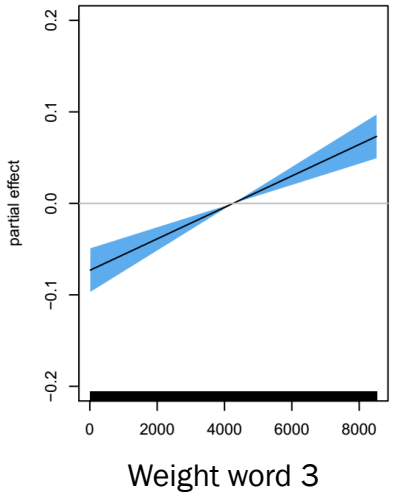


Second fixation durations

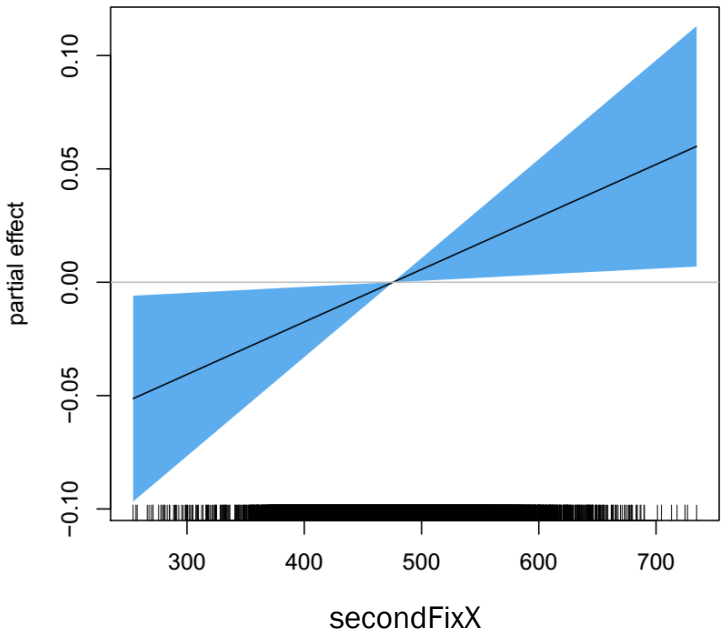


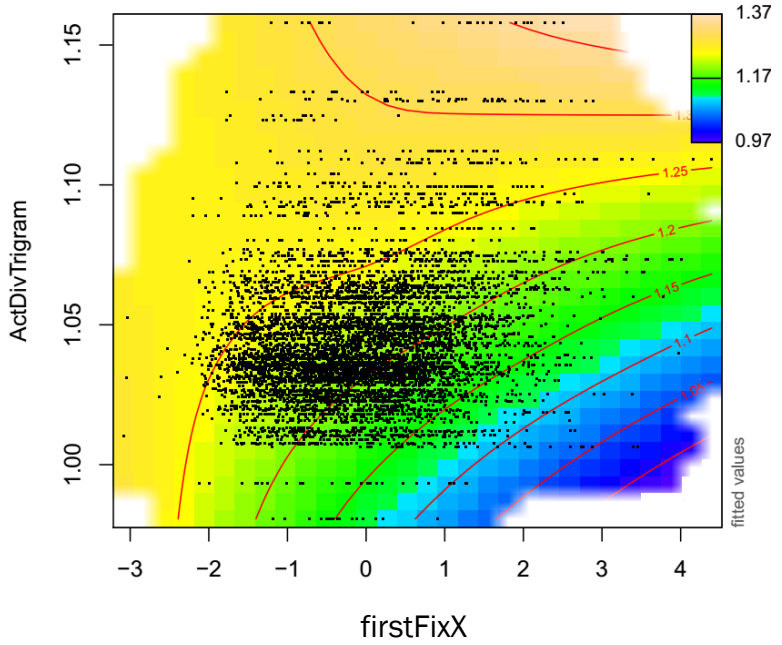






Number of fixations





Discussion eye-tracking data

- Already in the first fixation effects of the trigram frequencies and third word
- Processes of top down information (frequency effects), bottom-up information (activations) and **uncertainty reduction** (**activation diversity/neighborhood effects**)
- Knowledge verification (frequencies): a reader spends more time in early measures with higher frequencies and if enough information is available – if not, a new fixation is planned asap
- Bottom-up information (w3): when further into the trigram at your second fixation, it pays to spend more time to resolve things locally if the third word provides a lot of support for the trigram. If not, participants are faster to refixate
- **uncertainty reduction (neighborhood density):** if there are many competing trigrams, shorter looking times in first fixations and a higher number of fixations.

General discussion

- Multi-word units are relevant unit of storage (also in Dutch)
- Both single words and the full trigram play a role
- Adding measures from a discriminative model provides us with new insights into the processing of MWUs
- Considering neighborhood density effects provides us with more insights into the workings of MWU processing
- In processing of multi-word units, opposing forces of top-down information, bottom-up information and uncertainty reduction are at work

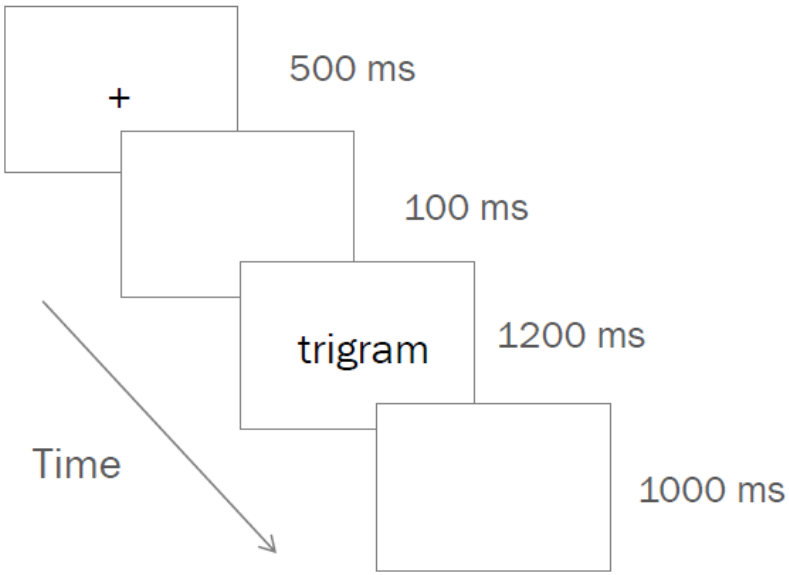
Questions?

Extra slides – production

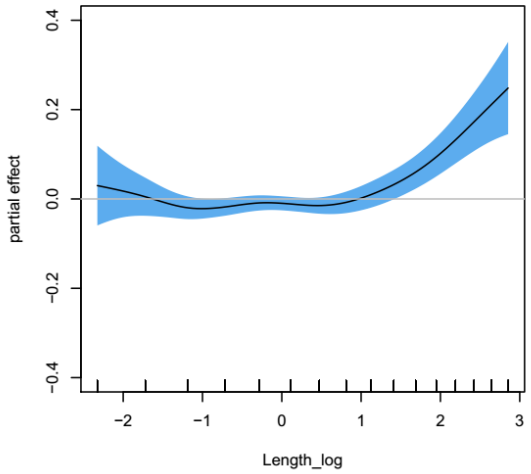
Production experiments

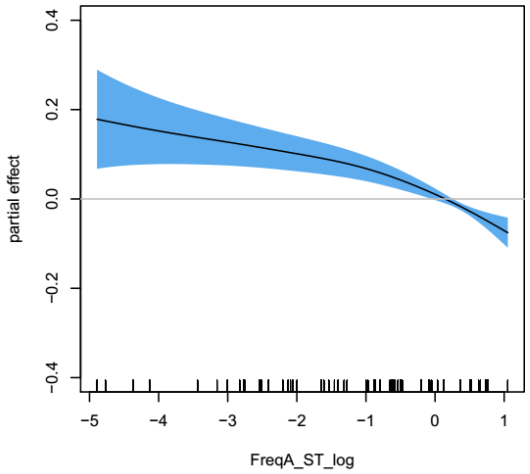
Procedure

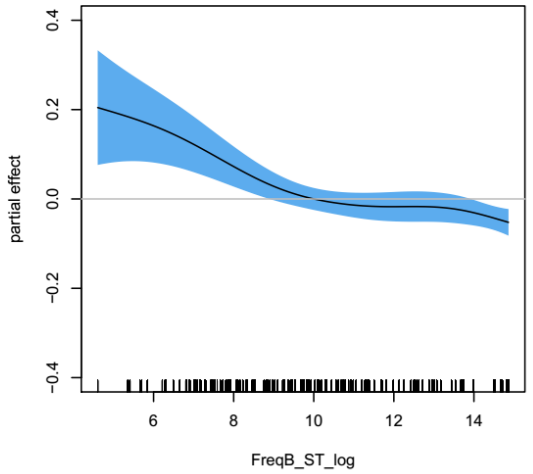
- Same stimuli as used in the eye-tracking study
- Word reading task
- 30 participants (8 male)
- Onsets and durations measured using Praat
- Analyzed using generalized additive mixed effect models (GAMMs)

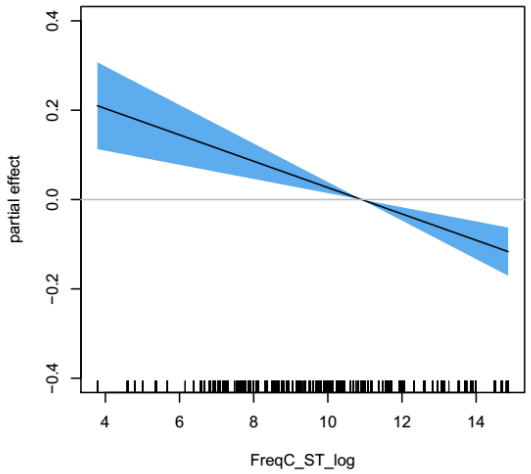


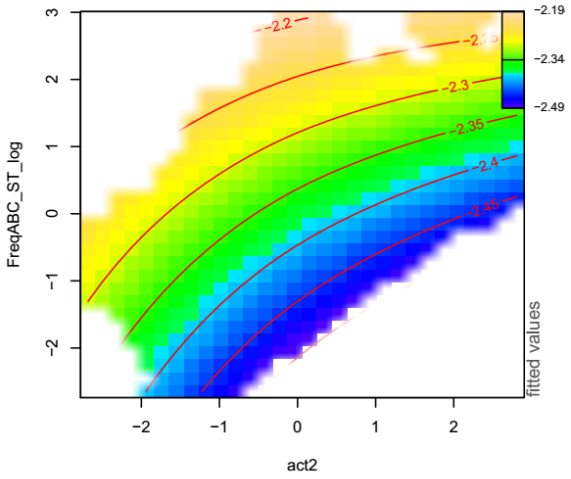
Production onsets



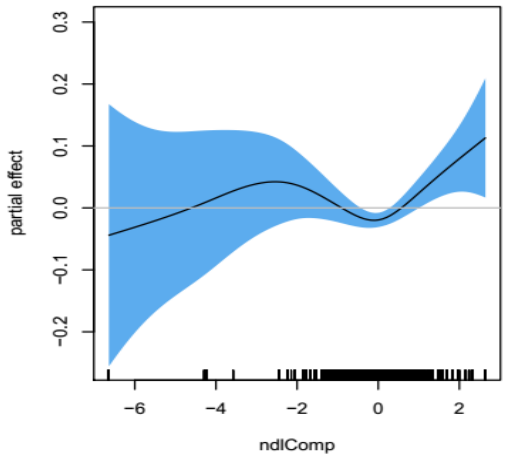


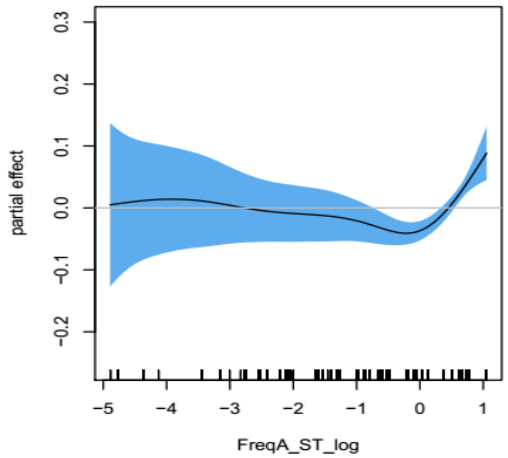


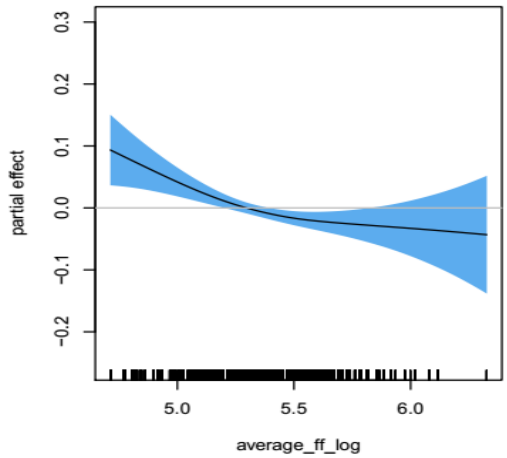


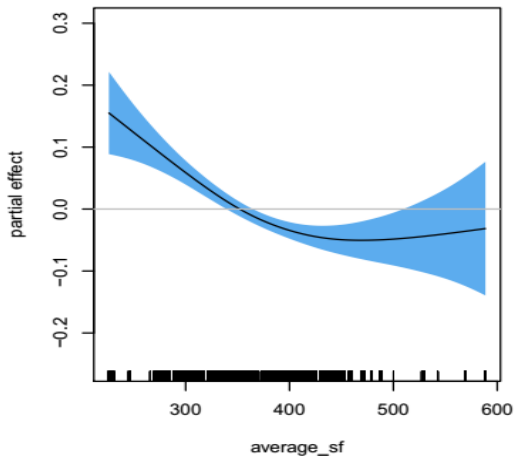


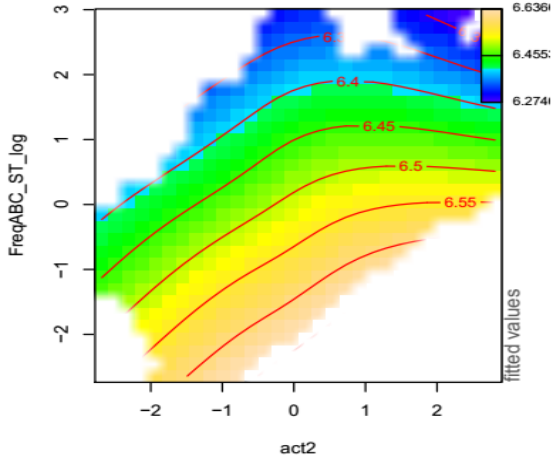
Production durations





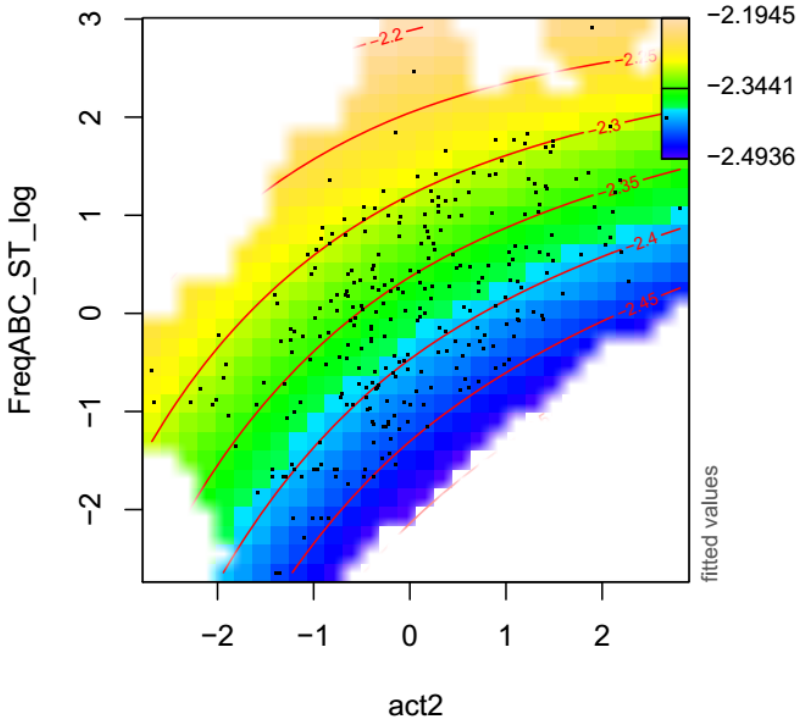




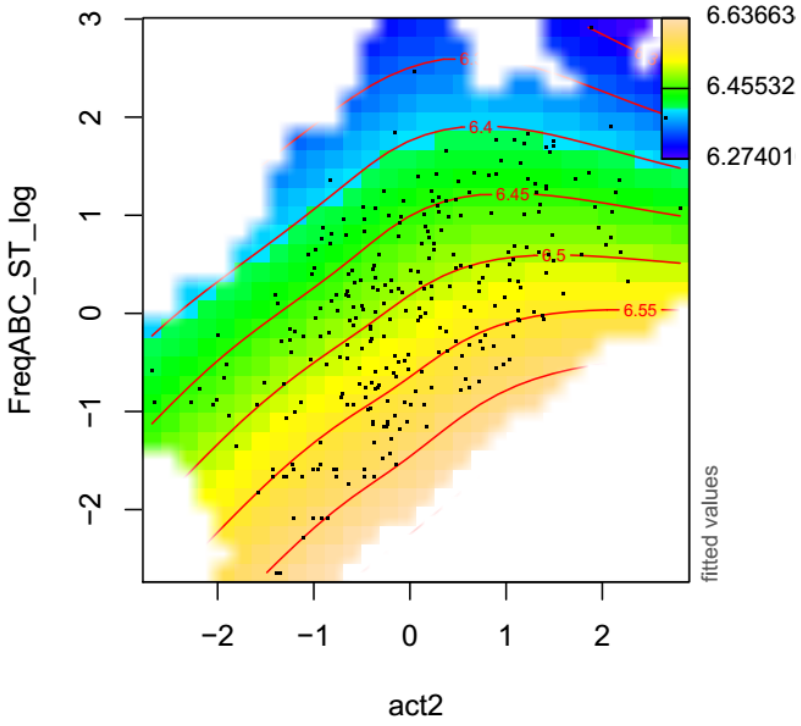


A trade-off

naming latencies



durations



Discussion production data

- Processes of top down information (frequency effects), bottom-up information (activations) and **uncertainty reduction** (**activation diversity/neighborhood effects**)
- There is a trade-off between starting early and being able to pronounce the trigram fast
- Top-down information slows you down at first, but makes total durations shorter (longer to plan, but easier motor program to execute)
- Bottom-up information gives you a quick start but slows you down later (shorter to plan, but harder motor program to execute)
- **Neighborhood effects** apparent in **production durations** – longer durations when the number of neighbors is different from the average (less motor practice)