JCSS 2026

さくら → らくだ → だいこん → ×

非同期L2しりとりにおける規則の曖昧性と分布的プレイ

Rule-ambiguity and distributional play in asynchronous L2 shiritori

Joseph L. Austerweil · Chiba Institute of Technology
Jumpei Nishikawa · Okayama Prefectural University
Junya Morita · Shizuoka University

CHIBA TECH · OKAYAMA PREF. UNIV. · SHIZUOKA UNIV.

THE GAME

しりとりというゲーム

Each word must begin with the final *mora* of the word before it. End on ん and you lose.

さくら → らくだ → だいこん → ×

sakura — rakuda — daikon — chain ends

MORA

Japanese timing unit — roughly a CV pair. ん counts as a mora of its own, and loses, because no Japanese word begins with it.

YŌON

Palatalised CV + small ゃゅょ — *one* mora written across *two* kana. きょ = *kyo*. This is where the rule gets interesting.

MOTIVATION

なぜしりとりを研究するのか

Shiritori is a naturalistic paradigm for *constrained lexical retrieval*.

動的な手がかり

The phonological cue changes every turn, set by the previous player's word.

移動する罠

Players must avoid endings that corner the next player — a strategic layer absent from fluency tasks.

曖昧な規則の縁

At yōon boundaries several continuations are equally legal. The rule has genuinely open edges.

semantic fluency	letter fluency	shiritori
cue is free	cue is fixed	**cue is self-conditioning**

Links search-on-a-semantic-network accounts (Abbott et al., 2015) to multiplex lexical networks (Vitevitch, 2008).

PRIOR WORK

先行研究：ACT-Rによるしりとりモデル

Shiritori already has a cognitive-architecture account. Ours sits at a different grain.

NISHIKAWA & MORITA 2022

An ACT-R model of phonological awareness: a child agent and an adult agent play shiritori. Retrieval errors arise from **partial matching** on mora similarity; repeated play plus adult scaffolding corrects them through **base-level learning**.

Nishikawa, Sasaki & Morita 2025 extend this to the activation- and association-based factors shaping *sequence length* and *coherence*.

粒度が違う

- **Theirs:** process-level simulation. Individual agents, trial by trial, with a mechanism for every step.
- **Ours:** population-level fit. 9,891 real moves, distributional structure, no commitment to within-trial process.

They are most informative read together — a claim this talk will try to earn rather than assert.

DATA

データ：Bunproフォーラム

Asynchronous play by L2 Japanese learners on a public Discourse forum.

9,891
valid game moves parsed

568
unique players

59
months, Oct 2019 – Feb 2026

118 min
median gap between turns

WHY THIS SETTING HELPS

Turns span minutes to days, so time pressure is removed. Players self-report a proficiency index (Bunpro level 1–101, loosely JLPT N5–N1). The forum *posts* three legal yōon continuations, which exposes rule-variant behaviour that synchronous play would drown out.

KNOWN CAVEATS

The proficiency level is a snapshot scraped at analysis time, not a post-time value. The corpus selects for motivated self-study learners. Data licensed CC-BY-NC-SA 3.0 from community.bunpro.jp.

THE CORPUS

コーパスの姿

A few core players, and a strong pull toward words nobody has played yet.

参加はごく少数に偏る

0.848
Gini coefficient

41.5%
of moves from the top 1%

2
moves by the median player

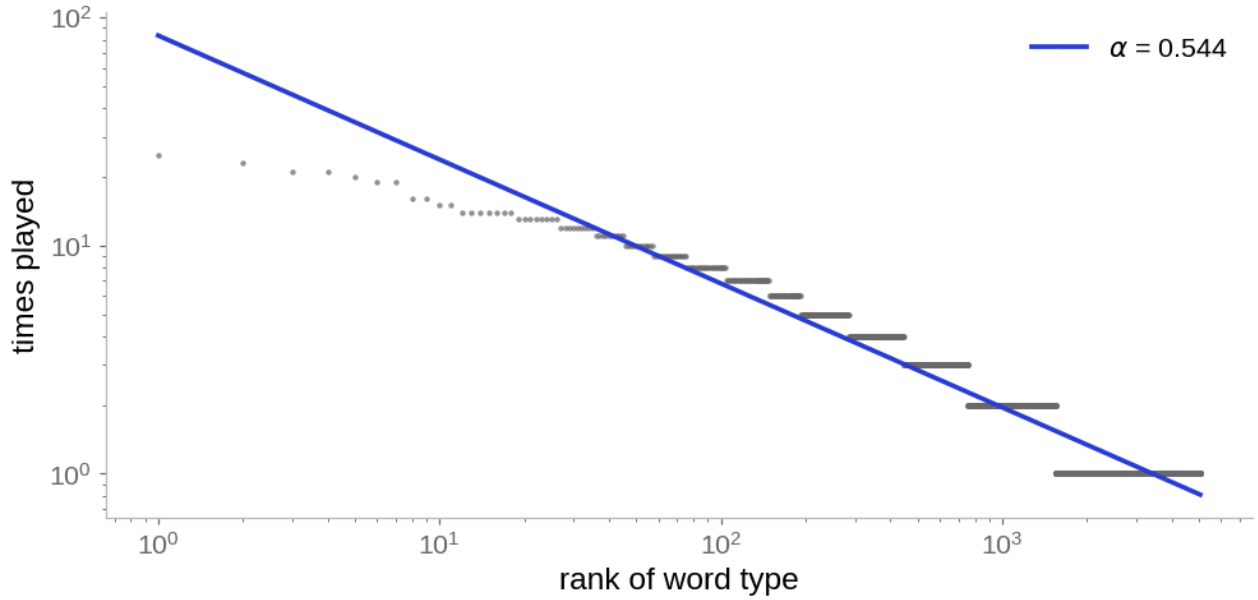
The most prolific player made 976. This is why every model below carries user-level random effects.

新しい語への圧力

69.4%
of word types played once

57.2%
of moves were new words

Heaps' β = 0.872. Players actively seek novel vocabulary rather than recycling common words — hold on to this, it returns at the model.



5,061 word types across the corpus. α = 0.544, well below Zipf's ideal ~1.0.

STRUCTURE

モーラ構造とトラップ

Some morae are easy to hand on. Some corner the next player.

TRAP MORAE — MANY WORDS END HERE, FEW START HERE

ん	n	0.016
ー	chōonpu	0.024
よ	yo	0.295

HUB MORAE — EASY TO CONTINUE FROM

ほ	ho	8.75
あ	a	4.80
は	ha	4.71

ratio = times used as initial ÷ times used as final

3.75

mean word length in morae (SD 1.82)

1.584 bits

mean transition entropy per mora

4.2%

of moves break the chain — median at move 6

Low-entropy endings (さ 0.90, み 0.94) are bottlenecks: few options for whoever plays next. These asymmetries are where the strategy lives.

CHIBA TECH OKAYAMA PREF. UNIV. SHIZUOKA UNIV.

THE CLAIM

音韻的制約は三つの水準で処理される

We argue the constraint is handled at three levels that come apart in the data.

H1

内容

Content. *Which* word is retrieved. Indexed by proficiency.

H2

方略

Strategy. *What it does to the next player.* Indexed by individual experience.

H3

規則解釈

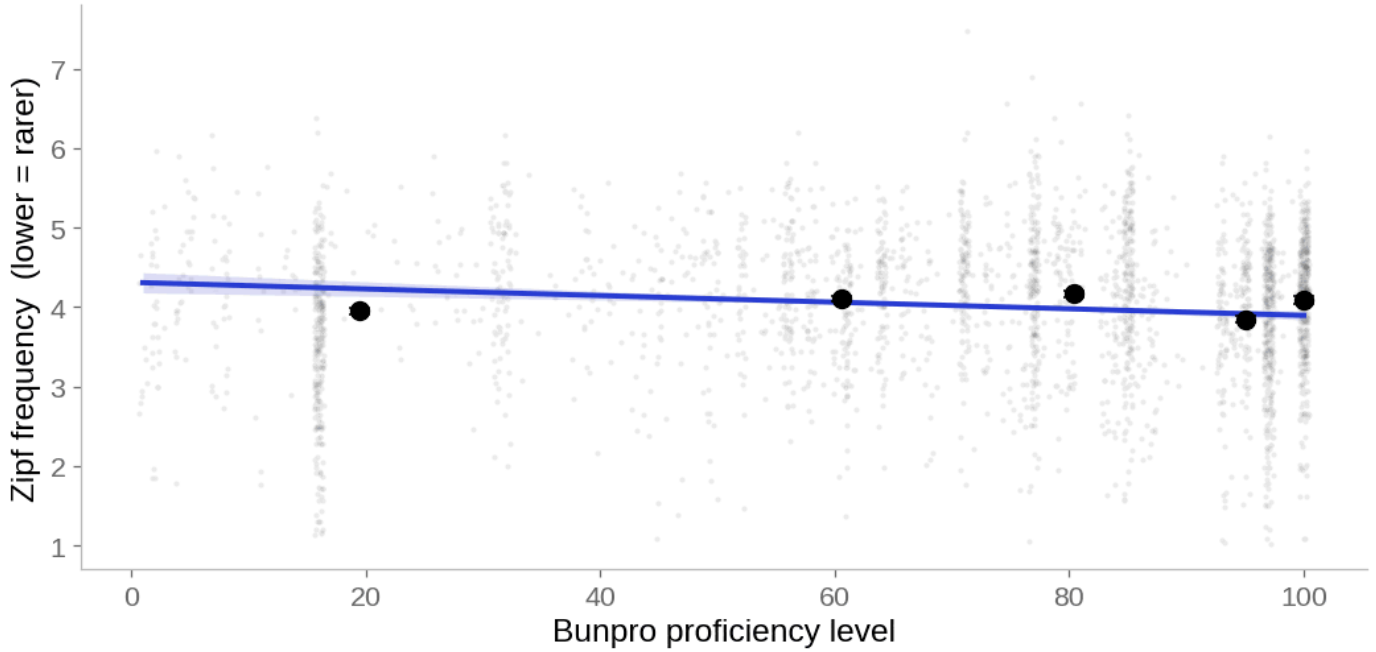
Rule interpretation. *What counts as legal.* Indexed by within-player variability.

Then **H4**: one multiplex random-walk model that has to accommodate all three at once.

H1 · CONTENT

熟達度は語の希少性を予測する

Higher-proficiency players produce measurably rarer words.



BAYESIAN LMM

β = −0.004

95% HDI [−0.006, −0.002]

P(β < 0) = 1.00

The entire posterior sits below zero. $\hat{R} \leq 1.01$.

Random intercepts and slopes by user; thread and position-in-game as covariates. Naturalistically recapitulates Laufer & Nation (1995).

The effect is small and *within*-player: raw between-player quintile means (black) are nearly flat, because who plays at each level differs. The LMM slope, not the binned means, carries the claim.

H2 · STRATEGY

個人か、共同体か

For any move-level outcome, we can ask two very different questions.

個人の経験

INDIVIDUAL EXPERIENCE

The player's own cumulative move index, min–max scaled within player, so the slope is identified *within* a person. Between-player differences are absorbed by user random intercepts and slopes.

> “Do *you* change over your own participation?”

共同体の水準

COMMUNITY EXPOSURE

The monthly mean of the outcome across *all* players' moves that month — the register a new move is entering.

> “What is the room doing right now?”

Two outcomes, two parallel models, following Barr et al. (2013): vocabulary sophistication (Zipf frequency) and trap-mora avoidance (binary).

H2 · STRATEGY

個人の経験が効くのは方略だけ

The dissociation lives in the individual-experience column.

	INDIVIDUAL EXPERIENCE	COMMUNITY
語彙の洗練 vocabulary sophistication	○ NO EFFECT β = −0.019 p = .84	● RELIABLE **β = 0.683** $p = 2.2\times10^{-10}$
トラップ回避 trap-mora avoidance	● RELIABLE **OR = 0.272** p = .010	● RELIABLE **also reliable** LPM $p = 1.2\times10^{-6}$

Players **individually learn** to avoid traps — roughly a 73% within-player reduction over their participation.

They do **not** individually become more sophisticated. Vocabulary comes from the room.

CHIBA TECH · OKAYAMA PREF. UNIV. · SHIZUOKA UNIV.

BRIDGE TO ACT-R

生得と経験、個人と共同体

The same shape of split appears one level down, inside the ACT-R account.

NISHIKAWA & MORITA 2022 · WITHIN THE AGENT

P_i — similarity, the **innate** term

B_i — base level, the **experiential** term

Their simulations sweep the balance between the two. Errors come from the innate side; correction comes from the experiential side.

HERE · ACROSS THE POPULATION

individual experience — **strategy**

community register — **vocabulary**

Same shape, different register: theirs is a within-agent split, ours is a within/between-person split. Worth putting side by side, not merging.

One striking contrast: their child agent, left to repeat the task, converges pathologically and repeats a single word — it needs adult intervention between sessions to escape. Our forum needs no such intervention: 57.2% of moves introduce a word nobody has played. The community supplies for free what their simulation had to be given.

CHIBA TECH · OKAYAMA PREF. UNIV. · SHIZUOKA UNIV.

H2 · STRATEGY

L1とL2でトラップは異なる

Trap structure is a property of the population's accessible lexicon, not of Japanese.

り

ratio ≈ 1.24

used *more* often as initial than as final

り is a classically rare-initial mora in Japanese (Murata & Isahara, 2002) — a trap for L1 players. It is **not** a trap here.

L2-accessible vocabulary reseeds it as an initial: 鳥 *tori*, 利用 *riyō*, 料理 *ryōri*. The rare-initial status in Murata & Isahara rests on a much broader native lexicon.

THE L2 TRAP SET INSTEAD

ん　よ　ー

These appear often word-finally but rarely word-initially in L2 play, structurally cornering the next player.

The L1/L2 divergence is itself a **prediction**: change the population, and the trap set should move with its accessible vocabulary.

CHIBA TECH　OKAYAMA PREF. UNIV.　SHIZUOKA UNIV.

H3 · RULE INTERPRETATION

拗音の境界は形式的に曖昧である

When the previous word ends in a yōon mora, several continuations are equally legal.

金魚

kingyo “goldfish”

ends in ぎょ — one mora, two kana

→

ぎょ the whole mora	よ the small kana alone
き the base kana	きょ modulo *dakuten* voicing

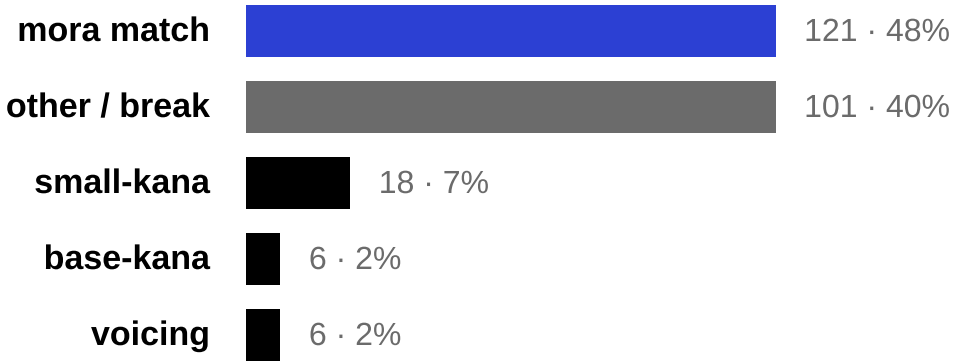
This is not us being permissive. The forum’s own posted rules explicitly license the first three. That permissiveness is atypical for native play — and it is exactly what makes rule-variant navigation measurable.

H3 · RULE INTERPRETATION

規則は分布として使われる

Mora-match is the plurality, but not the majority.

252 YŌON-CONTEXT TRANSITIONS



WORKED EXAMPLES

領収書・しょ → 証明・しょ

mora match · *ryōshūsho → shōmei*

金魚・ぎょ → 夜・よる

small-kana · *kingyo → yoru*

労働者・しゃ → 師匠・し

base-kana · *rōdōsha → shishō*

神社・じゃ → 社宅・しゃ

voicing · *jinja → shataku*

H3 × ACT-R

部分照合の誤りか、曖昧性の航行か

The ACT-R account predicts one of these cells. Does it explain them all?

THE PARTIAL-MATCHING ACCOUNT

In Nishikawa & Morita, retrieval can land on a chunk that does not exactly match the request when its activation is high enough. Similar morae are confusable, so **voicing confusions** し ↔ じ fall out of the mechanism directly.

Our voicing cell — 神社・じゃ → 社宅・しゃ, $n = 6$ — is exactly what that predicts. This is a real hit, not a coincidence.

BUT IT CANNOT BE THE WHOLE STORY HERE

90.5% of players contributing ≥ 3 yōon transitions used **two or more** distinct variants (19 of 21)

The same player is not consistently picking one reading. They sample from the variant distribution *within themselves* — inconsistent with a fixed-but-wrong internal rule.

The setting adjudicates: in a typed forum, players re-read before posting. The production errors that model is built for are largely suppressed — so what remains reads as ambiguity navigation.

CHIBA TECH · OKAYAMA PREF. UNIV. · SHIZUOKA UNIV.

JOINT SUMMARY

三水準は互いに還元できない

None of the three levels reduces to the others in these data.

LEVEL	INDEXED BY	NOT PREDICTED BY
内容 content · H1	proficiency	experience
方略 strategy · H2	individual experience	proficiency
規則解釈 rule interpretation · H3	within-player heterogeneity	either of the above

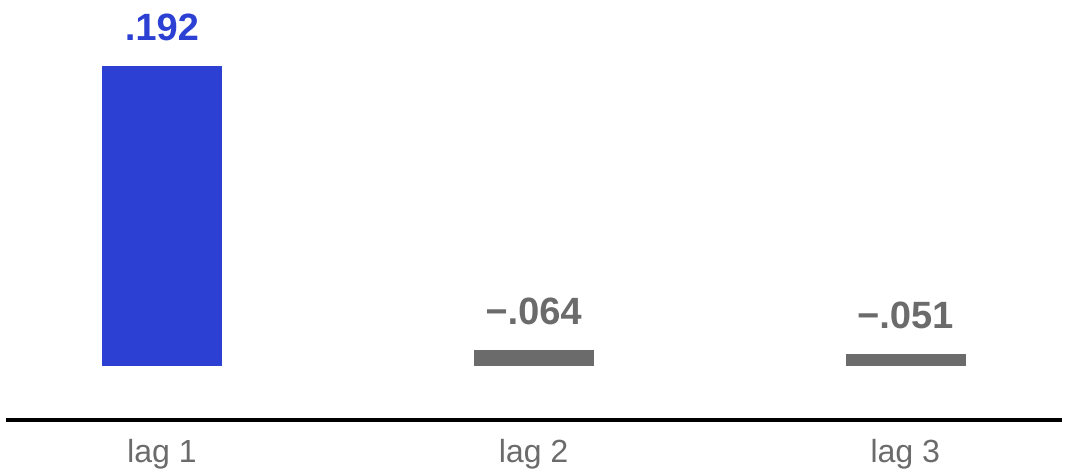
So: can a *single* computational account capture all three at once?

SEMANTIC STRUCTURE

意味構造は制約を生き延びるか

Only just, and only next door.

SEMANTIC AUTOCORRELATION BY LAG



Phonologically matched null ≈ −.05 at every lag. Lag 1 clears it; nothing after does.

局所的な意味のまとまりは残る

Consecutive moves are more similar than a matched null — .898 vs .896, $t = 7.13$, $p = 1.1\times10^{-12}$ — but the gap is 0.002 on a 0–1 scale. Reliable, negligible.

熟達度は意味空間の次元に効く

Proficiency does not predict local similarity ($p = .484$), but it does predict how many components a player's word choices need: $r = -.249$, $p = .017$. Higher proficiency spans a **higher-dimensional** semantic space.

Semantic structure is present but weak — which is exactly why the model needs a semantic layer *and* cannot lean on it alone.

H4 · MODEL

多重ネットワーク上の打ち切りランダムウォーク

Each move is one step on a walk over three superimposed lexical layers.

音韻層
kana edit-distance neighbourhoods

意味層
multilingual-e5 embedding *k*-NN

遷移層
the legal-next-move graph, induced by mora rules

FOUR FREE PARAMETERS

- ω how the layers are mixed
- ε residual probability on rule-violating moves
- α Zipf-frequency bias
- λ exponential memory decay

Five variants fit by MLE (L-BFGS-B) on 7,298 moves, compared by AIC.

H4 · MODEL

規則を構造として符号化すると勝つ

Model comparison, ΔAIC — lower is better.

Model A and Model B share the same semantic layer. The difference is *only* whether the mora rule is an external filter on the walk or is **built into the graph as edges**. That is worth 154 AIC points.

Adding the orthographic layer on top (Model C) **hurts**. The next two slides are about why that is not the phonology null it looks like.

H4 · MODEL

推定されたパラメータ

What the winning model says the players are doing.

$\omega_{transition} = 0.674$

The rule graph gets roughly twice the weight of meaning ($\omega_{semantic} = 0.326$). Retrieval is dominated by what is *legal*, but meaning is not vestigial.

$\alpha = -0.055$

A **negative** frequency bias — the model prefers rarer words. This is the community rare-word norm from slide 11, recovered independently by the fit.

$\lambda = 0.10$

Slow decay: the walk carries a long memory of where it has been.

$\varepsilon = 0.30$ — saturated

ε sits on its upper bound and acts as a uniform rule-error residual. It **absorbs** the whole yōon variant distribution of H3 as noise without representing any of its structure.

The model accommodates H3. It does not explain it. Making the variants real sub-edges would turn ε from a catch-all into a prediction.

H4 · LIMITATION

なぜ音韻層は効かなかったのか

Not a phonology null — a construct mismatch.

WHAT WE ACTUALLY BUILT

A **kana-string edit distance**: a direct import of the Vitevitch-style orthographic-neighbourhood construct from English work, ported to kana.

さくら — one kana — さくらん

It measures how close two words look as strings of characters. That is an orthographic claim wearing a phonological label.

WHY IT SHOULD NOT HAVE CARRIED MUCH LOAD HERE

The principled Japanese construct works at the **mora** level, on distinctive features — and it is designed to capture within-mora confusions that produce rule-violating *production* errors.

In a text forum, players type and re-read before posting. Production errors of that kind are rare. Neither construct should be expected to carry much predictive weight in *this* corpus.

So Model C's loss is narrower than it looks. The right response is not to drop phonology — it is to use the right construct.

THE RIGHT CONSTRUCT

西川・森田のモーラ類似度

Mora-level distinctive-feature similarity — the layer we should have had.

HOW IT IS BUILT

1. Each mora becomes a vector of **distinctive features** — vocalic, consonantal, high, back, low, anterior, coronal, voice, nasal, continuant, strident…
2. Blank / − / + are coded 0 / 1 / 2, so context-dependent features carry the least weight.
3. Vowel-only morae take the mean over all Japanese consonants in the consonant slot — every mora is one duration unit, consonant before vowel.
4. Cosine similarity over every pair: **103 morae → 10,506 pairs**.

WHAT CHANGES

き ↔ ぎ

one feature apart (voice). **Near neighbours** under features; unrelated under kana edit distance, which sees two different characters.

し ↔ じ

the voicing confusion behind our H3 *voicing* cell — representable in the feature layer, invisible in ours.

Concretely: replace the edit-distance adjacency with this 103 × 103 matrix as the phonological layer's edge weights, and refit.

READING THEM TOGETHER

二つのモデルは何を分担するか

Different commitments, and a place where the parameters line up.

	ACT-R PROCESS MODEL	MULTIPLEX NETWORK MODEL
grain	one agent, trial by trial	568 players, 7,298 moves
phonological similarity	P · feature-cosine mismatch penalty	ω_{phon} · layer mixture weight
memory / recency	base-level decay $d = 0.5$	$\lambda = 0.10$
word frequency	base-level activation from use	$\alpha = -0.055$

Neither yet has the piece the next slide is about. An ACT-R production system *could* carry it — a single-stage random walk structurally cannot.

H4 · GENERATIVE TEST

モデルが失敗するところ

Simulate games from the fitted model and the failure is diagnostic.

CHAINS ENDING IN ん

MEAN GAME LENGTH

16.3
simulated moves

26.2
observed moves

人は二段階でやっている

Real players appear to **generate** candidates by navigating the network, then **screen** them against ん-endings before committing. A single-stage walk has nowhere to put the second step.

And H2 says who does the screening: trap avoidance is the one thing players learn *individually*. The screening stage is the individually-learned stage.

Sequence length and coherence are exactly what Nishikawa, Sasaki & Morita (2025) model — the natural place to look next.

CHIBA TECH · OKAYAMA PREF. UNIV. · SHIZUOKA UNIV.

SUMMARY

まとめと今後の課題

WHAT WE FOUND

Three dissociable levels. Proficiency indexes content; individual experience indexes strategy; within-player heterogeneity indexes rule interpretation.

One model accommodates all three — transition + semantic, ω = 0.674, AIC weight .9995.

And visibly falls short. The generative test exposes a screening stage the walk cannot represent.

All three dissociations are post-hoc, from exploratory analyses. The proficiency measure is a snapshot, and the corpus selects for motivated self-study learners.

NEXT

Swap in **mora-level distinctive-feature similarity** as the phonological layer, and refit.

Represent the yōon variants as **sub-edges** in the transition graph, so ε becomes predictive rather than a catch-all.

Run the **matched-rule comparison with L1 players** — the cleanest test of the distributional account.

Read the layer weights against the **ACT-R activation dynamics** directly.

ありがとうございました

BACKUP · PARSING

パースの信頼性

Move extraction confidence by thread.

THREAD	High	Medium	Low
standard	5,991 80.4%	401	1,059
kanji	1,139 85.9%	44	142
original	10 13.2%	3	63

The original thread's early posts have inconsistent formatting, hence the low confidence. It contributes 76 of 9,891 moves, so it cannot drive any reported effect; thread is carried as a fixed covariate throughout.

BACKUP · TIMING

応答時間

Asynchronous by a wide margin.

118

min

median inter-post interval

339

min

mean — heavy right skew

8,308

intervals measured

Most responses arrive within a few hours; some gaps span days. Response latency is **not** interpretable as a retrieval-time measure here — which is the point. Removing time pressure is what separates content from cadence.

Proficiency does not predict response time: β ≈ 0, $p = .989$.

BACKUP · PLAYERS

ヘビーユーザーとライトユーザー

Engagement volume vs. proficiency.

GROUP	Users	Moves	Guiraud	Restart rate
heavy · > 50 moves	25	6,140	12.70	4.4%
moderate · 10–50	78	1,667	4.45	4.6%
light · < 10	465	1,045	1.39	6.1%

Guiraud's index scales with sample size, so the heavy/light gap is partly mechanical. An LMM separating engagement from proficiency finds engagement volume non-significant ($p = .904$) while proficiency holds ($\beta = 0.0032$, $p = .047$) — proficiency composition, not sheer volume.

BACKUP · MODEL SPEC

モデルの定式化

The censored mixture, and what each variant actually contains.

The walker's next-move distribution is a mixture over layers weighted by ω, censored by the shiritori rule: illegal next-moves receive probability $\varepsilon/|V|$, with $0 \le \varepsilon \le 0.3$.

A Zipf-frequency bias term α and an exponential memory decay λ complete the specification. Fit by MLE (L-BFGS-B) on 7,298 moves; all five variants converged.

Semantic layer: multilingual-e5-small *k*-NN, validated against JWSAN similarity norms.

THE FIVE VARIANTS, AS IMPLEMENTED

Model B	transition + semantic · rule as edges	0.0
Model C	phonological + transition + semantic	15.1
Uniform	semantic layer uniformised · rule as filter	45.4
Model A	phonological + semantic · rule as filter	153.9
Phon-only	phonological + uniformised semantic	253.5

BACKUP · ACT-R

ACT-Rの活性化式

Nishikawa & Morita (2022), for reference.

$$A_i = B_i + S_i + P_i + \varepsilon_i$$

chunk activation; the highest-activation match is retrieved

$$P_i = P \cdot M_i$$

partial matching. $M_i \in [-1, 0]$ is the mismatch penalty from mora feature-cosine similarity; P weights it. The **innate** term.

$$B_i = \ln\left(\sum_j t_j^{-d}\right) + \beta_i$$

base-level learning over n presentations. The **experiential** term.

SIMULATION SETTINGS

similarity weight $P \in \{1, 10, 30, 60\}$

base-level offset $\beta_i \in \{0.1, 10, 20\}$

decay $d = 0.5$ noise $\varepsilon_i = 0.5$

child lexicon 2,054 nouns sampled from 20,544

mora chunks 103 → 10,506 similarity pairs

runs 100 trials × 3,600 s of ACT-R time

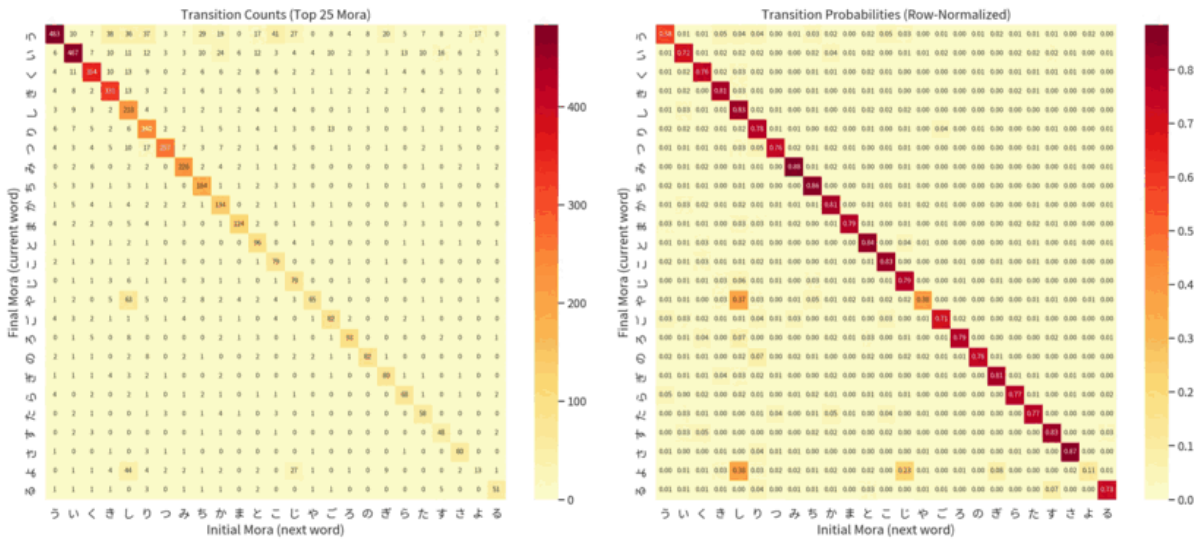
Low P yields mora confusions; raising P collapses the mora-pair entropy onto correct pairs. Repeated sessions converge pathologically until an between-session intervention broadens activation across the kana inventory.

CHIBA TECH OKAYAMA PREF. UNIV. SHIZUOKA UNIV.

BACKUP · BASELINE

遷移行列とJMdictベースライン

Observed play against a natural-language null.



Initial × final mora transitions, standard thread (7,523 transitions).

The null distribution is built from JMdict nouns: what mora transitions *would* look like if players drew from the dictionary rather than from what they can retrieve under pressure.

Mean Shannon entropy is 1.584 bits per mora. Low-entropy endings — さ 0.90, み 0.94, す 1.08 — are the bottlenecks; high-entropy ones — よ 2.84, う 2.66, や 2.41 — leave the next player plenty of room.

The JMdict reference is the baseline distribution only — it is never treated as observed game data.

CHIBA TECH OKAYAMA PREF. UNIV. SHIZUOKA UNIV.