Documentation of the data extraction for the meta-analysis published in "Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis", Journal of Memory and Language 94, 316–339, 2017.

Lena A. Jäger, Felix Engelmann, & Shravan Vasishth

March 1st, 2017

1 Introduction

Here, we document how we extracted the estimates that went into our Bayesian meta-analysis from the respective publications or — if available — from the (preprocessed) data obtained from the authors. For details of the meta-analysis, see:
https://github.com/vasishth/MetaAnalysisJaegerEngelmannVasishth2017

1.1 General decision criteria

1. Only reading experiments using eye-tracking or self-paced reading were included.

2. In the case of eye-tracking experiments, we always use first-pass reading time. The reason for this decision was that different studies report different eye-tracking dependent measures and different dependent measures came out significant in the various studies, making comparison difficult. Since all studies report first-pass reading time, we decided to use this measure.

3. Whenever we obtained data from the authors of a publications, these data already has been preprocessed by the authors (i.e., the regions of interest and reading measures have already been computed).

4. We always stick to interest area partitioning used by the authors of the respective publication.

5. With the term 'critical region' we refer to the region containing the verb or reflexive/reciprocal. Note that in some experiments, this region also contains material (characters or whole words) to the left or the right of the verb or reflexive/reciprocal.[1]

6. With the term 'post-critical region' we refer to the region directly following the critical region, no matter how many characters/words this region spanned.

7. We only considered the critical and the post-critical regions.

8. In case an effect was reported as significant at the post-critical region but not at the critical region, we use the effect size at the post-critical region. Otherwise, we always use the effect observed at the critical region.

9. The condition labels vary considerably between the different experiments. For example, the term 'match' sometimes refers to a feature match between the distractor and the target, whereas in other publications it refers to the target's and distractor's match with the manipulated retrieval cue. Moreover, the specific research question of the experiments varied, meaning that in some studies different conditions were compared to each other than the contrasts that we are modeling. We therefore recoded the condition labels and the comparisons in order to get the correct effect estimates for the purpose of our meta-analysis.

---

[1] It was impossible to partition the regions of interest of all experiments in the same way as the necessary information was not always available.

10. Whenever the authors provided the (preprocessed) data to us, we fit a linear mixed model with varying intercepts and slopes for both items and subjects. No correlations were fit between varying intercepts and slopes. We fit the model on non-transformed first-pass reading times (eye-tracking) or reading times (self-paced reading) in order to replicate the author's analysis as closely as possible. We coded the main effect of grammaticality (in case both target-match and target-mismatch conditions were tested) and pairwise comparisons comparing the distractor-match condition (coded as +0.5) with the distractor-mismatch condition (coded as −0.5) within target-match and target-mismatch conditions and (if applicable) target-type (match vs mismatch) as fixed effects. We use the coefficients of the pairwise comparisons and the standard errors associated with them for the meta-analysis.

11. When we use the numbers provided in the respective publication, we either directly use the corresponding coefficients from a linear mixed model (if provided by the authors) or calculate the difference between the respective condition means. The standard errors (if not directly reported in the paper) are derived from the confidence intervals or from the standard deviation and the sample size, dependent on whatever numbers are provided by the authors.

12. Whenever possible, we applied the same data trimming procedure as did the authors. Note that in several cases, this had a considerable impact on the estimates: the results are quite different depending on whether one applies trimming or not.

13. As mentioned above, all analyses were done on raw reading time in milliseconds. A better way would have been to work with log-transformed measures; but this was possible to do only when we had the raw data (see Appendix B of paper).

2 Data extraction of the individual experiments

2.1 Experiments on subject-verb agreement

1. **Dillon et al. (2013, Experiment 1, agreement conditions)**

   Source of the estimates:
   Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model on non-transformed first-pass reading times with varying intercepts and slopes for subjects and items. No correlations were fit between varying intercepts and slopes. We fit a single model for the agreement conditions and the reflexives conditions with the main effect of dependency type as predictor in addition to the relevant pairwise comparisons.

   Target-match (singular verb)
   Region: critical
   Effect: -14ms
   Standard Error: 16ms

   Target-mismatch (plural verb)
   Region: critical
   Effect: -7ms
   Standard Error: 22ms

2. **Franck et al. (2015, Experiment 1, conditions with complement clauses)**

Source of the estimates:
Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model on non-transformed reading times with varying intercepts and slopes for subjects and items. No correlations were fit between varying intercepts and slopes. We removed the same items as did the authors (personal communication), which resulted in a total of 11 items only.

Target-match (singular verb)
Region:	critical
Effect:	32ms
Standard Error:	33ms

Target-mismatch
not available

3. **Franck et al. (2015, Experiment 1, conditions with relative clauses)**

   Source of the estimates:
   Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model on non-transformed reading times with varying intercepts and slopes for subjects and items. No correlations were fit between varying intercepts and slopes. We removed the same items as did the authors (personal communication), which resulted in a total of 11 items only.

   Target-match (singular verb)
Region:	critical
Effect:	110ms
Standard Error:	48ms

   Target-mismatch
   not available

4. **Lago et al. (2015, Experiment 1)**

   Source of the estimates:
   Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model on non-transformed reading times with varying intercepts and slopes for subjects and items. No correlations were fit between varying intercepts and slopes. The same data trimming procedure was used as by the authors.

   Target-match (singular verb)
Region:	critical
Effect:	-4ms
Standard Error:	14ms

   Target-mismatch (plural verb)
Region:	post-critical
Effect:	-40ms
Standard Error:	14ms

5. **Lago et al. (2015, Experiment 2)**

Source of the estimates:
Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model on non-transformed reading times with varying intercepts and slopes for subjects and items. No correlations were fit between varying intercepts and slopes. The same data trimming procedure was used as by the authors.

Target-match (singular verb)
Region:	critical
Effect:	-7ms
Standard Error:	8ms

Target-mismatch (plural verb)
Region:	post-critical
Effect:	-36ms
Standard Error:	18ms

6. **Lago et al. (2015, Experiment 3a)**

   Source of the estimates:
   Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model on non-transformed reading times with varying intercepts and slopes for subjects and items. No correlations were fit between varying intercepts and slopes. The same data trimming procedure was used as by the authors.

   Target-match (singular verb)
Region:	post-critical
Effect:	-12ms
Standard Error:	6ms

   Target-mismatch (plural verb)
Region:	post-critical
Effect:	-15ms
Standard Error:	7ms

7. **Lago et al. (2015, Experiment 3b)**

   Source of the estimates:
   Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model on non-transformed reading times with varying intercepts and slopes for subjects and items. The same data trimming procedure was used as by the authors. No correlations were fit between varying intercepts and slopes.

   Target-match (singular verb)
Region:	critical
Effect:	12ms
Standard Error:	9ms

   Target-mismatch (plural verb)
Region:	post-critical
Effect:	-22ms
Standard Error:	11ms

8. **Pearlmutter et al. (1999, Experiment 1)**

Source of the estimates:
Numbers provided in the paper. The authors provide condition means and confidence intervals of residualized reading times by items and by participants. For the means, we take the average of both values. We compute the standard error from the larger CI. We are assuming that the confidence intervals that the authors provide on Table C1 of Appendix C is referring to $+/-2 \cdot SE$ (one-sided rather than two sided confidence interval). For the computation of the standard error, we therfore divide this number by 2 rather than by 4: $SE = 0.5 \cdot CI$

Target-match (singular verb)
Region:	critical
Effect:	-35ms
Standard Error:	10ms

Target-mismatch (plural verb)
Region:	critical
Effect:	19ms
Standard Error:	10ms

9. **Pearlmutter et al. (1999, Experiment 2)**

   Source of the estimates:
   Same procedure was applied as for Pearlmutter et al. (1999, Experiment 1) based on the numbers provided by the authors on Table C2 of Appendix C.

   Target-match (singular verb)
Region:	post-critical
Effect:	-36ms
Standard Error:	18ms

   Target-mismatch (plural verb)
Region:	post-critical
Effect:	-4ms
Standard Error:	18ms

10. **Pearlmutter et al. (1999, Experiment 3)**

    Source of the estimates:
    Same procedure was applied as for Pearlmutter et al. (1999, Experiment 1) based on the numbers provided by the authors on Table C5 of Appendix C.

    Target-match (singular verb)
Region:	critical
Effect:	-36ms
Standard Error:	10ms

    Target-match (plural verb)
Region:	post-critical
Effect:	24ms
Standard Error:	10ms

11. **Tucker et al. (2015)**

Source of the estimates:
Preprocessed data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed first-pass reading times. No correlations were fit between varying intercepts and slopes. Note that the authors fit a model with varying intercepts only and used many additional predictors, hence the difference in the outcome; moreover, the comparisons applied by the authors were different from ours. They report to have also run pairwise comparisons (i.e., the comparisons we are interested in), but they do not report any details of this model — they only describe in detail the first main effects and interaction model. We removed reading times larger than 2000ms and smaller than 100ms as there were extremely large values (40000ms) in the data. The authors used a different method for outlier removal (they excluded data points at the 5% extremes) and also excluded bad participants who scored below 70%, which we did not. For target-match, we took the estimates at the post critical region because the authors report a sign effect there while they did not find an effect at the critical region. In our analysis, however, this effect did not reach significance either.

Target-match (singular verb)

Region:	post-critical
Effect:	-7ms
Standard Error:	7ms

Target-mismatch (plural verb)

Region:	critical
Effect:	-29ms
Standard Error:	14ms

12. **Wagers et al. (2009, Experiment 2)**

Source of the estimates:
Preprocessed raw data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

Target-match (singular verb)

Region:	critical
Effect:	-8ms
Standard Error:	13ms

Target-mismatch (plural verb)

Region:	post-critical
Effect:	-51ms
Standard Error:	23ms

13. **Wagers et al. (2009, Experiment 3)**

Source of the estimates:
Preprocessed raw data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

Target-match (singular verb)

Region:	critical
Effect:	-1ms
Standard Error:	16ms

Target-match (plural verb)

Region:	critical
Effect:	13ms
Standard Error:	17ms

Target-mismatch (singular verb)

Region:	critical
Effect:	-33ms
Standard Error:	23ms

Target-mismatch (plural verb)

Region:	post-critical
Effect:	-31ms
Standard Error:	29ms

14. **Wagers et al. (2009, Experiment 4)**

   Source of the estimates:
   Preprocessed raw data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

   Target-match (singular verb)

Region:	critical
Effect:	-27ms
Standard Error:	13ms

   Target-mismatch (plural verb)

Region:	post-critical
Effect:	-42ms
Standard Error:	17ms

15. **Wagers et al. (2009, Experiment 5)**

Source of the estimates:
Preprocessed raw data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

Target-match (singular verb)

Region:	post-critical
Effect:	-11ms
Standard Error:	11ms

Target-mismatch (plural verb)

Region:	post-critical
Effect:	-37ms
Standard Error:	16ms

16. **Wagers et al. (2009, Experiment 6)**

Source of the estimates:
Preprocessed raw data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

Target-match (singular verb)

Region:	critical
Effect:	0ms
Standard Error:	12ms

Target-mismatch
not available

2.2 Experiments on non-agreement subject-verb dependencies

17. **Van Dyke (2007, Experiment 1, LoSyn conditions)**

Source of the estimates:
Numbers provided in the paper

Target-match (singular verb)

Region:	critical
Effect:	54ms
Standard Error:	34ms

Target-mismatch
not available

18. **Van Dyke (2007, Experiment 2, LoSyn conditions)**

Source of the estimates:
Numbers provided in the paper

Target-match (singular verb)
Region:	post-critical
Effect:	44ms
Standard Error:	19ms

Target-mismatch
not available

19. **Van Dyke (2007, Experiment 3, LoSyn conditions)**

Source of the estimates:
Numbers provided in the paper

Target-match (singular verb)
Region:	critical
Effect:	8ms
Standard Error:	8ms

Target-mismatch
not available

20. **Van Dyke and McElree (2006)**

Source of the estimates:
Numbers provided in the paper

Target-match (singular verb)
Region:	critical
Effect:	38ms
Standard Error:	20ms

Target-mismatch
not available

21. **Van Dyke and McElree (2011, Experiment 1b, proactive interference conditions)**

Source of the estimates:
Numbers provided in the paper

Target-match (singular verb)
Region:	critical
Effect:	5ms
Standard Error:	8ms

Target-mismatch
not available

22. **Van Dyke and McElree (2011, Experiment 1b, retroactive interference conditions)**

Source of the estimates:
Numbers provided in the paper

Target-match (singular verb)
Region: critical
Effect: -2ms
Standard Error: 11ms

Target-mismatch
not available

23. **Van Dyke and McElree (2011, Experiment 2b, proactive interference conditions)**

   Source of the estimates:
   Numbers provided in the paper

   Target-match (singular verb)
   Region: critical
   Effect: 7ms
   Standard Error: 9ms

   Target-mismatch
   not available

24. **Van Dyke and McElree (2011, Experiment 2b, retroactive interference conditions)**

   Source of the estimates:
   Numbers provided in the paper

   Target-match (singular verb)
   Region: critical
   Effect: -7ms
   Standard Error: 9ms

   Target-mismatch
   not available

25. **Van Dyke (2007, Experiment 1, LoSem conditions)**

   Source of the estimates:
   Numbers provided in the paper

   Target-match (singular verb)
   Region: critical
   Effect: 13ms
   Standard Error: 30ms

   Target-mismatch
   not available

26. **Van Dyke (2007, Experiment 2, LoSem conditions)**

Source of the estimates:
Numbers provided in the paper

Target-match (singular verb)
Region: critical
Effect: 37ms
Standard Error: 21ms

Target-mismatch
not available

27. **Van Dyke (2007, Experiment 3, LoSem conditions)**

Source of the estimates:
Numbers provided in the paper

Target-match (singular verb)
Region: critical
Effect: 20ms
Standard Error: 11ms

Target-mismatch
not available

28. **Van Dyke and Lewis (2003, Experiment 4, unambiguous high/low interference conditions)**

Source of the estimates:
Numbers provided in the paper. The authors provide raw and residualized trimmed reading times. We used the residualized trimmed reading times.

Target-match (singular verb)
Region: critical
Effect: 56ms
Standard Error: 25ms

Target-mismatch
not available

2.3 Experiments on reflexive/reciprocal-antecedent dependencies

29. **Jäger et al. (2015, Experiment 1)**

   Source of the estimates:
   Estimates were computed from the raw data. We fit a linear mixed model with varying slopes and intercepts for items and subjects on non-transformed first-pass reading time. No correlations were fit between varying intercepts and slopes. As in the original analysis reported in Jäger et al. (2015), log-frequencies of the target and the distractor were included as covariates in the model. Note that the original analysis was performed on log-transformed reading times; here we decided to model raw reading times in order to make the analysis more similar to the one of the other experiments included in the meta-analysis.

   Target-match
   Region: critical
   Effect: -3ms
   Standard Error: 5ms

   Target-mismatch
   Region: critical
   Effect: 22ms
   Standard Error: 7ms

30. **Jäger et al. (2015, Experiment 2, local conditions)**

   Source of the estimates:
   Estimates were computed from the raw data. Only the two conditions with a local antecedent were considered (non-locally bound reflexives were not included in this meta-analysis). We fit a linear mixed model with varying slopes and intercepts for items and subjects on non-transformed first-pass reading time. No correlations were fit between varying intercepts and slopes. As in the original analysis reported in Jäger et al. (2015), experimental session was included as a covariat in the model. Note that the original analysis was performed on log-transformed reading times; here we decided to model raw reading times in order to make the analysis more similar to the one of the other experiments included in the meta-analysis.

   Target-match
   Region: critical
   Effect: 17ms
   Standard Error: 8ms

   Target-mismatch
   not available

31. **Felser et al. (2009, Experiment 2b, native speakers, inaccessible-mismatch conditions)**

   Source of the estimates:
   Numbers provided in the paper

   Target-match
   Region: critical
   Effect: 4ms
   Standard Error: 9ms

   Target-mismatch
   not available

32. **Badecker and Straub (2002, Experiment 3)**

Source of the estimates:
Numbers provided in the paper

Target-match

Region:	post-critical	
Effect:	42ms	
Standard Error:	28ms	*SEs are provided for experimental conditions, not for the comparison. We use the larger of the two SEs (upper bound).*

Target-mismatch
not available

33. **Badecker and Straub (2002, Experiment 5)**

Source of the estimates:
Numbers provided in the paper

Target-match

Region:	critical	
Effect:	2ms	
Standard Error:	13ms	*SE is the larger of the two reported values.*

Target-mismatch
not available

34. **Badecker and Straub (2002, Experiment 6)**

Source of the estimates:
Numbers provided in the paper

Target-match

Region:	critical	
Effect:	0ms	*Effect size is assumed to be 0 here because no estimates are reported in the paper.*
Standard Error:	10ms	*SE is estimated by estimating standard deviation from Expt. 5: $sd = \sqrt{28} \cdot 13$, so Expt. 6: $SE = \sqrt{28} \cdot 13/\sqrt{48} \approx 10$.*

Target-mismatch
not available

35. **Cunnings and Felser (2013, Experiment 1, participants with high memory capacity)**

Source of the estimates:
Preprocessed raw data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed first-pass reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

Target-match

Region:	critical
Effect:	-2ms
Standard Error:	14ms

Target-mismatch

Region:	critical
Effect:	-2ms
Standard Error:	14ms

36. **Cunnings and Felser (2013, Experiment 1, participants with low memory capacity)**

Source of the estimates:
Preprocessed data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed first-pass reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

Target-match

Region:	critical
Effect:	-5ms
Standard Error:	22ms

Target-mismatch

Region:	critical
Effect:	-2ms
Standard Error:	16ms

37. **Cunnings and Felser (2013, Experiment 2, participants with high memory capacity)**

Source of the estimates:
Preprocessed raw data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed first-pass reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

Target-match

Region:	critical
Effect:	0ms
Standard Error:	18ms

Target-mismatch

Region:	critical
Effect:	4ms
Standard Error:	17ms

Exp2 HI 0 (18)4 (17)

38. **Cunnings and Felser (2013, Experiment 2, participants with low memory capacity)**

Source of the estimates:
Preprocessed raw data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed first-pass reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

Target-match

Region:	critical
Effect:	-47ms
Standard Error:	15ms

Target-mismatch

Region:	critical
Effect:	26ms
Standard Error:	15ms

39. **Cunnings and Sturt (2014, Experiment 1)**

Source of the estimates:
Preprocessed raw data provided by the authors. We fit a linear mixed model with varying slopes and intercepts for subjects and items on non-transformed first-pass reading times. No correlations were fit between varying intercepts and slopes. We removed the same outliers as did the authors for their published analysis.

Target-match

Region:	critical
Effect:	-1ms
Standard Error:	9ms

Target-mismatch

Region:	post-critical
Effect:	37ms
Standard Error:	17ms

40. **Felser et al. (2009, Experiment 2b, native speakers, no c-command conditions)**

Source of the estimates:
Numbers provided in the paper.

Target-match

Region:	critical
Effect:	3ms
Standard Error:	8ms

Target-mismatch
not available

41. **Patil et al. (2016)**

Source of the estimates:

Estimates were computed from the raw data. We fit a linear mixed model with varying slopes and intercepts for items and subjects on non-transformed first-pass reading time. No correlations were fit between varying intercepts and slopes.

Target-match

Region:	critical
Effect:	-13ms
Standard Error:	18ms

Target-mismatch

Region:	critical
Effect:	10ms
Standard Error:	12ms

42. **Sturt (2003, Experiment 1)**

Source of the estimates:

Numbers provided in the paper

Target-match

Region:	critical	
Effect:	-5ms	
Standard Error:	30ms	*The SE was estimated from Fig 1, p. 550. This is the SE for re-reading time, and is a reasonable upper bound for the SE for first-pass reading time as well (the SE in first-pass reading time is likely to be smaller than the one for re-reading time).*

Target-mismatch

Region:	critical	
Effect:	-7ms	
Standard Error:	30ms	*See comment for target-match above.*

43. **Sturt (2003, Experiment 2)**

Source of the estimates:
Numbers in provided in the paper

Target-match
Region: critical
Effect: 12ms
Standard Error: 10   ms   *SE estimated by looking at largest SE in first-pass reading time because on p. 557 Sturt says: "The gaze durations [i.e., first-pass reading times] on the critical reflexive, again [was] calculated using the leftward-shifting procedure." The smallest significant F value is 7.47, and so for any given difference in means, the smallest significant t-value is $\sqrt{7.47} \approx 2.73$. Since the effect of gender of the antecedent is $(268+280)/2-(292+307)/2 = -25.5$, we can estimate the SE as $SE = 25.5/2.73 = 9.34 \approx 10$.*

Target-mismatch
Region: critical
Effect: 15ms
Standard Error: 10ms   *See comment for target-match above*

44. **Dillon et al. (2013, Experiment 1, conditions with reflexives)**

Source of the estimates:
Source of the estimates:
Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model with varying intercepts and slopes for subjects and items on non-transformed first-pass reading times. No correlations were fit between varying intercepts and slopes. We fit a single model for the agreement conditions and the reflexives conditions with the main effect of dependency type as predictor in addition to the relevant pairwise comparisons.

Target-match
Region: critical
Effect: 1ms
Standard Error: 16ms

Target-mismatch
Region: critical
Effect: -7ms
Standard Error: 19ms

45. **Dillon et al. (2013, Experiment 2, conditions with *himself*)**

Source of the estimates:
Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model with varying intercepts and slopes for subjects and items on non-transformed first-pass reading times. No correlations were fit between varying intercepts and slopes.

Target-match

Region:	critical
Effect:	-14ms
Standard Error:	14ms

Target-mismatch

Region:	critical
Effect:	-10ms
Standard Error:	14ms

46. **Dillon et al. (2013, Experiment 2, conditions with *themselves*)**

Source of the estimates:
Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model with varying intercepts and slopes for subjects and items on non-transformed first-pass reading times. No correlations were fit between varying intercepts and slopes.

Target-match

Region:	critical
Effect:	-14ms
Standard Error:	16ms

Target-mismatch

Region:	critical
Effect:	30ms
Standard Error:	15ms

47. **Chen et al. (2012, local conditions)**

Source of the estimates:
Raw data. We removed all reading times larger than 2000ms (as Chen et al. (2012) did) and fit a linear mixed model with varying intercepts and slopes for subjects and items on non-transformed reading times. No correlations were fit between varying intercepts and slopes.

Target-match

Region:	critical
Effect:	5ms
Standard Error:	13ms

Target-mismatch
not available

48. **Badecker and Straub (2002, Experiment 4)**

Source of the estimates:
Numbers provided in the paper. As the authors base their conclusions on a sginificant effect observed on a post-hoc defined post-critical region containing four words and only provide numbers for this effect and not for the non-significant effect at the critical or the (originally smaller) post-critical region, we use the effect and standard error reported for this post-hoc collapsed spill-over region.

Target-match

Region:	post-critical
Effect:	48ms
Standard Error:	37ms

Target-mismatch
not available

49. **Kush and Phillips (2014)**

Source of the estimates:
Preprocessed data provided by the authors. Estimates were obtained by fitting a linear mixed model with varying intercepts and slopes for subjects and items on non-transformed reading times. No correlations were fit between varying intercepts and slopes. The same data trimming procedure was applied as by the authors.

Target-match

Region:	critical
Effect:	3ms
Standard Error:	54ms

Target-mismatch

Region:	post-critical
Effect:	21ms
Standard Error:	32ms

The following eye-tracking-while-reading and self-paced reading experiments were excluded from the meta-analysis:

- Acuña–Fariña et al. (2014)
  This eye-tracking-while-reading experiment was excluded from the meta-analysis as no first-pass reading times are provided in the paper.

- Pearlmutter (2000, Experiments 1 and 2)
  These experiments were excluded from the meta-analysis as the experimental design is slightly different from the other experiments in the meta-analysis. In these experiments, two distractors were manipulated independently and their influence was analyzed separately.

References

Acuña–Fariña, J. C., Meseguer, E., and Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua*, 143:108–128.

Badecker, W. and Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4):748–769.

Chen, Z., Jäger, L. A., and Vasishth, S. (2012). How structure-sensitive is the parser? Evidence from Mandarin Chinese. In Stolterfoht, B. and Featherston, S., editors, *Empirical Approaches to Linguistic Theory: Studies of Meaning and Structure*, Studies in Generative Grammar, pages 43–62. Mouton de Gruyter, Berlin, Germany.

Cunnings, I. and Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, 28(1-2):188–219.

Cunnings, I. and Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, 75:117–139.

Dillon, B. W., Mishler, A., Sloggett, S., and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69:85–103.

Felser, C., Sato, M., and Bertenshaw, N. (2009). The on-line application of Binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, 12:485–502.

Franck, J., Colonna, S., and Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in Psychology*, 6(349).

Jäger, L. A., Engelmann, F., and Vasishth, S. (2015). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, 6(617).

Kush, D. and Phillips, C. (2014). Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology*, 5(1252).

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., and Phillips, C. (2015). Agreement processes in Spanish comprehension. *Journal of Memory and Language*, 82:133–149.

Patil, U., Hanne, S., Burchert, F., Bleser, R. D., and Vasishth, S. (2016). A computational evaluation of sentence comprehension deficits in aphasia. *Cognitive Science*, 40:5–50.

Pearlmutter, N. J. (2000). Linear versus hierarchical agreement feature processing in comprehension. *Journal of Psycholinguistic Research*, 29(1):89–98.

Pearlmutter, N. J., Garnsey, S. M., and Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41:427–456.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48:542–562.

Tucker, M. A., Idrissi, A., and Almeida, D. (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology*, 6(347).

Van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 33(2):407–430.

Van Dyke, J. and Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49:285–316.

Van Dyke, J. and McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55(2):157–166.

Van Dyke, J. and McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65(3):247–263.

Wagers, M., Lau, E. F., and Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61:206–237.