Sentence Comprehension as a Cognitive Process

Sentence comprehension – the way we process and understand spoken and written language – is a central and important area of research within psycholinguistics. This book explores the contribution of computational modelling to the field, showing how computational models of sentence processing can help scientists in their investigation of human cognitive processes. It presents the leading computational model of retrieval processes in sentence processing, the Lewis and Vasishth cue-based retrieval model, and proposes a principled methodology for parameter estimation and model comparison/evaluation using benchmark data to enable researchers to test their own models of retrieval against the present model. It also provides readers with an overview of the last 20 years of research on the topic of retrieval processes in sentence comprehension, along with source code that allows researchers to extend the model and carry out new research. Comprehensive in its scope, this book is essential reading for researchers in cognitive science.

SHRavan vAsishth is Professor of Linguistics at the University of Potsdam. He is also a chartered statistician (Royal Statistical Society).

Felix Engelmann obtained his PhD in linguistics from the University of Potsdam (2016). He is a data scientist and entrepreneur, currently based in Berlin. His published research applies diverse computational methods to the modelling of human language processing and language acquisition.
Sentence Comprehension
as a Cognitive Process

A Computational Approach

Shravan Vasishth
University of Potsdam

Felix Engelmann
University of Potsdam
# Contents

**List of Figures**  
page ix

**List of Tables**  
xxvii

**Foreword by Richard L. Lewis**  
xx

**Preface**  
xxiii

**Acknowledgements**  
xxiv

1 Introduction  
1.1 Working Memory in Theories of Sentence Comprehension  
1.2 Prediction in Sentence Processing  
1.3 Working Memory and Prediction as Explanations for Processing Difficulty  
1.4 Current Beliefs about Constraints on Sentence Comprehension  
1.5 Some Gaps in the Sentence Processing Literature
  1.5.1 The Relative Scarcity of Computationally Implemented Models  
  1.5.2 A Focus on Average Behaviour and Neglect of Individual-Level Differences  
  1.5.3 The Absence of High-Precision Studies  
  1.5.4 Unclear Desiderata for a Good Model Fit  
1.6 The Goals of This Book  
  1.6.1 Providing Open Source Model Code  
  1.6.2 Modelling Average Effects as Well as Individual Differences  
  1.6.3 Developing a Set of Modelling and Empirical Benchmarks for Future Model Comparison  
1.7 Looking Ahead  

2 Dependencies in Sentence Comprehension  
2.1 Memory Processes in Sentence Comprehension  
2.2 Dependency Completion in Sentence Processing  
2.3 Subject-Verb Non-Agreement Dependencies  
2.4 Subject-Verb Number Agreement  
2.5 Reflexives and Reciprocals  
  2.5.1 Individual-Level Effects in the Dillon et al. Design  
  2.5.2 A Sensitivity Analysis on the Ungrammatical Agreement and Reflexives Conditions Using Informative Priors  
2.6 Concluding Remarks  

3 The Core ACT-R-Based Model of Retrieval Processes  
3.1 ACT-R  
3.2 The Lewis and Vasisht (2005) Model  
  3.2.1 A Priori Predictions of the Model  

© in this web service Cambridge University Press  
www.cambridge.org
## Contents

3.2.2 Comparison of the LV05 Prediction Space with the Results of the Jäger et al. Meta-analysis 60

3.3 A More Principled Approach to Parameter Estimation 63
  3.3.1 Bayesian Parameter Estimation 64
  3.3.2 Approximate Bayesian Computation 66

3.4 Concluding Remarks 69

4 An Extension of the Core Model: Modelling Prominence and Multi-associative Cues 71
  4.1 Incorporating Prominence and Multi-associative Cues 72
    4.1.1 Item Prominence 74
    4.1.2 Multi-associative Cues 84
    4.1.3 Implementation of Item Prominence and Multi-associative Cues 89
    4.1.4 Multi-associative Cues 90
    4.1.5 Prominence 93
  4.2 A Simulation of the Meta-analysis Studies 94
    4.2.1 Data 95
    4.2.2 Method 95
    4.2.3 Results 98
  4.3 Discussion 103
    4.3.1 Distractor Prominence 107
    4.3.2 Multi-associative Cues 108

Appendices
  4.A Key Terms and Concepts 111
  4.B List of Experiments Included in the Simulations 113
  4.C Model Specifications 114

5 An Extension of the Core Model: Modelling the Interaction of Eye-Movement Control and Parsing 116
  5.1 The EMMA/ACT-R Reading Model 118
  5.2 Replication of Salvucci (2001) 119
    5.2.1 Data 119
    5.2.2 Model 120
    5.2.3 Analysis 120
    5.2.4 Results 122
    5.2.5 Discussion 122
  5.3 The Extended EMMA/ACT-R Model 122
    5.3.1 Surprisal 124
  5.4 Simulations on the Potsdam Sentence Corpus 125
    5.4.1 Data 126
    5.4.2 Model 127
    5.4.3 Results 128
    5.4.4 Discussion 131
  5.5 General Discussion 132
    5.5.1 Comparison with E-Z Reader 132
    5.5.2 Future Prospects 134
6 Reanalysis and Underspecification in Sentence Comprehension: Modelling Eye Movements

6.1 Introduction

6.2 Modelling Reanalysis: Memory and Expectation Processes in Parsing
   6.2.1 Memory and Expectation in Relative Clauses
   6.2.2 Simulation: Modelling the Staub (2010) Data
   6.2.3 Results
   6.2.4 Discussion

6.3 Modelling Underspecification: The Adaptive Interaction between Parsing, Eye-Movement Control, and Working Memory Capacity
   6.3.1 Good-Enough Parsing
   6.3.2 Simulation: Modelling the von der Malsburg and Vasishth (2013) Experiment
   6.3.3 Results
   6.3.4 Discussion

6.4 General Discussion

7 Competing Accounts of Interference in Sentence Processing

7.1 The Direct-Access Model

7.2 Comparing the Predictive Performance of the Models
   7.2.1 Inhibitory Interference
   7.2.2 Relative Clauses in Chinese
   7.2.3 Discussion

7.3 Encoding Interference in Agreement Attraction
   7.3.1 An Evaluation of the Nairne Proposal
   7.3.2 Model Comparison
   7.3.3 Discussion

7.4 Summary

8 Modelling Sentence Comprehension Deficits in Aphasia

8.1 Theories and Models of Sentence Comprehension Deficits
   8.1.1 Timing Deficit
   8.1.2 Reduction in Memory
   8.1.3 Intermittent Deficiency
   8.1.4 Weakened Syntax
   8.1.5 Slow Syntax
   8.1.6 Lexical Integration Deficit
   8.1.7 Lexical Access Deficits
   8.1.8 A Comparison of Theories of Impaired Processing, and Their Relation to Theories of Unimpaired Processing

8.2 Modelling Individual-Level Differences
   8.2.1 Mapping ACT-R Parameters to Sources of Deficits
   8.2.2 Simulations
   8.2.3 Results
   8.2.4 Discussion
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.3 Competing Models of Retrieval in Aphasia</td>
<td>197</td>
</tr>
<tr>
<td>8.3.1 Materials</td>
<td>197</td>
</tr>
<tr>
<td>8.3.2 Results and Discussion</td>
<td>198</td>
</tr>
<tr>
<td>8.4 Concluding Remarks</td>
<td>198</td>
</tr>
<tr>
<td>9 Future Directions</td>
<td>200</td>
</tr>
<tr>
<td>9.1 Developing Implemented Computational Models</td>
<td>200</td>
</tr>
<tr>
<td>9.2 An Excessive Focus on Average Behaviour</td>
<td>200</td>
</tr>
<tr>
<td>9.3 Creating Higher-Precision Benchmark Data-Sets for Model Evaluation and Comparison</td>
<td>201</td>
</tr>
<tr>
<td>9.4 Developing Better Criteria for Evaluating Model Fit</td>
<td>202</td>
</tr>
<tr>
<td>9.5 In Closing</td>
<td>202</td>
</tr>
<tr>
<td>Bibliography</td>
<td>203</td>
</tr>
<tr>
<td>Index</td>
<td>221</td>
</tr>
</tbody>
</table>
Figures

1.1 A schematic summary of the Roberts and Pashler (2000) discussion regarding what constitutes a good fit of a model to data. The data are represented by the circle (the estimated mean) and the vertical uncertainty interval, and the model predictions by the diagonal parallel lines. If a model predicts a positive correlation between two variables \(x\) and \(y\), strong support for the model can only be argued for if both the data and the model predictions are highly constrained: the model must make predictions over a narrow range, and the data must have low uncertainty associated with it. 12

1.2 A demonstration of Type M and S error. Low power studies will yield overestimates and/or incorrect signs whenever a result is significant. 14

1.3 The five possible outcomes when using the null region or “region of practical equivalence” method for decision-making (Kruschke, 2015). Outcomes A and B are inconsistent with the quantitative predictions of the theory; C and D are inconclusive; and E is consistent with the quantitative theoretical prediction. Figure reproduced from Vasishth and Gelman (2019). 16

2.1 A schematic illustration of the fan effect. Searching for an object that is grey and a square (the target item) is more difficult when competing items have one or more features matching cues used for identifying the target item. 22

2.2 Inhibitory interference effects (sorted in increasing order by magnitude) in reading studies by Van Dyke and colleagues. The grey vertical lines show the 95% credible interval for the meta-analysis estimate of the effect. 28

2.3 Distribution of power (paired, two-sided \(t\)-test) assuming that the effect has normal distribution with mean 13 and standard deviation 6, the standard deviation ranges from 75 to 100 ms, and the subject sample size is 60. 29
List of Figures

2.4 Visualization of two conditions in the Cunnings and Sturt (2018) experiment and the predictions of the cue-based retrieval model. The verb *shattered* attempts to retrieve an item in memory that is a direct object and has the property “is shatterable”. In both the (a) and (b) conditions shown, the direct object (which is the target noun that should be retrieved) matches the direct object retrieval cue. However, in (b), the distractor noun matches the “is shatterable” cue. As a consequence, in (b), both the target and distractor nouns enter into a race, and whichever item is non-deterministically retrieved is the winner of the race. This race process leads to a faster reading time at the verb *shattered* in (b) vs. (a).

2.5 Subject-verb number agreement effects in ungrammatical sentences (reading studies). Shown are the means (sorted by increasing magnitude of the effect) and 95% confidence intervals that were either computed from publicly available data or derived from published estimates.

2.6 The role of case marking in agreement attraction configurations. The figure is reused here under a CC-BY4.0 license and is available from https://doi.org/10.6084/m9.figshare.11440854.v1.

2.7 Target match number agreement effects in reading studies.

2.8 Summary for total reading time of the Dillon et al. (2013) comparisons for ungrammatical sentences involving agreement and reflexives. The sample size was 40 participants. The upper plot shows the posterior distributions of the facilitatory interference effect in agreement and reflexives, and the lower plots show the individual-level estimates of the effect, with 80% and 95% credible intervals.

2.9 Summary for total reading time of the Jäger et al. (2020) comparisons for ungrammatical sentences involving agreement and reflexives. The sample size was 181 participants. The upper plot shows the posterior distributions of the facilitatory interference effect in agreement and reflexives, and the lower plots show the individual-level estimates of the effect, with 80% and 95% credible intervals.

3.1 Figure 1 of Lewis and Vasishth (2005). The figure shows the representation of chunks (maximal projections of phrases) that constitute a syntactic tree. The figure is copyrighted by Wiley, and is reused with permission, license number 4782371233287.

3.2 Figure 2 of Lewis and Vasishth (2005). The figure shows the processing cycle of the parsing algorithm. The figure is
3.3 Spreading activation according to ACT-R/LV05 in the four conditions shown in Example 21. Line weights indicate the amount of spreading activation from a cue to an item. Black oval boxes represent a feature match. Grey oval boxes indicate features matching an “overloaded” cue (MASC in b), and white boxes indicate a mismatch. The figure is by Engelmann and Vasishth (2019); available at https://doi.org/10.6084/m9.figshare.9305456 under a CC-BY4.0 license.

3.4 An illustration of a race process involving two distributions that represent retrieval time distributions of two items. When the two distributions have similar means (Figure A), the distribution of the retrieval times of the winner (which may differ from trial to trial) will have a distribution with a mean that is lower than the mean of the two distributions involved in the race (statistical facilitation). When one distribution has a much smaller mean than the other distribution’s mean (Figure B), the distribution of the winner’s retrieval times will have the same mean as that of the distribution of the item with the smaller mean.

3.5 Prediction space for the interference effect in ACT-R in target-match (circles, solid line) and target-mismatch configurations (triangles, broken line). Interference is plotted in terms of the difference in mean retrieval latencies between the interference (labelled distractor-match) and the no-interference (labelled distractor-mismatch) condition, and as a function of the latency factor $F$. Positive values indicate longer mean retrieval latencies in the interference condition (inhibitory interference) due to cue-overload (fan effect) from a partially matching distractor; negative values indicate shorter mean retrieval latencies in the interference condition (facilitatory interference) due to retrievals of the partially matching distractor on trials where the distractor is highly activated and hence fast. Each individual data point represents the mean interference effect of 6,000 iterations with one out of 10,980 different parameter settings (each in target-match and target-mismatch configurations; i.e., there are 21,960 data points plotted in total). Each parameter setting is a combination of the following parameter values: latency factor $F \in \{0, 0.01, \ldots, 0.6\}$, noise parameter $ANS \in \{0.1, 0.2, 0.3\}$, maximum associative strength $MAS \in \{1, 2, 3, 4\}$, mismatch penalty $MP \in \{0, 1, 2\}$, retrieval threshold $\tau \in \{-2, -1.5, \ldots, 0\}$.

3.6 A $\text{Beta}(2, 6)$ prior on the latency factor.
xii List of Figures

3.7 The posterior distributions of the latency factor parameters for agreement and reflexive conditions using the original Dillon et al. (2013) data (40 participants, 48 items) and our own Jäger et al. (2019) replication data (181 participants, 48 items).

3.8 The posterior predictive distributions of the facilitatory interference in ungrammatical agreement and reflexive conditions, derived using the posterior distributions of the latency factor parameter.

4.1 Predictions of ACT-R for the four conditions shown in Example (22). Line weights indicate the amount of spreading activation from a cue to an item. Black oval boxes represent a feature match. Grey oval boxes indicate features matching an “overloaded” cue (MASC in b), and white boxes indicate a mismatch.

4.2 Predicted target-match and target-mismatch interference effects (distractor-match minus distractor-mismatch) as a function of distractor prominence \(p_{distr}\) ranging from \(-3, \ldots, 5\) when target prominence is zero (mean of 10,000 iterations with parameters \(F = 0.2, ANS = 0.2, MAS = 2, MP = 0\)). Positive values indicate longer mean retrieval latencies (inhibition) in the interference condition due to cue overload (fan effect). Negative values indicate shorter mean retrieval latencies (facilitation) in the interference condition due to retrievals of the distractor on trials where the distractor is highly activated and hence fast. The points where the vertical line intersects with the curves represent standard LV05 predictions.

4.3 Mechanisms underlying the effect of distractor prominence in target-mismatch configurations. The x-axis in each panel shows increasing distractor prominence (with target prominence = 0). The panels from top left are: (1) Mean activation of target and distractor at retrieval event in interference (distractor-match) and no-interference (distractor-mismatch) condition (the sign of the activation value – negative or positive – has no special meaning in ACT-R). (2) Proportion of distractor retrievals over multiple iterations (retrieval probability). Values above 0.5 indicate higher retrieval probability for the distractor than the target (misretrievals). (3) Mean retrieval latencies (of most activated item at retrieval). (4) Mean interference effect as the difference in retrieval latencies between interference and no-interference condition. Positive values mean inhibitory interference (longer latencies when distractor matches); negative values mean facilitatory interference (short latencies due to misretrievals when...
distractor mismatches). The vertical lines mark locations of (a) low interference due to low prominence; (b) LV05 equivalence at prominence $= 0$ (equal activation of target and distractor in interference condition); (c) maximal facilitatory interference effect due to misretrievals; (d) low interference due to latencies close to zero.

4.4 Mechanisms underlying the effect of distractor prominence in target-match configurations. The x-axis in each panel shows increasing distractor prominence (with target prominence $= 0$). The panels from top left are: (1) Mean activation of target and distractor at retrieval event in interference (distractor-match) and no-interference (distractor-mismatch) condition (the sign of the activation value – negative or positive – has no special meaning in ACT-R). (2) Proportion of distractor retrievals over multiple iterations (retrieval probability). Values above 0.5 indicate higher retrieval probability for the distractor than the target (misretrievals). (3) Mean retrieval latencies (of most activated item at retrieval). (4) Mean interference effect as the difference in retrieval latencies between interference and no-interference condition. Positive values mean inhibitory interference (longer latencies when distractor matches); negative values mean facilitatory interference (short latencies due to misretrievals when distractor mismatches). The vertical lines mark locations of (a) low interference due to low prominence; (b) maximal inhibitory interference effect due to increased fan; (c) equal activation of target and distractor in interference condition: lower fan effect due to statistical facilitation; (d) zero interference effect because of equal strength of fan effect and facilitation due to misretrievals of the highly activated distractor; (e) maximal facilitatory interference effect due to misretrievals; (f) low interference due to latencies close to zero.

4.5 Spreading activation in conditions labelled distractor-match (c) and distractor-mismatch (d) conditions in target-mismatch configurations when cues are cross-associated. Line weight and box shading indicate the amount of spreading activation added to an item due to a feature match. Dashed lines represent spreading activation to a cross-associated feature.

4.6 Predicted target-match and target-mismatch interference effects (distractor-match minus distractor-mismatch) as a function of the cross-association level c. Lines and shaded area show mean and range of the effect, respectively, for parameter values of the latency factor $F$ ranging from 0.2 to 0.4, and distractor prominence...
ranging from $-0.5, 0, 0.5$, running 5,000 iterations each; other parameters were fixed as $\text{ANS} = 0.2, \text{MAS} = 2, \text{MP} = 0$. Positive values indicate longer mean retrieval latencies (inhibition) in the interference condition due to cue-overload (fan effect). Negative values indicate shorter mean retrieval latencies (facilitation) in the interference condition due to misretrievals of the distractor. 87

4.7 Standard target-mismatch/distractor-match condition without cross-associated cues. 90

4.8 Target-mismatch/distractor-match condition when cues are cross-associated. 92

4.9 Number of studies included in the Jäger et al. (2017) meta-analysis and in the simulations, grouped by dependency type and distractor prominence status (studies are listed in Table 4.6 in the Appendix). 96

4.10 Mean interference effects from simulations with LV05 and the extended model, labelled LV05+IP+MAC, for target-match and target-mismatch configurations of the meta-analysis, grouped by dependency type (studies are listed in Table 4.6 in the Appendix). The behavioural data is shown as mean effect estimates with 95% credible intervals as reported in Jäger et al. (2017). 99

4.11 Mean interference effects from simulations with LV05 and the extended model, labelled LV05+IP+MAC, for target-match (top panel) and target-mismatch configurations (bottom panel) in the Jäger et al. (2017) meta-analysis, grouped by distractor prominence level within dependency types (studies are listed in Table 4.6 in the Appendix). The behavioural data is shown as raw means with additional smaller points representing individual studies. The target-mismatch plot in non-agreement subject-verb dependencies does not contain data because no data were available at the time of the meta-analysis. However, Cunnings and Sturt (2018) have recently found evidence consistent with the predictions of the model; in two experiments, they obtained an estimated mean of $-22$ ms with a 95% credible interval of $[-4, -42]$, and in a second experiment, a mean of $-19$ ms, $[-40, 1]$. 100

4.12 Reading time data and simulation results of LV05 and the extended model, labelled LV05+IP+MAC, for interference effects in target-match and target-mismatch configurations of four individual studies: Kush and Phillips (2014), (Jäger et al., 2015, Exp. 1), (Sturt, 2003, Exp. 1), and (Cunnings and Felser, 2013, Exp. 2, participants with low working memory). 102

5.1 Replication of Salvucci (2001) on the Schilling Corpus. Effects of word frequency on gaze, first, and single fixation duration, and on
List of Figures

the rate of skipping a word, fixating it once, and fixating it more than once. Grey solid lines represent experimental data, black dotted lines show Salvucci’s simulation results, and black dashed lines show the replication results. Lexical frequency is divided into classes 1 (lowest) to 5 (highest).

5.2 A schematic figure illustrating the Time Out mechanism in the case of object vs. subject relatives. In an object relative, if the integration of the relative clause verb is still in progress while the encoding of the word following it has already completed, time out initiates an attention shift to the word to the left of the currently fixated one (Time Out regression). Once the integration of the relative clause verb has finished, the exit-time-out rule returns the model into the state of continuing fixating in the reading direction.

5.3 Shown are the predictions of Model 2 (EMMA, dotted lines) vs. Model 7 (EMMA+rs2, dashed lines) vs. experimental data (grey solid lines) for the Potsdam Sentence Corpus. The figure shows means of early (first row) and late measures (second row) as a function of frequency class. Each row shows reading time durations on the left and probabilities on the right side.

5.4 Coefficients and 95% confidence intervals for predictors surprisal and retrieval estimated by linear regression. Predictors were log frequency, length, log retrieval, and surprisal. Coefficients are plotted along the y-axis for surprisal on the left side and retrieval on the right side. Regressions were carried out on the simulated data of all six EMMA models (shown on the x-axis); 95% confidence intervals that do not cross 0 indicate statistical significance at $\alpha = 0.05$.

6.1 Model predictions for reading times in subject- and object-relative clauses.

6.2 Predicted first-pass regressions from the model for subject- and object-relative clauses.

6.3 Proportions of sentence rereading by working memory capacity in the data of von der Malsburg and Vasishth (2013).

6.4 Predicted gaze durations by source activation at ambiguous and unambiguous attachments.

6.5 Predicted time out proportions by source activation at ambiguous and unambiguous attachments.

6.6 Predicted proportions of sentence rereading by source activation at ambiguous and unambiguous attachments.

6.7 Predicted attachment proportions by source activation at ambiguous and unambiguous attachments.
xvi List of Figures

7.1 A schematic illustration of the direct-access model. For sentences like (38a), the model assumes that once a search is initiated in memory using a set of retrieval cues (here, subject and animate), one of two events can happen. Either the correct item is retrieved from memory, or the incorrect item, which matches some of the retrieval cues, is misretrieved. In the case of a misretrieval, either processing ends with a misretrieval, or a reanalysis step is initiated that leads to a correct retrieval. This reanalysis step costs time, and therefore leads to slowdowns in processing on average. 163

7.2 A comparison of observed sample means with the posterior predictive distributions of the activation-based model, and the direct-access model. The figure is adapted from the online materials available from a StanCon 2017 conference talk by Bruno Nicenboim and Shravan Vasishth, which are under a CC-BY 4.0 licence. 166

8.1 Marginal distributions of each of the three parameters for subject relatives in controls (solid lines) vs. IWA (dotted lines). The vertical line shows the default setting for the respective parameter. 193

8.2 Marginal distributions of each of the three parameters for object relatives in controls (solid lines) vs. IWA (dotted lines). The vertical line shows the default setting for the respective parameter. 193

8.3 Shown are the differences between the two models in expected pointwise log density for each data-point. Points above the zero line show an advantage for the activation-based model, and points below the zero line an advantage for the direct-access model. The darkness in the hexagons represents density, with darker hexagons representing more dense data-points. The figure is under a CC-BY 4.0 license, https://doi.org/10.6084/m9.figshare.12114075.v1. 199
Tables

2.1 Nested contrast coding to investigate the effect of intrusion in grammatical and ungrammatical agreement and reflexive constructions. The contrast dep is the main effect of dependency type (agreement or reflexive). The abbreviation intr.au means intrusion (interference effect) in agreement dependencies, ungrammatical; intr.ag stands for intrusion (interference effect) in agreement dependencies, grammatical; intr.ru refers to intrusion (interference effect) in reflexive dependencies, ungrammatical; intr.rg stands for intrusion (interference effect) in reflexive dependencies, grammatical.

2.2 Summary of the sensitivity analysis investigating the effect of incorporating prior knowledge from mildly uninformative priors; a meta-analysis of existing reading data on ungrammatical agreement and reflexives; and the model predictions in Engelmann et al. (2020). The dependent measure in the analysis is total fixation time and the posterior estimates are back-transformed to the ms scale from log ms. The priors are shown in the ms scales.

3.1 Results of the Jäger et al. (2017) meta-analysis showing mean effect estimates $\hat{b}$ with Bayesian 95% credible intervals in the Estimates column. The range specified by a 95% credible interval contains the true value of the estimated parameter with 95% certainty, given the model and the data. A positive interference effect means inhibition, a negative one facilitation. Results are compared with the predictions of cue-based retrieval as implemented in the LV05 ACT-R model, and the additional contributions of the extensions item prominence (IP) and multi-associative cues (MAC), which are discussed in Chapter 4.

4.1 Possible feature combinations exhibited by correct antecedents of English reflexives, reciprocals, and Chinese ziji.

4.2 Root-mean-square deviation between modelling results and observed data, averaged within dependency type and model (best
List of Tables

values in bold). The superscript no dec means that the decay parameter is set to 0.

4.3 Estimated values for prominence parameter in the LV05+IP+MAC model with decay for three prominence levels.

4.4 Shown here is the terminology used in the present chapter in relation to cue-based retrieval and interference in dependency resolution.

4.5 Shown here is the terminology used in the extension of the cue-based retrieval model (continued from previous page).

4.6 List of experiments included in the simulations.

4.7 List of experiments included in the simulations (continued from previous page).

4.8 Model parameters, their default values, and the values used in the simulation of the studies in the meta-analysis.

5.1 Frequency classes used in the analyses of the Schilling Corpus (SC) and Potsdam Sentence Corpus (PSC).

5.2 Fit and parameter estimates for all simulations. The interpretation of the data are discussed in the Results and Discussion sections.

5.3 Linear regression results for predictors retrieval and surprisal

6.1 ACT-R/EMMA parameter values.

7.1 Model comparison using K-fold cross-validation for the Gibson and Wu 2013 data. Shown are the differences in elpd, along with standard errors of the differences. In a comparison between a model A vs B, a positive Δelpd favours model A.


7.3 Comparison of the 10 sets of hierarchical models. Shown are the differences in elpd between (a) the standard hierarchical model and the homogeneous variance mixture model; (b) the feature percolation model and the homogeneous variance mixture model; and (c) the homogeneous vs. heterogeneous variance mixture model. Also shown are standard errors for each comparison. If the difference in elpd is positive, this is evidence in favour of the second model. The pairwise model comparisons are transitive. These comparisons show that the heterogeneous variance mixture model has the best predictive performance.

8.1 A matrix showing how the models relate to each other along dimensions of the three working-memory related events – delays, forgetting (or failure to retrieve), and misretrieval – that have been investigated in sentence comprehension research.

8.2 The number of participants in subject/object relatives (SR/OR) for which nondefault parameter values were predicted, in the subject
List of Tables

| vs. object relative tasks, respectively; for goal activation (GA), default action time (DAT), and noise (ANS) parameters. | 194 |
| 8.3 Discrimination ability of hierarchical clustering on the combined data for subject/object relatives. Numbers in bold show the number of correctly clustered data points. The bottom row shows the percentage accuracy. | 195 |
Foreword by Richard L. Lewis

In reading a draft of this remarkable book, I was reminded of the title that Allen Newell chose for his contribution to a meeting celebrating the scientific contributions of Herbert Simon: “Putting It All Together”. Newell was both acknowledging Simon’s putting so much together under bounded rationality, but also looking forward to the theoretical integration made possible by models of cognitive architecture. What Shravan Vasishth and Felix Engelmann have offered us is perhaps the most comprehensive and integrated attempt yet to put it all together in sentence processing in a way that begins to do justice to its rich, cross-linguistic empirical details.

I’ll point out below a few of my favourite contributions to integration that appear in the book. But I first want to draw the reader’s attention to another thread running through all of the chapters – and one that is perhaps even more important than the specific details of the models, explanations, and empirical analysis that is focus of each chapter.

That thread is a sharp critique of our current practices in empirical and theoretical psycholinguistics. Indeed, the first two chapters do not read like the expected triumphant summary of 20 years of empirical research confirming effects of similarity-based interference and other predictions of our early sentence processing models. On the contrary, it is a sobering taking-stock of the empirical record and current methodological practice through the lens of what we have too slowly come to understand about what is required to make progress. And what is required is quite often much larger amounts of carefully collected data, rigorous statistical analysis, and multiple alternative model testing. In short, this book and the work it reports is part of the larger movement throughout the psychological and cognitive sciences that is helping us to wake up to the reality of just how challenging our science really is. In the case of psycholinguistics, we take for granted that we can infer internal cognitive and linguistic structure from movements of the eyes and hands and tongue and lips or fluctuations in electrical potentials on the scalp. Why did we think that task would be easier than it in fact is?

But along with this sobering critique, the book also takes us on a kind of joy ride – letting us experience the joy of a few real advances and interesting
Foreword by Richard L. Lewis

ideas that help us see a little bit further ahead. My own favourites include the
detailed comparison of different retrieval models (concluding with a rejection
of the specific model in Lewis and Vasishth, 2005), the model-based accounts
of individual differences and pathologies (paralleling a renaissance across the
field ranging from areas such as computational psychiatry to cognitive aging),
and the beginnings of explicit models of adaptive eye movements that start
to do justice to the flexible and highly adaptive nature of human language
comprehension.

On a more personal note, it is a unique privilege as a scientist to be able to
look back to a collaboration that started over 20 years ago, and to see how far
the work has come. I cannot believe the good fortune I had to cross paths with
Shravan when he was a prodigious graduate student in linguistics and I was a
young professor in computer science at Ohio State. The ideas we explored in
the early ACT-R models of parsing were really an evolution and combination
of insights of George Miller, Noam Chomsky, and John Anderson. And by now
what is generously referred to in the book as the Lewis and Vasishth model is
really the Vasishth and colleagues model. But in the end, scientific ideas do not
belong to any of us – they belong to the field, and individuals and teams are but
stewards.

Of course there are many gaps, weaknesses, and shortcomings in the pages
that follow. But unlike most scientific books, a great many of them are docu-
mented by the authors themselves! And so I am reminded of another colourful
quote (attributed to Warren McCulloch), one that Allen Newell enjoyed using
when advancing his candidate integrated theories: “Don’t bite my finger, look
where I’m pointing.” Vasishth and Engelmann are pointing the way to a better
science of sentence processing, and we’d do well to take a look in their direction.
I especially hope that new students joining the field will do so, and be inspired
to take us down ambitious and imaginative new paths towards integrated and
deeply explanatory theory.
Preface

The early work of Richard L. Lewis in the 1990s set the stage for the work reported in the present book. Rick’s research on developing a language processing model within the SOAR architecture (Lewis, 1993) evolved into a sharper focus on developing process models of dependency completion in sentence comprehension. He initiated the use of the cognitive architecture ACT-R to model proactive and retroactive interference effects (Lewis, 1996). The first major elaboration of these ideas appeared in Lewis and Vasishth (2005) and Lewis et al. (2006). In the late 1990s and early 2000s, both Julie Van Dyke and the first author of the present book were Rick’s PhD students. Since then, quite a lot of evidence has accumulated that is consistent with Rick’s original insight that dependency completion time (retrieval time in ACT-R parlance) in sentence processing is affected by similarity-based interference. However, some important counterexamples to this proposal have also emerged, and there are some important empirical details relating to retrieval processes that may not be explainable by the general mechanisms posited within ACT-R (Anderson et al., 2004) or other memory architectures. We discuss several of these counterexamples in detail in the present book. More generally, the present book takes stock of the computational modelling done in this context and situates the modelling within some (but not all) of the important scientific questions in sentence processing research that are actively under consideration today. We hope that this book will be useful to researchers seeking to build on the work presented here and to develop the next generation of computational models of sentence processing.
Acknowledgements

The following researchers made important contributions to the computational modelling presented in this book during the time that they were doing their PhD or postdoctoral research with the first author.

- Ms. Paula Lissón was the lead on the aphasia modelling work on the comparison of the models of retrieval processes in aphasia (Lissón et al., 2021).
- Ms. Dorothea Pregla developed the German data-set on individuals with aphasia and controls as part of her PhD dissertation; in future work, this will serve as benchmark data for evaluating some of the models discussed in this book.
- Ms. Daniela Mertzen carried out several large-sample experiments on interference as part of her PhD dissertation work. These will serve as benchmark data for model evaluation in future work.
- Mr. Himanshu Yadav, Dr. Garrett Smith, and Dr. Dario Paape helped develop extensions of the Approximate Bayesian Computation approach reported here.
- Dr. Garrett Smith developed the new principled approach for determining lexical features (Smith and Vasishth, 2020).
- Prof. Dr. Lena Jäger: Co-authored the meta-analysis (Jäger et al., 2017) that forms the empirical basis for some of the model evaluations reported in this book, and co-developed the prominence and multi-associative cues extension of the core model, as reported in Engelmann et al. (2020).
- Prof. Dr. Bruno Nicenboim: Developed the implementation of the direct-access model, as reported in Nicenboim and Vasishth (2018).
- Prof. Dr. Titus von der Malsburg: Provided the empirical basis for modelling underspecification and reanalysis (von der Malsburg and Vasishth, 2013), as discussed in Chapter 6.
- Mr. Paul Mätzig: Carried out the model development and simulations reported in Mätzig et al. (2018).
Acknowledgements

• Prof. Dr. Pavel Logačev implemented and tested several models relating to attachment ambiguities and underspecification (Logačev and Vasishth, 2015; Logačev and Vasishth, 2016).
• Dr. Umesh Patil: Carried out the model development and simulations reported in Patil et al. (2016a).

For comments, helpful feedback, and advice over the years, we would like to thank the following people: Serine Avetisyan, Douglas Bates, Michael Betancourt, Adrian Brasoveanu, Bob Carpenter, Pyeong Wang Cho, Ian Cunnings, Brian Dillon, Jakub Dotlačil, Ralf Engbert, Julie Franck, Hiroki Fujita, Andrew Gelman, Matt Goldrick, Robert Grant, Sandra Hanne, Lena Jäger, Reinhold Kleigl, Dave Kush, Sol Lago, Anna Laurinavichyute, Richard L. Lewis, Tal Linzen, Paula Lissón, Pavel Logačev, Paul Mätzig, Daniela Mertzzen, Mitzi Morris, Bruno Nicenboim, Dario Paape, Dan Parker, Umesh Patil, Colin Phillips, Dorothea Pregla, Maxmilian Rabe, Milena Rabovsky, Daniel Schad, Pia Schoknecht, Scott Sisson, Garrett Smith, Patrick Sturt, Whitney Tabor, Matt Tucker, Julie Van Dyke, Mick van het Nederend, Hedderik van Rijn, Sashank Varma, Titus von der Malsburg, Matt Wagers, Jan Winkowski, and Himanshu Yadav. Elna Haffner and Junilda Petriti helped prepare the index, and Junilda Petriti checked for formatting and other errors and inconsistencies. Some of the data that formed the basis for modelling came from our own lab, but a lot of other data came from Adrian Staub, Brian Dillon, Sol Lago, Matt Wagers, Matt Tucker, Julie Franck, Patrick Sturt, Ian Cunnings, Roger Levy, Frank Keller, and Ted Gibson (among others). The openness and transparency of these scientists is greatly appreciated. Colin Phillips and Ted have also been very supportive colleagues over the years, despite many scientific disagreements between us. Sashank Varma reviewed the final draft of this book and made very important suggestions for improvement, which we have tried to implement in the final version of this book. Our apologies if we have forgotten anyone.

This book was generously funded by the Volkswagen Foundation through an Opus Magnum award (grant number 89 953) to Shravan Vasishth. The grant allowed the first author to take a two-year sabbatical from teaching to concentrate on research and writing. Shravan Vasishth also thanks his wife and son for their patience over the last few years.