



# Any questions?

- Practicalities?
- Any open issues from yesterday?

# Lecture 3: *prediction*

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say -thee uh- you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 914–930. doi:[10.1037/0278-7393.33.5.914](https://doi.org/10.1037/0278-7393.33.5.914)

**Hans Rutger Bosker**

Speech Perception in Audiovisual Communication [SPEAC] lab

*Donders Institute, Radboud University, Nijmegen, The Netherlands*

<https://hrbosker.github.io>

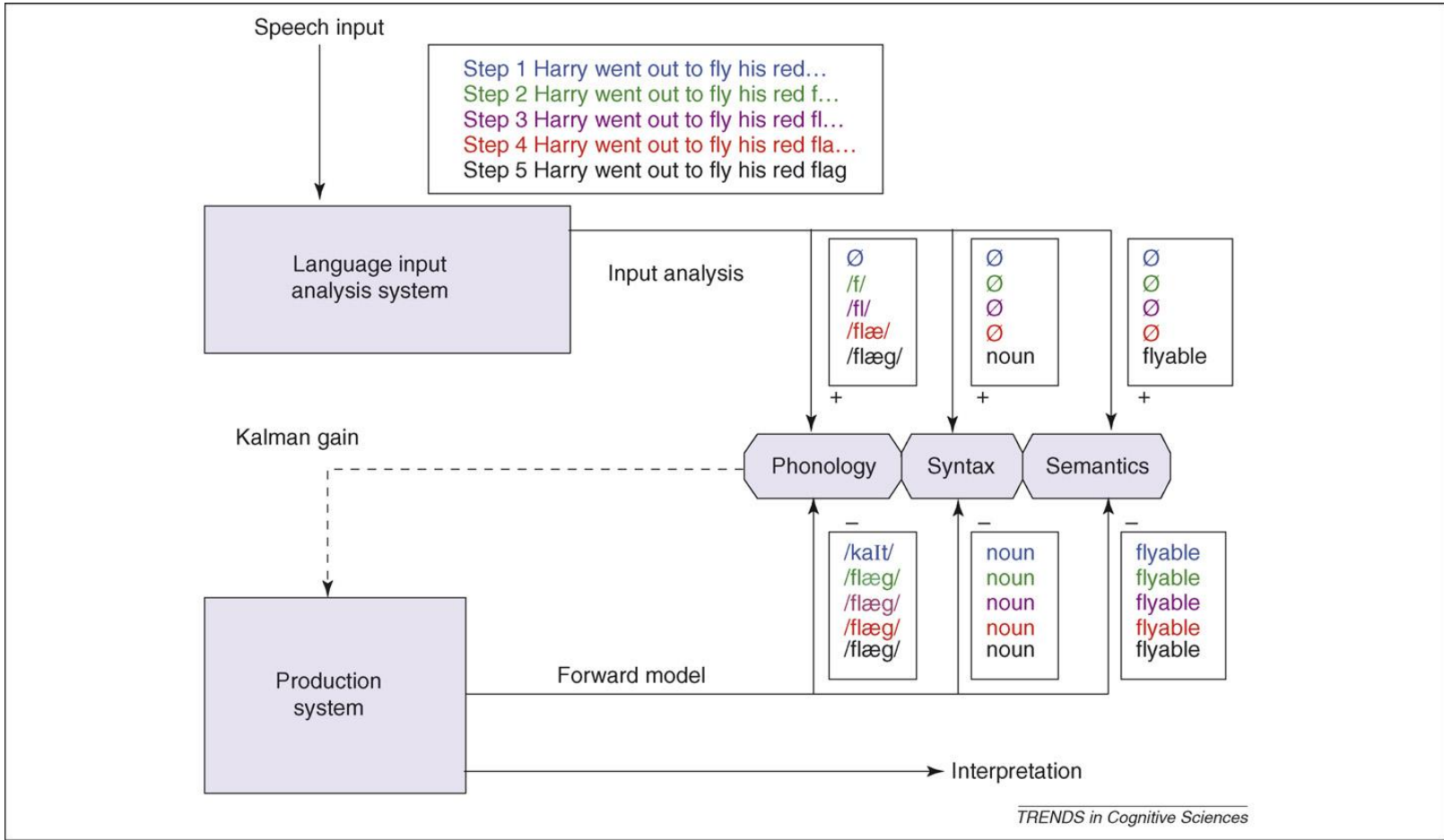
[hansrutger.bosker@donders.ru.nl](mailto:hansrutger.bosker@donders.ru.nl)





# Why prediction?

- Predictive Processing (e.g., Andy Clark)
  - ‘the human brain is a prediction machine’  
Clark, A. (2015) “Surfing Uncertainty: Prediction, Action, and the Embodied Mind”
- Predictive Coding (e.g., Karl Friston)
  - minimizing prediction error  
Friston, K. (2010, *Nature Rev Neuroscience*)





# Why prediction?

- Bayesian inference
  - Priors and posteriors
  - Speech perception:
    - Kleinschmidt & Jaeger (2015)
    - Norris et al. (2016)
    - McQueen & Dilley (2021)
    - Kurumada & Roettger (2021)



# Why prediction?

- If prosody could serve as a source for predictive processes, this has the potential to speed up processing, explaining comprehension efficiency and robustness, and fast turn-taking in conversation.
- Could reveal a critical role for prosody in speech comprehension

What does prosody predict?



# What does prosody predict?

- ...where words begin and end [speech segmentation]
- ...syntactic boundaries [syntactic structure]
- ...when a talker will finish speaking [turn-taking]
- ...when important information will arrive [information structure]



# Predicting speech segmentation

- Prosodic rhythm ‘sets the stage’ for following words
  - distal rhythm effect



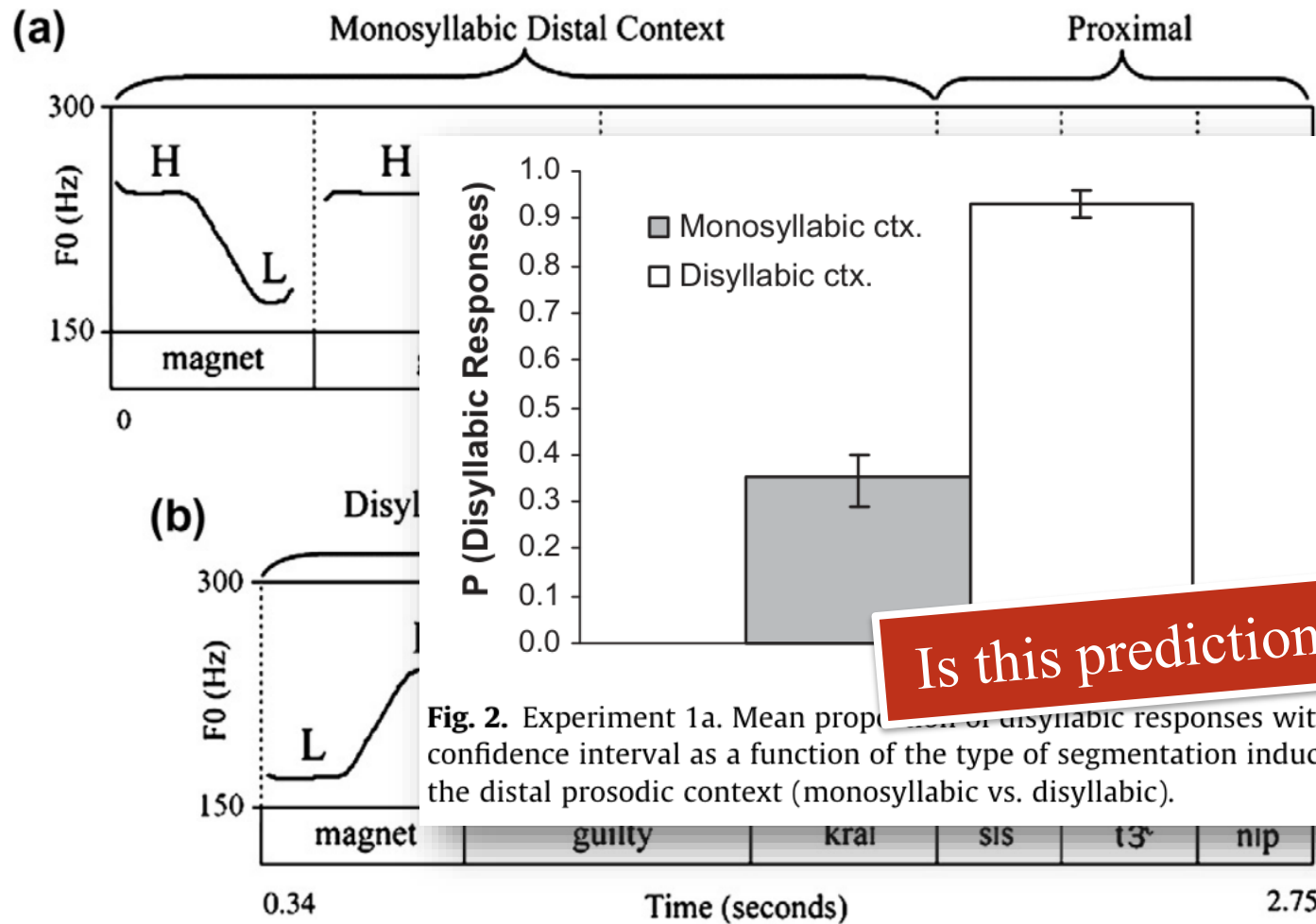
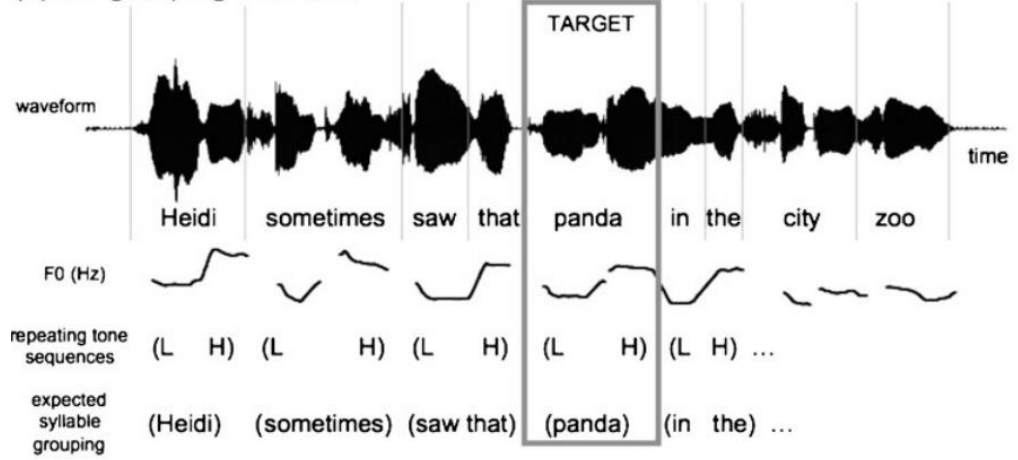


Fig. 2. Experiment 1a. Mean proportion of disyllabic responses with 95% confidence interval as a function of the type of segmentation induced by the distal prosodic context (monosyllabic vs. disyllabic).

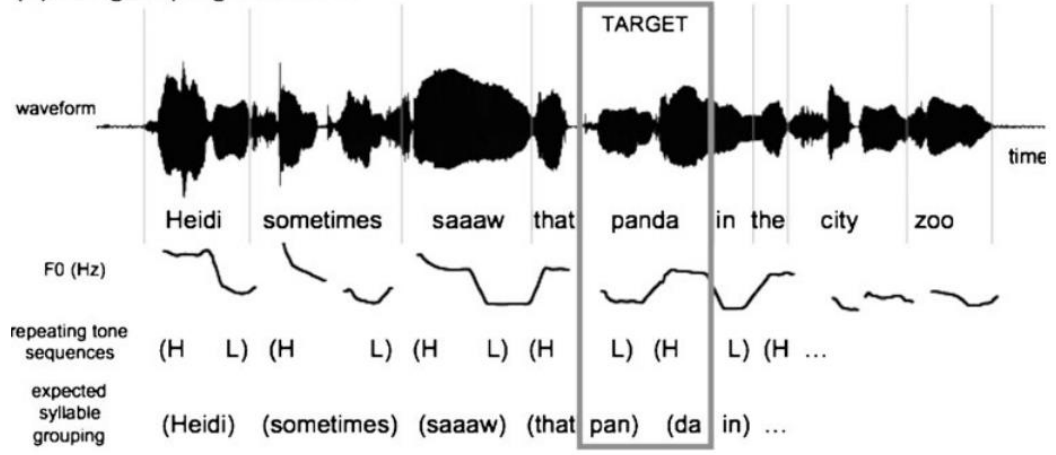
Is this prediction?



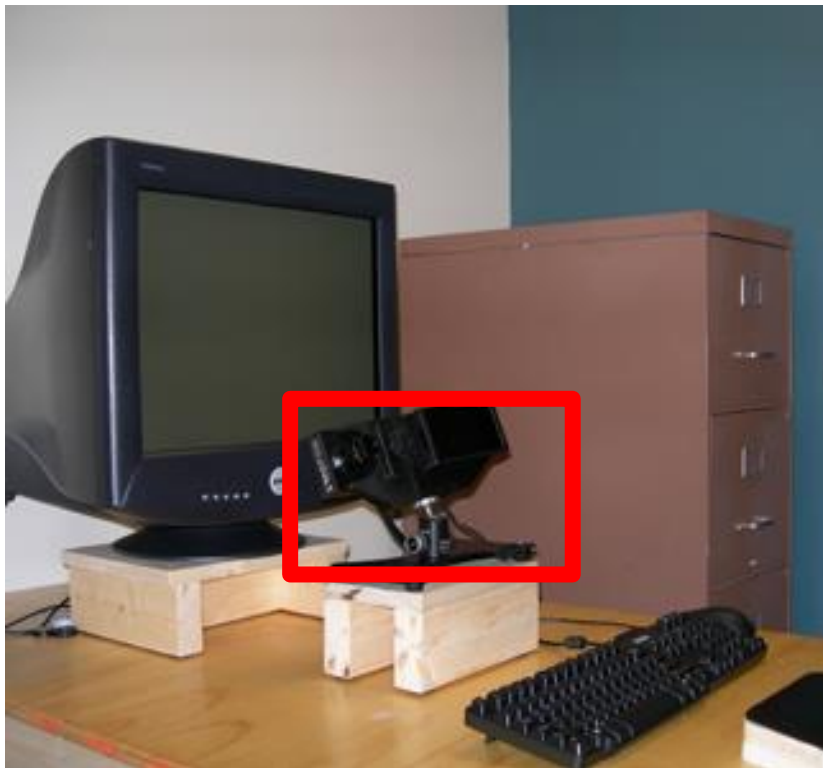
(a) LH-grouping condition



(b) HL-grouping condition



# Visual World Paradigm in eye-tracking





**SPEAC** 



*“Klik op de*



TARGET



COMPETITOR



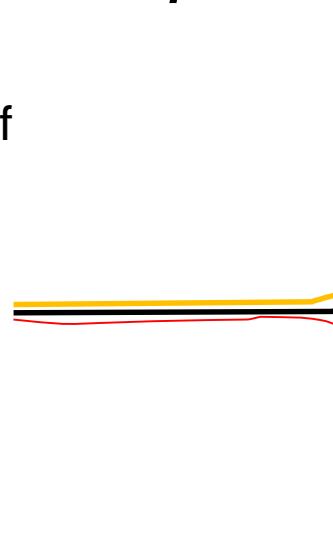
DISTRACTORS





*“Klik op de*

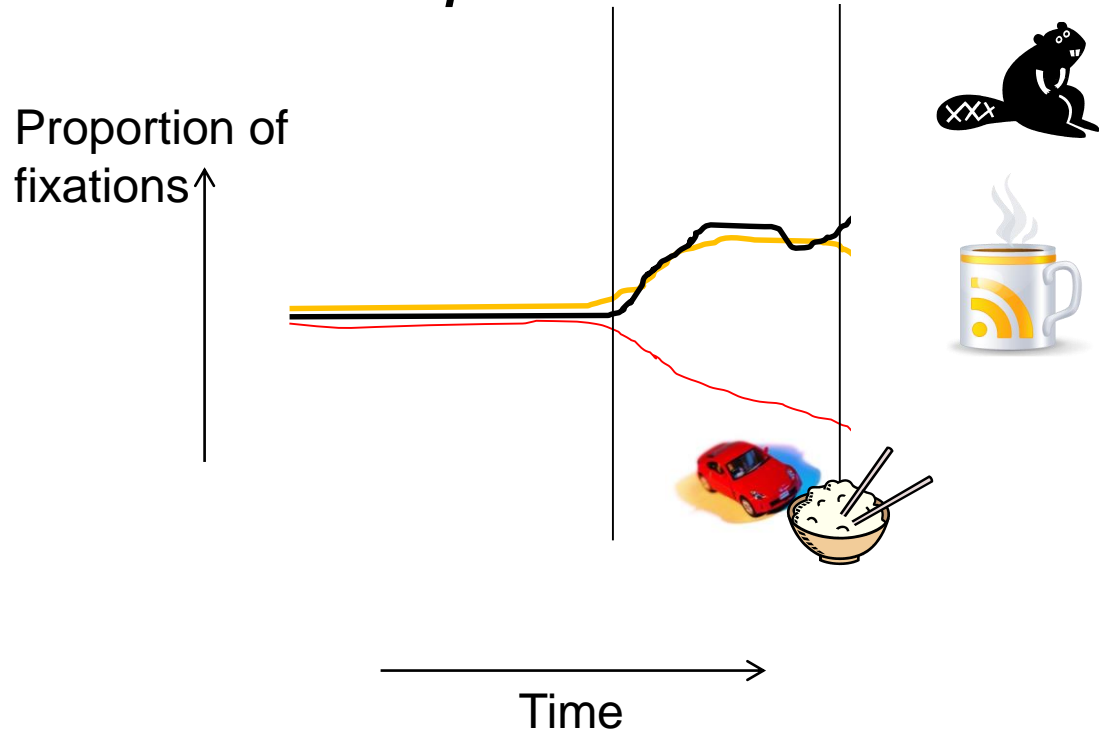
Proportion of fixations ↑



Time →



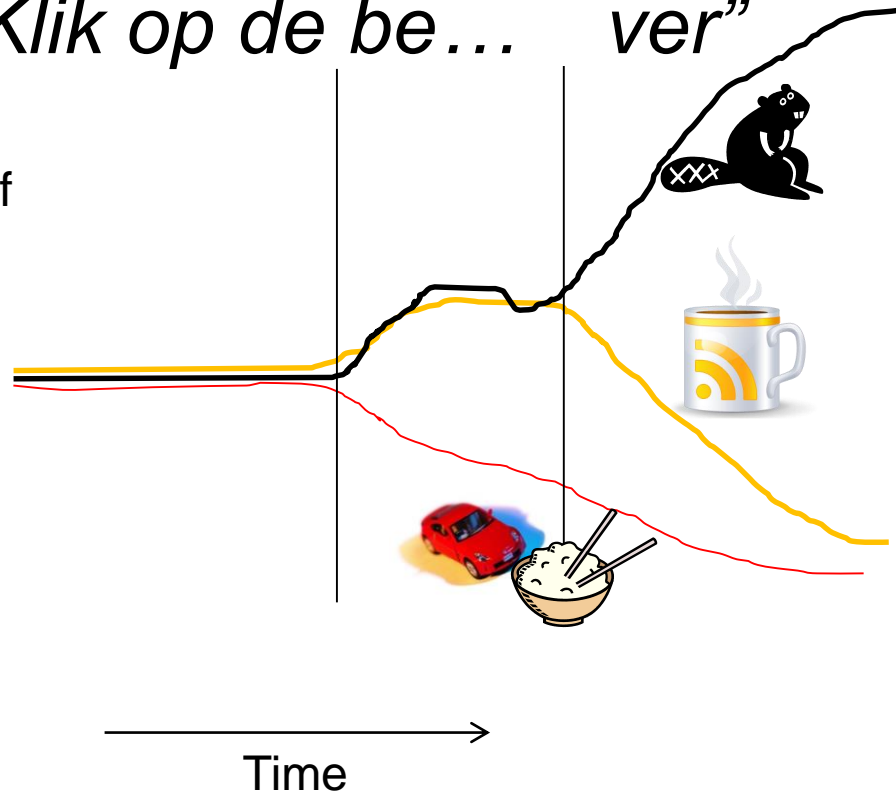
*“Klik op de be...*





“Klik op de be... ver”

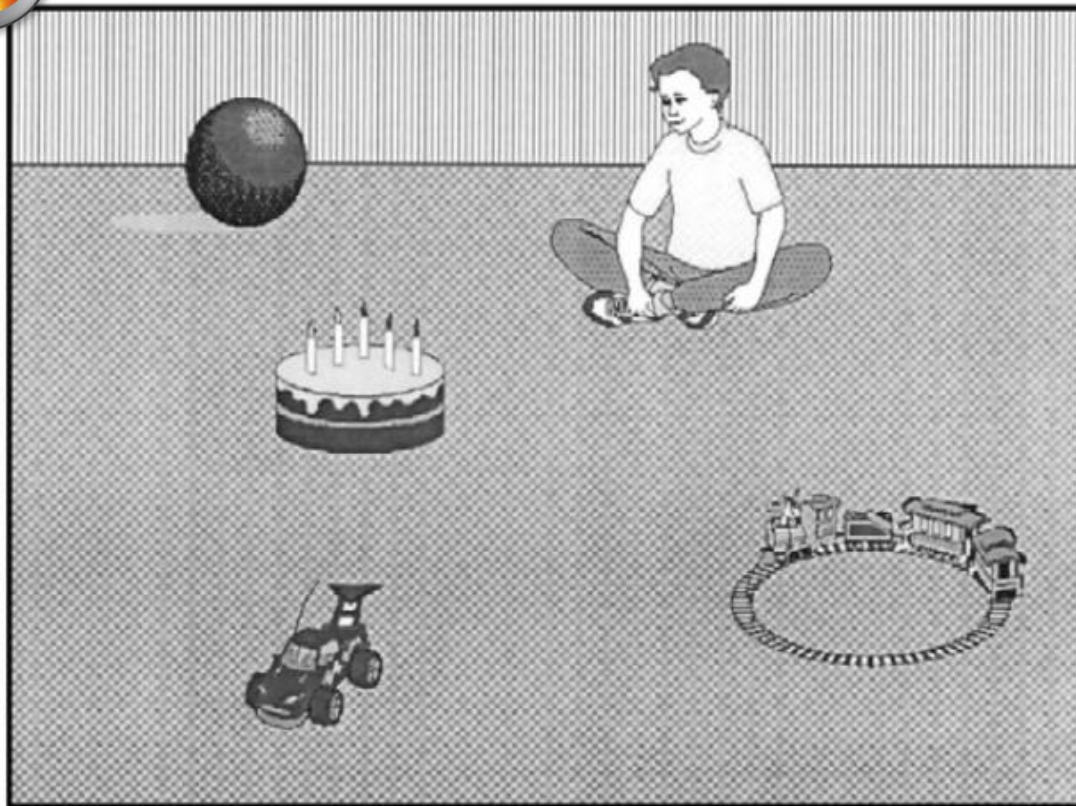
Proportion of fixations ↑







“De jongen eet / pakt de cake”



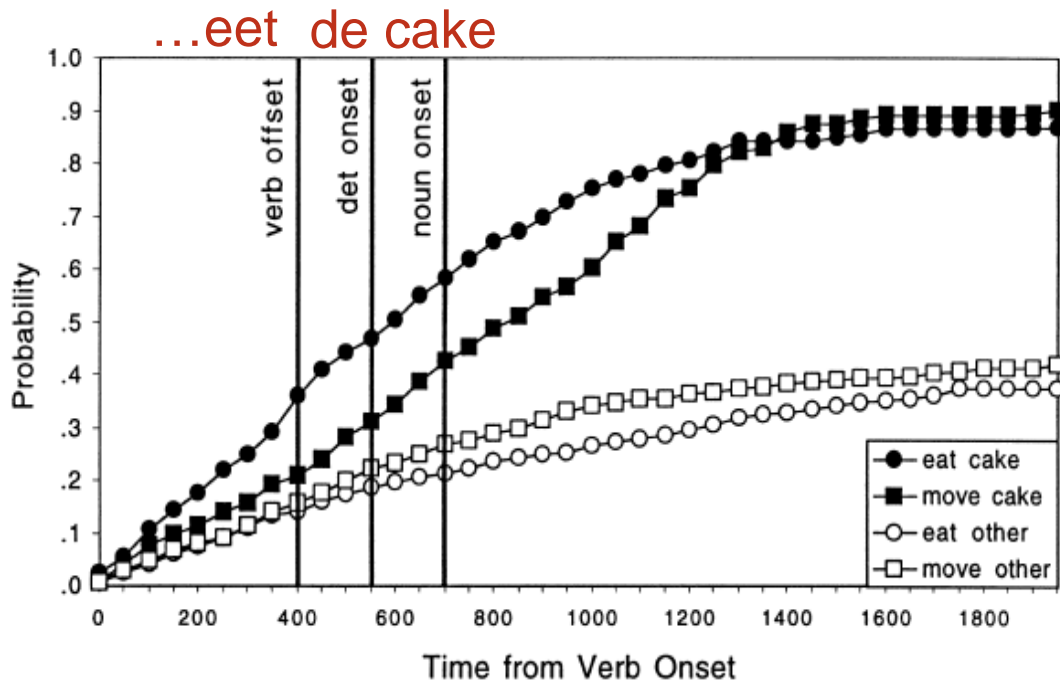
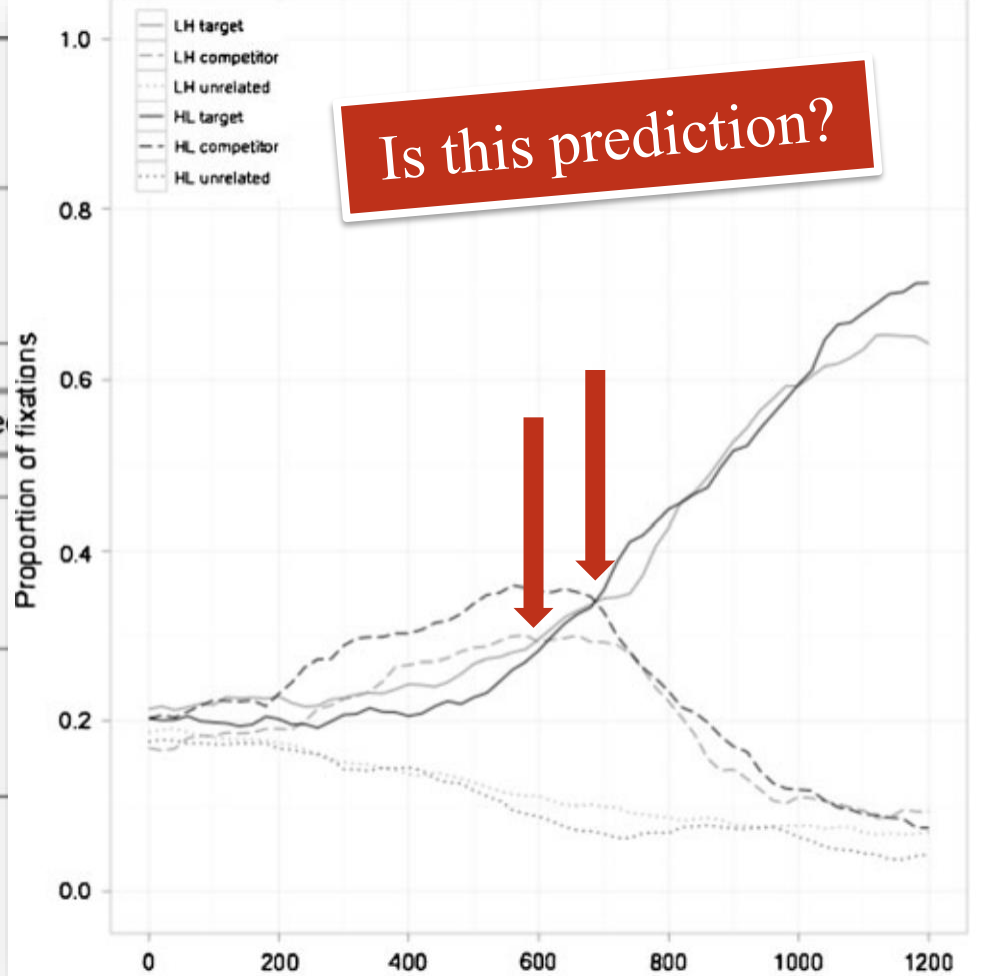
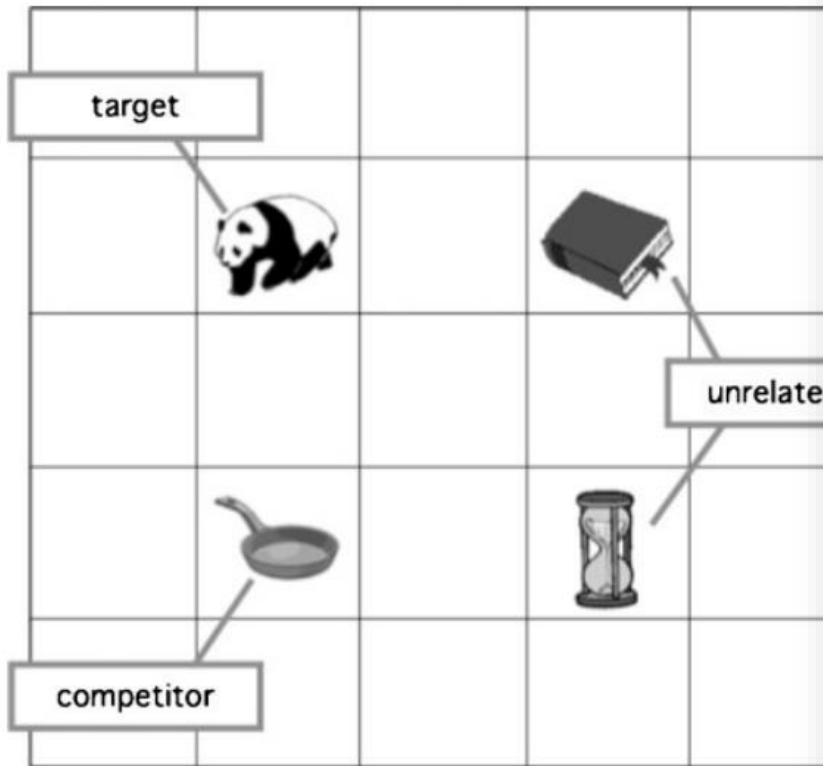


Fig. 2. The cumulative probability of fixating the target object (cake) or a distractor object (other) as a function of condition ('eat' vs. 'move') in Experiment 1 (Section 2). Note: The verb offset, determiner onset, and noun onset are shown, for display purposes, averaged across trials, and are aligned to the 50ms bin within which they fall.

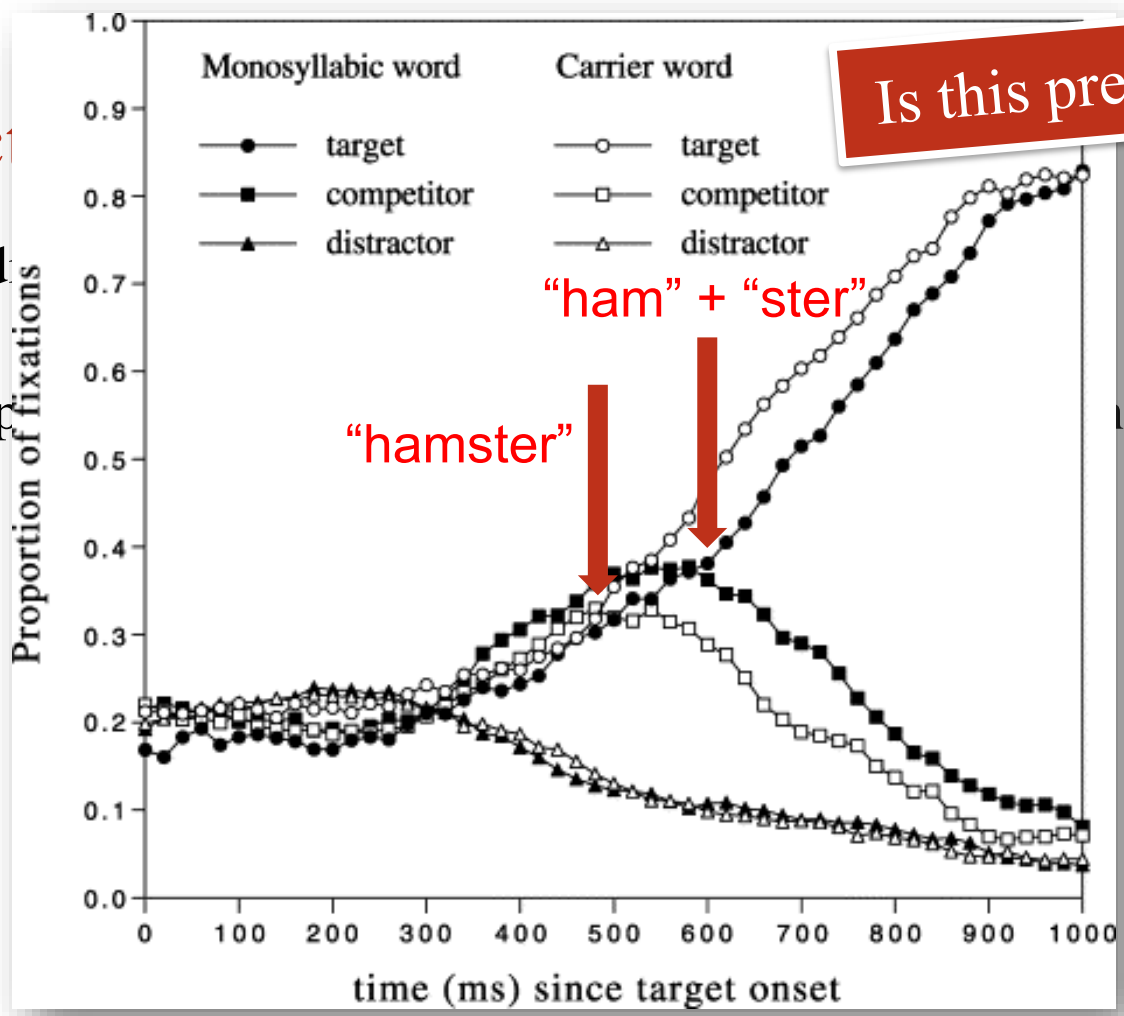




Is this prediction?

# Prediction

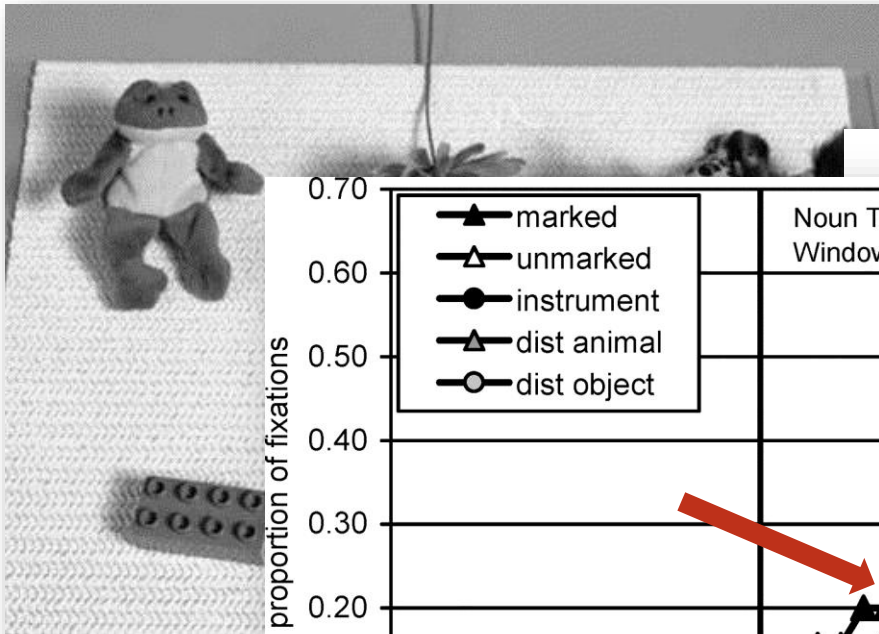
- Prosodic
  -
- Anticipatory
  -



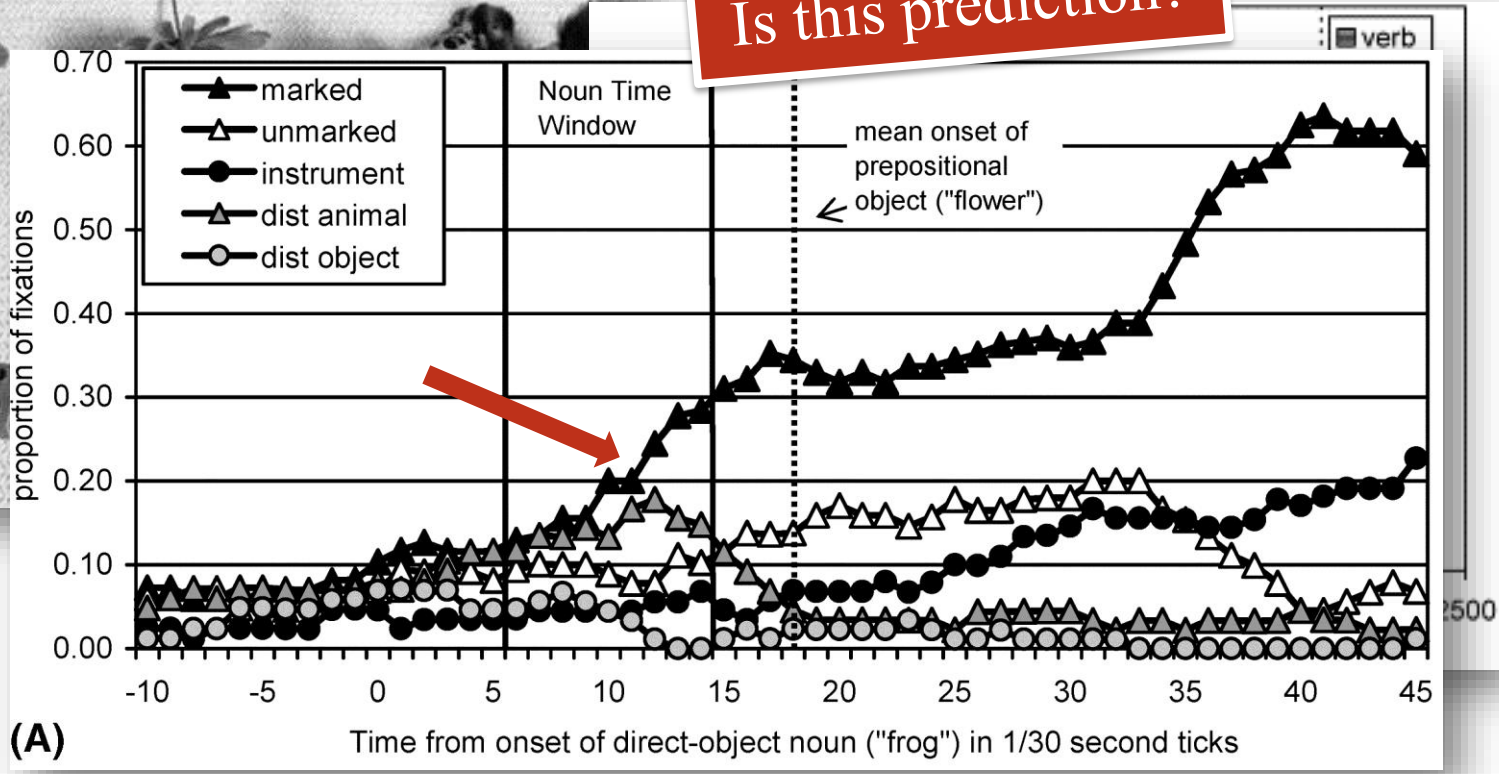


# Predicting syntax

- Prosody is closely tied to syntactic structures
- Prosody is used as a predictive cue about how to interpret structurally ambiguous sentences
  - “Tap [the frog] [with the flower]” > instrument
  - “Tap [the frog with the flower]” > modifier



Is this prediction?



(A)



# Predicting turn-taking

- Prosody is closely tied to turn-taking (turn-yielding & turn-keeping)

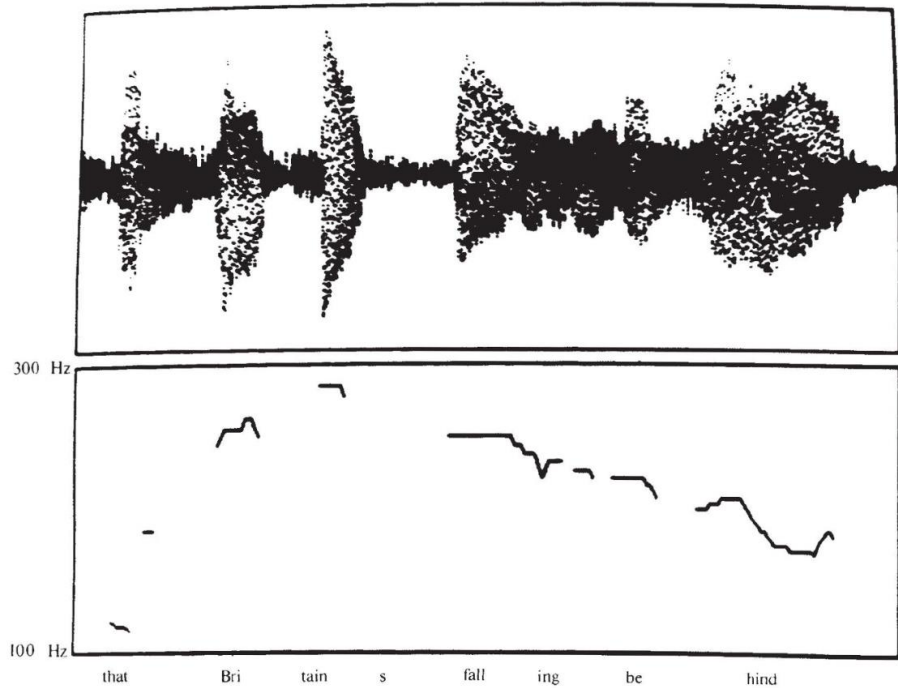


Fig. 1 Amplitude trace and pitch contour of turn-medial utterance, showing slow fall in utterance pitch.

turn-medial

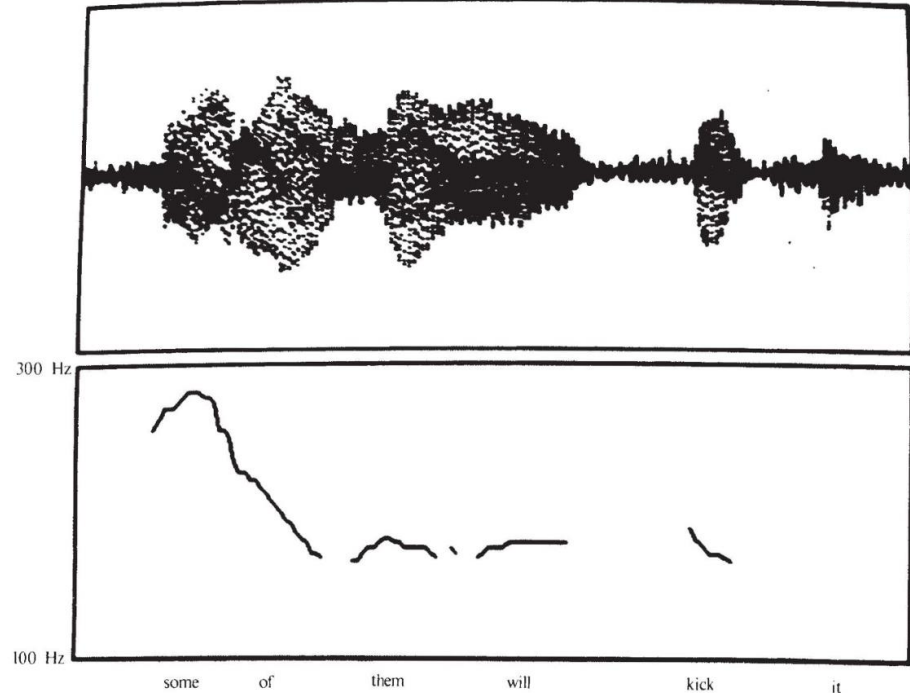


Fig. 2 Amplitude trace and pitch contour of turn-final utterance, showing fast fall in utterance pitch.

turn-final





# Predicting turn-taking

- Prosody is closely tied to turn-taking (turn-yielding & turn-keeping)

*Nature Vol. 300 23/30 December 1982*

*Letters to  $\Lambda$*

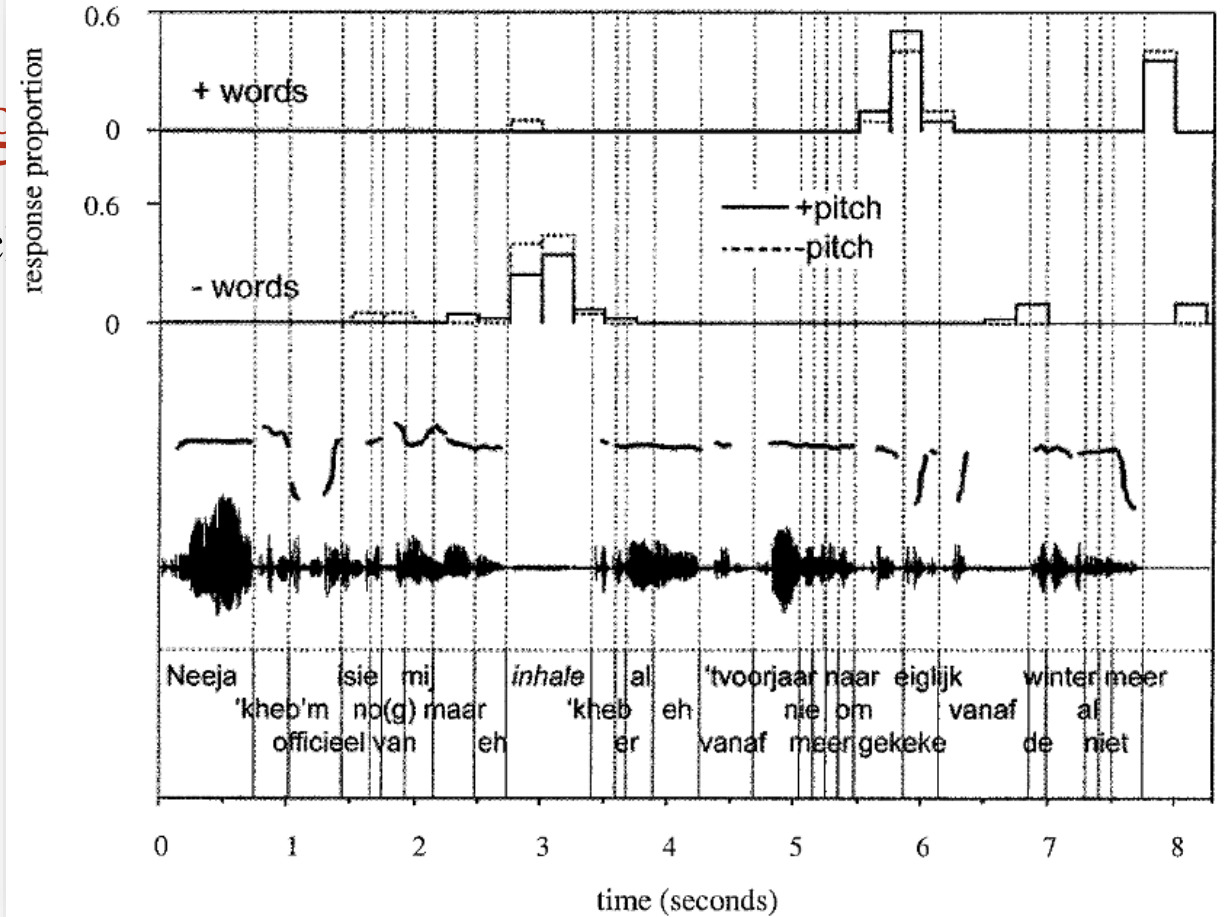
**Table 2** Mean peak and trough values for three types of utterance

	Peak	Trough	$\Delta$	Span
Turn-final	238	141	97	693
Turn-disputed	263	167	96	463
Turn-medial	275	161	114	953



# Predicting

- Prosody is c
- ...or is it?
  - Press

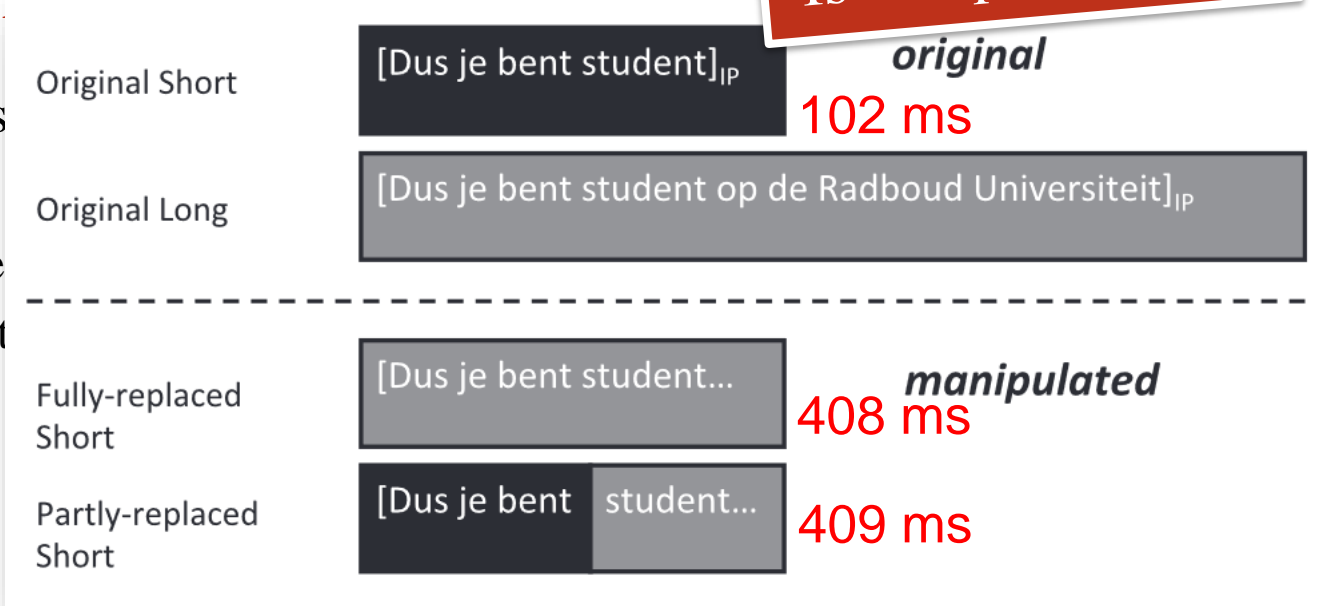




# Predicting turn-taking

Is this prediction?

- Prosody is
- ...or is it?
  - Pre
  - But
    - Fully-replaced
    - Short
    - Partly-replaced
    - Short
  - Truncated stimuli ending in a syntactic completion point but lacking an intonational phrase boundary led to significantly delayed button presses

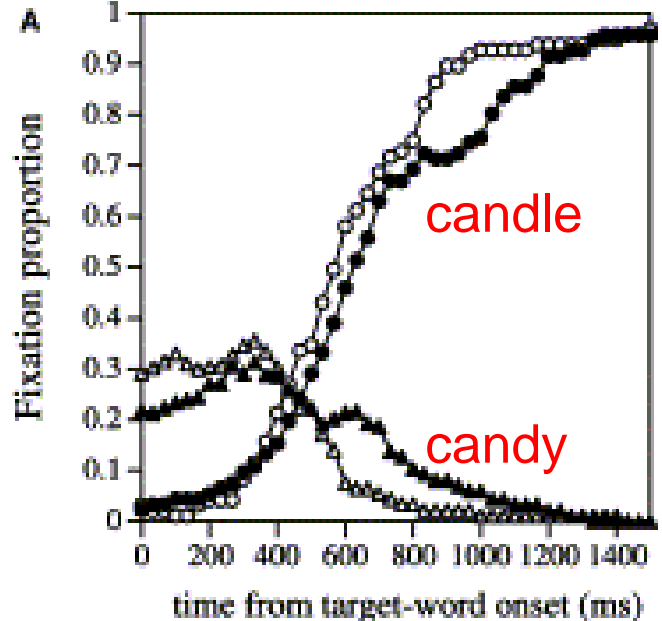


ary.  
phrase



# Predicting information structure

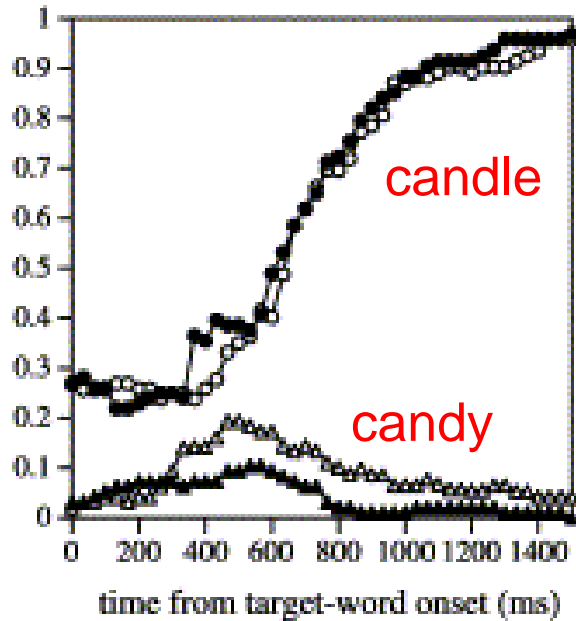
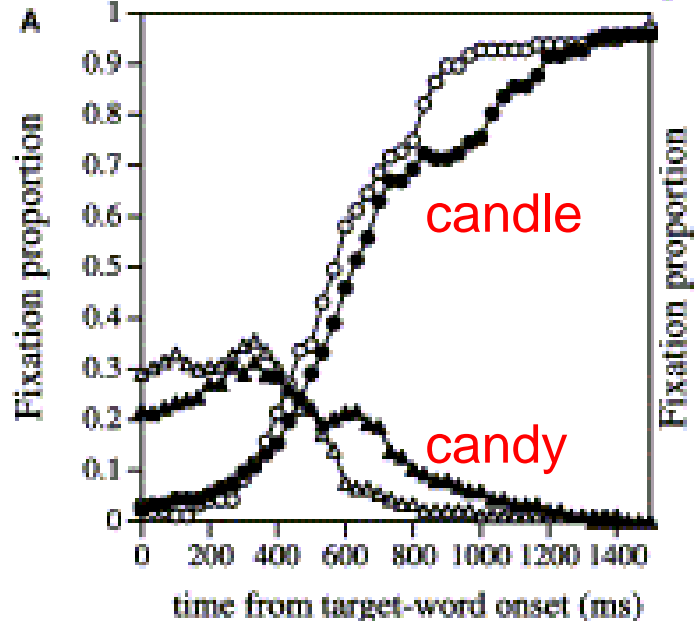
- Prosody cues information structure through...
  - intonational cues (contrastive focus: given/new)



- Target, Accented
- ▲— Competitor, Accented
- Target, Deaccented
- △— Competitor, Deaccented

Put the **candle** below the triangle

- Now put the **candle** (...ABOVE THE SQUARE)
- Now put the **CANDLE** (...above the square)



- Target, Accented
- ▲— Competitor, Accented
- Target, Deaccented
- △— Competitor, Deaccented

Put the **candy** below the triangle

- Now put the **candle** (...ABOVE THE SQUARE)
- Now put the **CANDLE** (...above the square)



# Predicting information structure

- Prosody cues information structure through...
  - intonational cues (contrastive focus: given/new)
  - disfluencies
    - ...affect ca. 6 per 100 words (Fox Tree, 1995; Bortfeld et al., 2001)
    - ...don't happen randomly
    - ...tend to occur before 'more complex' referents, such as low-frequency words, open-class words, unpredictable words, etc.

Levelt (1983); Maclay & Osgood (1959); Hartsuiker & Notebaert (2010); Beattie & Butterworth (1979)

Is this prosody?



# Predicting information structure

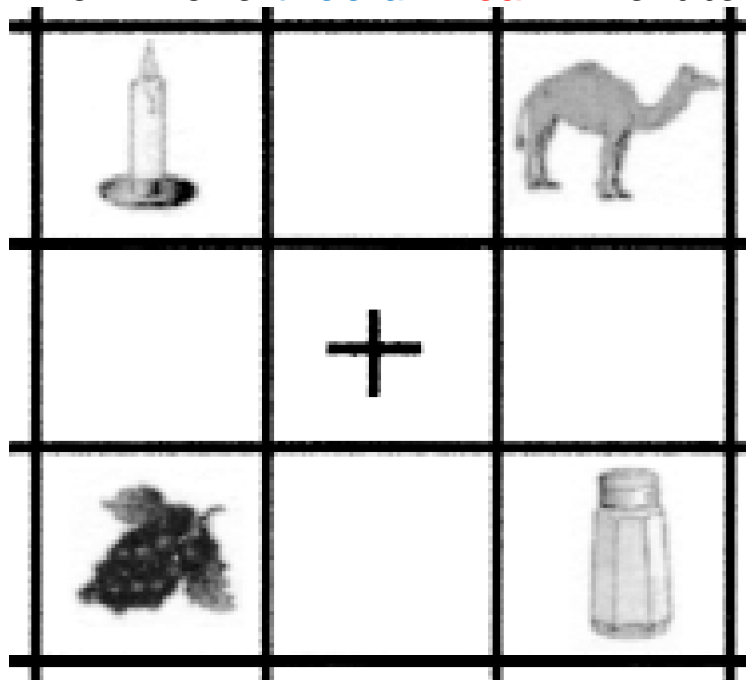
- Disfluences induce prediction of *new* referents



# Predicting information structure

- Disfluences induce prediction of *new* referents  
 "... candle. Now move *thee uh* *ca...* next to the grapes."

given



new



# Predicting information structure

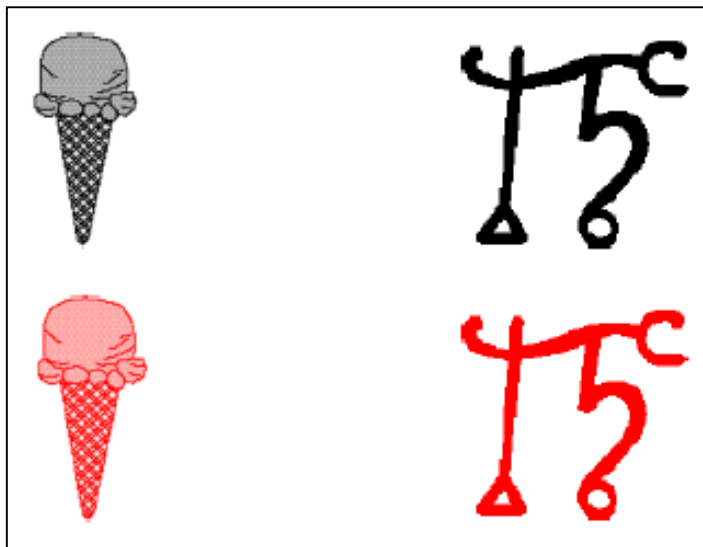
- Disfluences induce prediction of *novel* referents

# Predicting information structure

- Disfluences induce prediction of *novel* referents

“Click on *thee* ..uhm.. red funny squiggly shape”

known

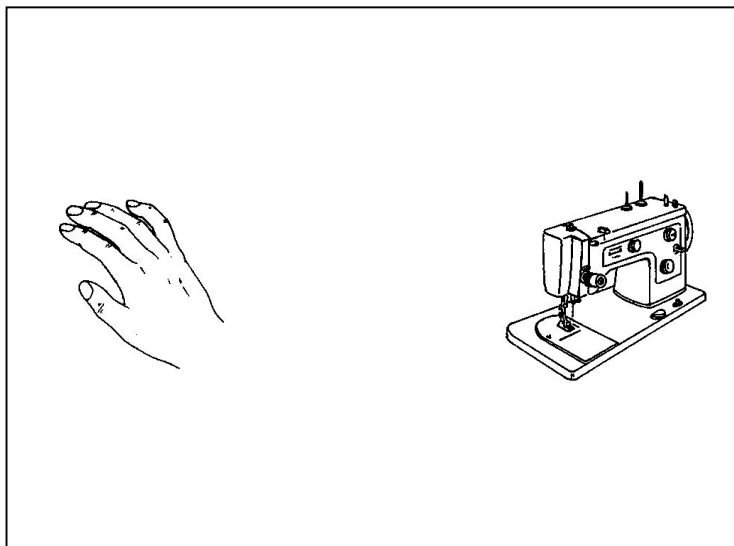


unknown

# Predicting information structure

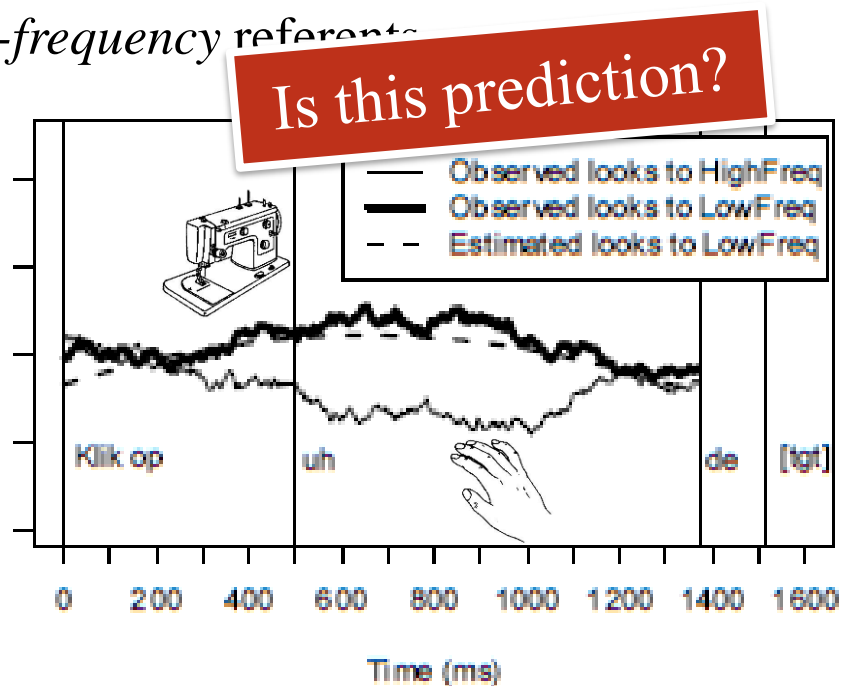
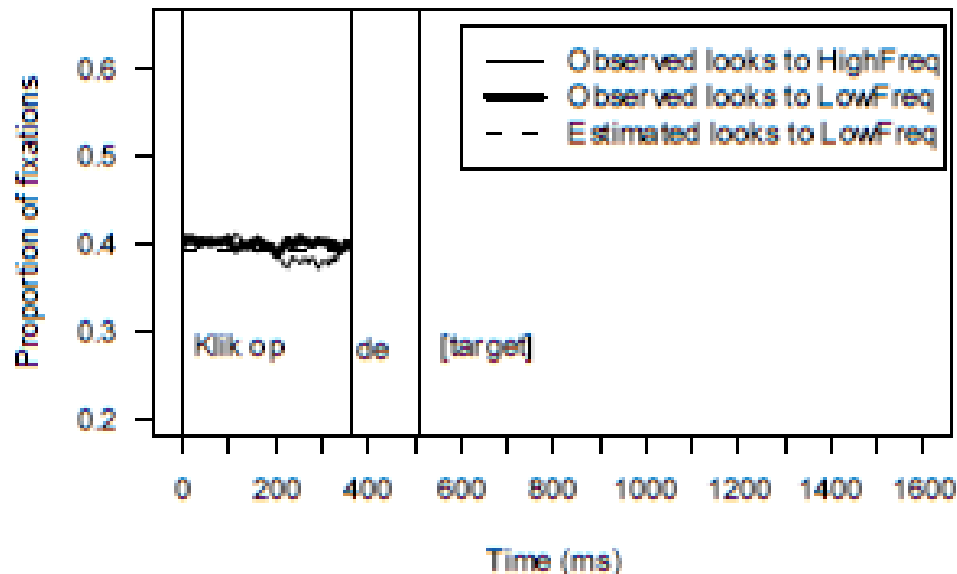
- Disfluencies induce prediction of *low-frequency* referents

"Now click on *thee uh...*"



# Predicting information structure

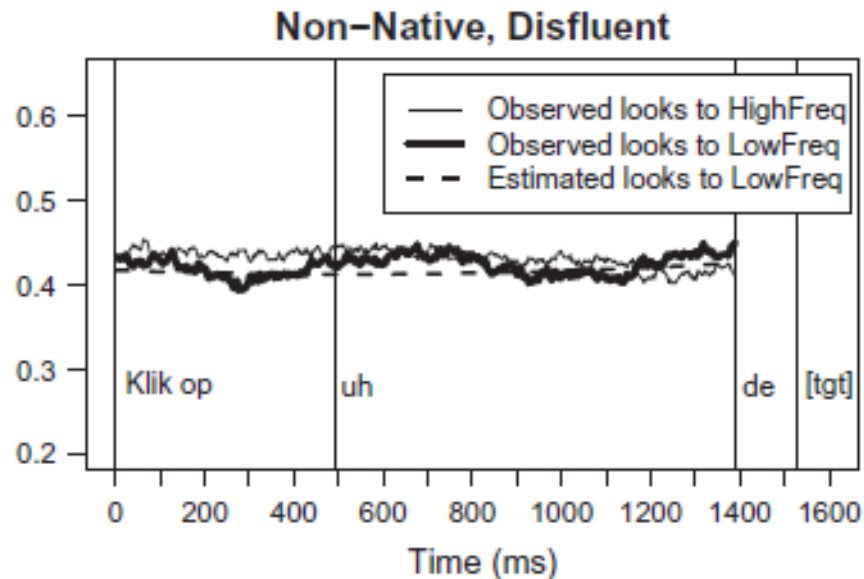
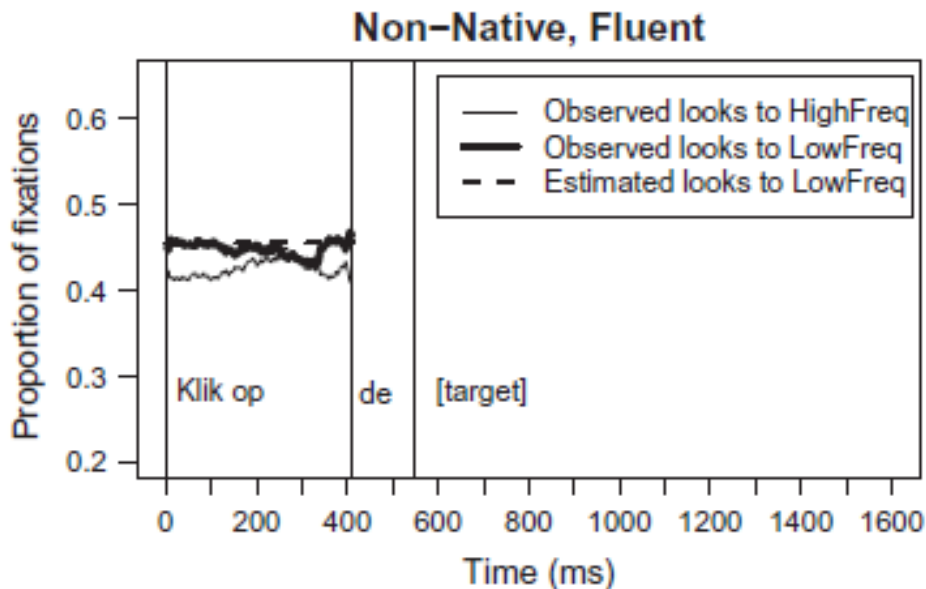
- Disfluencies induce prediction of *low-frequency* referents





# Predicting information structure

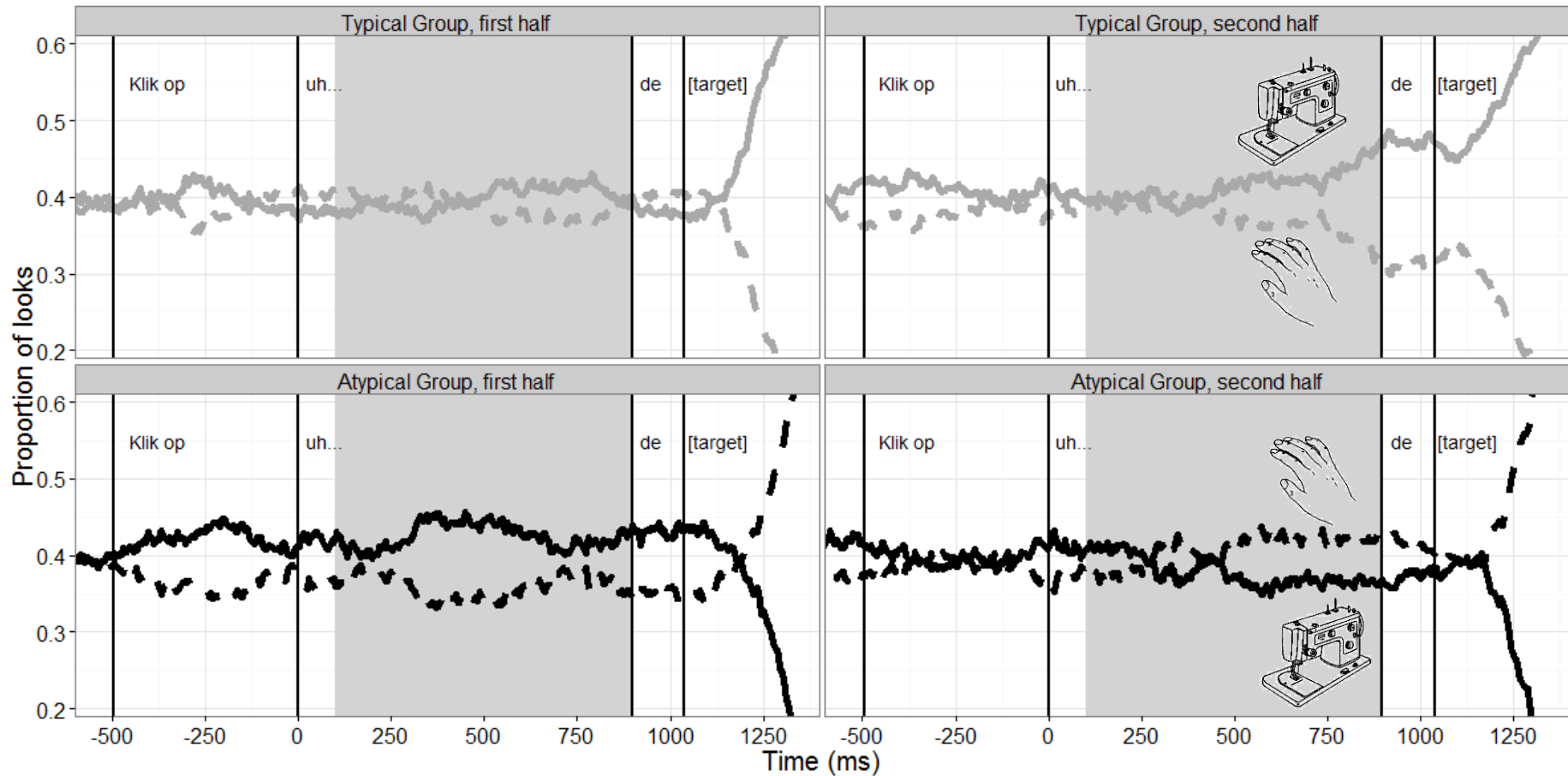
- Disfluencies induce prediction of *low-frequency* referents





# How do people know what to predict?

- Statistical/distributional learning
- Can we change people's priors?





# Wrap-up of today

- Prosody predicts
- ...where words begin and end [speech]
- ...syntactic boundaries
- ...when a talker will finish [turn-taking]
- ...when important information is presented

How real is this?

Do we predict all the time?

How much of this is language-specific?

How quickly do we adjust our predictions?

How robust are prosody-guided predictions?



# Competing predictions

- Prosody is not the only source inducing predictions...
- ...so is syntax.
- What happens when these two provide conflicting cues?
- Which of these two sources is ‘the strongest’?

# Competing predictions

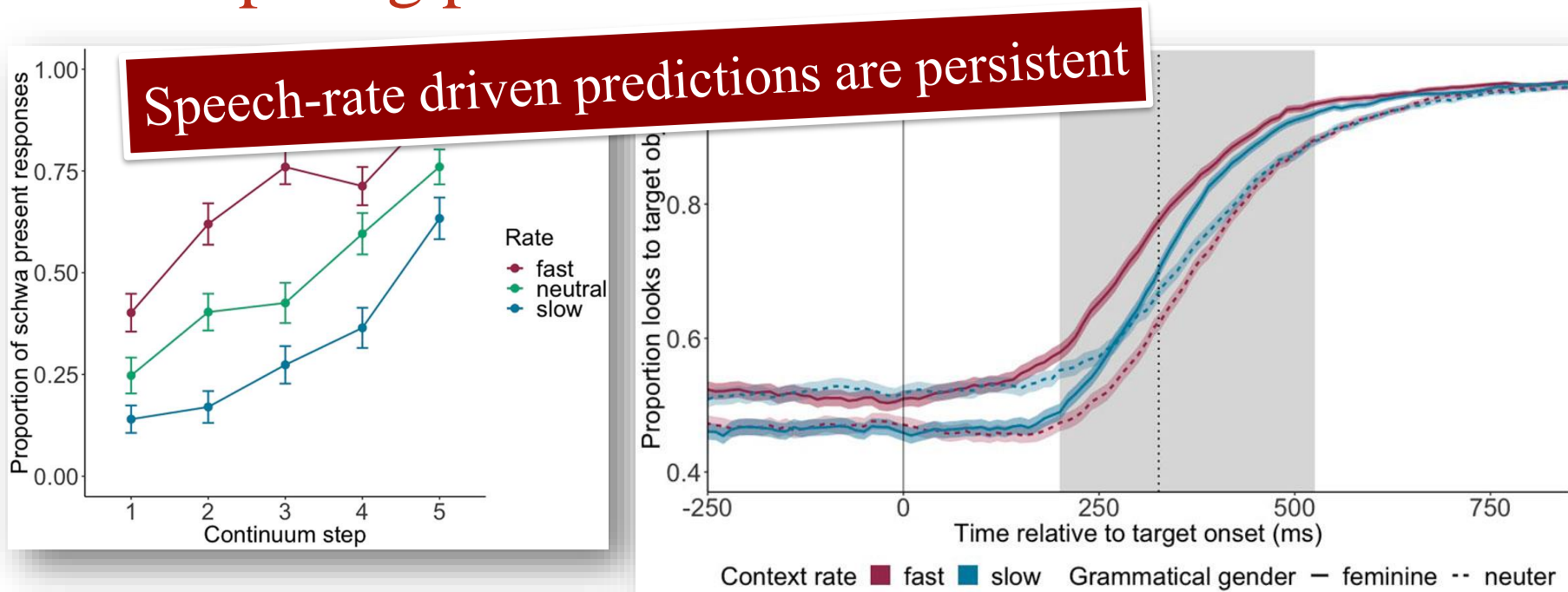
- German:
  - *Schauen Sie jetzt sofort auf ein(e) außergewöhnlich liebe(s) Katze<sub>FEM</sub>/Reh<sub>NEU</sub>*



- Fast speech rate > + “eine” > + **Katze<sub>FEM</sub>**
- Fast speech rate > + “eine” > + **Reh<sub>NEU</sub>**
- Slow speech rate > + “ein” > + **Katz<sub>NEU</sub>**
- Slow speech rate > + “ein” > + **Reh<sub>NEU</sub>**



# Competing predictions





# Competing predictions

- Dutch:

- *Kijk nu eens naar de<sub>COMMON</sub>/het<sub>NEUTER</sub> ontzettend vuile vat<sub>COMMON</sub>/vaat<sub>NEUTER</sub>*

speech rate

article

buffer

V length

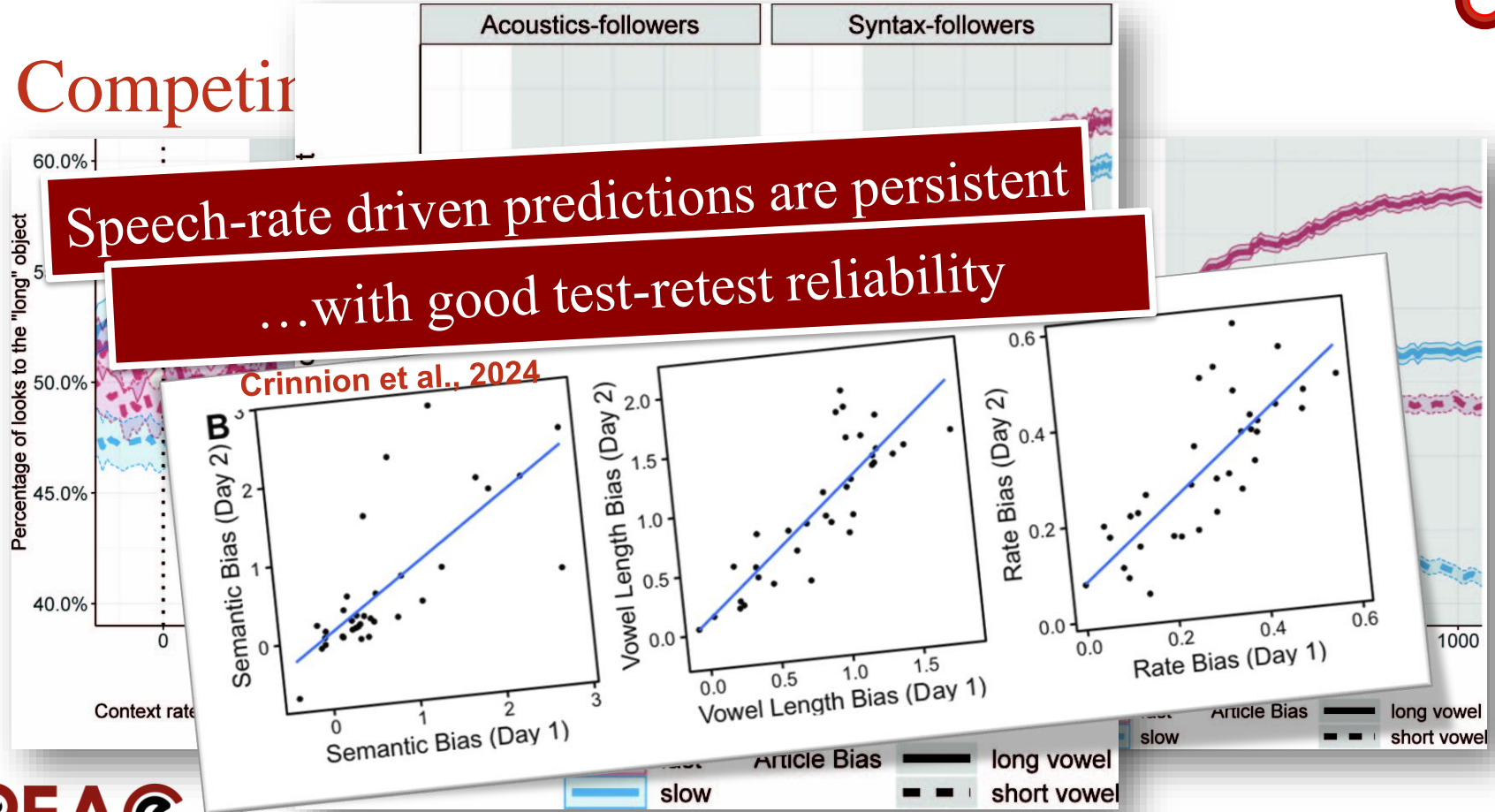
- “de” + slow rate > vat<sub>COMMON</sub>
- “de” + slow rate > ??
- “het” + fast rate > ??
- “het” + fast rate > vaat<sub>NEUTER</sub>



# Competit

Speech-rate driven predictions are persistent  
...with good test-retest reliability

Crinnion et al., 2024



# Wrap-up of today

- Prosody predicts
- ...where words begin and end [speech]
- ...syntactic boundaries
- ...when a talker will finish [turn-taking]
- ...when important information is presented

How real is this?

Do we predict all the time?

How much of this is language-specific?

How quickly do we adjust our predictions?

How robust are prosody-guided predictions?

Is it always helpful to predict?



## Discourse markers active their, *like*, cohort competitors

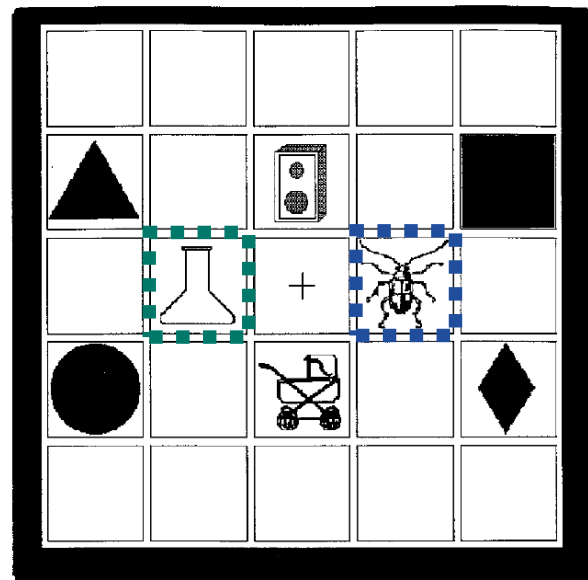
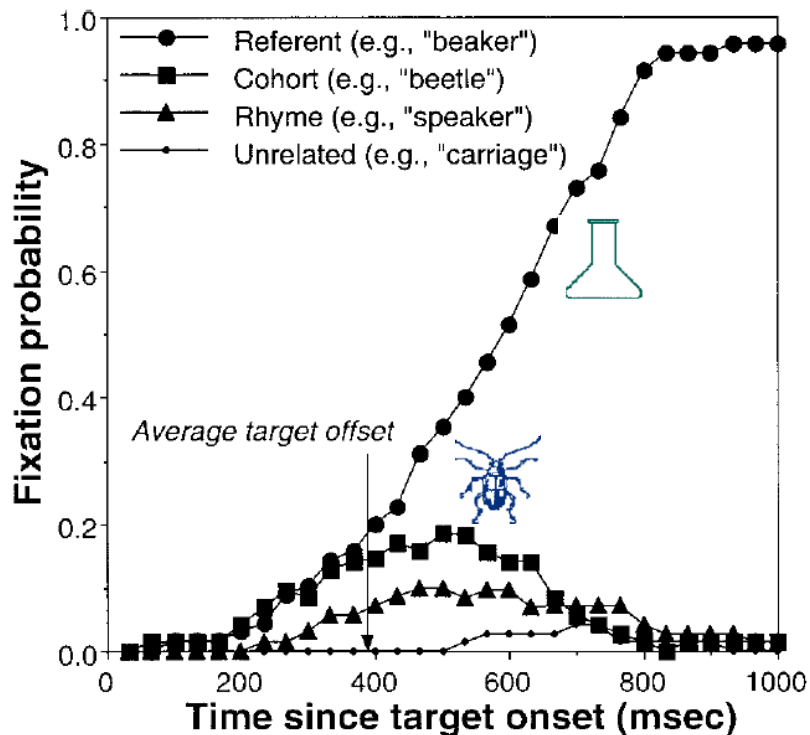
- Spontaneous speech is riddled with discourse markers (DMs), such as: *well, you know, like, uhm...*
- However, little is known about their role in speech perception since lab-based studies have primarily used ‘clean’ sentence stimuli
- This study focuses on the online processing of DM ‘like’:  
“*Then I saw a... like... unicycle*”





# Discourse markers active their, *like*, cohort competitors

- When hearing a **target word**, listeners also activate its **cohort competitors**.



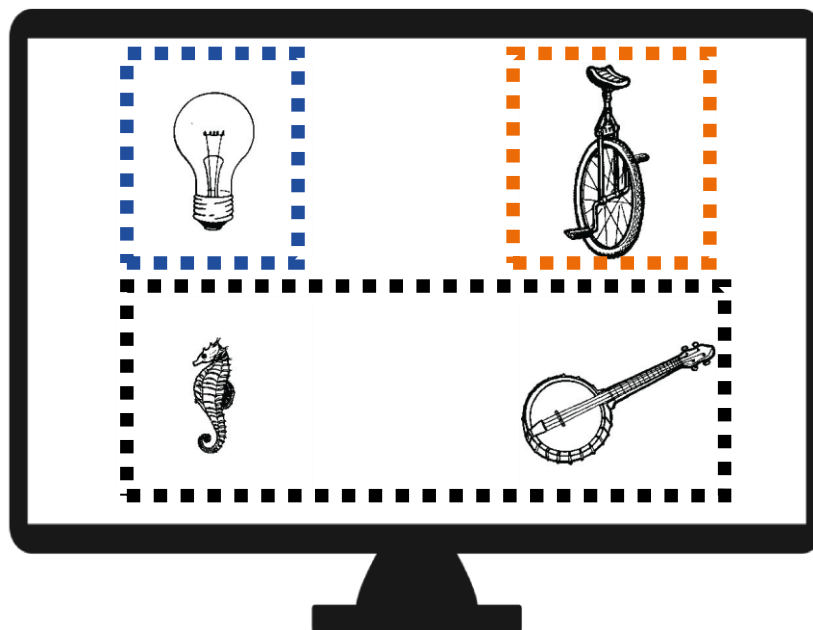


## Discourse markers activate their, *like*, cohort competitors

- Do DMs activate their cohort competitors?
- Does ‘...like’ activate *lightbulb*?

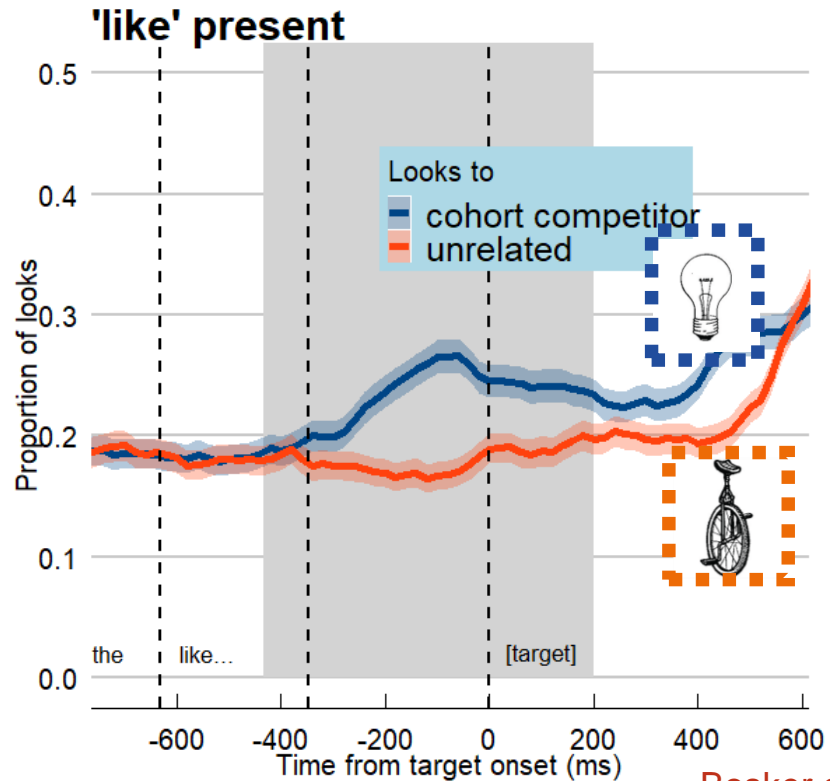
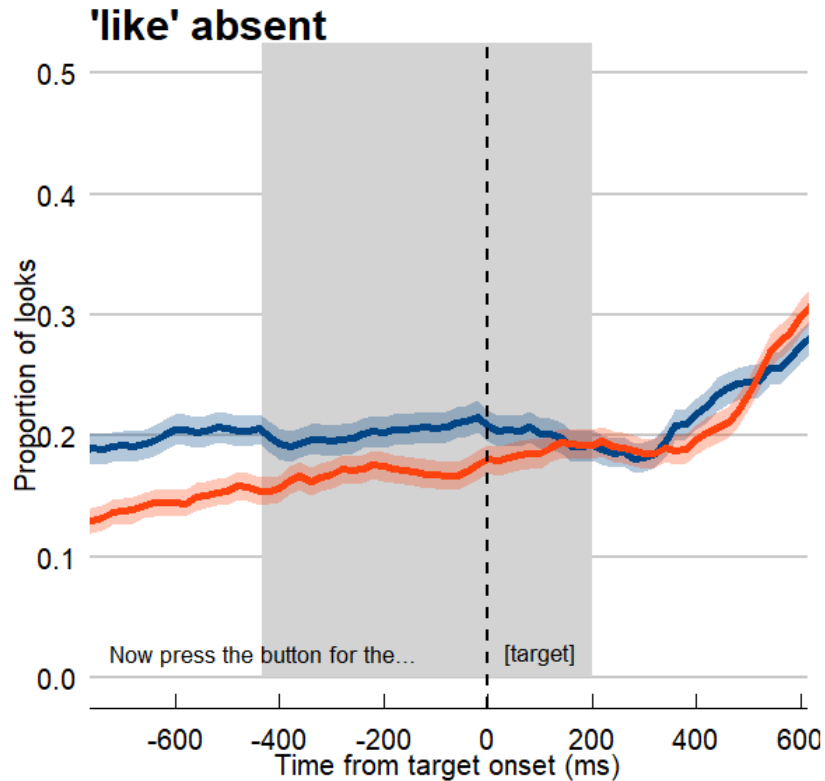
# Discourse markers active their, *like*, cohort competitors

- Eye-tracking study (N=64)
  - like: “Now press the button for the lightbulb”
  - +like: “Now press the button for the, *like*, lightbulb”
- Cohort competitor
- Unrelated  
*matched in freq, AoA, length*
- Distractors





# Discourse markers active their, *like*, cohort competitors





## Discourse markers activate their, *like*, cohort competitors

- DM ‘like’ activates its cohort competitors
- DM ‘like’ speeds up RTs for target words that are cohort competitors:  
“*the... like... lightbulb*”
- Prosody can be misleading?!



## Next up:

- Lecture 4: *Learning about talker-specific prosody*

## Hans Rutger Bosker

Speech Perception in Audiovisual Communication [SPEAC] lab

*Donders Institute, Radboud University, Nijmegen, The Netherlands*

<https://hrbosker.github.io>

[hansrutger.bosker@donders.ru.nl](mailto:hansrutger.bosker@donders.ru.nl)

