No scope for planning – language pre-planning as mixture process

Jens Roeser

Mark Torrance   Mark Andrews   Thom Baguley

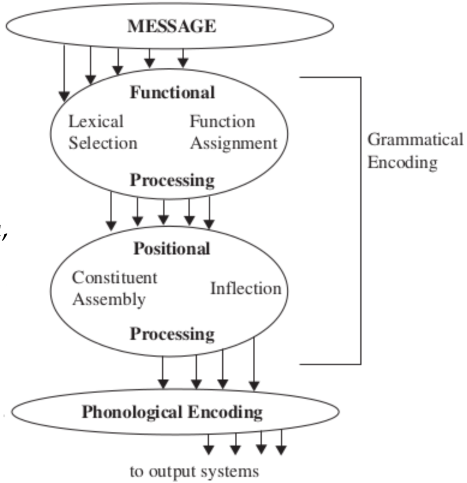
Nottingham Trent University, UK
jens.roeser@ntu.ac.uk

26th AMLaP, University of Potsdam

Sept 3, 2020

Turning ideas into language (Bock & Levelt, 1994)

- Unordered message units
- Output has linear order
- Linearisation via pragmatic, lexical and / or syntactic factors

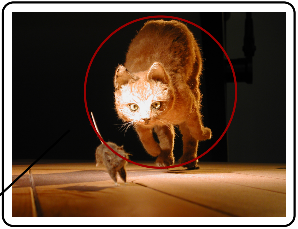


To what extent does syntax affect the linearisation of the output before production onset?

Syntax in language production (Bock & Ferreira, 2014)

1. Syntax is emergent property of lexically-driven planning
2. Syntax is generated from message representation
   a. **Deterministic:** syntax determines size of planning unit
   b. **Non-deterministic:** multiple candidate structures (Kempen & Hoenkamp, 1987)
3. Both routes (relational and non-relational) are available (at the message level; see Konopka & Meyer, 2014).

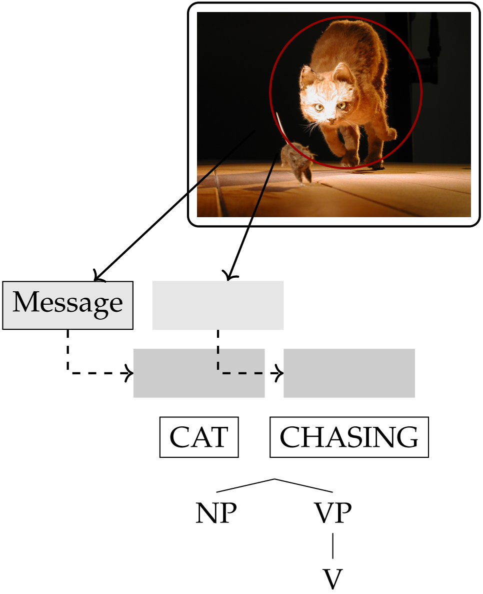
Consider the following evidence for possibility (2a).

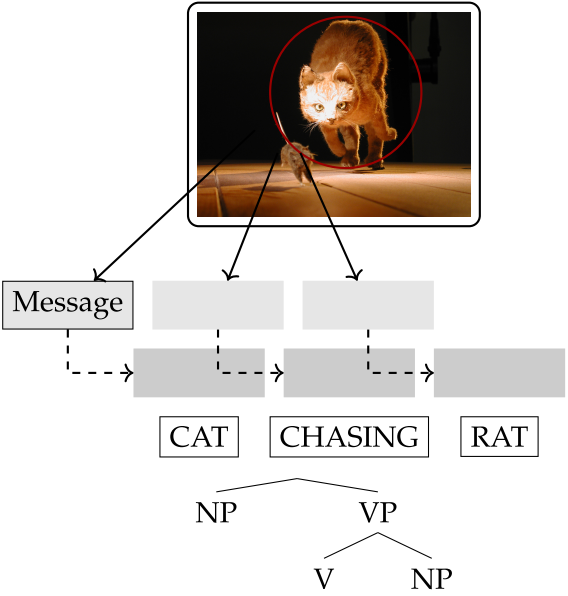


Message

CAT

NP

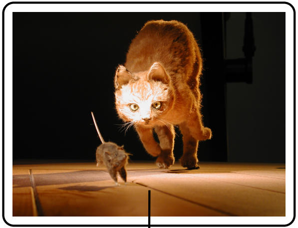

Message
CAT
CHASING
NP
VP
V


Message
CAT
CHASING
RAT
NP
VP
V
NP

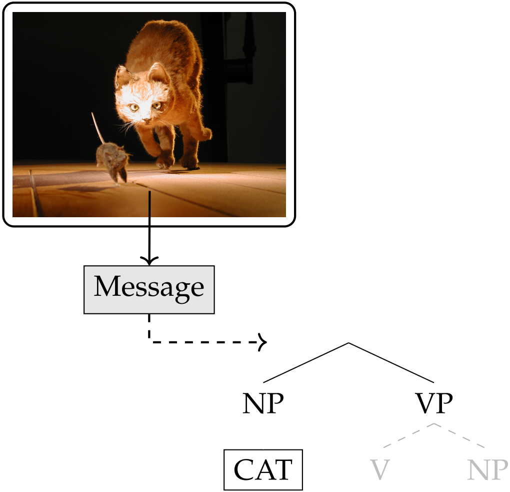
Syntax in language production (Bock & Ferreira, 2014)

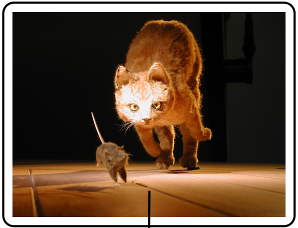
1. Syntax is emergent property of lexically-driven planning
2. Syntax is generated from message representation
   a. **Deterministic:** syntax determines size of planning unit
   b. **Non-deterministic:** multiple candidate structures (Kempen & Hoenkamp, 1987)
3. Both routes (relational and non-relational) are available (at the message level; see Konopka & Meyer, 2014).

Consider the following evidence for possibility (2a).



Message


Message
NP
VP
CAT
V
NP

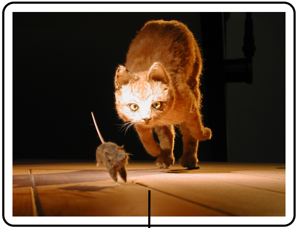


Message

NP VP

V NP

CHASE



Message

NP VP

V NP

RAT

Syntax in language production (Bock & Ferreira, 2014)

1. Syntax is emergent property of lexically-driven planning
2. Syntax is generated from message representation
   a. **Deterministic:** syntax determines size of planning unit
   b. **Non-deterministic:** multiple candidate structures (Kempen & Hoenkamp, 1987)
3. Both routes (relational and non-relational) are available (at the message level; see Konopka & Meyer, 2014).

Consider the following evidence for possibility (2a).

Syntax in language production (Bock & Ferreira, 2014)

1. Syntax is emergent property of lexically-driven planning
2. Syntax is generated from message representation
   a. **Deterministic:** syntax determines size of planning unit
   b. **Non-deterministic:** multiple candidate structures (Kempen & Hoenkamp, 1987)
3. Both routes (relational and non-relational) are available (at the message level; see Konopka & Meyer, 2014).

Consider the following evidence for possibility (2a).

Syntax in language production (Bock & Ferreira, 2014)

1. Syntax is emergent property of lexically-driven planning
2. Syntax is generated from message representation
   a. **Deterministic:** syntax determines size of planning unit
   b. **Non-deterministic:** multiple candidate structures (Kempen & Hoenkamp, 1987)
3. Both routes (relational and non-relational) are available (at the message level; see Konopka & Meyer, 2014).

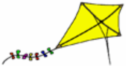
Consider the following evidence for possibility (2a).

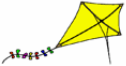
Syntax in language production (Bock & Ferreira, 2014)

1. Syntax is emergent property of lexically-driven planning
2. Syntax is generated from message representation
   a. **Deterministic:** syntax determines size of planning unit
   b. **Non-deterministic:** multiple candidate structures (Kempen & Hoenkamp, 1987)
3. Both routes (relational and non-relational) are available (at the message level; see Konopka & Meyer, 2014).

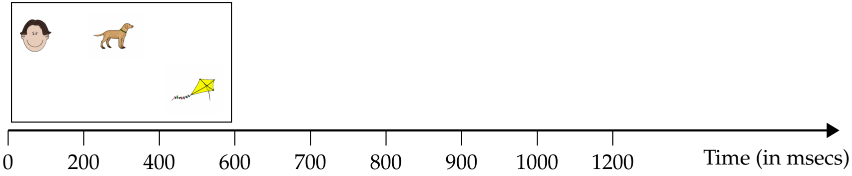
Consider the following evidence for possibility (2a).

+

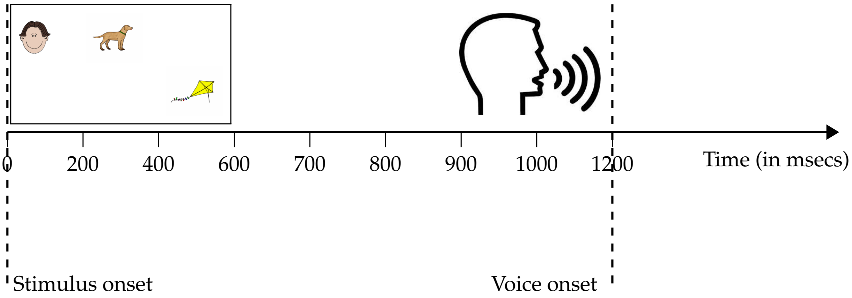




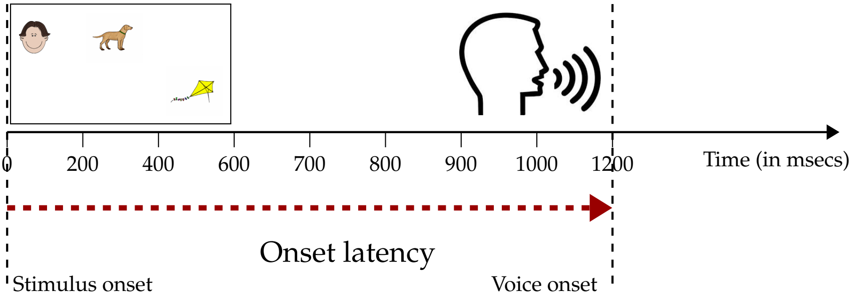
The boy and the dog moved above the kite.



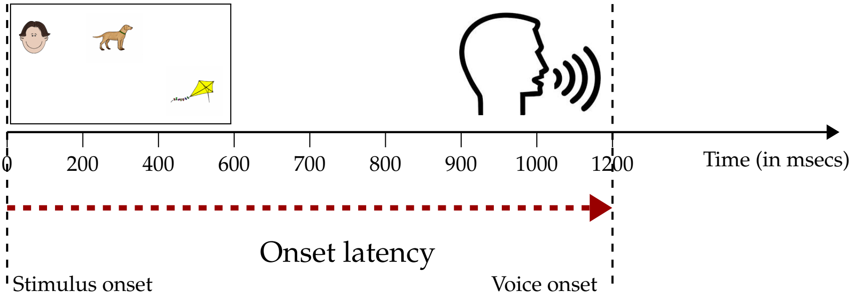
1. **The boy and the dog** moved above the kite.
2. **The boy** moved above the dog and the kite.



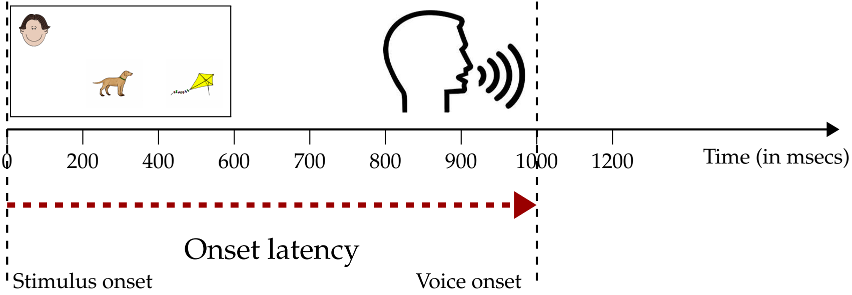
1. **The boy and the dog** moved above the kite.
2. **The boy** moved above the dog and the kite.



1. **The boy and the dog** moved above the kite.
2. **The boy** moved above the dog and the kite.



1. **The boy and the dog** moved above the kite.
2. **The boy** moved above the dog and the kite.



1. **The boy and the dog** moved above the kite.
2. **The boy** moved above the dog and the kite.

Preplanning is guided by syntax

- Frequently reproduced systematic slowdown for conjoined NPs (e.g. Martin et al., 2014; Smith & Wheeldon, 1999; Wagner et al., 2010; Wheeldon et al., 2013).
- "Phrase as default planning scope" (Martin et al., 2010)
- NP syntax must be planned before onset; hence determines planning scope.

Alternative hypothesis

- **Preplanning beyond the first noun is more likely but not obligated by the phrase syntax** because, for example, …

1. Fluency requires preplanning of B in *The A and the B moved …* when there isn't enough time to plan B in parallel to articulation (Allum & Wheeldon, 2007; Griffin, 2003).
2. Correction of incorrectly activated candidate syntax.
3. Relational message-level route activated syntactic route.

Alternative hypothesis

- **Preplanning beyond the first noun is more likely but not obligated by the phrase syntax** because, for example, …

1. Fluency requires preplanning of B in *The A and the B moved* … when there isn't enough time to plan B in parallel to articulation (Allum & Wheeldon, 2007; Griffin, 2003).
2. Correction of incorrectly activated candidate syntax.
3. Relational message-level route activated syntactic route.

Alternative hypothesis

- **Preplanning beyond the first noun is more likely but not obligated by the phrase syntax** because, for example, …

1. Fluency requires preplanning of B in *The A and the B moved* … when there isn't enough time to plan B in parallel to articulation (Allum & Wheeldon, 2007; Griffin, 2003).
2. Correction of incorrectly activated candidate syntax.
3. Relational message-level route activated syntactic route.

Alternative hypothesis

- **Preplanning beyond the first noun is more likely but not obligated by the phrase syntax** because, for example, ...

1. Fluency requires preplanning of B in *The A and the B moved* ... when there isn't enough time to plan B in parallel to articulation (Allum & Wheeldon, 2007; Griffin, 2003).
2. Correction of incorrectly activated candidate syntax.
3. Relational message-level route activated syntactic route.

Research focus

- Direct comparison of two hypotheses re onset-latency slowdown for conjoined NPs.
  i. Phrase scope obligated by the production system, leading to a systematic slowdown.
  ii. Preplanning beyond the first noun is more likely for conjoined NPs but not obligated by the production system.
- Implementation of both hypotheses as statistical models in Stan (Carpenter et al., 2017); code based on Sorensen et al. (2016) and Vasishth, Chopin et al. (2017); also Vasishth, Jäger et al. (2017).

Research focus

- Direct comparison of two hypotheses re onset-latency slowdown for conjoined NPs.

  i. Phrase scope obligated by the production system, leading to a systematic slowdown.

  ii. Preplanning beyond the first noun is more likely for conjoined NPs but not obligated by the production system.

- Implementation of both hypotheses as statistical models in Stan (Carpenter et al., 2017); code based on Sorensen et al. (2016) and Vasishth, Chopin et al. (2017); also Vasishth, Jäger et al. (2017).

Research focus

- Direct comparison of two hypotheses re onset-latency slowdown for conjoined NPs.

i. Phrase scope obligated by the production system, leading to a systematic slowdown.

ii. Preplanning beyond the first noun is more likely for conjoined NPs but not obligated by the production system.

- Implementation of both hypotheses as statistical models in Stan (Carpenter et al., 2017); code based on Sorensen et al. (2016) and Vasishth, Chopin et al. (2017); also Vasishth, Jäger et al. (2017).

Research focus

- Direct comparison of two hypotheses re onset-latency slowdown for conjoined NPs.
  i. Phrase scope obligated by the production system, leading to a systematic slowdown.
  ii. Preplanning beyond the first noun is more likely for conjoined NPs but not obligated by the production system.
- Implementation of both hypotheses as statistical models in Stan (Carpenter et al., 2017); code based on Sorensen et al. (2016) and Vasishth, Chopin et al. (2017); also Vasishth, Jäger et al. (2017).

Pooled re-analysis of 8 experiments

- Hardy et al. (2019): 90 ppts; 36 items
- Hardy et al. (2020): 105 ppts; 80 items
- Martin et al. (2010): 3×12 ppts; 2×48 and 1×64 items
- Roeser et al. (2019): 3×32 ppts; 96 items

- 2 conditions

a. **Conjoined NPs:** *The boy and the dog moved above the kite.*

b. **Simple NPs:** *The boy moved above the dog and the kite.*

Pooled Linear Mixed Effects Model (null hypothesis)

$$y_{ijk} \sim LogNormal(\mu_{ijk}, \sigma^2_{e_k})$$
$$\mu_{ijk} = \alpha_k + u_i + w_j$$

- Non-centred mean α_k each for $k = 1, \ldots, K$ where K is the number of studies (Gelman et al., 2014) with pooled latency α_μ.
- Participants: $u_i \sim Normal(0, \sigma^2_u)$
- Items: $w_j \sim Normal(0, \sigma^2_w)$
- Pooled error variance σ^2_e

Pooled Linear Mixed Effects Model

$$y_{ijk} \sim LogNormal(\mu_{ijk}, \sigma^2_{e_k})$$
$$\mu_{ijk} = \alpha_k + \boxed{\beta_k \cdot x_{[0,1]}} + u_i + w_j$$

- $x = 0$ for simple NPs; $x = 1$ for conjoined NPs.
- β_k: by-study NP difference with pooled effect β_μ.
- **Conceptual idea:**
  - Underlying process can be characterised as **one distribution**.
  - Deterministic syntax-driven model: conjoined NPs slow down preplanning by β msecs.

Pooled Mixture of Gaussians

$$\begin{aligned} y_{ijk} \sim \theta_{NP_k} \cdot LogNormal(\mu_{ijk} + \delta_k, \sigma^2_{e'_k}) + \\ (1 - \theta_{NP_k}) \cdot LogNormal(\mu_{ijk}, \sigma^2_{e_k}) \\ \mu_{ijk} = \alpha_k + u_i + w_j \end{aligned}$$

- **Conceptual idea:**
- Underlying process is **mixture of two distributions**:

i. onset latency with variance $\sigma^2_{e_k}$

ii. slowdown δ with larger variance $\sigma^2_{e'_k}$

- θ captures the probability of long latencies by NP type.

Pooled Mixture of Gaussians

$$y_{ijk} \sim \theta_{NP_k} \cdot LogNormal(\mu_{ijk} + \delta_k, \sigma^2_{e'_k}) +$$
$$(1 - \theta_{NP_k}) \cdot \boxed{LogNormal(\mu_{ijk}, \sigma^2_{e_k})}$$
$$\mu_{ijk} = \alpha_k + u_i + w_j$$

- **Conceptual idea:**
- Underlying process is **mixture of two distributions**:

i. onset latency with variance $\sigma^2_{e_k}$

ii. slowdown δ with larger variance $\sigma^2_{e'_k}$

- θ captures the probability of long latencies by NP type.

Pooled Mixture of Gaussians

$$y_{ijk} \sim \theta_{NP_k} \cdot \boxed{LogNormal(\mu_{ijk} + \delta_k, \sigma^2_{e'_k})} +$$
$$(1 - \theta_{NP_k}) \cdot LogNormal(\mu_{ijk}, \sigma^2_{e_k})$$
$$\mu_{ijk} = \alpha_k + u_i + w_j$$

- **Conceptual idea:**
  - Underlying process is **mixture of two distributions**:
    i. onset latency with variance $\sigma^2_{e_k}$
    ii. slowdown δ with larger variance $\sigma^2_{e'_k}$
  - θ captures the probability of long latencies by NP type.

Pooled Mixture of Gaussians

$$y_{ijk} \sim \boxed{\theta_{NP_k}} \cdot LogNormal(\mu_{ijk} + \delta_k, \sigma^2_{e'_k}) +$$
$$(1 - \theta_{NP_k}) \cdot LogNormal(\mu_{ijk}, \sigma^2_{e_k})$$
$$\mu_{ijk} = \alpha_k + u_i + w_j$$

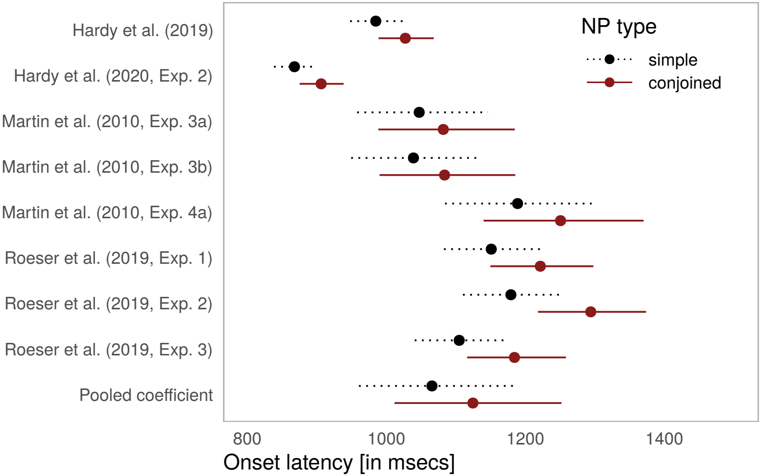
- **Conceptual idea:**
  - Underlying process is **mixture of two distributions**:
    - i. onset latency with variance $\sigma^2_{e_k}$
    - ii. slowdown δ with larger variance $\sigma^2_{e'_k}$
  - θ captures the probability of long latencies by NP type.

Pooled Mixture of Gaussians

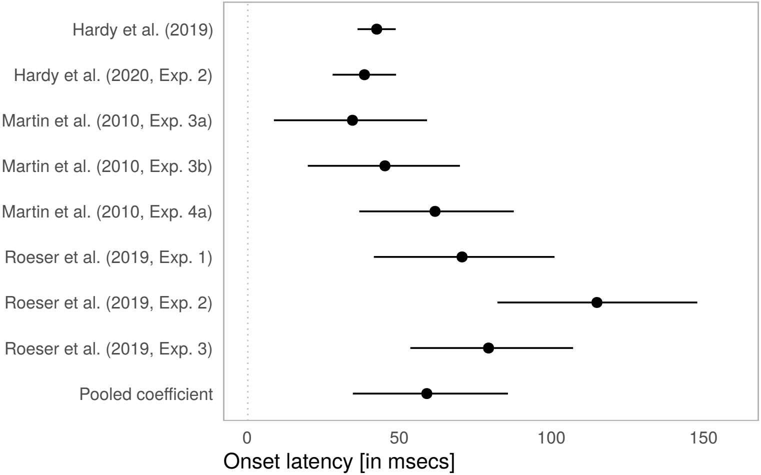
$$
\begin{aligned}
y_{ijk} \sim \boxed{\theta_{NP_k}} \cdot LogNormal(\mu_{ijk} + \delta_k, \sigma^2_{e'_k}) + \\
(1 - \theta_{NP_k}) \cdot LogNormal(\mu_{ijk}, \sigma^2_{e_k}) \\
\mu_{ijk} = \alpha_k + u_i + w_j
\end{aligned}
$$

- **Conceptual idea:**
- Planning beyond first noun is possible for both NP types resulting in a slowdown δ.
- Non-deterministic model: planning beyond the first noun is more likely for conjoined NPs reflected in larger probability of long onset latencies θ.

NP effect (LMM)



NP effect (LMM)



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.

Models	$\Delta\widehat{elpd}$	$\widehat{elpd}$	Description
MoG-1			Mixing proportions by NP-type
MoG-0			Null model (no NP difference)
LMM-1			Slowdown for conjoined NPs
LMM-0			Null model (no NP difference)
LMM-2			Larger variance for conjoined NPs

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.

Models	$\Delta\widehat{elpd}$	$\widehat{elpd}$	Description
MoG-1	–	-201,486 (176)	Mixing proportions by NP-type
MoG-0	-15 (8)	-201,500 (176)	Null model (no NP difference)
LMM-1	-1,006 (97)	-202,492 (214)	Slowdown for conjoined NPs
LMM-0	-1,192 (92)	-202,678 (212)	Null model (no NP difference)
LMM-2	-3,537 (214)	-205,022 (307)	Larger variance for conjoined NPs

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

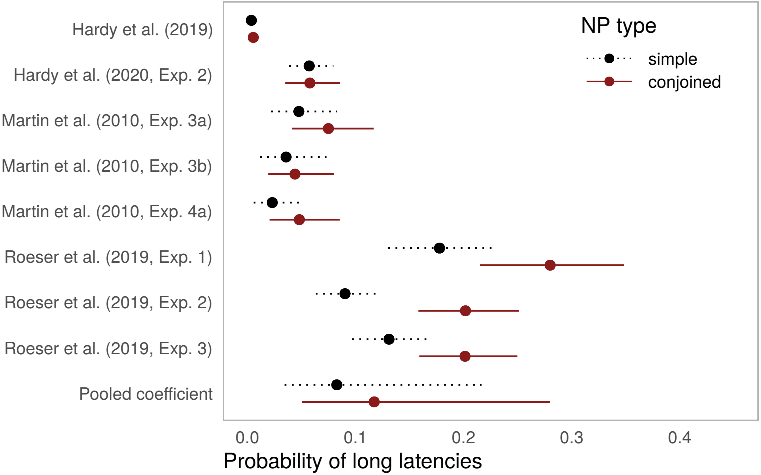
Model comparisons

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.

Models	$\Delta\widehat{elpd}$	$\widehat{elpd}$	Description
MoG-1	–	-201,486 (176)	Mixing proportions by NP-type
MoG-0	-15 (8)	-201,500 (176)	Null model (no NP difference)
LMM-1	-1,006 (97)	-202,492 (214)	Slowdown for conjoined NPs
LMM-0	-1,192 (92)	-202,678 (212)	Null model (no NP difference)
LMM-2	-3,537 (214)	-205,022 (307)	Larger variance for conjoined NPs

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

Probability of long latencies (MoG)



Bridge to follow-up experiments

- Slowdown for conjoined NPs is better explained by a larger, yet relatively small, probability of long latencies.
- Most studies in our pool included other manipulations.

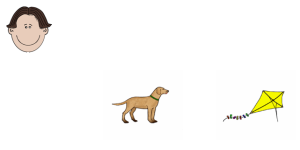
- **Experiment 1:**
  - Reproduce analysis after …

  i. Reducing the manipulation to simple and conjoined NPs.

  ii. Controlling image names
- **Experiment 2:**
  - Assess impact of visual manipulation (as in Martin et al., 2010).

Methods

Condition 1: A and B moved above C



Condition 2: A moved above B and C

- 48 items; 96 fillers; 6 practice trials
- First noun: *Peter* or *Tania*
- Image names: high frequency and naming agreement.

- **Experiment 1:**
  - 1a. **Peter and the dog** moved above the kite
  - 1b. **Peter** moved above the dog and the kite
  - 78 ppts (after cleaning)
- **Experiment 2:**
  - 2. Peter, the dog, the kite
  - 45 ppts (after cleaning)

Methods

Condition 1: A and B moved above C

Condition 2: A moved above B and C

- 48 items; 96 fillers; 6 practice trials
- First noun: *Peter* or *Tania*
- Image names: high frequency and naming agreement.

- **Experiment 1:**
  1a. **Peter and the dog** moved above the kite
  1b. **Peter** moved above the dog and the kite
  - 78 ppts (after cleaning)
- **Experiment 2:**
  2. Peter, the dog, the kite
  - 45 ppts (after cleaning)

Methods

Condition 1: A and B moved above C

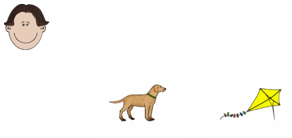
Condition 2: A moved above B and C

- 48 items; 96 fillers; 6 practice trials
- First noun: *Peter* or *Tania*
- Image names: high frequency and naming agreement.

- **Experiment 1:**
  1a. **Peter and the dog** moved above the kite
  1b. **Peter** moved above the dog and the kite
  - 78 ppts (after cleaning)
- **Experiment 2:**
  2. Peter, the dog, the kite
  - 45 ppts (after cleaning)

Methods

Condition 1: A and B moved above C

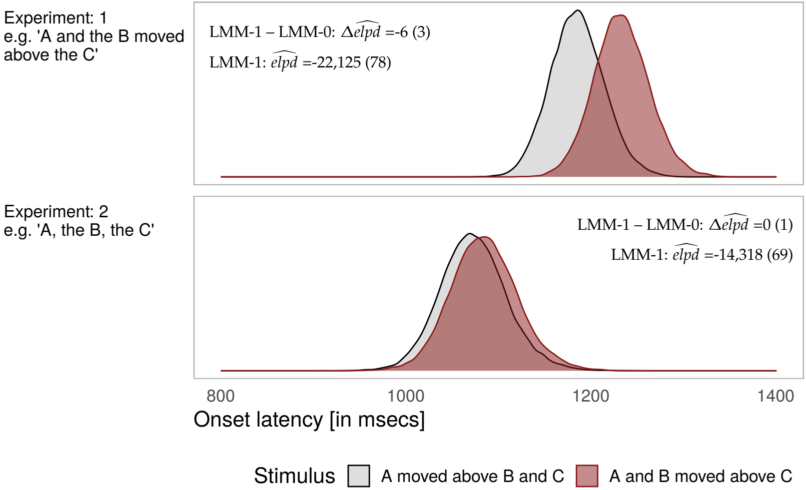


Condition 2: A moved above B and C

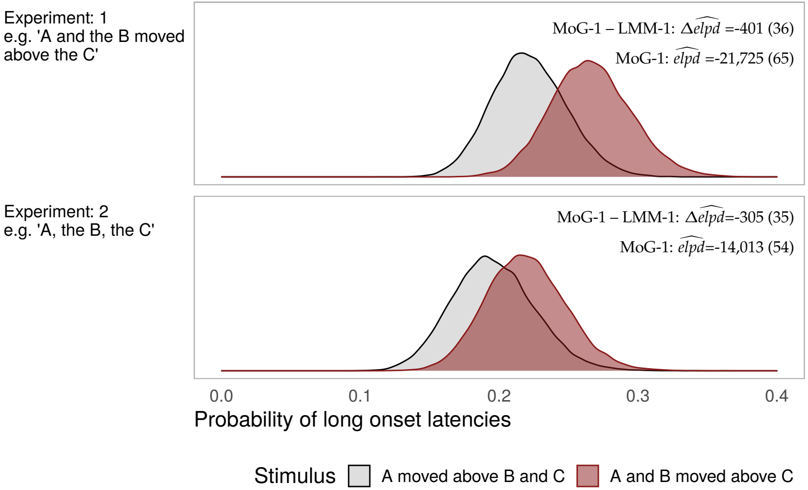
- 48 items; 96 fillers; 6 practice trials
- First noun: *Peter* or *Tania*
- Image names: high frequency and naming agreement.

- **Experiment 1:**
  1a. **Peter and the dog** moved above the kite
  1b. **Peter** moved above the dog and the kite
  - 78 ppts (after cleaning)
- **Experiment 2:**
  2. Peter, the dog, the kite
  - 45 ppts (after cleaning)

NP-type effect (LMM)



Probability of long latencies (MoG)



Summary

- No evidence for phrase-as-planning-unit hypothesis: NP syntax isn't obligated by production system.
- Instead, the slowdown for conjoined NPs (as in Martin et al., 2010; Smith & Wheeldon, 1999) is better explained by a larger probability of long latencies which, however, remained in a minority.
- Syntax in language production must result from a non-deterministic planning mechanism.

Thank you for listening!

…and **Randi Martin**, **Jason Crowther**, and **Sophie Hardy** (et al.) for sharing their data,
…and **Dora Kramar** and **Andra Tanasescu** for supporting the data collection.

email: jens.roeser@ntu.ac.uk
nottinghamtrent.academia.edu/JensRoeser
R and *Stan* code: github.com/jensroes/NP-effect

References I

Allum, P. H. & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 791–810.

Bock, J. K. & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). Academic Press.

Bock, J. K. & Ferreira, V. S. (2014). Syntactically speaking. In M. Goldrick, V. S. Ferreira & M. Miozzo (Eds.), *The Oxford Handbook of Language Production* (pp. 21–46). Oxford University Press.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Chapman; Hall/CRC.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review*, *10*(3), 603–609.

Hardy, S. M., Segaert, K. & Wheeldon, L. (2019). Age-related disruption in the use of lexical information during sentence production, despite preserved syntactic planning. *PsyArXiv*.

References II

Hardy, S. M., Segaert, K. & Wheeldon, L. (2020). Healthy aging and sentence production: Disrupted lexical access in the context of intact syntactic planning. *Frontiers in Psychology*, *11*, 257.

Kempen, G. & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive science*, *11*(2), 201–258.

Konopka, A. E. & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1–40.

Martin, R. C., Crowther, J. E., Knight, M., Tamborello II, F. P. & Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, *116*(2), 177–192.

Martin, R. C., Yan, H. & Schnur, T. T. (2014). Working memory and planning during sentence production. *Acta Psychologica*, *152*, 120–132.

Roeser, J., Torrance, M. & Baguley, T. (2019). Advance planning in written and spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(11), 1983–2009.

Smith, M. & Wheeldon, L. R. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*, 205–246.

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.

References III

Swets, B., Jacovina, M. E. & Gerrig, R. J. (2014). Individual differences in the scope of speech planning: Evidence from eye-movements. *Language and Cognition*, *6*(1), 12–44.

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*.

Vasishth, S., Jäger, L. A. & Nicenboim, B. (2017). Feature overwriting as a finite mixture process: Evidence from comprehension data. *arXiv preprint arXiv:1703.04081*.

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.

Wagner, V., Jescheniak, J. D. & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 423–440.

Wheeldon, L. R., Ohlson, N., Ashby, A. & Gator, S. (2013). Lexical availability and grammatical encoding scope during spoken sentence production. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1653–1673.

Meta LMM

$$
\begin{aligned}
y_{ijk} &\sim LogNormal(\mu_{ijk}, \sigma^2_{e_k}) \\
\mu_{ijk} &= \alpha_k + \beta_k \cdot x_{[0,1]} + u_i + w_j \\
\alpha_k &= \alpha_\mu + \alpha_\tau \cdot \alpha_{\eta_k} \\
\beta_k &= \beta_\mu + \beta_\tau \cdot \beta_{\eta_k} \\
\text{constraint: } \sigma^2_{e_k} &> 0
\end{aligned}
$$

- $x = 0$ for simple NPs; $x = 1$ for conjoined NPs.

Meta LMM (unequal variance)

$$y_{ijk} \sim \begin{cases} LogNormal(\mu_{ijk}, \sigma^2_{e_k}), & \text{if NP}_{ijk} = \text{simple} \\ LogNormal(\mu_{ijk} + \beta_k, \sigma^2_{e'_k}) & \text{else if NP}_{ijk} = \text{conjoined} \end{cases}$$

$$\mu_{ijk} = \alpha_k + u_i + w_j$$

$$\alpha_k = \alpha_\mu + \alpha_\tau \cdot \alpha_{\eta_k}$$

$$\beta_k = \beta_\mu + \beta_\tau \cdot \beta_{\eta_k}$$

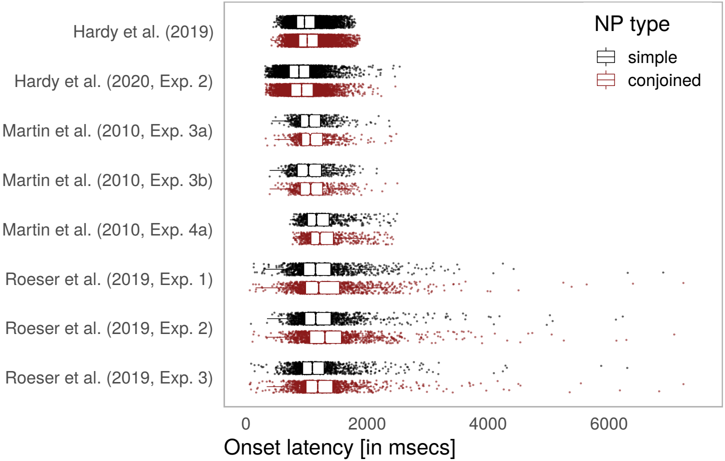
$$\text{constraint: } \sigma^2_{e_k} > 0$$

$$\sigma^2_{e'_k} > \sigma^2_{e_k}$$

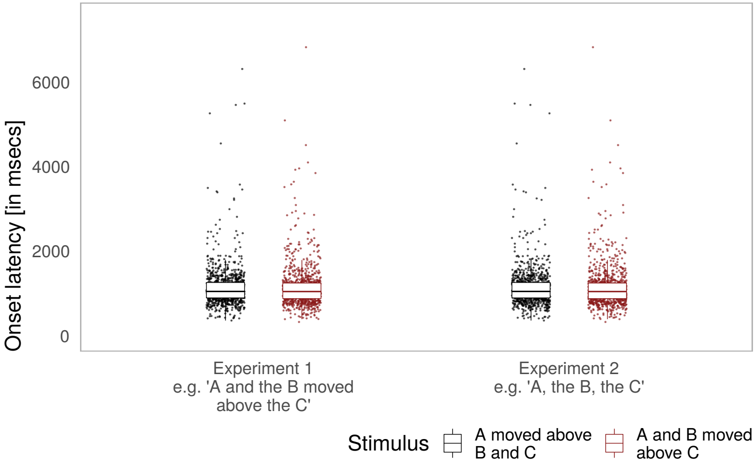
Meta mixture model (alternative hypothesis)

$$
\begin{aligned}
y_{ijk} \sim \theta_{NP_k} \cdot LogNormal(\mu_{ijk} + \delta_k, \sigma^2_{e'_k}) + \\
(1 - \theta_{NP_k}) \cdot LogNormal(\mu_{ijk}, \sigma^2_{e_k}) \\
\mu_{ijk} = \alpha_k + u_i + w_j \\
\alpha_k = \alpha_\mu + \alpha_\tau \cdot \alpha_{\eta_k} \\
\theta_{NP_k} = Logit^{-1}(\phi_{NP_k}) \\
\theta_{\mu_{NP}} = Logit^{-1}(\phi_{\mu_{NP}}) \\
\phi_{NP_k} \sim Normal(\phi_{\mu_{NP}}, \phi^2_\tau) \\
\delta_k \sim Normal(\delta_\mu, \delta^2_\tau) \\
\text{constraint: } \delta_k, \sigma^2_{e_k}, \phi_\tau, \delta_\tau > 0 \\
\sigma^2_{e'_k} > \sigma^2_{e_k}
\end{aligned}
$$

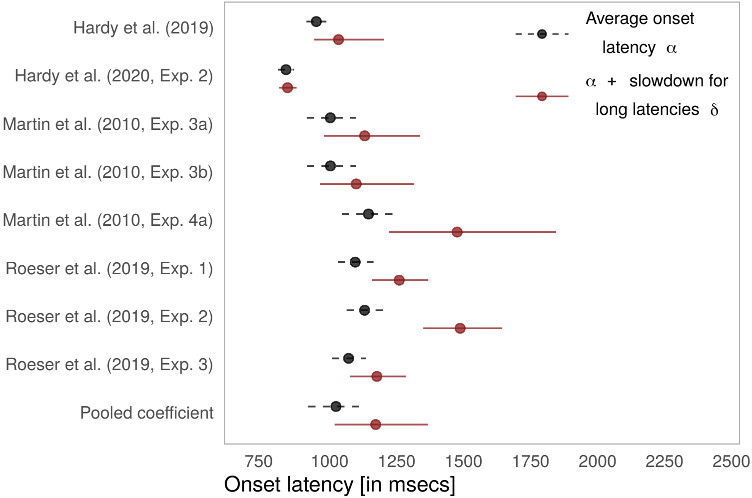
Data overview (pooled data)



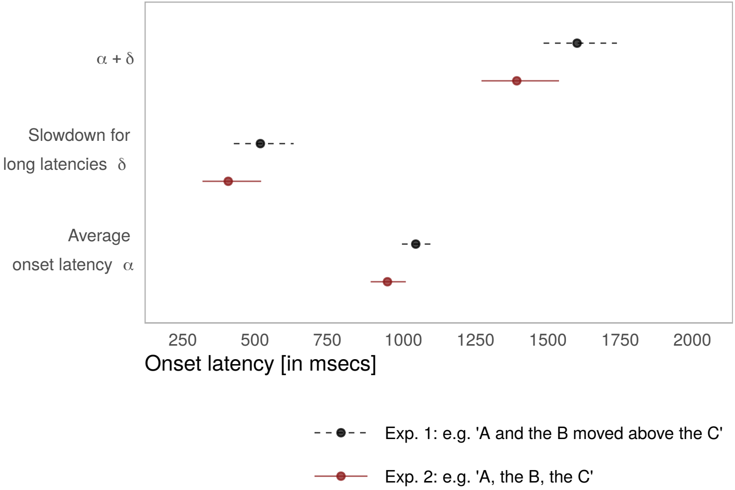
Data overview (Experiments 1& 2)



Onset-latency coefficients (pooled mixture components)



Onset-latency coefficients (mixture components)



Model comparisons

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

	Model	$\Delta\widehat{elpd}$	$\widehat{elpd}$
Expeirment 1	MoG-1	–	-21,724 (65)
	MoG-0	-1 (2)	-21,725 (65)
	LMM-1	-401 (36)	-22,125 (78)
	LMM-2	-402 (36)	-22,126 (78)
	LMM-0	-406 (36)	-22,131 (78)
Expeirment 2	MoG-0	–	-14,012 (54)
	MoG-1	-1 (1)	-14,013 (54)
	LMM-0	-306 (35)	-14,318 (69)
	LMM-1	-306 (35)	-14,318 (69)
	LMM-2	-308 (35)	-14,319 (69)

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