Analysing inter-key intervals: Beyond means, medians and pause frequencies

Jens Roeser   Sven De Maeyer   Mark Torrance

Mariëlle Leijten   Luuk Van Waes

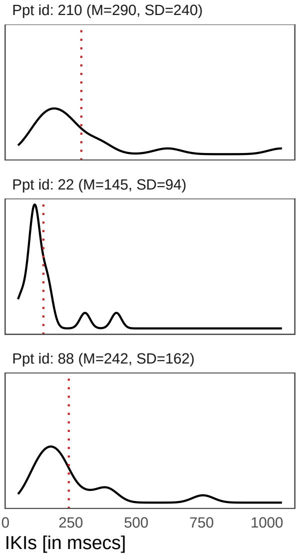
jens.roeser@ntu.ac.uk

SIG 27 Conference
University of Antwerp

Dec 15, 2020

The problem: what's a pause?

- Keystroke data are heavily skewed.
- Skew reflects cognitive processes.
- How can we distinguish between fluent and disfluent key transitions?
- Fixed thresholds: 0.5 or 2 secs?
- Key transitions normal for a learner might be longer than pauses of experienced writers.



The problem: what's a pause?

- Keystroke data are heavily skewed.
- Skew reflects cognitive processes.
- How can we distinguish between fluent and disfluent key transitions?
- Fixed thresholds: 0.5 or 2 secs?
- Key transitions normal for a learner might be longer than pauses of experienced writers.

- Pause sizes depend on:
  - writing skills / style
  - position in text, sentence, word
  - experience with target language (in L2)
  - cognitive process (lexical, motor, semantic)
  - writing task

Research focus

- How do we deal with the heavy tail without loosing data or imposing pause thresholds?
- Finding a principled way to extract:

1. by-ppt typing estimates
2. by-ppt pause frequencies

Method

- Implementation of the copy-typing process as statistical model in Stan (Carpenter et al., 2016); code based on Sorensen et al. (2016) and Vasishth et al. (2017).
- Key-interval data: Dutch subset (N=250) of copy-task corpus (Van Waes et al., 2019; Van Waes et al., 2020).
- Lexical vs non-lexical task

Consonants task

tjxgfl pgkfkq dtdrgt npwdvf

Low-Frequency (LF) bigrams task

een chaotische cowboy

Low-Frequency (LF) bigrams task

een chaotische cowboy

$\Downarrow$

e^e^n c^h^a^o^t^i^s^c^h^e c^o^w^b^o^y

Low-Frequency (LF) bigrams task

een chaotische cowboy

$\Downarrow$

e^e^n c^h^a^o^t^i^s^c^h^e c^o^w^b^o^y

$\Downarrow$

162 97 107 141 800 148 278 132 199 94 154 177 870 88 274 611

Low-Frequency (LF) bigrams task

een chaotische cowboy

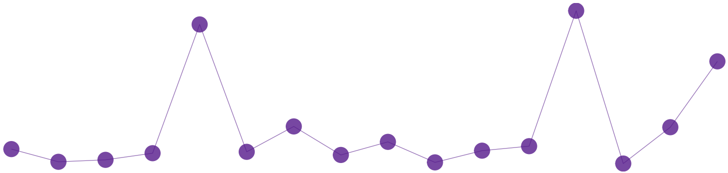
$\Downarrow$

e^e^n c^h^a^o^t^i^s^c^h^e c^o^w^b^o^y

$\Downarrow$

162 97 107 141 800 148 278 132 199 94 154 177 870 88 274 611

$\Downarrow$



Standard analysis: Mixed-Effects Model

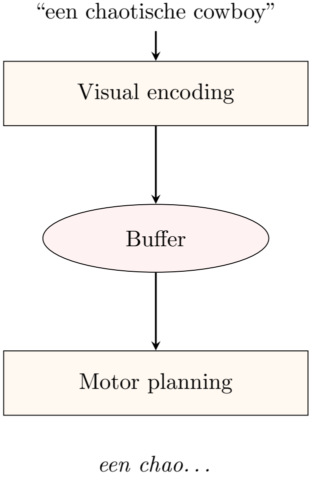
$$y_{ij} \sim LogNormal(\alpha + u_i + w_j, \sigma_e^2)$$

Standard analysis: Mixed-Effects Model

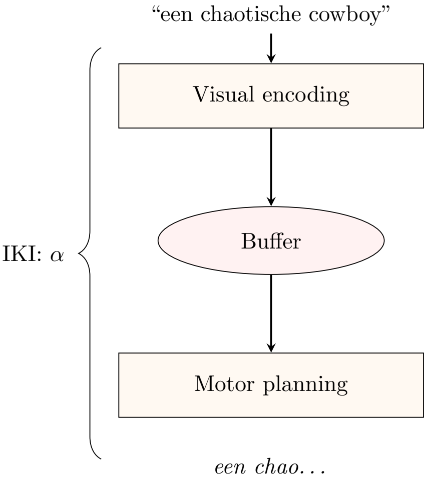
$$y_{ij} \sim LogNormal(\alpha + u_i + w_j, \sigma_e^2)$$

- α: population-level IKI
- σ_e^2: error variance
- Participants: u_i
- Bigrams: w_j

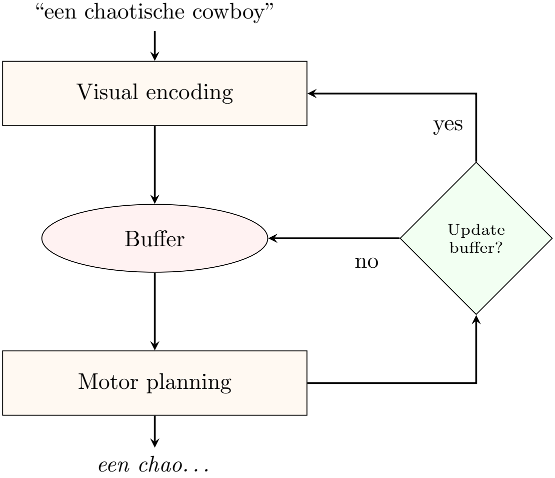
Model of copy-typing: standard analysis



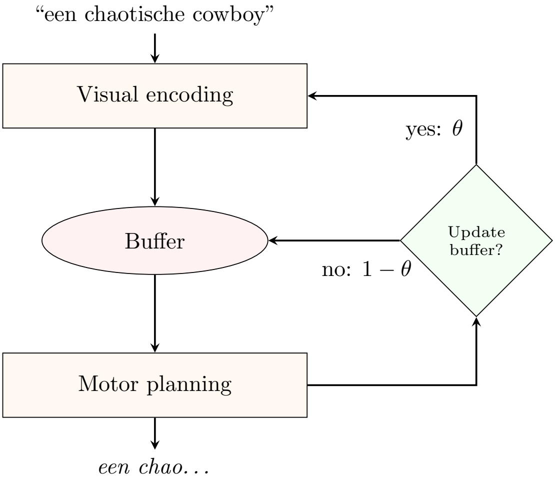
Model of copy-typing: standard analysis



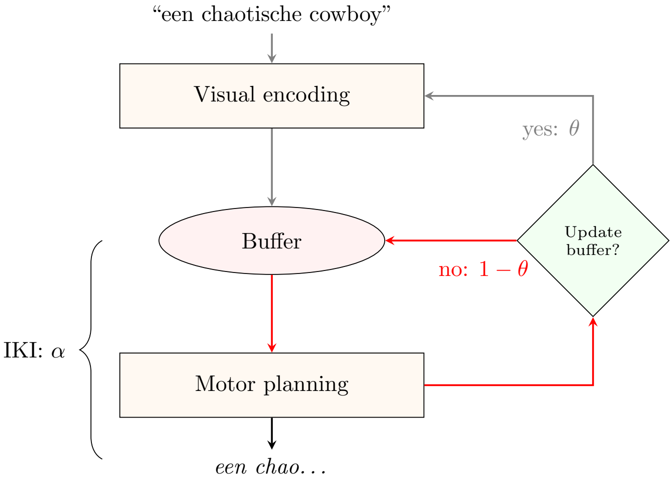
Model of copy-typing: disfluency model



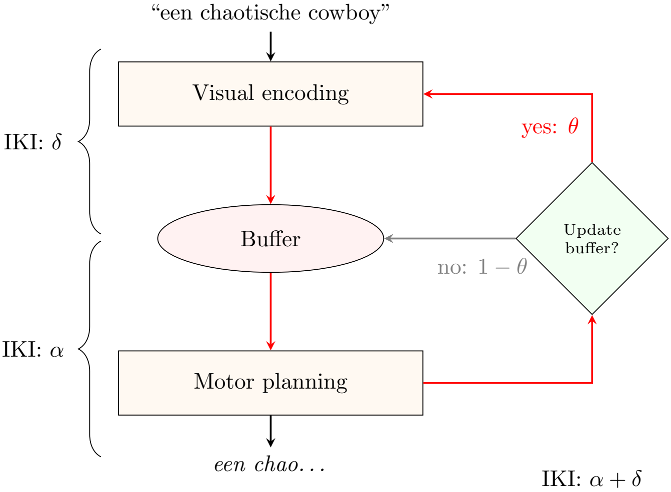
Model of copy-typing: disfluency model



Model of copy-typing: disfluency model



Model of copy-typing: disfluency model



Finite Mixture of two log-Gaussians

$$y_{ij} \sim \theta_i \cdot LogNormal(\alpha + \delta + u_i + w_j, \sigma^2_{e'}) +$$
$$(1 - \theta_i) \cdot LogNormal(\alpha + u_i + w_j, \sigma^2_e)$$

- α: fluent IKI (e.g. no buffer update; no difficulty)
- δ: buffer update; other difficulty (finding correct key)
- θ: disfluency probability (by ppt *i*)
- $\sigma^2_{e'}$: variance larger than σ^2_e

Model comparisons

Predictive performance estimated as the *expected log predictive density* ($\widehat{elpd}$) (Vehtari et al., 2015, 2017). Models are ordered by predictive performance (model with highest predictive performance in top row). Standard error in parentheses.

		Consonants task		LF-bigrams task	
Models	Distribution	$\Delta\widehat{elpd}$	$\widehat{elpd}$	$\Delta\widehat{elpd}$	$\widehat{elpd}$
MoG	$2 \times$ Log-normal				
LMM	Log-normal				

Note. LMM = Linear mixed effects model; MoG = Mixture of Gaussians

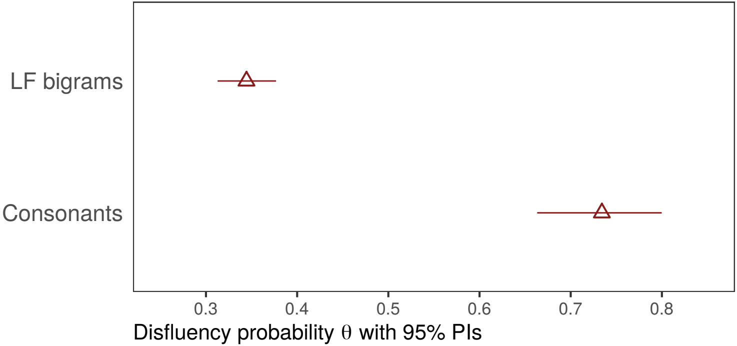
Model comparisons

Predictive performance estimated as the *expected log predictive density* ($\widehat{elpd}$) (Vehtari et al., 2015, 2017). Models are ordered by predictive performance (model with highest predictive performance in top row). Standard error in parentheses.

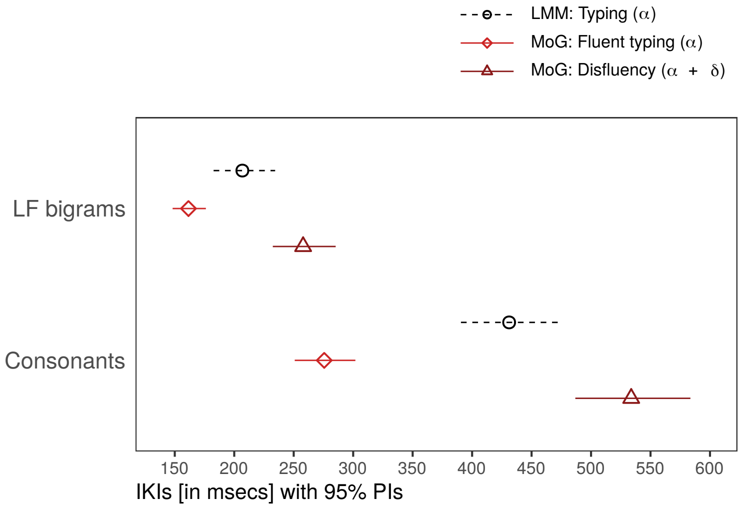
		Consonants task		LF-bigrams task	
Models	Distribution	$\Delta\widehat{elpd}$	$\widehat{elpd}$	$\Delta\widehat{elpd}$	$\widehat{elpd}$
MoG	2 $\times$ Log-normal	–	-37,069 (101)	–	-33,178 (113)
LMM	Log-normal	-281 (25)	-37,350 (99)	-994 (63)	-34,173 (121)

Note. LMM = Linear mixed effects model; MoG = Mixture of Gaussians

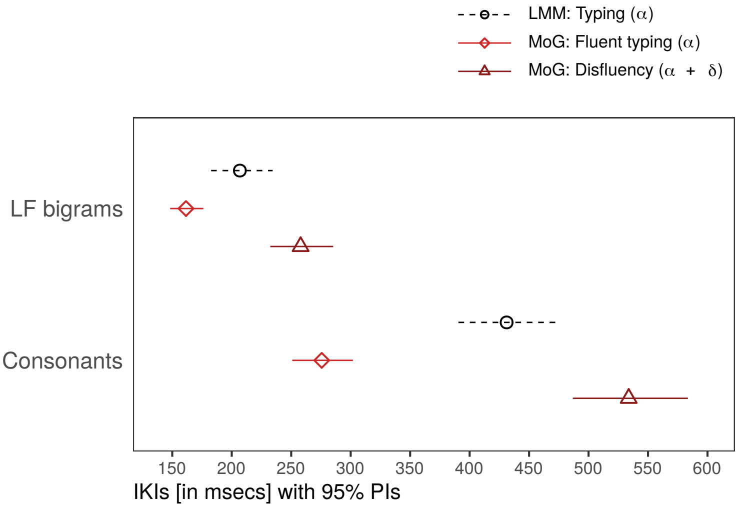
Population estimates



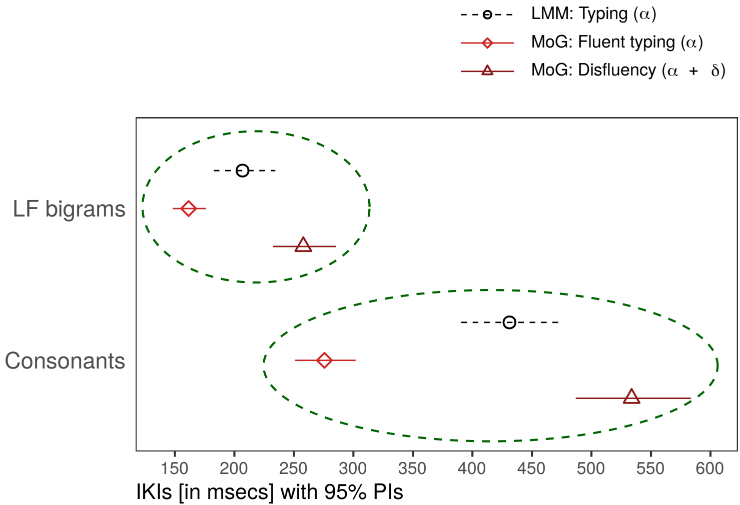
Population estimates



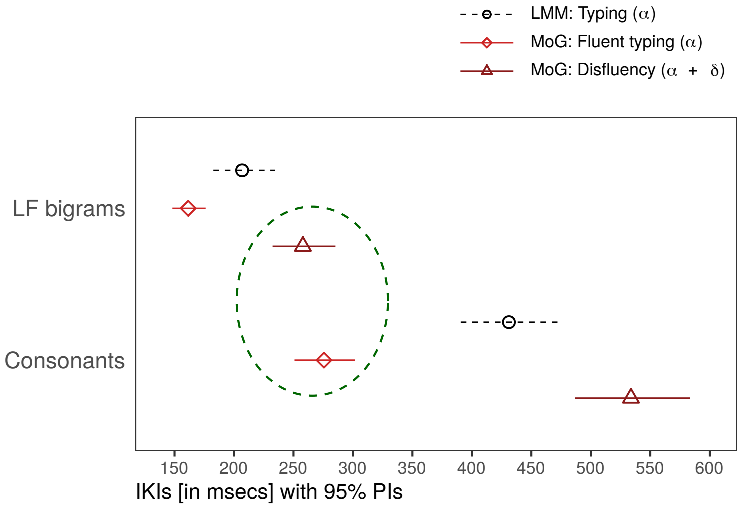
Population estimates



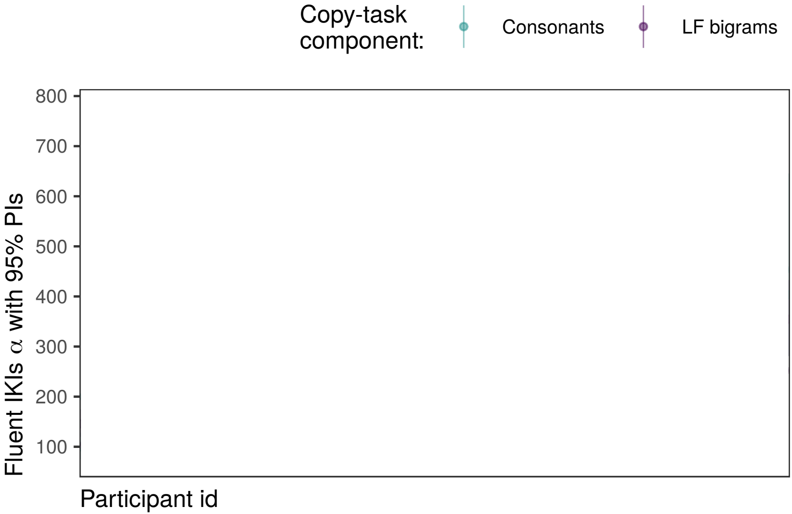
Population estimates



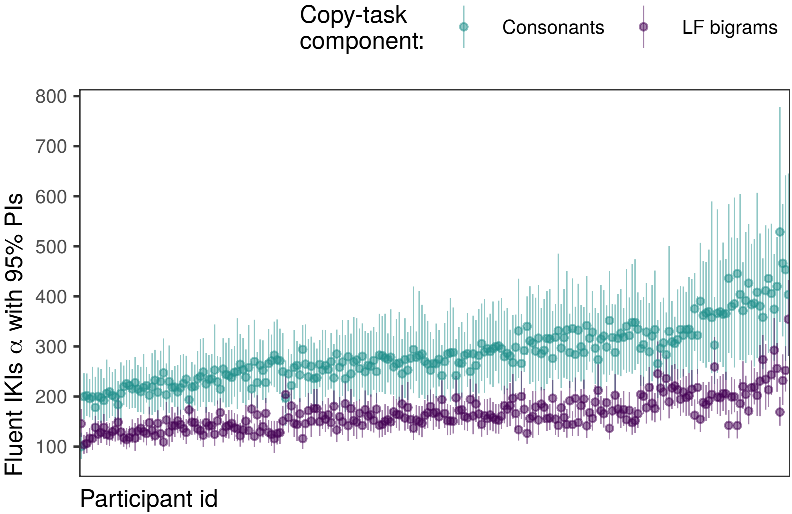
Population estimates



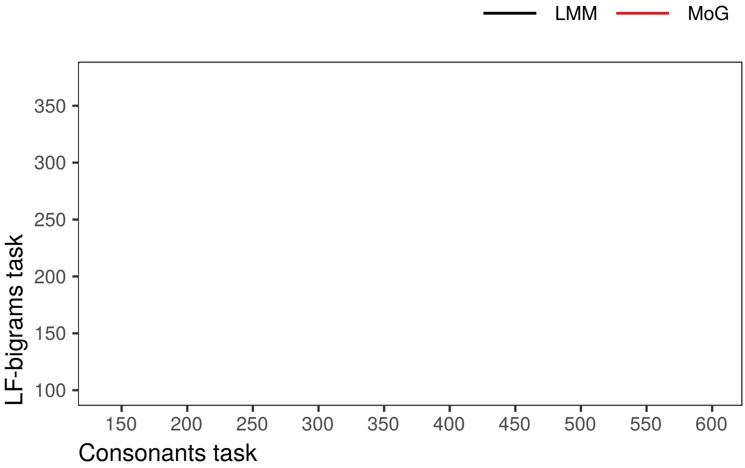
By-participant fluent-typing intervals



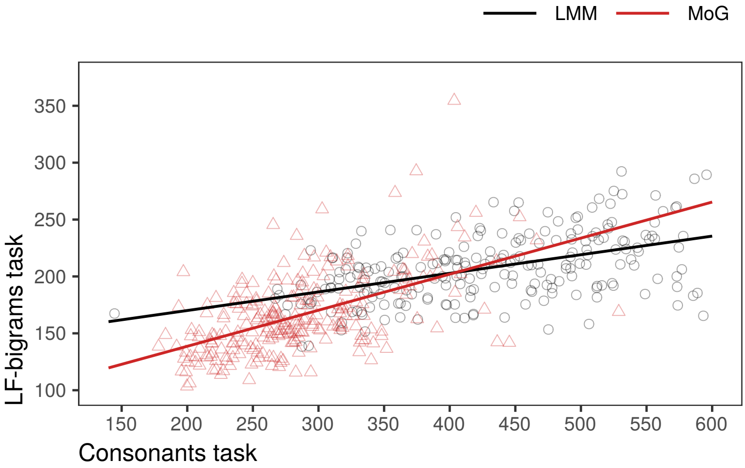
By-participant fluent-typing intervals



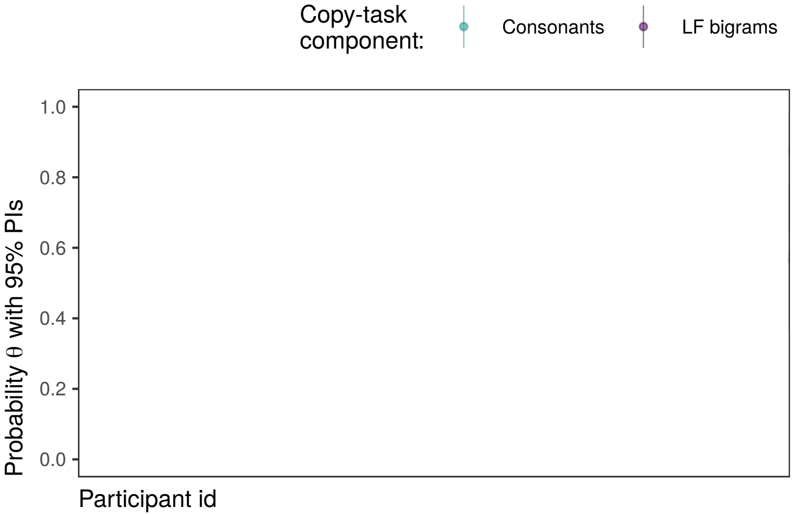
Estimated (fluent) keystroke transitions



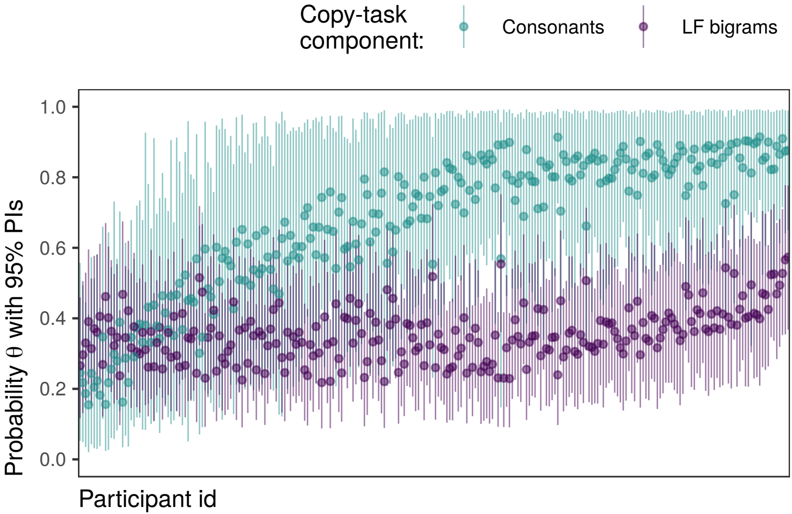
Estimated (fluent) keystroke transitions



By-participant disfluency probability



By-participant disfluency probability



Conclusion

- Better fit for mixture models over standard analysis.
- Capture writing process as a mixture of fluent and disfluent key transitions.
- Advantages of mixture models for writing research:
  1. map on cascading models of writing.
  2. capture disfluencies in a principled way.
  3. represent the probabilistic nature of disfluencies.
  4. provide reliable typing estimates and pause frequencies.

Thanks for listening!

email: jens.roeser@ntu.ac.uk
R-scripts, *Stan*-code, slides, preprint:
https://github.com/jensroes/Typing-disfluency

References I

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. & Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, *20*.

Sorensen, T., Hohenstein, S. & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *Quantitative Methods for Psychology*, *12*(3), 175–200.

Van Waes, L., Leijten, M., Pauwaert, T. & Van Horenbeeck, E. (2019). A multilingual copy task: Measuring typing and motor skills in writing with inputlog. *Journal of open research software*, *7*(30), 1–8.

Van Waes, L., Leijten, M., Roeser, J., Olive, T. & Grabowski, J. (2020). Designing a copy task to measure typing and motor skills in writing research [submitted]. *Journal of Writing Research*.

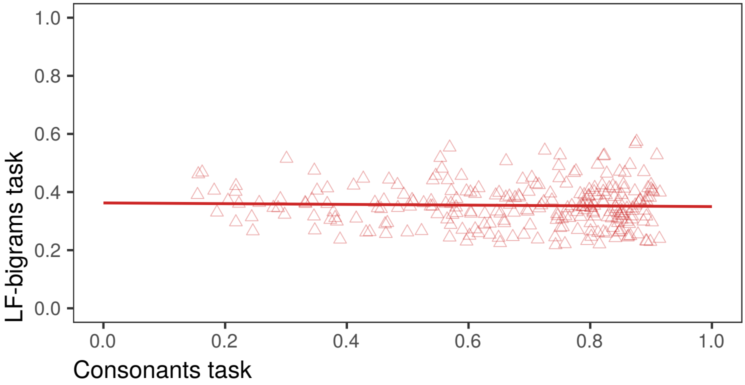
Vasishth, S., Chopin, N., Ryder, R. & Nicenboim, B. (2017). Modelling dependency completion in sentence comprehension as a Bayesian hierarchical mixture process: A case study involving Chinese relative clauses. *ArXiv e-prints*.

References II

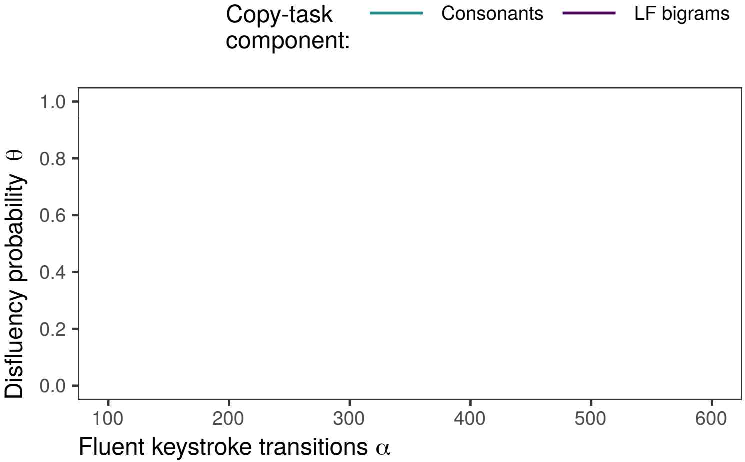
Vehtari, A., Gelman, A. & Gabry, J. (2015). Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*.

Vehtari, A., Gelman, A. & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.

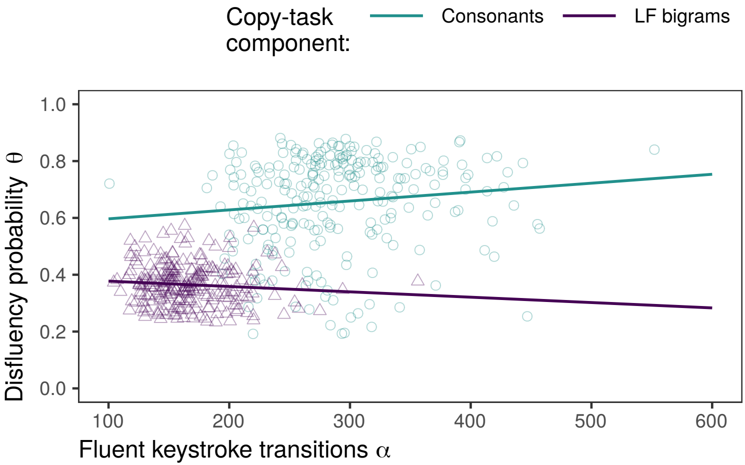
Estimated disfluency probability



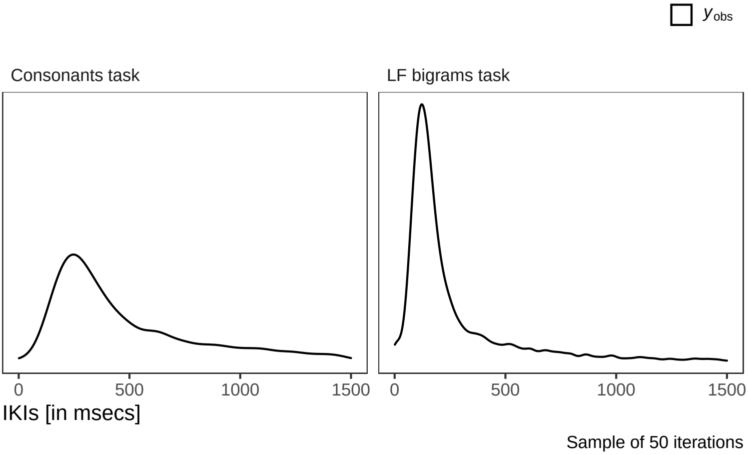
Disfluency typing-speed trade-off



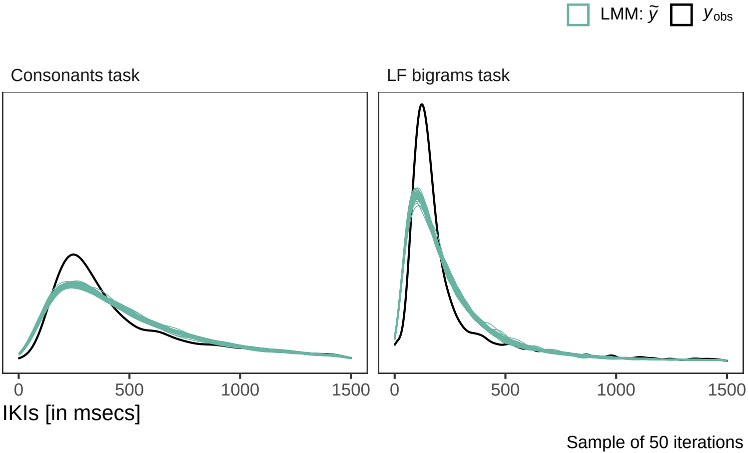
Disfluency typing-speed trade-off



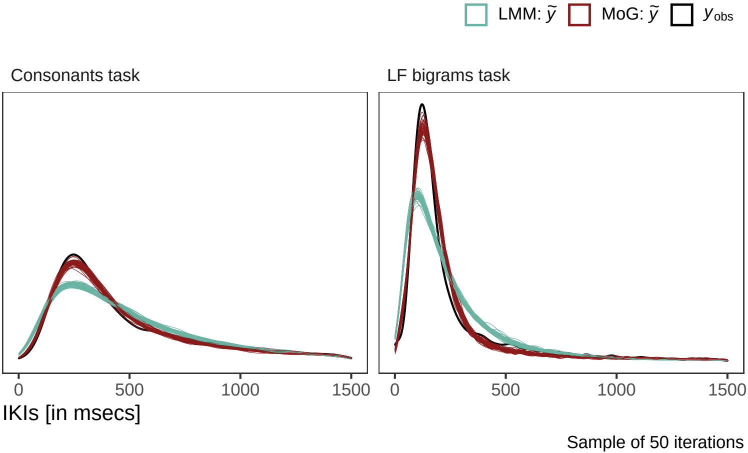
Observed vs. predicted IKIs



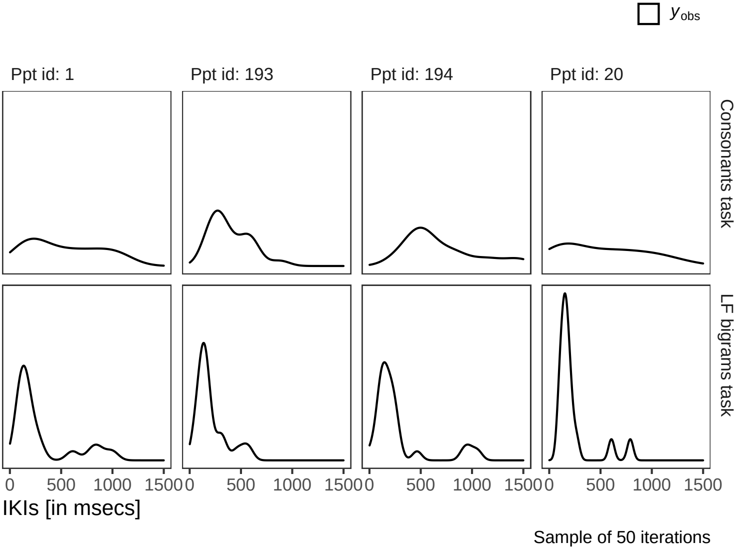
Observed vs. predicted IKIs



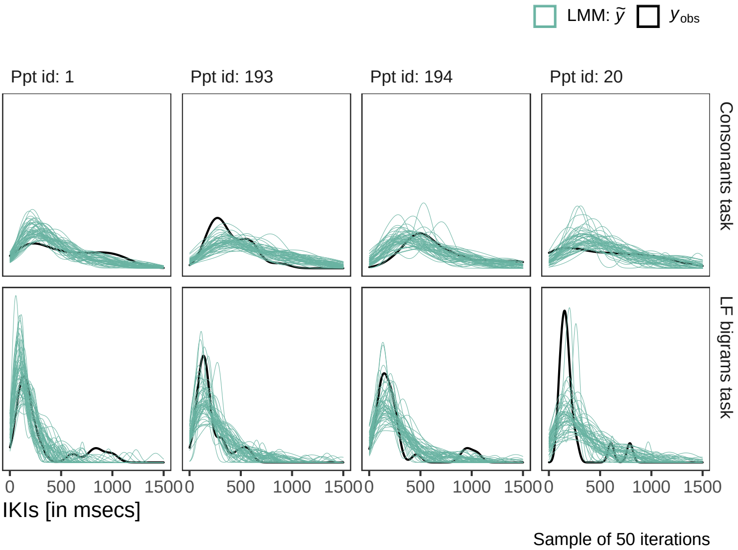
Observed vs. predicted IKIs



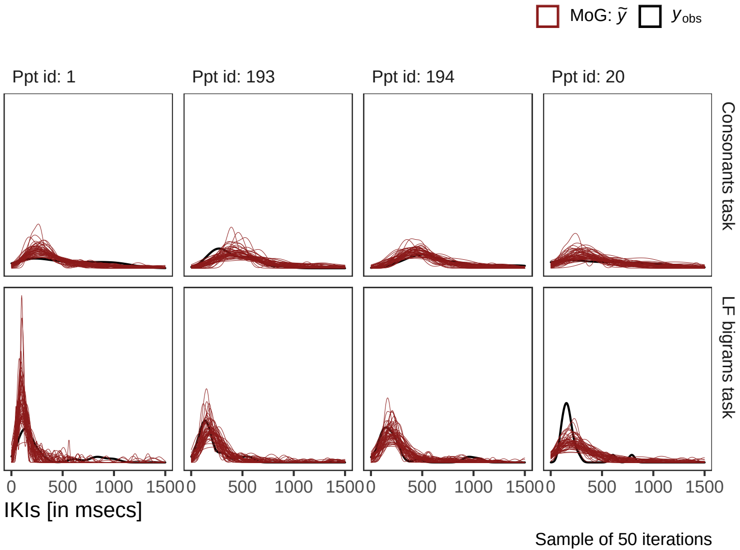
Observed vs. predicted IKIs



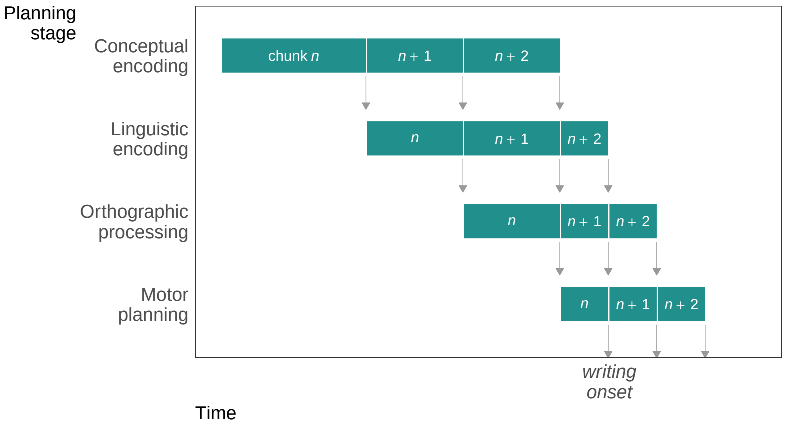
Observed vs. predicted IKIs



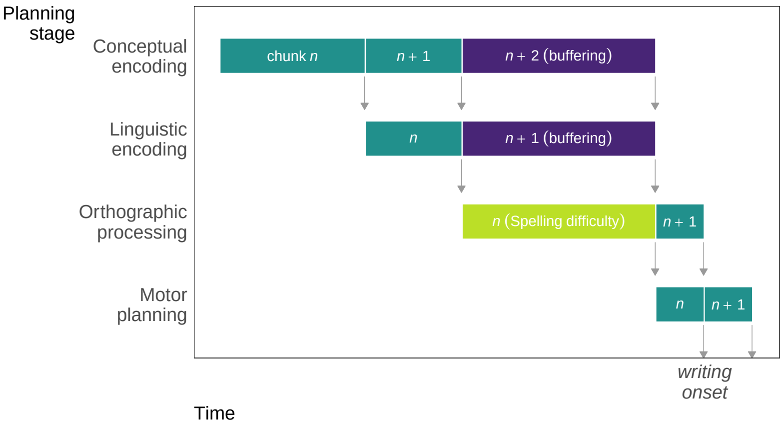
Observed vs. predicted IKIs



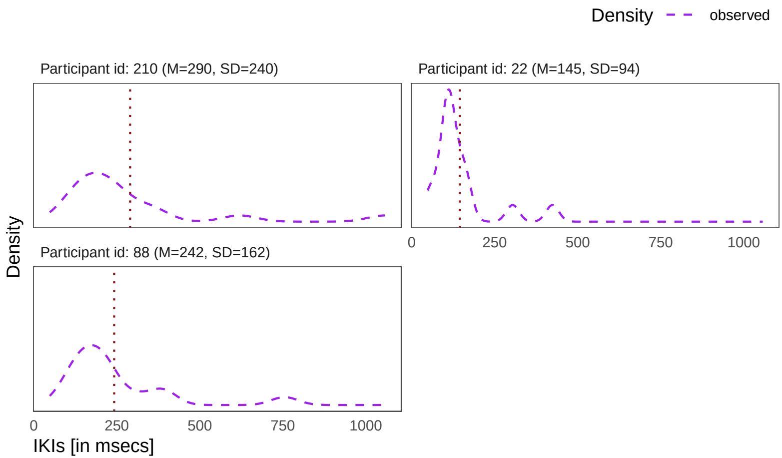
Planning cascade in writing



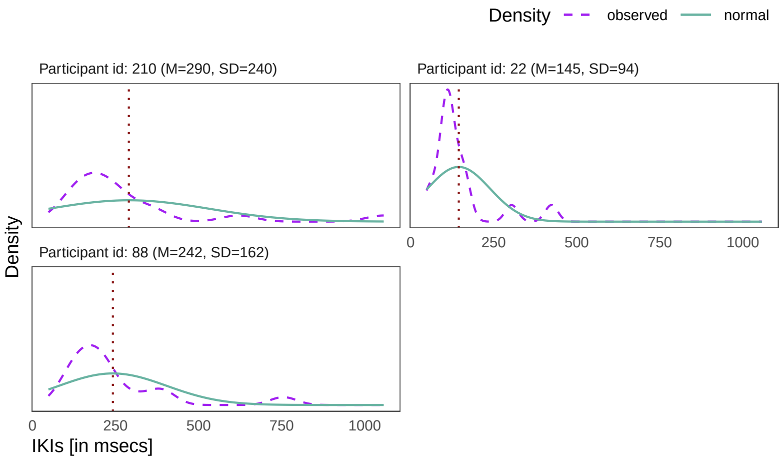
Planning cascade in writing



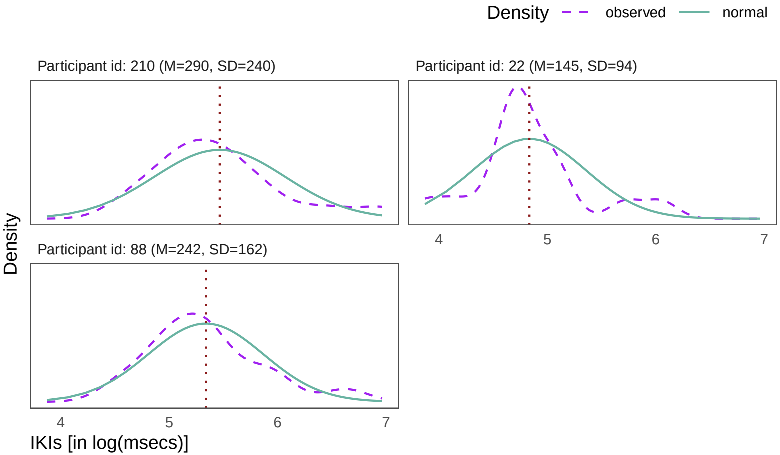
Keystroke transitions are not normal distributed



Keystroke transitions are not normal distributed



Keystroke transitions are not normal distributed



Long intervals are not bigram specific