

Optimal daily allocation between vocabulary and grammar in second- language study: a normative control model

Tian-Yu Wang

wangty0709@hotmail.com

Guangdong Peizheng College

Research Article

Keywords: optimal control, vocabulary learning, grammar learning, text coverage, Zipf–Mandelbrot distribution, sensitivity analysis

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11383493/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

1 **Optimal daily allocation between vocabulary and grammar in second-**
2 **language study: a normative control model**

3 **Tian-Yu Wang***

4 Medical Office, General Affairs Office, Guangdong Peizheng College, Guangzhou
5 510830, Guangdong, China

6 Corresponding author: Tian-Yu Wang (wangty0709@hotmail.com); ORCID 0000-0002-
7 5124-7965.

8 **Abstract**

9 Foreign-language learners with a fixed daily study budget divide their time between
10 vocabulary and grammar and decide when to stop memorising words. Published guidance
11 often places that stopping point near 98% coverage of the running words in a text. This
12 paper treats the decision as a finite-horizon optimal-control problem. The state is
13 mastered word families and a grammar-mastery index. A Zipf–Mandelbrot frequency
14 distribution converts the first into text coverage, and comprehension is a constant-
15 elasticity-of-substitution aggregate of coverage and grammar mastery, so the two skills
16 are complementary but not perfectly substitutable. Forgetting is exponential; reading
17 maintains some vocabulary at no cost, while grammar rehearsal is discounted by a
18 modelling parameter. Four coverage conventions are fitted and carried through every
19 analysis, each solved over a Latin-hypercube parameter box. Three findings hold across
20 all four conventions. A 98% target is optimal in only 1.67% to 4.17% of the parameter
21 box. The drivers of the optimal target are ranked identically every time: the free rehearsal
22 that reading supplies for vocabulary, then vocabulary retention, then the horizon, with
23 grammar-side parameters and the substitution elasticity nearly inert. One negative result

24 is reported: the grammar rehearsal weight shifts the objective level, not the policy
25 structure. Because the level of the optimal target depends on the coverage convention
26 while its drivers do not, no point estimate is offered. The analysis is a normative model,
27 not a prescription for individual learners.

28 **Keywords**

29 optimal control; vocabulary learning; grammar learning; text coverage; Zipf–Mandelbrot
30 distribution; sensitivity analysis

31 **1. Introduction**

32 **1.1 The problem and the received rule of thumb**

33 A learner of a foreign language who studies for a fixed time each day faces two decisions
34 at once: how to divide that time between vocabulary and grammar, and when to stop
35 memorising words. The literature offers a confident answer to the second in the form of a
36 coverage threshold. Knowing roughly 95% of the running words in a text is generally
37 taken as sufficient for adequate comprehension, and about 98% is often treated as the
38 level at which reading becomes unassisted; these thresholds have been converted into
39 vocabulary-size targets of a few thousand word families (Nation, 2006; Hu & Nation,
40 2000; Laufer & Ravenhorst-Kalovski, 2010). The dose-response relation between the
41 proportion of known words and comprehension is well replicated (Schmitt, Jiang &
42 Grabe, 2011). Because word frequencies are strongly skewed, a few thousand high-
43 frequency families already cover most tokens, which makes a frequency-ordered syllabus
44 attractive (Piantadosi, 2014). All of these are estimates rather than constants. The anchors
45 vary across sources, the counting unit is not applied identically everywhere, and word-
46 level frequency data deviate from the idealised Zipf form (Williams et al., 2015).

47 **1.2 Two literatures that have stayed apart**

48 Two mature bodies of work bear on this problem, and they have not been joined. The first
49 optimises the schedule of study. Novikoff, Kleinberg and Strogatz (2012) formalise the
50 tension between introducing new material and reviewing old material as a set of spacing
51 constraints, and bound the rate at which new material can be introduced. Tabibian et al.
52 (2019) cast spaced repetition as an optimal-control problem over a marked point process,
53 and show that the optimal review times follow from the recall probability itself. Pavlik
54 and Anderson (2008) compute practice schedules directly from a memory model. This
55 line answers when a given item should be revisited. Its material is homogeneous, a single
56 knowledge type, so the model carries no decision about dividing effort between two
57 skills, no distribution of item frequencies, and no production function combining them
58 into an outcome.

59 The second literature measures relative contribution. The simple view of reading treats
60 comprehension as the product of word-level decoding and language-level understanding
61 (Hoover & Gough, 1990), and meta-analyses assign separate weights to vocabulary
62 breadth and syntactic knowledge in second-language reading (Jeon & Yamashita, 2014;
63 Shiotsu & Weir, 2007). Work in the economics of skill formation supplies the
64 complementary structure and the vocabulary of dynamic complementarity (Cunha &
65 Heckman, 2007). These studies are static in a specific sense: the two skills enter as fixed
66 endowments whose joint contribution is estimated, and neither accumulates, decays, or is
67 abandoned, so stopping and switching do not arise.

68 **1.3 The gap**

69 Between those questions lies the gap this paper addresses. Vocabulary and grammar are
70 both needed for comprehension and are not perfect substitutes, and they differ in one
71 respect that neither literature has formalised: the probability that an encounter during
72 ordinary reading actually counts as rehearsal. Vocabulary is met continually in text, and
73 reading forces a retrieval of meaning at each re-encounter, so much of vocabulary
74 maintenance is supplied by the input itself. Grammar is met just as often, yet a reader
75 who recognises the words can often bypass the syntax and reach the meaning without
76 rehearsing the rule. We treat this asymmetry as a single rehearsal weight, applied in full
77 to vocabulary and discounted by a factor for grammar. The asymmetry has a qualitative
78 basis in the declarative/procedural account of memory, in which vocabulary behaves as
79 declarative fact and grammar as procedural rule (Ullman, 2004; Ullman, 2001). That
80 account supplies a direction rather than a rate, so the weight is a modelling assumption
81 and is scanned rather than set.

82 Once rehearsal is type-dependent, forgetting becomes skill-dependent, and the allocation
83 problem acquires a stopping dimension that has not been posed: at what coverage should
84 the learner stop building vocabulary and move to grammar, given that reading will keep
85 some words alive for free? A second, methodological gap follows from this.

86 Recommendations of the form “learn this many word families” rest on coverage anchors
87 drawn from different sources. It has not been established whether one frequency-based
88 coverage curve can satisfy those anchors at the same time, or whether the optimal target
89 is stable when the coverage curve is varied. To our knowledge no study has posed the
90 allocation question with the coverage curve treated as an object of sensitivity analysis.

91 **1.4 What this paper does**

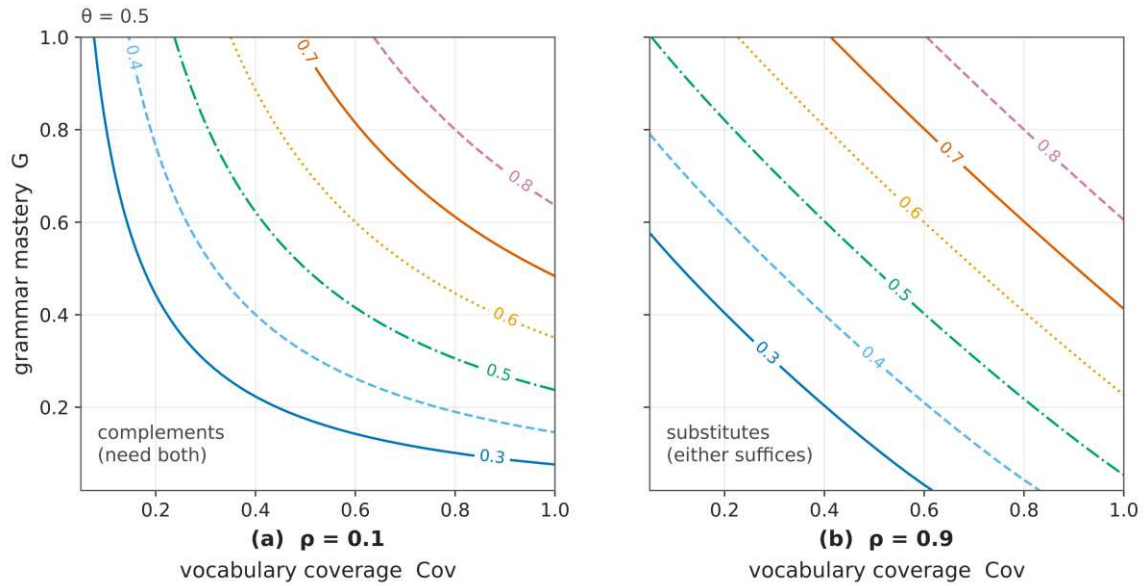
92 We build a parsimonious optimal-control model of a single learner who divides a fixed
93 daily budget among new-word study, vocabulary review, grammar study, and free
94 reading. The state is the number of mastered word families together with a grammar-
95 mastery index; a Zipf–Mandelbrot frequency distribution converts the first into text
96 coverage; comprehension is a constant-elasticity-of-substitution aggregate of coverage
97 and grammar mastery, so the two are complementary but not perfectly substitutable;
98 forgetting is exponential, with vocabulary rehearsal supplied by reading at no cost and
99 grammar rehearsal discounted by the rehearsal weight; and the objective is discounted
100 comprehension over a finite horizon plus a terminal comprehension term. Four coverage
101 curves, spanning the plausible range of tail exponents, are fitted and carried through
102 every analysis, and a 120-point Latin-hypercube parameter box is solved under each.
103 The paper makes five contributions. First, it gives an optimal stopping criterion for
104 vocabulary in decision-theoretic form, as the point at which the marginal comprehension
105 value of another word family falls below the marginal value of the next unit of grammar.
106 Second, it ranks the drivers of the optimal vocabulary target, and the ranking is identical
107 under all four coverage curves: the gain from free rehearsal (+0.71 to +0.79), the
108 vocabulary half-life (+0.34 to +0.39), and the horizon (+0.29 to +0.37) lead, while every
109 grammar-side parameter and the substitution elasticity lie at or below 0.13 in absolute
110 value. Third, the 98% rule occupies a corner of the parameter space rather than a typical
111 optimum: the share of the box at which a 98% target is optimal is 1.67%, 0.83%, 4.17%
112 and 1.67% across the four curves. Fourth, the level of the optimal target depends on the
113 coverage curve, which is why no point estimate is reported. Fifth, the model yields a set

114 of implications that existing data or a targeted experiment could test, set out in the
 115 Discussion.

116 2. Methods

117 2.1 Model overview

118 We model a learner who allocates a fixed daily study budget between vocabulary and
 119 grammar over a finite horizon. The state has two components: the number of mastered
 120 word families $N(t)$ and a grammar-mastery index $G(t) \in [0, 1]$. Vocabulary enters
 121 comprehension through text coverage $\text{Cov}(N)$ (2.2). Comprehension is a constant-
 122 elasticity-of-substitution (CES) aggregate (Figure 1),



123 Fig. 1. CES comprehension isoquants at low and high substitution. (a) $\theta = 0.5$, $\rho = 0.1$:
 124 complements. (b) $\theta = 0.5$, $\rho = 0.9$: substitutes. Curves are isoquants of $C = [\theta \text{Cov}^\rho +$
 125 $(1-\theta) G^\rho]^{1/\rho}$ at $C = 0.3\text{--}0.8$.

$$127 \quad C_t = \left(\theta \cdot \text{Cov}^{(N(t))} + (1 - \theta) \cdot G^{(t)} \right)^{\frac{1}{\rho}}, \quad \rho \leq 1 \quad (1)$$

128 so the two skills are complementary but not perfectly substitutable ($\rho \rightarrow 1$ is near-linear
 129 substitution, $\rho \rightarrow 0$ is the Leontief case). The structure follows the simple view of
 130 reading, which treats comprehension as a multiplicative function of word-level and
 131 language-level skill (Hoover & Gough, 1990), and the skill-formation tradition of
 132 dynamic complementarity (Cunha & Heckman, 2007).

133 One normalised daily unit, $T = 1$ (concretely, one hour), is split among four activities
 134 summing to one: new-word study x_1 , vocabulary review x_2 , grammar study x_3 , and free
 135 reading x_4 . High-frequency vocabulary is maintained partly for free, because reading
 136 forces meaning retrieval at each re-encounter. Writing $E = x_4 \cdot T$ for reading exposure,
 137 the state evolves as

$$138 \quad \frac{dN}{dt} = \frac{x_1 \cdot}{T / c_{\text{new}}} - \delta_V \cdot (1 - \lambda \cdot x_2) \cdot N + R_V, \quad R_V = \kappa_V \cdot E \cdot \text{Cov}(N) \quad (2)$$

$$139 \quad \frac{dG}{dt} = \frac{x_3 \cdot}{T / c_G} - \delta_G \cdot G + \varphi \cdot R_G, \quad R_G = \kappa_G \cdot E \cdot G \quad (3)$$

140 with $\delta_V = \ln 2 / \text{half_life_V}$ and $\delta_G = \ln 2 / \text{half_life_G}$. The asymmetry between the two
 141 equations is the object of interest: vocabulary rehearsal is granted in full, whereas
 142 grammar rehearsal is discounted by $\varphi \in [0, 1]$ (2.3). The objective is discounted
 143 comprehension plus a terminal term,

$$144 \quad J = \int_0^H e^{-rt} \cdot C_t \cdot x_4(t) \cdot T \, dt + e^{-rH} \cdot C_H \quad (4)$$

145 All results are reported as distributions over a parameter box (2.6), not as single point
 146 values.

147 **2.2 Coverage conventions**

148 Two quantities must be kept apart. The first is the word-frequency distribution: the
 149 frequency of the word of rank r follows a Zipf–Mandelbrot form proportional to $(r +$
 150 $\beta)^{-\alpha}$, with α near 1 in the token/type literature (Piantadosi, 2014). The second is

151 cumulative text coverage $Cov(N)$: the share of running words covered by a reader who
 152 knows the N most frequent word families, an integral of the first quantity truncated at a
 153 vocabulary size. Its empirical anchors are Nation (2006), who places 3,000 families at
 154 about 95% and 5,000 at about 98%, and Laufer & Ravenhorst-Kalovski (2010), who
 155 place a 95% threshold at 4,000–5,000 families and an optimal 98% threshold at 8,000.
 156 The unit is the word family (Nation, 2001); the operational coverage computation follows
 157 Cobb (2007). Because word-level Zipf behaviour has recognised deviations (Williams et
 158 al., 2015), α and β are scanned rather than fixed.

159 We calibrate coverage directly with three parameters,

$$160 \quad Cov(N) = \frac{\sum_{r \leq N} (r + \beta)^{-\alpha}}{\sum_{r \leq R_{MAX}} (r + \beta)^{-\alpha}} \quad (5)$$

161 where raising R_{MAX} adds low-frequency tail mass and lowers $Cov(N)$ at any fixed N . We
 162 fit four curves, C1–C4, and carry all four through every analysis:

Curve	α	β	R_{MAX}	residuals (3000 / 4000 / 5000 / 8000)	RMSE
C1 (v1 primary)	3.1708	1024.64	20000	0.000 / +0.0197 / 0.000 / +0.0125	0.01165
C2 (four- anchor joint fit)	2.0	170.87	80000	−0.0020 / +0.0110 / −0.0111 / +0.0011	0.00788
C3 ($\alpha = 1$, heavy tail)	1.0	−0.6395	6000	−0.0099 / +0.0150 / +0.0043 / +0.0200	0.01360
C4 ($\alpha = 3$, steep tail)	3.0	953.83	80000	−0.0081 / +0.0130 /	0.00927

				-0.0056 / +0.0088	
--	--	--	--	----------------------	--

163 A single curve would mislead. No curve with $\alpha \in [1, 3]$ satisfies all four anchors at once:
164 the RMSE floor is about 0.008, and the worst α -family member reaches 0.0136. An
165 exact-Nation two-anchor fit leaves residuals of +0.0198 at 4,000 families and +0.0121 at
166 8,000, RMSE 0.0116; enlarging R_{MAX} from 20,000 to 100,000 barely moves the 8,000
167 residual (0.0125 to 0.0121), so the third degree of freedom does not resolve the tension.
168 The tension is substantive: Nation locates the 95% threshold at 3,000 families, Laufer &
169 Ravenhorst-Kalovski at 4,000–5,000. This is a word-family measurement-unit difference,
170 not a failure of the frequency law, and we report it rather than smooth it away.

171 **2.3 Dynamics and the re-encounter parameter ϕ**

172 The parameter ϕ is a modelling assumption, not a measurement. Its qualitative basis is
173 the declarative/procedural account of memory, in which vocabulary behaves as
174 declarative fact and grammar as procedural rule (Ullman, 2004). That account supplies a
175 direction, not a rate: it gives no quantitative estimate of the probability that a grammar
176 rule met in reading is actually rehearsed. We therefore do not calibrate ϕ ; we scan it over
177 $[0, 1]$, with a finer scan on $[0, 0.2]$, and report its effect in full in 3.6.
178 The vocabulary free-rehearsal channel has an empirical basis: a single encounter yields a
179 modest acquisition gain (Swanborn & de Glopper, 1999; Nagy et al., 1985), and gains
180 grow with the number of encounters (Webb, 2007); the adaptive-memory literature
181 supplies the mechanism by which environmental occurrence statistics set retention
182 (Anderson & Schooler, 1991). We assume vocabulary re-encounters count as rehearsal in
183 full, and that grammar re-encounters count with weight ϕ . Setting $\phi = 1$ collapses the
184 model to a single-pool case in which grammar is rehearsed as readily as vocabulary.

185 2.4 Objective and baselines

186 We compare the optimised allocation against four fixed strategies: all-vocabulary ($x_1 =$
187 1), all-grammar ($x_3 = 1$), a fixed 50/50 uniform split ($x_1 = x_2 = x_3 = x_4 = 0.25$), and a
188 myopic greedy rule that maximises current comprehension one step ahead. In the central
189 scenario the optimised objective is $J = 74.0$, against $J = 18.8$ for the best fixed baseline, a
190 relative gain of +294%; in the faster-learning scenario the gain over the best baseline is
191 +251.5%. Every fixed baseline, including the myopic rule, lies far below the optimum.
192 This establishes that the control problem is well posed and that fixed allocations leave
193 substantial comprehension on the table; it is not, by itself, a statement about the level of
194 optimal coverage.

195 2.5 Numerical solution

196 Controls are parameterised by a softmax over the four activity shares, and the resulting
197 smooth, box-constrained problem is solved with SLSQP. The reference box uses $K = 10$
198 control intervals, $n_sub = 2$ integration substeps per interval, $maxiter = 120$, and two
199 starting points (uniform and a vocabulary-then-grammar ramp). The state equations are
200 integrated with a fourth-order Runge–Kutta step; N is clipped to the module constant
201 20,000, an operation identical across the four curves and therefore not a source of cross-
202 curve differences.

203 Discretisation was checked for convergence: the box uses $K = 10$, central-case
204 trajectories use $K = 24$, earlier runs used $K = 16$ and $K = 32$, and the model default is $K =$
205 48. Re-solving the first 20 box points at $K = 12$ reproduced the $K = 10$ medians (C1
206 0.465, C2 0.763, C3 0.806, C4 0.448), so the coarser box discretisation does not drive the
207 medians. One limitation applies here: at $N = 20$ the cross- K check is noisy for secondary

208 parameters, so this evidence supports the medians and the separation of the leading
209 drivers from the rest, but not the ordering of the middle parameters.

210 **2.6 Sensitivity design**

211 The parameter box is sampled with a Latin hypercube over discrete grids on nine
212 parameters: $c_G \in \{50, 100, 200, 400, 800\}$; grammar half-life $\in \{30, 60, 120, 240, 480,$
213 $1000\}$ days; vocabulary half-life $\in \{60, 180, 365, 730\}$ days; $\kappa_V \in \{2, 5, 10, 20\}$; $\kappa_G \in$
214 $\{0.01, 0.05, 0.2\}$; $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$; $\theta \in \{0.3, 0.5, 0.7\}$; $H \in \{180, 365, 730\}$
215 days; and $r \in \{0, 0.0005, 0.001\}$. The seed is fixed at 20260919. The original design used
216 $N = 300$ points.

217 For the cross-curve analysis each of the four coverage curves was solved on the first 120
218 points of that same design, so the four sets are nested and directly comparable. The
219 down-sampling is deliberate: at about 2.7 s per solve on a shared CPU, $300 \text{ points} \times 4$
220 curves would exceed 45 min, whereas 120×4 is reproducible and a genuine subset of the
221 full box. All solves completed without failure (120/120 per curve).

222 **2.7 Parameters and their provenance**

223 The model separates three kinds of input, and no free or assumed parameter is presented
224 as a measured value.

- 225 • Literature-anchored (functional form or anchor): the coverage anchors (Nation, 2006;
226 Laufer & Ravenhorst-Kalovski, 2010; Hu & Nation, 2000; Schmitt, Jiang & Grabe,
227 2011); the Zipf–Mandelbrot form (Piantadosi, 2014); the exponential-forgetting
228 family (Murre & Dros, 2015; Averell & Heathcote, 2011; Bahrack, 1984); the
229 incidental-acquisition channel (Swanborn & de Glopper, 1999; Nagy et al., 1985); the

230 grammar relearning threshold of at least three re-learnings (Serfaty & Serrano, 2024);
231 and the CES structural form (Hoover & Gough, 1990; Cunha & Heckman, 2007).

232 • Free (no reliable peer-reviewed value; scanned): c_G , half_life_V , half_life_G , κ_V ,
233 κ_G , and ρ . The absence of a peer-reviewed second-language vocabulary half-life is
234 precisely why the half-life is scanned over a wide range rather than fixed.

235 • Assumed (modelling choices; flagged and scanned): θ , H , r , ϕ , and λ .

236 The $\phi = 1$ corner is a modelling assumption used only to establish the single-pool limit; it
237 is not a claim that grammar re-encounters are fully rehearsed in practice.

238 **3. Results**

239 Every subsection closes with an explicit statement of what the reported numbers do and
240 do not support. Results are reported across the four coverage curves C1–C4 (2.2); no
241 single curve is treated as canonical.

242 **3.1 The optimal coverage target is driven by the vocabulary-side free-rehearsal** 243 **channel**

244 Across all four curves the rank correlation between the optimal terminal coverage Cov^*
245 and each parameter was stable. The leading three, in identical order under every curve,
246 are the free-rehearsal gain κ_V (+0.71, +0.78, +0.79, +0.71), the vocabulary half-life
247 (+0.38, +0.34, +0.37, +0.39), and the horizon H (+0.37, +0.33, +0.29, +0.37). Every
248 grammar-side parameter (c_G , κ_G , grammar half-life) and the substitution elasticity ρ has
249 $|\rho_s| \leq 0.13$ under all four curves; the largest absolute value among these four is 0.127. The
250 smallest correlation among the leading three (0.286) comfortably exceeds the largest
251 among the remaining four (0.127) under every curve. The same ordering appeared on the

larger 300-point box, where κ_v was +0.679, the vocabulary half-life +0.418, and H
+0.328. Figure 2 shows the nine correlations side by side for the four curves.

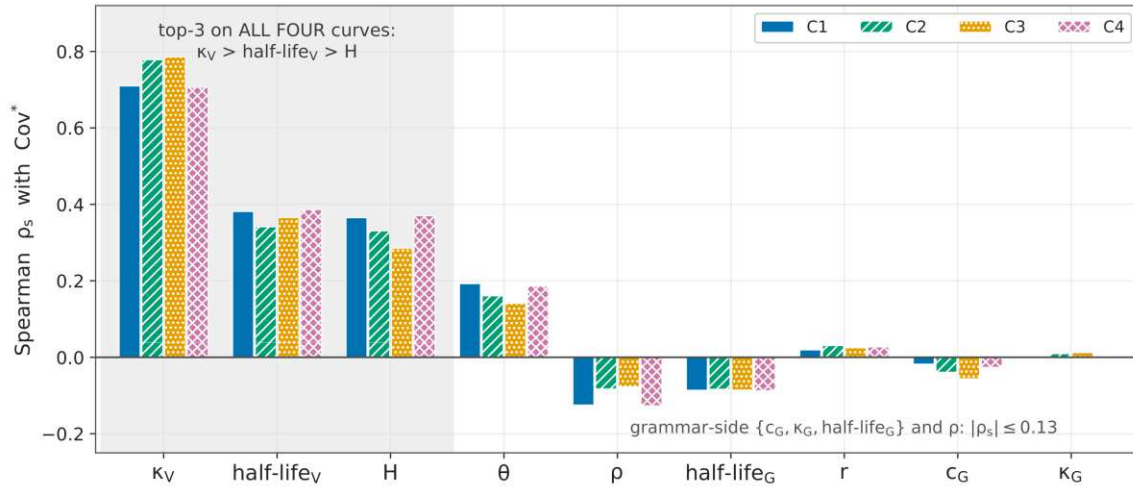


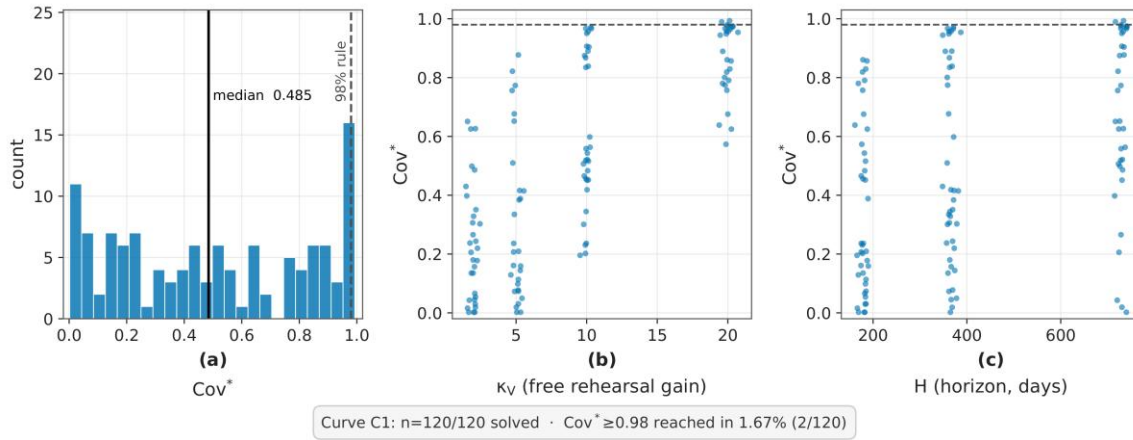
Fig. 2. Spearman association of each box parameter with optimal coverage, four curves.

Grouped bars: for each of the nine box parameters (ordered by $|\rho_s|$ under C1), the
Spearman correlation with Cov^* is shown for curves C1–C4.
The ordering that survives is stable: what the learner should do depends chiefly on the
free vocabulary rehearsal that reading supplies, on how long vocabulary is retained, and
on the length of the horizon. The same estimates give no support to the view that
grammar-side parameters or the substitution elasticity ρ drive the optimal vocabulary
target. The central role the design stage expected for ρ is absent, and the grammar-side
parameters are not merely weak but effectively irrelevant on this metric.

3.2 A 98% coverage target is a corner of the parameter box

The literature's 98% reading threshold is optimal only rarely. The share of the box with
 $\text{Cov}^* \geq 0.98$ is 1.67%, 0.83%, 4.17%, and 1.67% for C1–C4; on the larger 300-point box
the corresponding share is 3.33%. On every curve the share stays below about 4.2%.
Loosening the threshold does not change the picture qualitatively: at 95% coverage the

269 share of the box is still small, about 14.2% under C1. The corners that do reach 98%
 270 share a common signature: the free-rehearsal gain at its box maximum ($\kappa_v = 20$), the
 271 longest horizon ($H = 730$ days), and a slow-forgetting vocabulary (half-life of at least 365
 272 days). Figure 3 shows the Cov^* distribution and the 98% corner on curve C1; Figure 4
 273 reproduces the pattern for all four curves.



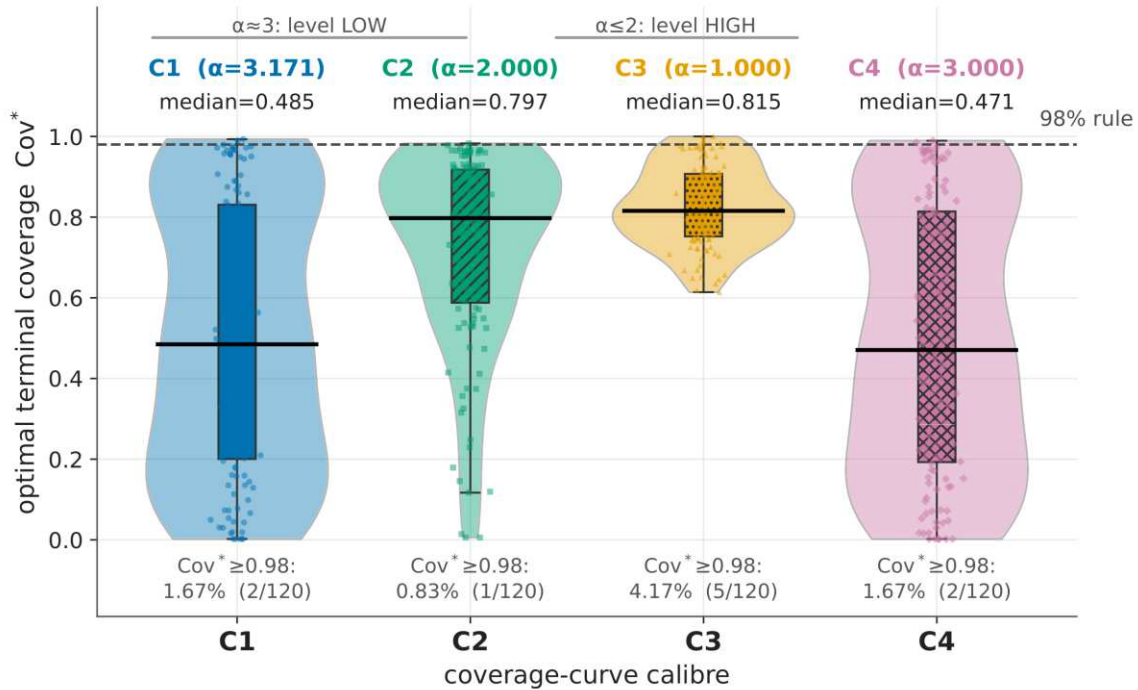
274 Fig. 3. Sensitivity of optimal coverage to the main drivers under the primary curve. Curve
 275 C1. (a) distribution of Cov^* ; (b) Cov^* against κ_v ; (c) Cov^* against H . Dashed line = 98%
 276 rule; solid line = median.

278 The claim that survives is conditional: 98% coverage becomes the optimal target when
 279 reading provides maximal free rehearsal, the horizon is long, and vocabulary decays
 280 slowly. Nothing here shows that 98% is never optimal, and nothing here yields a point
 281 estimate of the target.

282 3.3 The level of the target depends on the coverage curve

283 *The level of the optimal target is not invariant across curves. The box medians of Cov^**
 284 *are 0.485, 0.797, 0.815, and 0.471 for C1–C4, with interquartile ranges [0.200, 0.831],*
 285 *[0.588, 0.918], [0.752, 0.907], and [0.193, 0.814]. The steep-tail curves ($\alpha \approx 3$, C1 and*
 286 *C4) place the median near 0.47–0.49 and show wide spreads, whereas the flatter curves*

287 ($\alpha \leq 2$, C2 and C3) place it near 0.80–0.82 and show narrower spreads. The cross-K
 288 check reproduced these medians (C1 0.465, C2 0.763, C3 0.806, C4 0.448 at $K = 12$).
 289 Figure 4 displays the four distributions together.

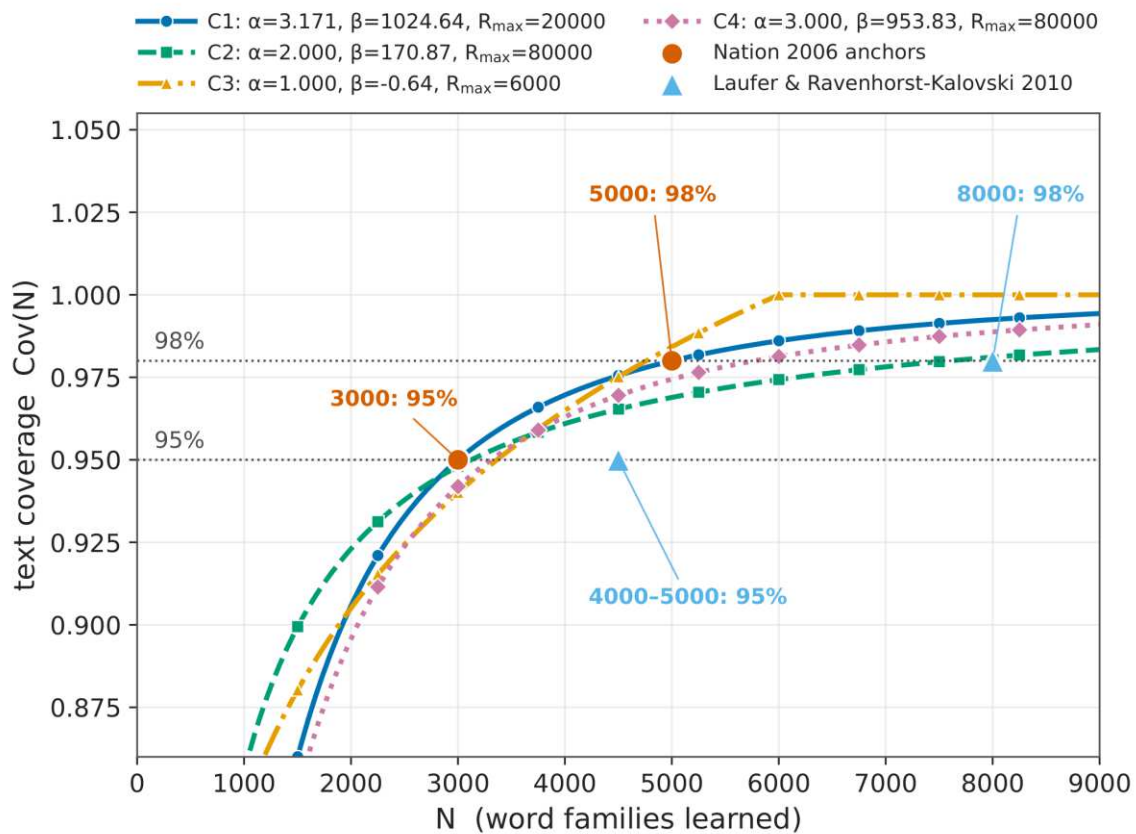


290
 291 Fig. 4. Optimal terminal coverage distribution flips with the coverage curve. Violin and
 292 box plots of the optimal terminal coverage Cov^* on the same 120-point sample space, one
 293 per coverage-calibre curve; per-curve median and the share with $\text{Cov}^* \geq 0.98$ are
 294 annotated. Dashed line = 98% rule.
 295 This is why the present study reports no point estimate of optimal coverage. What it
 296 reports instead is the ranking of drivers (3.1) and the rarity of the 98% corner (3.2), both
 297 of which hold across curves, with the level given only as curve-dependent. That the level
 298 moves with the coverage curve while the drivers do not is itself one of the study's
 299 findings. What survives the change of convention is the two statements above together

300 with the diagnosis of curve dependence; no single median, such as 0.485, can be quoted
 301 as *the* optimal coverage.

302 3.4 The four coverage anchors cannot be satisfied by one curve

303 No Zipf–Mandelbrot coverage curve with $\alpha \in [1, 3]$ reproduces all four literature anchors
 304 simultaneously; the residual floor over the α -family is about 0.008, and the worst case
 305 reaches 0.0136. Figure 5 plots the four calibrated curves against the four anchor points
 306 and shows that each curve misses at least one anchor. As set out in 2.2, the mismatch is
 307 between sources that disagree about where the 95% threshold lies (3,000 families in
 308 Nation, 2006; 4,000–5,000 in Laufer & Ravenhorst-Kalovski, 2010), not a failure of the
 309 frequency law.



310

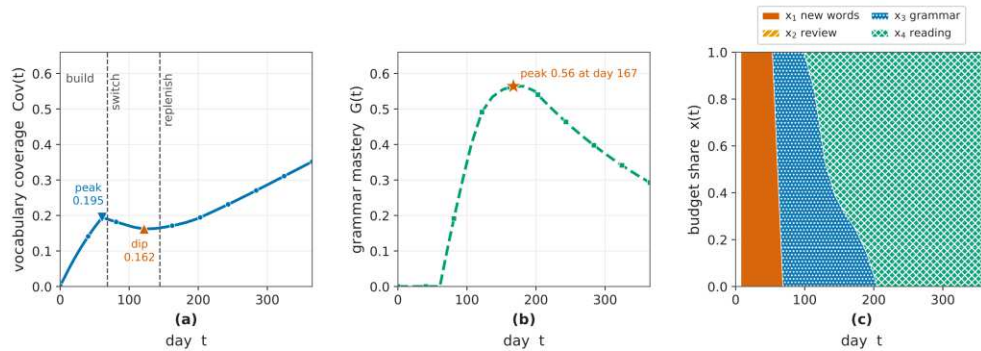
Fig. 5. Four coverage-calibre curves against four empirical vocabulary anchors. Single panel. Curves C1–C4 are Zipf–Mandelbrot cumulative text-coverage functions with the $(\alpha, \beta, R_{\max})$ values given in the legend (open markers every 750 word families). Filled markers are the four literature anchors: Nation 2006 (3,000→95%, 5,000→98%) and Laufer & Ravenhorst-Kalovski 2010 (4,000–5,000→95%, 8,000→98%). Dotted lines mark 95% and 98% coverage.

The methodological finding is this: any point recommendation of the form “memorise this many word families” is dependent on the coverage curve chosen, because the anchors on which such a recommendation would rest are mutually incompatible within the family of curves tested here. Nothing in the results fixes a specific vocabulary-size target, and nothing licenses the reverse claim that coverage is uninformative; the ranking of drivers survives the incompatibility (3.1).

3.5 A single build–switch–replenish episode, not a repeating structure

A single build–switch–replenish episode (vocabulary is built up, the learner switches to grammar, coverage is partly eroded by forgetting, and vocabulary is briefly replenished) occurs in 13.3%, 10.0%, and 14.2% of the box under C1, C2, and C4 respectively, but in 0% of the box under C3. No repeating structure is detected on any curve: the repeat detector returns 0% under all four curves, and the larger 300-point box also returned 0.0%. In the central scenario the episode appears and the maximum coverage drawdown over the grid is about 0.036. The central-scenario trajectory, including the budget shares across the four activities, is shown in Figure 6. Earlier bifurcation work located the episode in an intermediate grammar-half-life band (about 90–180 days) that is largely

insensitive to ρ . Figure S1 shows the existence boundary and is annotated with the C3 exception. Figure S2 repeats the three sensitivity panels for curves C2-C4.



One build-switch-replenish episode (not a repeating cycle). Episode share over the 120-point box: C1 13.3% · C2 10.0% · C4 14.2% · C3 0.0% (absent under C3). Repeating share: 0% on every curve.

Fig. 6. Optimal vocabulary and grammar trajectories with activity allocation. (a) vocabulary coverage $Cov(t)$; (b) grammar mastery $G(t)$; (c) budget shares $x(t)$ of the four activities (new words, review, grammar, reading). Central $\phi = 0.1$ scenario; dashed vertical lines separate the build, switch and replenish phases, and the build-phase peak and the switch-phase dip are marked. The terminal note reports the episode share over the 120-point box on each curve. Scope. The single build-switch-replenish episode is calibre-dependent: it is present on C1, C2 and C4 and absent under C3; no repeating episode occurs on any curve.

Because the structure never repeats, we do not describe it as periodic; the only description used is “a single build-switch-replenish episode”. One such episode does exist in the mid-half-life band, under three of the four curves. Periodicity is not supported, and neither is treating the episode as curve-invariant: it is undetectable under the heavy-tail curve C3, which offers no drawdown corridor.

3.6 Negative results: what the model did not deliver

Several effects predicted at the design stage were not reproduced, and are reported as negative results.

352 First, the re-encounter parameter ϕ changes only the level of the objective, not the
 353 structure of the policy. In the central scenario the optimised objective is $J \approx 74.0$ at $\phi =$
 354 0.1. Raising ϕ to 1 raises the objective to about $J \approx 154.2$, an increase of about 108%,
 355 while leaving the qualitative pattern of the allocation unchanged. The scan returns $\phi^* =$
 356 1.0 as the largest grid value at which the broad episode detector still fires, so ϕ never acts
 357 as a threshold bifurcation parameter.
 358 Second, the expected sign-flip theorem does not hold. The derivative of the optimal target
 359 with respect to reading exposure, $\partial \text{Cov}^* / \partial E$, is negative over the whole range $\rho \in [0.1,$
 360 1.0]: -0.10 at the lowest ρ and -0.23 at the highest. No critical ρ^* was found; the search
 361 returns no sign change.
 362 Third, a “mid-frequency band is optimal” result is not endogenous. The model does not
 363 generate a choice to study only one frequency band; the conditional pattern reported at
 364 the design stage depends on $\rho < 1$ and does not survive as a general result.
 365 What the mechanism does and does not do is now clear: the grammar rehearsal rate ϕ is a
 366 level shifter rather than a structural driver; the vocabulary-coverage choice is not a
 367 threshold phenomenon in ϕ ; and reading exposure does not flip the sign of the optimal
 368 vocabulary target. These results give no basis for presenting ϕ or ρ as structural drivers of
 369 the optimal allocation, and they are why the comparative-statics headline expected at the
 370 design stage was replaced by the driver-ranking result in 3.1.

371 **4. Discussion**

372 **4.1 Main findings**

373 The central result is a negative one about a widely repeated number. A 98% coverage
 374 target is optimal only in a small part of the parameter space. Across the four coverage

375 curves, the share of the box whose optimal target reaches or exceeds 0.98 is 1.67%,
 376 0.83%, 4.17% and 1.67%, and 3.33% on the larger 300-point box. The corners that do
 377 reach it share a signature: the free-rehearsal gain sits at its grid maximum, the horizon is
 378 at its longest, and vocabulary forgets slowly. All ten such samples in the larger box
 379 require $\kappa_v = 20$, $H = 730$ days, and a vocabulary half-life of at least 365 days. The
 380 condition is interpretable, but narrow.

381 The ranking of drivers is stable where the level is not. Under every coverage curve the
 382 optimal target is associated most strongly with the free rehearsal reading provides
 383 (Spearman +0.71 to +0.79), then with the vocabulary half-life (+0.34 to +0.39), then with
 384 the horizon (+0.29 to +0.37). Each grammar-side parameter and the substitution elasticity
 385 lies at or below 0.13 in absolute value, and the weakest of the leading three is more than
 386 twice the strongest of the remaining four under every curve. The elasticity of substitution,
 387 which the design stage expected to govern the comparative statics, is effectively inert on
 388 this metric.

389 The level of the target does move with the coverage curve. The box medians are 0.485,
 390 0.797, 0.815 and 0.471 from C1 to C4, separating the steep-tail curves near 0.47 to 0.49
 391 from the flatter curves near 0.80 to 0.82. Because the level is calibre-dependent while the
 392 drivers are not, we report the ranking and the rarity of the 98% corner, and decline to
 393 report a point estimate of optimal coverage.

394 **4.2 Expectations and reality**

395 The design stage anticipated several results that the model did not deliver. We report
 396 them as negative results rather than omit them.

397 The rehearsal weight for grammar shifts the level of the objective without changing the
398 structure of the policy. Raising it from 0.1 to 1.0 lifts the optimised objective from about
399 74.0 to about 154.2, an increase of roughly 108%, while leaving the qualitative allocation
400 pattern intact. It does not act as a threshold parameter.

401 A sign change in the comparative statics of reading exposure was expected but is absent.
402 The derivative of the optimal target with respect to reading exposure is negative
403 throughout the range of the substitution elasticity tested, from -0.10 at the lowest value
404 to -0.23 at the highest, and the search returns no critical value at which the sign turns. In
405 this model, reading more always lowers the optimal deliberate vocabulary target; it never
406 raises it.

407 A “mid-frequency band is optimal” result is not endogenous. The model does not
408 generate a choice to study only one band of the frequency distribution. The pattern seen
409 at the design stage depends on the substitution elasticity being strictly below one, and it
410 does not survive as a general result.

411 No repeating excursion exists. The repeat detector returns zero under all four curves, on
412 the 300-point box, and across the 90-cell bifurcation grid. A single build–switch–
413 replenish episode does occur, in which vocabulary is built up, the learner switches to
414 grammar, coverage is partly eroded by forgetting, and vocabulary is briefly replenished.
415 It appears in 13.3%, 10.0% and 14.2% of the box under three curves but in 0% under the
416 heavy-tail curve C3, so even this weaker structure is calibre-dependent. Because it never
417 repeats, we do not describe it as periodic.

418 **4.3 Relation to the nearest work**

419 Two neighbours deserve explicit contrast. Novikoff, Kleinberg and Strogatz (2012) treat
420 a schedule of instruction for a single, homogeneous knowledge type and bound the rate at
421 which new material can be introduced. Their model has no distribution of item
422 frequencies, so nothing is maintained free of charge by the input, and no production
423 function, so there is no outcome in which two skills combine. The present model reduces
424 their scheduling decision to a fixed daily budget and adds the asymmetry their framework
425 cannot express: vocabulary rehearsal is supplied at no cost by reading, whereas grammar
426 rehearsal must be purchased. That asymmetry, not the spacing constraint, drives the
427 allocative results here, and it is why the comparative statics of reading exposure have no
428 counterpart in their setting.

429 Zylich and Lan (2021) is the closest published attempt to model vocabulary and syntax
430 together, and it differs from the present work in kind rather than degree. It is a prediction
431 model: the aim is to forecast learner outcomes from features that include frequency-
432 dependent re-encounter, not to choose an allocation or a stopping point. It contains no
433 optimal control, no state dynamics with forgetting, and no division of effort across skills.
434 The mechanism claimed here, that the rehearsal value of a re-encounter is knowledge-
435 type-dependent, is not present there.

436 **4.4 Methodological contribution**

437 Distinct from what the model says about the learner, the calibration exercise carries a
438 methodological finding. No frequency-based coverage curve with a tail exponent in the
439 range tested reproduces all four literature anchors at once. The residual floor over the
440 family of curves is about 0.008, and the worst-fitting member of the family reaches
441 0.0136. An exact fit to the two Nation anchors still leaves a residual of about 0.020 at

442 4,000 families and of about 0.012 at 8,000, and enlarging the assumed total vocabulary
443 size from 20,000 to 100,000 barely moves the latter, from 0.0125 to 0.0121. The tension
444 is between sources that disagree about where the 95% threshold falls, at 3,000 families in
445 one and at 4,000 to 5,000 in the other, and it reflects a difference in counting unit rather
446 than a failure of the frequency law. The implication is that any point recommendation of
447 the form “memorise this many word families” is calibre-dependent, because the anchors
448 on which it would rest cannot be satisfied by a single curve. This does not make coverage
449 uninformative: the ranking of drivers survives the incompatibility.

450 **4.5 Practical implications**

451 The results do not license a prescription. The model is normative, and its free parameters
452 are not calibrated to any population, so it cannot tell a particular learner to aim at a
453 particular coverage level; the analysis also shows that any such level would depend on
454 the coverage curve chosen. What the model does support is weaker and, we think, more
455 useful. A high coverage target is defensible mainly when reading supplies abundant free
456 rehearsal, when the horizon is long, and when vocabulary decays slowly. Absent those
457 conditions, the model places the optimum well below the headline number of the
458 threshold literature. The practical content is conditional: the target should be set from the
459 learner’s exposure and retention, not read off a fixed percentage.

460 **4.6 Limitations**

461 Several limitations bound the claims. The model has many free parameters, and the box is
462 scanned rather than calibrated; in particular there is no peer-reviewed second-language
463 vocabulary half-life, which is why it is treated as free. The conclusions depend on the
464 coverage curve, as the cross-curve analysis shows, and we have not established which of

465 the four curves is correct. The analysis is restricted to English; coverage curves for other
466 languages lack comparable peer-reviewed values, so nothing here should be transferred
467 without recalibration. The learner is a single agent with no motivation or attention
468 constraints, and the budget is exogenous. The rehearsal weight for grammar rests on a
469 qualitative account of memory and has no quantitative estimate behind it, which makes
470 the grammar-side results the most fragile part of the model. Finally, the study is analytic;
471 no empirical test of its implications was carried out.

472 **4.7 Testable implications**

473 Five implications follow, each stated so that existing data or a targeted experiment could
474 test it.

- 475 1. If free rehearsal is the leading driver, learners with high reading exposure should, other
476 things equal, set a lower deliberate vocabulary target than learners with low exposure.
477 The model's comparative static on exposure is negative throughout, so a study relating
478 measured exposure to self-selected vocabulary targets should recover a negative
479 association once retention and horizon are held fixed.
- 480 2. If the 98% corner requires maximum free rehearsal and a long horizon, then short
481 intensive reading programmes should not, on average, settle on a 98% target; only long
482 programmes with sustained reading should.
- 483 3. If the optimal target tracks vocabulary retention, learners with better measured word
484 retention should invest relatively more in vocabulary and less in grammar. A design that
485 varies retrieval practice and then measures both retention and subsequent allocation
486 would test this.

487 4. If grammar-side parameters are near-irrelevant to the target, an intervention that
488 changes grammar-instruction efficiency should leave the chosen vocabulary target
489 roughly unchanged while moving the grammar allocation. The model predicts a
490 dissociation between the two.

491 5. If the single build–switch–replenish episode is real, longitudinal logs of study
492 allocation should show at most one shift from vocabulary-heavy to grammar-heavy study
493 followed by one short return to vocabulary, and no regular alternation. The repeat
494 detector returns zero across all four curves.

495 **5. Conclusion**

496 This study is normative. The model specifies how a learner should allocate a fixed daily
497 study budget once a particular objective and a particular set of mechanisms are accepted;
498 it does not estimate how any population behaves, and its free parameters are not
499 calibrated to any learner. Its conclusions are therefore confined to the model class
500 examined here, and two of them carry explicit qualifications. The level of the optimal
501 vocabulary target depends on the coverage convention, so the analysis reports no point
502 estimate; what survives across the conventions tested is the rarity of the 98% corner and
503 the ranking of drivers. The grammar rehearsal weight is a modelling assumption with no
504 quantitative estimate behind it, which makes the grammar-side results the most fragile
505 part of the model. Several expectations from the design stage were not reproduced and
506 are reported as negative results. Nothing here should be read as a prescription for
507 individual learners.

508 **Declarations**

509 **Declaration of competing interest**

510 The author declares that he has no known competing financial interests or personal
511 relationships that could have appeared to influence the work reported in this paper.

512 **Funding**

513 This research did not receive any specific grant from funding agencies in the public,
514 commercial, or not-for-profit sectors.

515 **CRedit authorship contribution statement**

516 Tian-Yu Wang: Conceptualization, Methodology, Software, Formal analysis,
517 Investigation, Writing – original draft, Writing – review & editing.

518 **Data availability**

519 This is a purely theoretical and numerical study; no primary data were collected or
520 analysed. The analysis scripts, the derived result files (JSON ledgers and result tables),
521 and the figure-generation scripts that reproduce every number and figure reported here
522 are available from the author on reasonable request.

523 **Declaration of generative AI and AI-assisted technologies in the manuscript** 524 **preparation process**

525 During the preparation of this work the author used Kimi K3 (Moonshot AI) in order to
526 carry out code development and language editing. After using this tool/service, the author
527 reviewed and edited the content as needed and takes full responsibility for the content of
528 the published article.

529 **Ethics statement**

530 Not applicable. This work is a mathematical modelling and numerical exercise. It uses no
531 human participants, no animal subjects, and no self-collected or personally identifiable
532 data, so no ethics approval, institutional review, or informed consent was required.

533 **References**

- 534 Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory.
535 *Psychological Science*, 2, 396–408. <https://doi.org/10.1111/j.1467-9280.1991.tb00174.x>
- 536 Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of
537 memories. *Journal of Mathematical Psychology*, 55, 25–35.
538 <https://doi.org/10.1016/j.jmp.2010.08.009>
- 539 Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory
540 for Spanish learned in school. *Journal of Experimental Psychology: General*, 113, 1–29.
541 <https://doi.org/10.1037/0096-3445.113.1.1>
- 542 Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning*
543 *& Technology*, 11, 38–63. <https://doi.org/10.64152/10125/44117>
- 544 Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. *American*
545 *Economic Review*, 97, 31–47. <https://doi.org/10.1257/aer.97.2.31>
- 546 Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and*
547 *Writing*, 2, 127–160. <https://doi.org/10.1007/bf00401799>
- 548 Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading
549 comprehension. *Reading in a Foreign Language*, 13, 403–430.
550 <https://doi.org/10.64152/10125/66973>
- 551 Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A
552 meta-analysis. *Language Learning*, 64, 160–212. <https://doi.org/10.1111/lang.12034>
- 553 Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical
554 text coverage, learners' vocabulary size and reading comprehension. *Reading in a*
555 *Foreign Language*, 22, 15–30. <https://doi.org/10.64152/10125/66648>

556 Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting
 557 curve. *PLOS ONE*, 10, e0120644. <https://doi.org/10.1371/journal.pone.0120644>
 558 Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context.
 559 *Reading Research Quarterly*, 20, 233. <https://doi.org/10.2307/747758>
 560 Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University
 561 Press. <https://doi.org/10.1017/cbo9781139524759>
 562 Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The*
 563 *Canadian Modern Language Review*, 63, 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
 564 Novikoff, T. P., Kleinberg, J., & Strogatz, S. H. (2012). Education of a model student.
 565 *Proceedings of the National Academy of Sciences*, 109, 1868–1873.
 566 <https://doi.org/10.1073/pnas.1109863109>
 567 Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of
 568 practice. *Journal of Experimental Psychology: Applied*, 14, 101–117.
 569 <https://doi.org/10.1037/1076-898x.14.2.101>
 570 Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review
 571 and future directions. *Psychonomic Bulletin & Review*, 21, 1112–1130.
 572 <https://doi.org/10.3758/s13423-014-0585-6>
 573 Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and
 574 reading comprehension. *The Modern Language Journal*, 95, 26–43.
 575 <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
 576 Serfaty, J., & Serrano, R. (2024). Practice makes perfect, but how much is necessary?
 577 The role of relearning in second language grammar acquisition. *Language Learning*, 74,
 578 218–248. <https://doi.org/10.1111/lang.12585>

579 Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and
580 vocabulary breadth in the prediction of reading comprehension test performance.
581 *Language Testing*, 24, 99–128. <https://doi.org/10.1177/0265532207071513>

582 Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A
583 meta-analysis. *Review of Educational Research*, 69, 261–285.
584 <https://doi.org/10.3102/00346543069003261>

585 Tabibian, B., Upadhyay, U., De, A., Zarezade, A., Schölkopf, B., & Gomez-Rodriguez,
586 M. (2019). Enhancing human learning via spaced repetition optimization. *Proceedings of*
587 *the National Academy of Sciences*, 116, 3988–3993.
588 <https://doi.org/10.1073/pnas.1815156116>

589 Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second
590 language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4,
591 105–122. <https://doi.org/10.1017/s1366728901000220>

592 Ullman, M. T. (2004). Contributions of memory circuits to language: The
593 declarative/procedural model. *Cognition*, 92, 231–270.
594 <https://doi.org/10.1016/j.cognition.2003.10.008>

595 Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*,
596 28, 46–65. <https://doi.org/10.1093/applin/aml048>

597 Williams, J. R., Lessard, P. R., Desu, S., et al. (2015). Zipf’s law holds for phrases, not
598 words. *Scientific Reports*, 5, 12209. <https://doi.org/10.1038/srep12209>

599 Zylich, B., & Lan, A. (2021). Linguistic skill modeling for second language acquisition.
600 In *LAK21: 11th International Conference on Learning Analytics and Knowledge*.
601 <https://doi.org/10.1145/3448139.3448153>

602 **Figure legends**

603 Fig. S1. Switch-episode existence boundary on the grammar half-life plane. (a) coverage
604 drawdown over (ρ , grammar half-life); (b) broad-detector episode existence over the
605 same plane; (c) critical grammar half-life against ρ ; (d) $K = 16$ half-life sweep at $\rho = 0.5$.
606 The bottom note reports the episode share over the 120-point box on each curve. Scope.
607 The episode is present on C1, C2 and C4, undetectable under C3; no repeating episode
608 occurs on any curve.

609 Fig. S2. Sensitivity panels for the remaining three coverage curves. (a)–(i): the Fig. 3
610 layout reproduced for curves C2 (a–c), C3 (d–f) and C4 (g–i).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1Bifurcation.tif](#)
- [FigS2SensitivityAppendix.tif](#)