Discovering Phonological Representations: The Case of French Liaison

Annika Heuser

Overview

In this talk, we'll:

1. Specify a quantitative theory of how children learn abstract underlying representations (URs)

2. Validate it on the problem of French liaison

Underlying representations (URs) in phonology

- How should we evaluate different UR alternatives?
  - Whether they allow for fewer/simpler rules? Similarity to surface representations (SRs)?

- Useful insight can come from language acquisition

- E.g. Alternation Condition (Kiparsky, 1968)

Alternation Condition

- Ensures that features of SRs that are always the same are also the same in the UR and features that are always different are also different in the UR

- Motivated by greater learnability

- Difficult to to verify without quantitative theory for how easily a child could learn different URs

Proposal of UR discovery

- The Surface True Hypothesis (STH)
- Big idea: Children will not posit abstract URs unless **necessary**
- Starting rule: URs are identical to SRs
- Quantification of necessary: Too many alternations to be tolerable to the rule, as defined by the Tolerance Principle (TP; Yang, 2016)

Reformulation of an old idea

- Alternation Condition: restricts URs from being too different from SRs
- Faithfulness constraints of OT
- Invariant Transparency: learners will project SRs into URs unless there’s an alternation to account for (Ringe & Eska, 2013)

Quantification of abstract UR discovery

- When will children posit abstract URs that are different from SRs?

→ With enough evidence of an alternation

What counts as evidence?

→ Multiple phonological representations for the same concept

How do we define enough?

→ Tolerance Principle (TP; Yang, 2016)

What is French liaison?

a. un ami
ãe nami
a friend

b. les ami
le zami
the friends

c. petit ami
pəti tami
little friends

d. joli ami
ʒɔli ami
pretty friend

- Prepending a consonant to a vowel-initial word in specific syntactic and lexical contexts
- Obligatory (vs optional): ungrammatical not to produce liaison in these contexts (80.45% of realized liaisons are obligatory; Meinschaefer et al., 2015)

Why French liaison as a case study?

- Empirical evidence of different stages of liaison acquisition during which children have different URs

- Debate over whether linguistic representations of liaison are phonological or morphological (e.g. Tranel, 1981; Encrevé, 1988; Smolensky & Goldrick 2016 vs. Klausenburger 2001; Storme 2024)

    → Acquisition account can weigh in on this debate

Empirical stages of French liaison acquisition

	Stage 1	Stage 2	Stage 3
Age	Before ~2 years	~2 – ~4 years	After ~4 years
Description	• Heard [nami] so word must be /nami/ • Lexically/consonant-specific	• Heard [nami] but word could be /nami/ or /ami/ (or /tami/) • Know liaison forms are related	• Learning a phonological rule for liaison • Abstract liaison representation
Evidence	• 20-month-olds do not recognize vowel-initial nonce words in isolation after hearing them in liaison contexts	• 24-month-olds recognize the vowel-initial nonce words in isolation • 30-month-olds quickly recognize both frequent liaison forms + vowel-initial form for a word in non-liaison contexts	• Regularization errors on words like “nombril”: produce “zombril” in /z/-liaison context
Study	Babineau & Shi, 2014	Babineau & Shi, 2014; Babineau et al., 2021	Chevrot et al., 2009

Models of acquisition

- Constructionist (Chevrot et al., 2009; Nicoladis & Paradis, 2011)
  - Memorize chunks of speech and segment them based on transition probabilities
  - Transition probabilities favor consonant-initial segmentation for liaison-participating words because of French CV-dominance
  - Segmentations can be inserted into the X slot of “schemas”
  - Example schema: les + X

Construction model stages

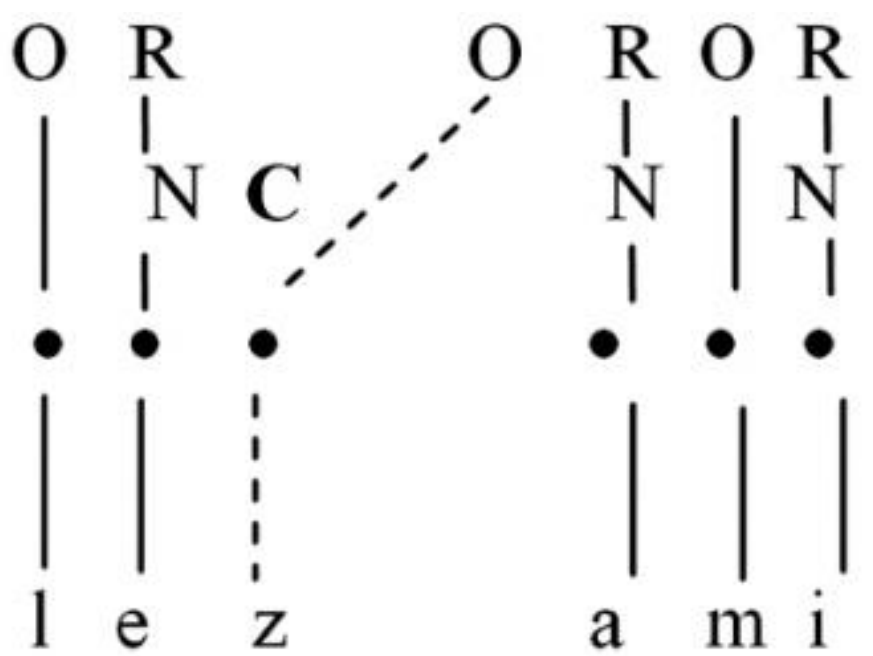
	1st stage	2nd stage
Age	2-4 years	4-5 years
Description	Only have general schema	Construct liaison-specific schema
Example schema	les + X	les + z/X/ z/X/ = /z/-initial phonological variant

- Does not predict regularization errors
  - Explanation is that regularization errors are due to analogy to liaison allomorphs
  - Need a theory of learning by analogy to verify this

- Does not explain why 1st and 2nd stage transition occurs at 4 years

Models of acquisition

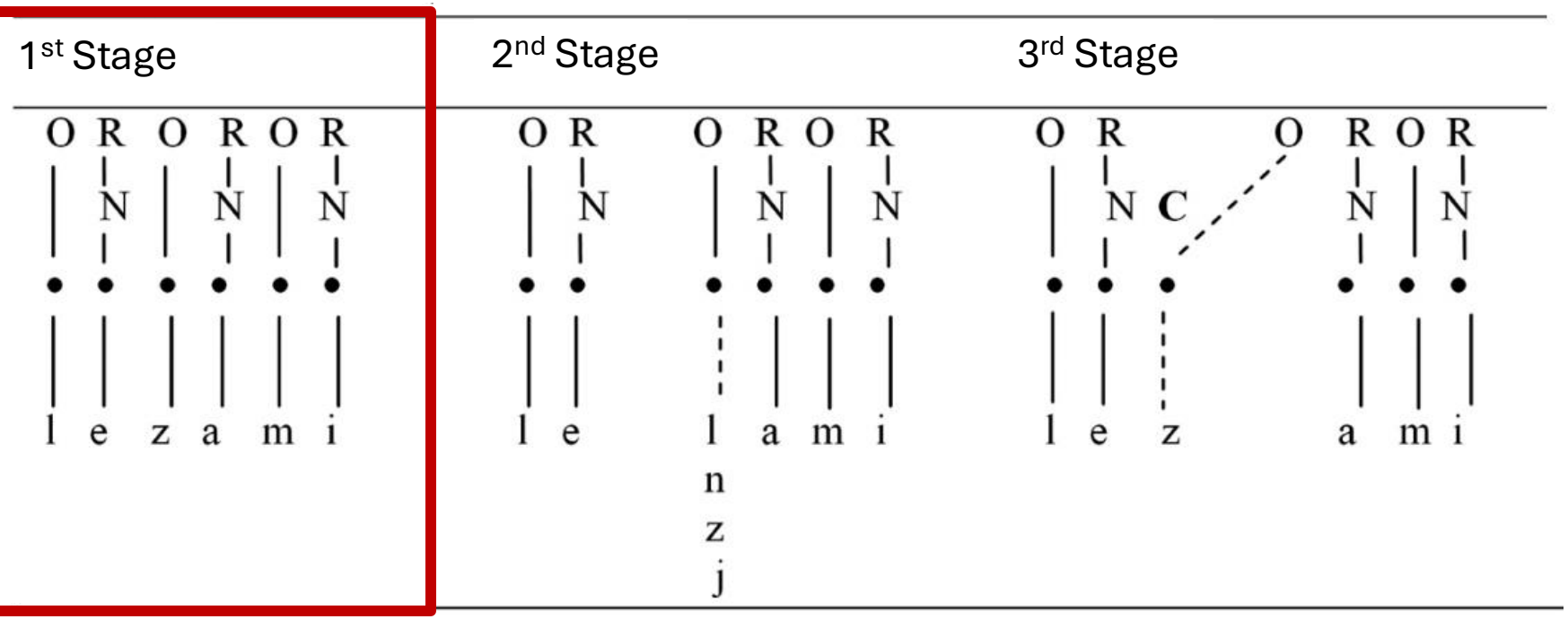
- Phonological (Wauquier-Gravelines & Braud, 2005; Wauquier, 2009)
  - Multilinear URs
  - Liaison consonant in coda of liaison-triggering word becomes anchored to onset position of vowel-initial word



Syllable tier

Segmental tier

Phonological model stages

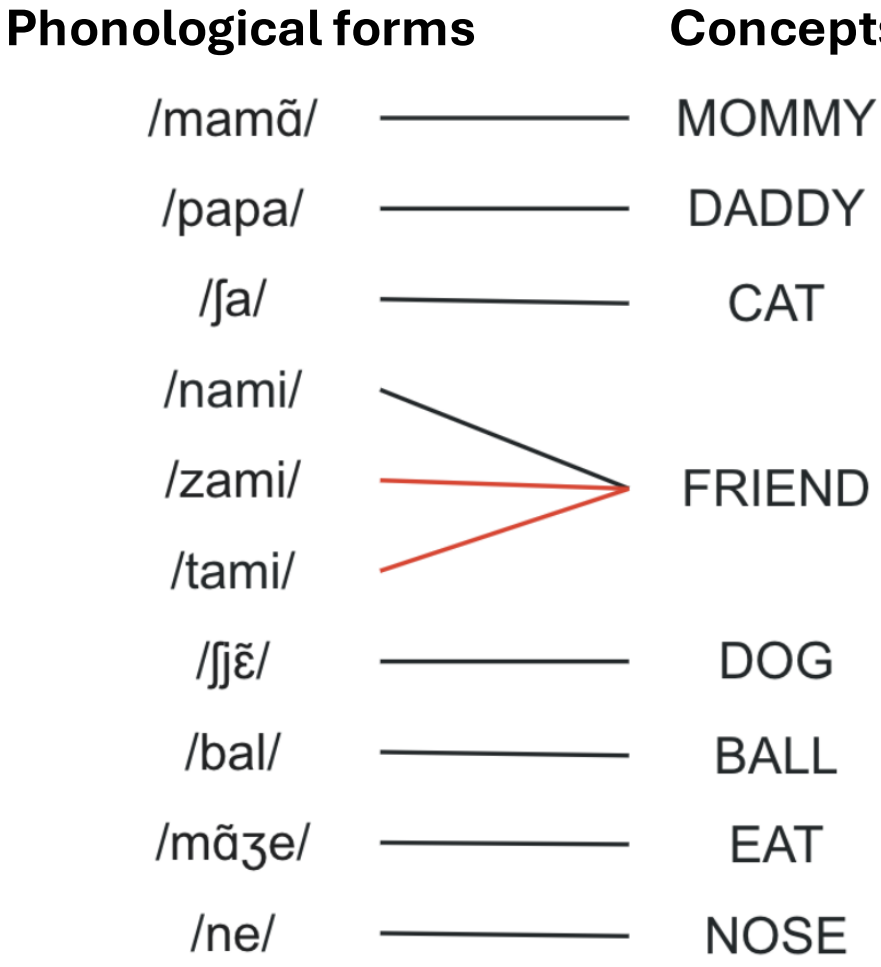


From Chevrot et al. 2013

Consistent with STH!

- Does not predict when children will transition between stages

Applying our proposal to liaison

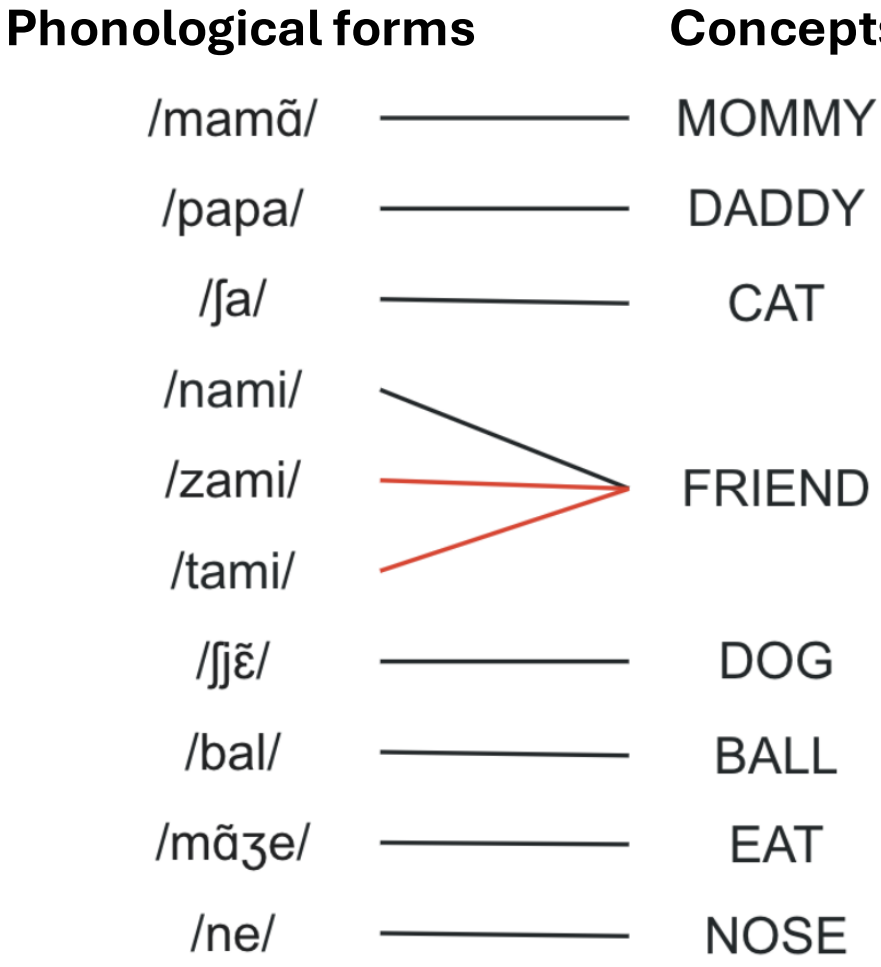


Model depends on:

- Ability to tell whether words are the same or different from words already in the lexicon (i.e. essentially lemmatize)
  - Empirical evidence of this: Infants as young as 6 months old can relate inflected nonce words to their stems

(Shi & Cyr, 2008; Marquis & Shi, 2012; Mintz, 2013; Kim & Sundara, 2021)

Applying our proposal to liaison



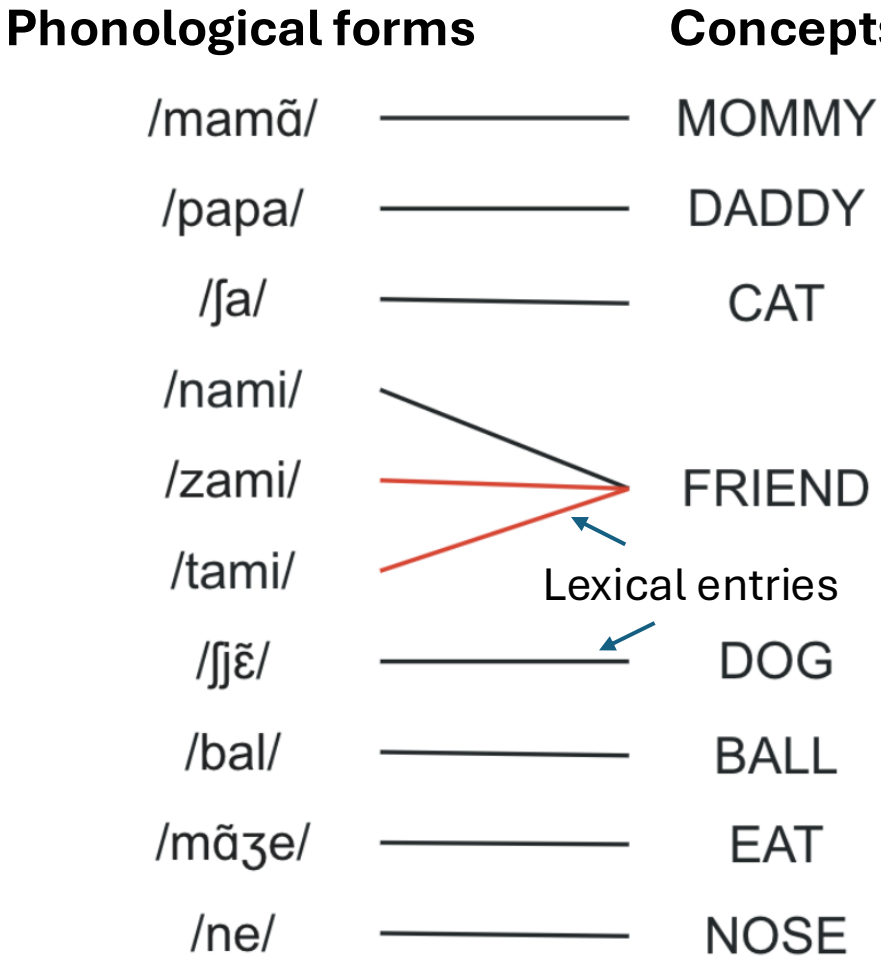
- STH encoded as 1-to-1 mapping constraint on phonological form to concept mappings
- Multiple liaison forms result in many-to-1 mappings
  - Empirical evidence that children expect 1-to-1 mappings: Infants have mutual exclusivity bias (against many-to-1 mappings)

(Markman et al., 2003; Halberda, 2003)

  - Phonological similarity makes storing liaison forms even more difficult

(Swingley & Aslin, 2007)

Applying our proposal to liaison



What counts as evidence against the STH?

- Exceptions (e): Additional phonological forms beyond the first for any lemma, =2 for this example

How much evidence is enough evidence to reject the STH?

- When the e > TP threshold (N/ln(N))
- N = number of lexical entries (phonological form to concept mappings) = 10 for this example lexicon
- 2 < 10/ln(10) ≈ 4 → don't reject STH

Simulation of French liaison acquisition

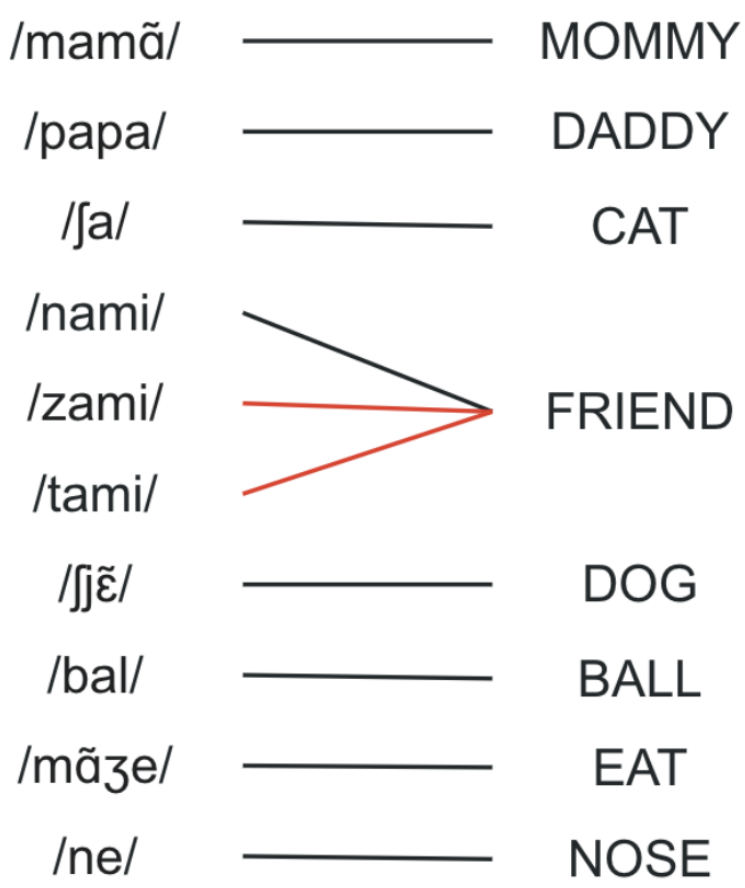
- To answer the question: How many words does a French-learning child need in their (receptive) vocabulary to abandon the STH?

- Sampled words from the top 1000 most frequent words across15 French CHILDES (MacWhinney, 2014) corpora
  - Word frequency is a good predictor of a word's age of acquisition

  (Braginsky et al., 2016; Swingley & Humphrey, 2018; Braginsky et al., 2019)

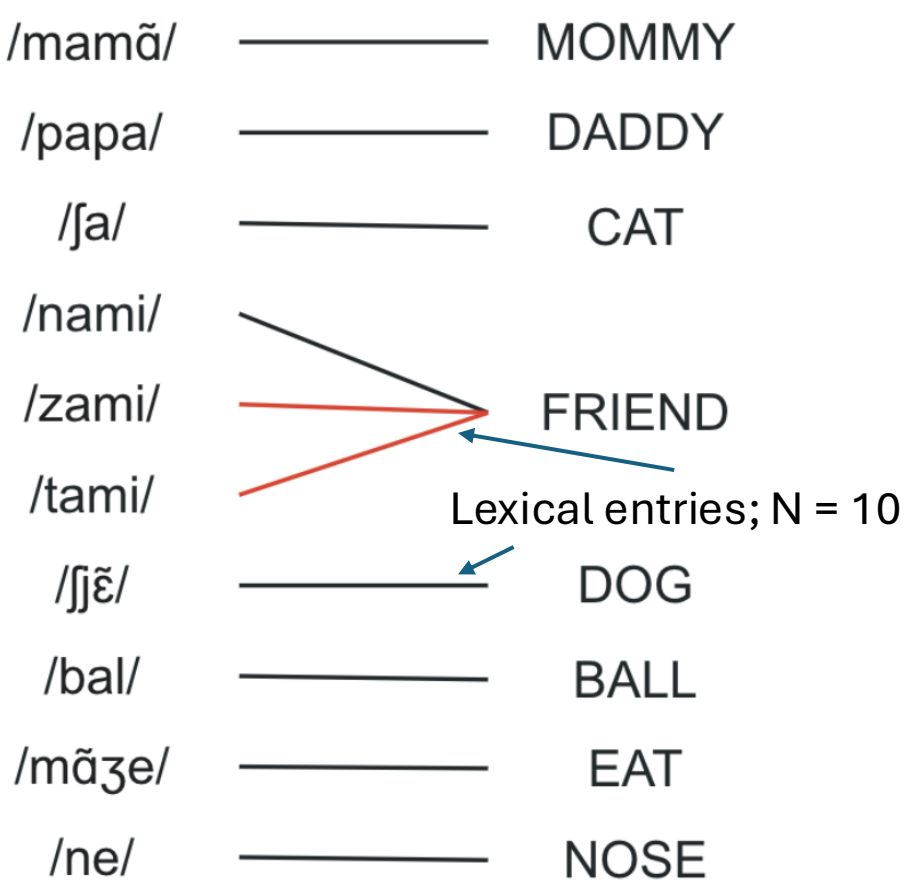
Simulation of French liaison acquisition

- Sampled W words until we had M lemmas (W ≥ M)
- Checked transcript bigrams of any of the W words for obligatory liaison conditions
  - E.g. sampling “un” and “ami” would result in adding /nami/
  - Also sampling “les” would give us /zami/, resulting in a STH exception



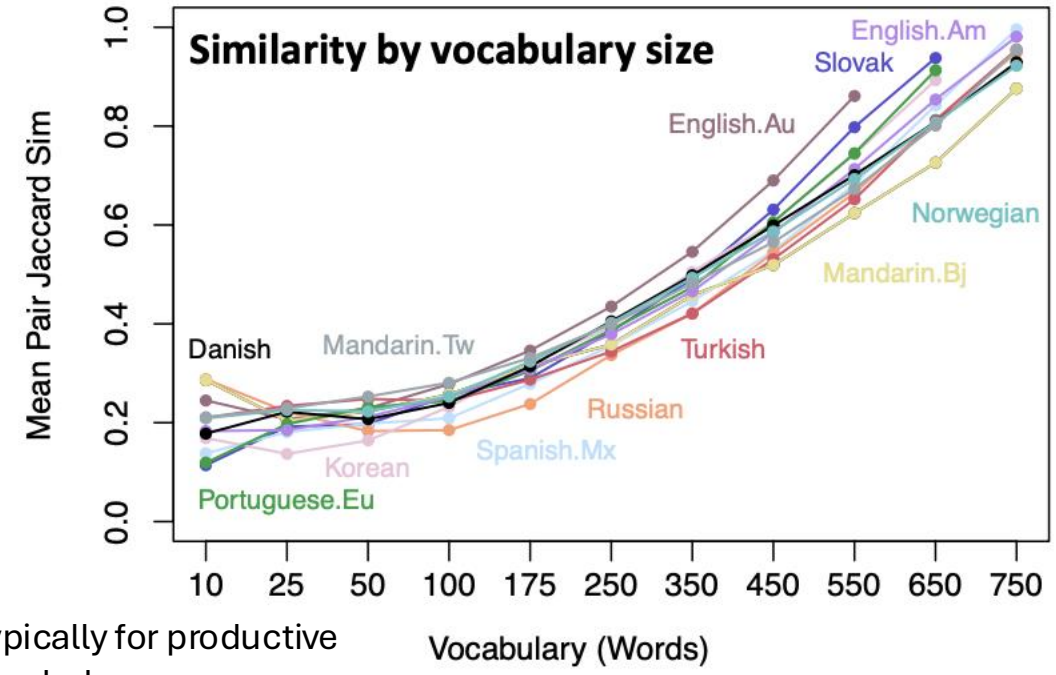
Simulation of French liaison acquisition

- Built 100 lexicons for each of M lemmas
- M-sized lexicons have different numbers of lexical entries (N) depending on the number of liaison forms
- Calculated proportion of lexicons for which we exceed the TP threshold based on their value of N



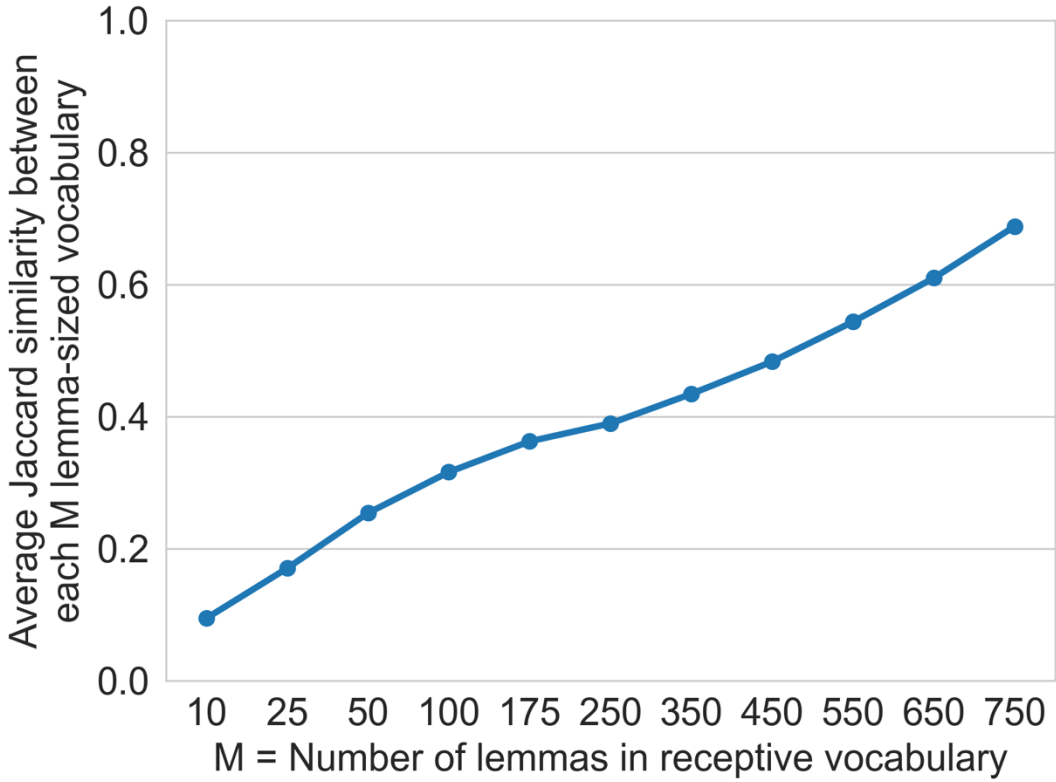
Validation of the sampling procedure

Reference Jaccard similarity on children's vocabularies estimated via CDI-checklists* (from Richter, 2021a)

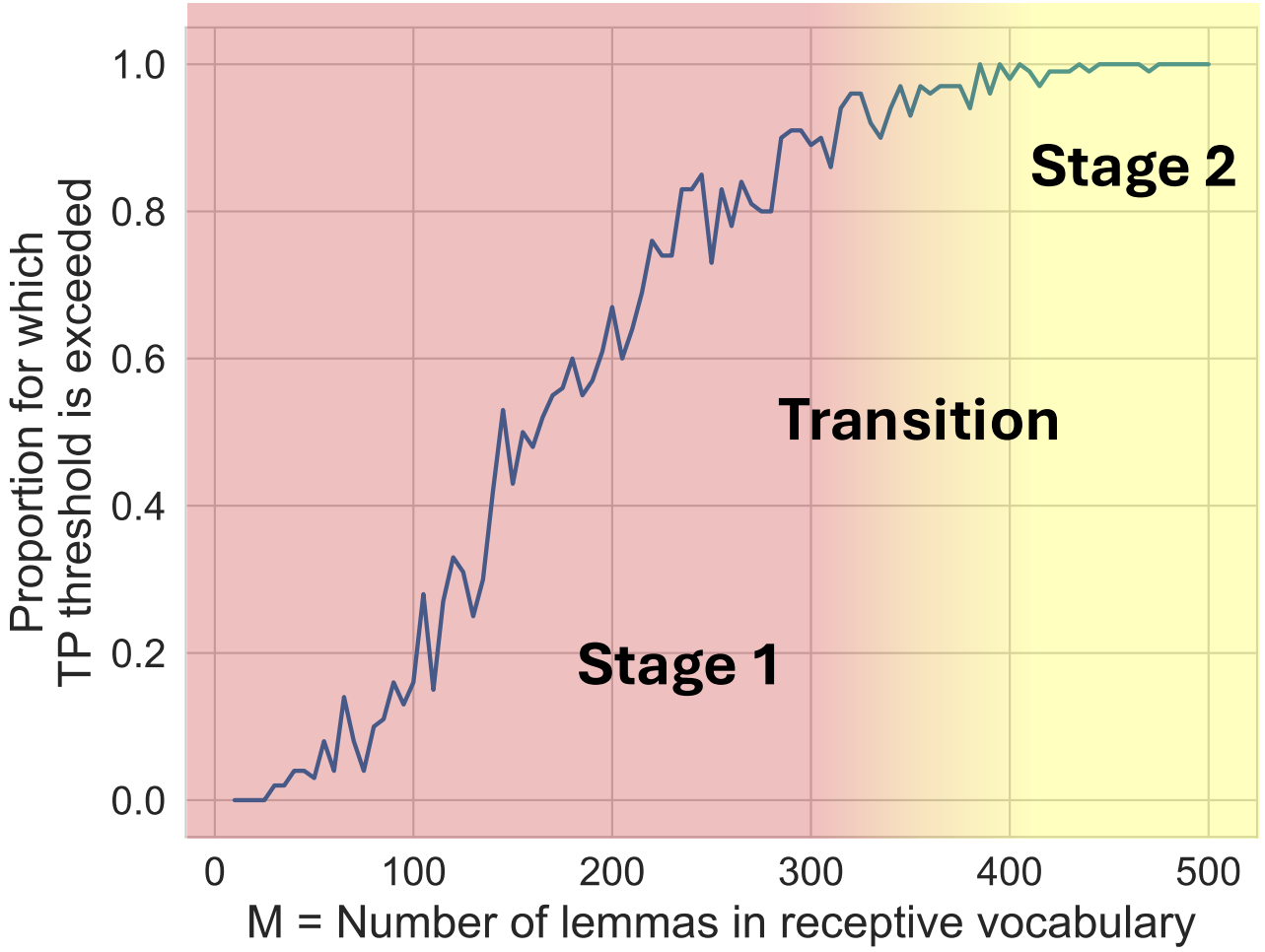


*Typically for productive vocabulary

Average Jaccard similarity of 100 vocabularies with M lemmas



Results



- Stage 1 to stage 2 transition occurs at about 24 months

- Expressive vocabulary of 250-275 at 24 months (Bouchard et al., 2009; Trudeau & Sutton, 2011)

→ Corresponds to predicted vocabulary size of 350-400 at time of transition

Discussion

- Surface True Hypothesis model predicts age of transition between stage 1 and stage 2 of French liaison acquisition

- First model to predict **when** and **why** children progress through these stages
  - Major advantage over constructionist and phonological models

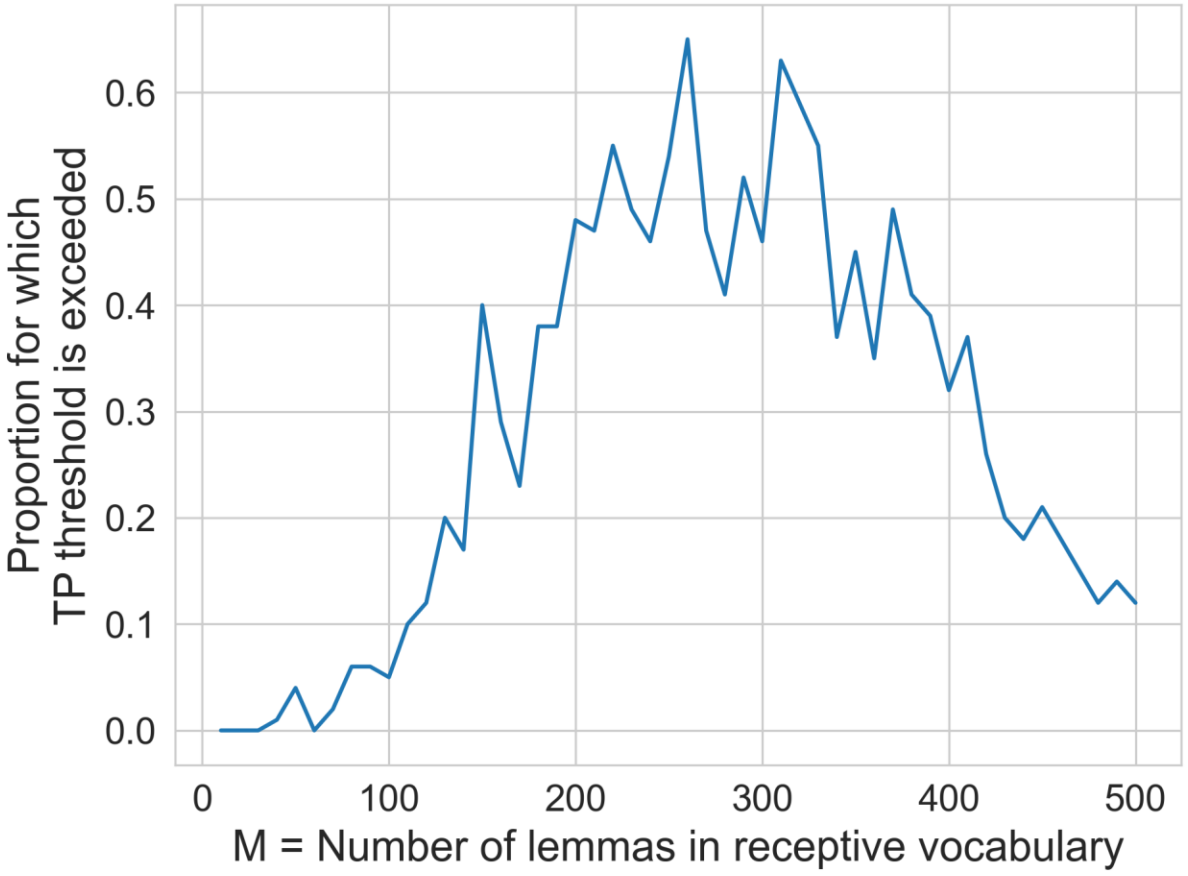
Discussion

- Able to discover that abstract URs are necessary via alternations in the input data
  - Assuming minimal morphological knowledge
- Acquisition of other morphophonological phenomena explained via a similar learning story
  - American English medial flap allophony (Richter, 2021b)
  - Dutch voicing alternation (Belth, 2024)
- Nothing about the model/method is specific to phonology, could also be used for morphological/syntactic acquisition

Implications and next steps

- Next children need to discover what the abstract URs are after deeming them necessary
- What this might look for liaison: Having 2 phonological forms for liaison triggering-words (e.g. /lez/ and /le/ for “les”)
- Modeled vocabulary acquisition with this representation (vs the multiple representations for the liaison-participating words)

Implications and next steps



- Does not become intolerable so this is a viable representation
- Suggests that children learn liaison as allomorphy AND as a phonological process

Acknowledgements

I would like to thank the Child Language Acquisition Lab at Penn, as well as Brittany Zykoski, Andrew Zhu, Professors Achim Stein, Katie Schuler, and Professor Charles Yang for all their help with this project.

Any errors are my own.

References

Babineau, M., Legrand, C., & Shi, R. (2021). Variable forms in french-learning toddlers' lexical representations. *Developmental Psychology*, 57(4), 457.

Babineau, M., & Shi, R. (2014). Distributional cues and the onset bias in early word segmentation. *Developmental psychology*, 50(12), 2666.

Belth, C. (2024). A Learning-Based Account of Non-Productivity in Dutch Voicing Alternations.

Bouchard, C., Trudeau, N., Sutton, A., Boudreault, M.-C., & Deneault, J. (2009). Gender differences in language development in French Canadian children between 8 and 30 months of age. *Applied psycholinguistics*, 30(4), 685–707.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. (2016). From uh-oh to tomorrow: Predicting age of acquisition for early words across languages. In *Cogsci* (Vol. 6).

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, 3, 52–67.

Chevrot, J.-P., Dugua, C., & Fayol, M. (2009). Liaison acquisition, word segmentation and construction in French: a usage-based account. *Journal of child language*, 36(3), 557–596.

Chevrot, J.-P., Dugua, C., Harnois-Delpiano, M., Siccardi, A., & Spinelli, E. (2013). Liaison acquisition: debates, critical issues, future research. *Language Sciences*, 39, 83–94.

Encrevé, P. (1988). La liaison avec et sans enchaînement: phonologie tridimensionnelle et usages du français.

References continued

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1), B23–B34

Kim, Y. J., & Sundara, M. (2021). 6–month–olds are sensitive to english morphology. *Developmental science*, 24(4), e13089.

Kiparsky, P. (1968). *How abstract is phonology?* Indiana University Linguistics Club.

Klausenberger, J. (2001). The Morphologization and Grammaticalization of French Liaison. In *Features and Interfaces in Romance: Essays in Honor of Heles Contreras* (pp. 175-188).

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs*. Psychology Press.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive psychology*, 47(3), 241–275.

Marquis, A., & Shi, R. (2012). Initial morphological learning in preverbal infants. *Cognition*, 122(1), 61–66.

Mintz, T. H. (2013). The segmentation of sub-lexical morphemes in English-learning 15-month-olds. *Frontiers in psychology*, 4, 24.

Nicoladis, E., & Paradis, J. (2011). Learning to liaise and elide comme il faut: Evidence from bilingual children. *Journal of Child Language*, 38(4), 701–730.

Richter, C. 2021a. *Consequences of lexical variability in toddlers' vocabularies*. Poster presented at the 45th Penn Linguistics Conference, Philadelphia.

References continued

Richter, C. (2021b). *Alternation-Sensitive Phoneme Learning: Implications for Children's Development and Language Change*. University of Pennsylvania.

Ringe, D., & Eska, J. F. (2013). *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge University Press.

Shi, R., & Cyr, M. (2008). Processing of morphological variations in toddlers. In *BUCLD* (Vol. 34, pp. 363–374).

Smolensky, P., & Goldrick, M. (2016). Gradient symbolic representations in grammar: The case of french liaison. *Rutgers Optimality Archive*, 1552, 1–37.

Storme, B. (2024). French liaison is allomorphy, not allophony: evidence from lexical statistics. *Morphology*, 1–41.

Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific English words. *Child development*, 89(4), 1247–1267.

Trudeau, N., & Sutton, A. (2011). Expressive vocabulary and early grammar of 16-to 30-month-old children acquiring Quebec French. First language, 31(4), 480–507.

Wauquier, S. (2009). Acquisition de la liaison en L1 et L2: stratégies phonologiques ou lexicales? *Acquisition et interaction en langue étrangère* (Aile... Lia 2), 93–130.

Wauquier-Gravelines, S., & Braud, V. (2005). Proto-déterminant et acquisition de la liaison obligatoire en français. *Langages*, 158(2), 53–65.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.