BUCLD 49 Proceedings
To be published in 2025 by Cascadilla Press
Rights forms signed by all authors

Discovering Phonological Representations: The Case of French Liaison

Annika Heuser

1. Introduction

How similar are the abstract underlying representations (URs) that children learn to those posited by phonologists? There are many URs that could possibly account for the phonological alternations in a language. Why certain URs were preferred over others has changed over the development of the field of phonology. In the 1950s and 1960s, many phonologists valued the rule simplicity that a UR allowed for most. This resulted in highly convoluted URs sometimes being preferred over simpler alternatives that required more complicated rules to derive the surface representation (SR).

In response, Kiparsky (1968) proposed the Alternation Condition, which ensures a certain degree of alignment between URs and SRs. Part of the motivation behind the Alternation Condition is greater learnability (Kiparsky, 2012, p. 59). However, Kiparsky does not quantify how much more learnable URs that obey the Alternation Condition are than those that do not. Without a quantitative theory for how children can learn abstract URs from SRs, we can only evaluate alternative URs subjectively.

The Optimality Theory (OT) framework (Prince, 2004) provides a means of evaluating potential SRs, in part with respect to the input (i.e. URs). In OT, languages having different SRs is a function of having different constraint rankings, rather than different URs. This idea is called richness of the base. However, even assuming children are endowed with the set of every possible UR, they would still need to learn the specific URs for the morphemes of their language. In the OT framework, this learning problem is solved by optimizing the set of possible URs based on the current constraint ranking (Tesar and Smolensky, 1998). Nonetheless, without a definition of the universal set of URs, we cannot verify this via the SRs that children hear.

*Annika Heuser, University of Pennsylvania, aheuser@sas.upenn.edu. I thank my advisor, Professor Charles Yang, for providing invaluable support, advice, and feedback on this work, Brittany Zykoski for answering numerous French liaison-related questions, often even eliciting judgments from native French speakers to do so, Professor Achim Stein for providing me with the POS-tagged and lemmatized data, Andrew Zhu for conducting a code review and for latex assistance, Professor Katie Schuler for all her helpful input on this project, and the Child Language Acquisition Lab at Penn for their feedback on how to present this work.

Can abstract URs emerge from the SRs alone, without needing to presume a substantial amount of innate knowledge? In this work, we argue that children learn to construct abstract URs due to evidence from their input data. We propose a concrete psychological mechanism to explain how, and verify it through a case study of French liaison.

French liaison is the process by which a vowel-initial word is pronounced with a consonant onset in specific syntactic and phonological/lexical contexts (Cŏté, 2011). We use the word "ami" (meaning "friend") to illustrate this process:

(1) a. un ami
ɜ̃ nami
a friend

b. les ami
le zami
the friends

c. petit ami
pəti tami
little friends

d. joli ami
ʒɔli ami
pretty friend

In examples 1a and 1b, liaison consonants /n/ and /z/ are prepended to /ami/ after the indefinite singular and definite plural determiners, respectively. In 1c, an adjective adds the liaison consonant /t/ to /ami/. Only in example 1d, where a different adjective precedes it, is it pronounced as its underlying form, /ami/.

Not all theories agree that the UR for "ami" should be /ami/. Recently, Smolensky and Goldrick (2016) proposed that "ami" is underlyingly /$\mathscr{L}$ami/ where $\mathscr{L}$ is a weak, partially-active consonant. At the same time, liaison-triggering words, like "un," "les," and "petit," also have weak, partially active $\mathscr{L}$s word-finally in their URs. Liaison then results from the simultaneous presence of both $\mathscr{L}$s. Another phonological account is that the URs of liaison-triggering words include a floating final consonant that anchors to the following vowel-initial word's onset position (Encrevé, 1988). The alternative morphological account is that liaison is the result of liaison-triggering words having two allmorphs. One allomorph does not have a final consonant while the other does. The two allomophs of "les" are then /le/ and /lez/. The /z/ is realized as the onset of the following vowel-initial word because of enchaînement. A recent argument in favor of this account is made by Storme (2024). The best linguistic representation for liaison is still an open question.

It is important to note that there are two types of liaison: obligatory and optional. It is ungrammatical not to pronounce a liaison consonant in an obligatory liaison context. In optional contexts, the pronunciation of a liaison consonant often signals a more formal register (Durand and Lyche, 2008). While the number of obligatory and optional liaison contexts are comparable (e.g. 642 vs 577 over 10,000 words), realization of optional liaison is significantly less frequent (642 vs 156 over 10,000 words, Meinschaefer et al., 2015). We focus on obligatory liaison

in this paper, as acquisition-related research on liaison tends to (e.g. Babineau and Shi, 2014; Babineau et al., 2021; Chevrot et al., 2013). Unless otherwise stated, "liaison" throughout this paper is referring to obligatory liaison.

In modeling the acquisition of French liaison, we posit a quantitative theory of how much evidence is necessary to postulate abstract URs. In section 2, we present the empirical evidence of the liaison acquisition timeline and some models of acquisition. We propose our model, described in section 3, as an alternative that is able to account for the empirical evidence. We explain how we validate our model in section 4 and provide the validation results in section 5. We thereby provide a falsifiable learning account for URs of the generative tradition (assuming they are in line with the Alternation Condition), that depends on minimal innate knowledge.

2. Related Work

2.1. Stages of Liaison Acquisition

Empirical results indicate three stages of liaison acquisition. In stage 1, children always parse liaison-participating words as being consonant-initial in liaison contexts. Babineau and Shi (2014) provide the evidence for this stage: 20-month-old French-learning infants who heard a vowel-initial nonce word in 4 different liaison contexts did not recognize the vowel-initial form in isolation. This means that if a 20-month infant heard /n/onche, /z/onche, /t/onche and /r/onche in liaison contexts, the infant would look longer after hearing the vowel-initial "onche." The phonetic cues of liaison onsets support this segmentation. Native French-speaking adults often cannot consistently tell liaison onsets apart from those of consonant-initial words (Babineau et al., 2017). Therefore, phonetic cues would be difficult to leverage for segmenting liaison-participating words.

Nonetheless, children transition to stage 2, in which they consider both a consonant- and vowel-initial parse for a liaison-participating word. Babineau and Shi (2014) found that while 20-month-old infants fail to recognize a nonce vowel-initial word after only hearing it in liaison contexts, 24-month-olds succeed. Further evidence that children begin considering multiple parses in this stage comes from Babineau et al. 2021.

In this eye-tracking study, Babineau et al. (2021) tested 30-month-old French-learning infants on familiar liaison-participating words preceded by "joli." "Joli" does not trigger liaison, but the plural-inflected form, "jolis," does. Crucially, "joli" and "jolis" have the same pronunciation. The infants were presented with images of singular objects, making a /z/-liaison form pragmatically incorrect. The infants also heard lexically incorrect /t/- and /g/-forms of familiar words after "joli". Note that /t/ is a common liaison consonant but /g/ is not. The 30-month-olds recognized the vowel-initial form, the /z/-form, and the /t/-form in the "joli" context, even though the vowel-initial form is the only one that is pragmatically and lexically correct. These results indicate that in stage 2, French-learning infants' lexical representation of liaison-participating words in-

cludes multiple phonemic variants.

We must emphasize that some phonemic variants are probably only in the child's lexicon due to abstraction. For example, in Babineau and Shi 2014, 24-month-olds had never heard the vowel-initial variant of the word, but inferred it. Babineau et al. 2023 provides further evidence of this: 24-month-olds segmented the second word of "un nonche" as "onche," assuming the /n/-onset was from liaison, despite no concrete evidence that the word was vowel-initial. Lastly, the 30-month-olds in Babineau et al. 2021 may not have heard the vowel-initial forms of all the familiar words being tested, but seemed to reliably have them stored in their lexicons nonetheless.

In stage 3, children form a phonological rule for liaison. Evidence for this comes from the liaison regularization errors reported by Chevrot et al. (2009) and Tessier et al. (2022). Chevrot et al. (2009) found that children will replace the initial consonant of a consonant-initial word with a liaison consonant. For example, "nombril," pronounced /nɔ̃bʁil/ should start with /n/ in every context it is said. However, when children make a regularization error, they replace the /n/ with /z/, for example, pronouncing the word as /zɔ̃bʁil/ in appropriate liaison contexts (for vowel-initial words). Tessier et al. (2022) found that adults will do liaison with h-aspiré words, even though they are not supposed to participate in liaison. However, it is important to note that neither of these regularization errors have, to our knowledge, been reported in child speech in naturalistic settings. Therefore, further research is needed to better understand stage 3 and its implications for the adult representation of liaison.

2.2. Models of Liaison Acquisition

The constructionist model (Chevrot et al., 2009; Nicoladis and Paradis, 2011) is the formulation of how children would learn liaison assuming a usage-based framework. This model posits that children memorize chunks of speech, some of which will contain liaison. They segment these chunks based on transitional probabilities, which favor a consonant-initial segmentation due to French being CV-syllable dominant. These segmentations can be placed into the slots of what Chevrot et al. (2013) refer to as schemas. An example of a schema is les + X, where X is a segmented noun and les is pronounced /le/, as in example 1b. Children learn liaison by first forming general schema, like les + X, and then liaison-specific schema, like les + zX where the zX slot must be filled by a z-initial phonological variant.

The phonological model (Wauquier-Gravelines and Braud, 2005; Wauquier, 2009) consists of the multilinear representations proposed by Encrevé (1988). In an adult's representation, the liaison consonant is floating in the final coda position of the liaison-triggering word. During liaison, it anchors to the vowel-initial word's onset. In children's earliest liaision URs, the liaison consonant belongs to the second word's onset. This means that children's initial URs include /nami/ and /zami/. Across three stages, the child learns that the liaison consonant is actually

associated with the first word (e.g. “un” and “les”) to acquire the adult representation. However, how and why the child transitions through the stages is not well specified. The same is true for the constructionist model. In the following section, we delineate a model which explains the transition between stage 1 and stage 2 of liaison acquisition in reference to children’s input data.

3. Proposal

3.1. The Surface True Hypothesis (STH)

We propose that children do not posit abstract URs that differ from SRs unless strictly necessary. Therefore, they must start with what we call the Surface-True Hypothesis (STH). Under the STH, URs are identical to SRs. This is not a new idea: Hale (1973) stated that children tend to minimally, “postulate underlying phonological representations of morphemes which violate the universal surface canonical patterns of the language” (p. 420). Kiparsky’s Alternation Condition is strongly related, presupposing that children will not presuppose underlying distinctions that do not map to surface distinctions (Kiparsky, 2012). The faithfulness constraints of OT are also based on this the principle that underlying and surface representations should be as similar as possible. More recently, Ringe and Eska (2013) proposed Invariant Transparency, which posits that learners will project SRs into URs if there is no alternation to account for. Richter (2021a) built a quantitative model of allophone acquisition based on this idea.

The major contribution of Richter (2021a)’s and our work is formulating this old idea from an acquisition perspective. The STH is the resulting formulation. Under the STH, children expect a direct mapping between SRs and URs and need to memorize alternations as exceptions. This significantly reduces the hypothesis space and simplifies the learning problem: Children only need to learn abstract URs that allow them account for surface alternations because the STH is otherwise accurate.

The STH is consistent with stage 1 of liaison acquisition, as well as the initial stage of the phonological model. 20-month-old children do not yet have enough evidence of the liaison alternation to stray from the STH. More concretely, under the STH, “un /n/onche,” “les /z/onche,” and “petit /t/onche,” must be parsed as /n/onche, /z/onche, and /t/onche, respectively. Therefore, 20-month old children do not recognize “onche” in isolation because the STH never results in that parse for any of the words. The 24-month-old children in stage 2, on the other hand, are sufficiently aware of the distributional regularity of liaison to consider a vowel-initial parse or otherwise know that a vowel-initial variant belongs to the alternation. This begs the questions: When then does a child have enough evidence to stray from the STH? And what constitutes evidence against the STH?

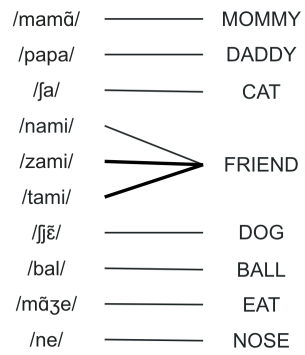

Figure 1: Model lexical entry representations. Links resulting in many-to-one mappings (bolded) are considered exceptions.

3.2. What constitutes evidence against the STH?

To answer this question, we formalize the STH as a one-to-one mapping constraint between phonological forms and concepts. We refer to these mappings between phonological forms and concepts as lexical entries. A lexical entry encodes the semantic and morphological information available to a child at a given point in their linguistic development. Our model's lexical entries are represented by lemmas. We argue that 20-month-old infants have the same morphological and semantic knowledge that is encoded in lemmas for all the words that they know. The acquisition of this knowledge only presupposes that 20-month-olds know that different phonological words of the same lemma can be grouped together.

3.2.1. Model Assumptions

The only assumption underlying the model is that the learner can classify new words as the same as or different from the words that they already know. To illustrate what this means, we refer to figure 1. The learner needs to know that "dog" and "ball" are different, for example. The learner also needs to realize that /mɑ̃ʒɔ̃/ (meaning "we eat") is the same as /mɑ̃ʒe/ so no new mapping is required.

In short, we assume the learner can essentially lemmatize. This is is justified by evidence that at 20 months, French-learning infants demonstrate the ability to relate rule-following morphological variants of nonce verbs, despite the stem having undergone a phonemic change (Shi and Cyr, 2008). Even earlier, at 11 months, French-learning infants treat nonce bare verb roots and their inflected forms as related to one another (Marquis and Shi, 2012). More generally, English-learning infants as young as 6 months old have been shown to be able to relate inflected nonce words with their stems (Kim and Sundara, 2021). We are therefore confident that at 20 months, French-learning infants can lemmatize to the limited extent that our model assumes.

Importantly, we do not assume that a child can supplement incomplete knowl-

edge of a lexical entry with morphological knowledge. For example, if the child has learned "manges" ("I/he/she/it eat(s)", pronounced /mãʒe/) as in figure 1, we do not assume that they will know "mangeons" ("we eat", pronounced /mãʒõ/).

3.2.2. STH Formulation: 1-to-1 Mapping Constraint

We now turn to the justification for and implications of formulating the STH as a 1-to-1 mapping constraint between phonological forms and concepts. We argue that children expect 1-to-1 mappings because of evidence that they resist 1-to-many mappings and many-to-1 mappings. 1-to-many mappings are simply homophones. Peters and Zaidel (1980) found that 20-month-old French-learning infants resist homophones that are too semantically similar. The mutual exclusivity bias in word learning is a bias against many-to-1 mappings (Markman and Wachtel, 1988). Infants as young as 15 months old are shown to have this bias (Markman et al., 2003; Halberda, 2003).

French liaison results in many-to-1 mappings, as shown in figure 1. /nami/, /tami/, and /zami/ all map to same concept. The phonological similarity of these alternating forms is also problematic because 18-month-old infants have been shown to resist learning words that are phonologically similar to words that they already know (Swingley and Aslin, 2007). Multiple French liaison forms for the same concept should therefore be challenging to learn with the 1-to-1 mapping constraint imposed by the STH on the lexicon. Thus, any liaison forms beyond the first to map to a concept constitutes as evidence against the STH.

In thoroughly outlining the formulation of the STH, it is important to clarify that inflected forms of a lexical entry do not count as additional phonological forms like liaison forms do. This is because inflected forms consistently fit within the paradigm corresponding to the syntactic category of the lexical entry. Liaison forms, on the other hand, do not. For example, /zami/ is often the result of "ami" being plural and definite because it is preceded by the determiner "les." However, consonant-initial nouns typically do not have different plural+definite phonological forms. Additionally, /zami/ could instead be the result of "ami" being preceded by the adjective "gros" (meaning "big"), for example. Liaison also occurs with words of many different syntactic categories. Because of the resulting semantic and syntactic inconsistency, we doubt that children could consider liaison forms as inflected forms that fit into the paradigm of a syntactic category. We therefore argue that inflected forms must be treated differently than liaison forms in the context of our model.

3.3. When does a child have enough evidence against the STH to abandon it?

We defined evidence against the STH as many-to-1 phonological form to concept mappings. This allows us to precisely quantify such evidence as the number of additional phonological word mappings beyond the first to the same concepts. The toy lexicon in figure 1 helps us illustrate what this means. There is just one

phonological word for each of the CAT and DOG concepts. They therefore provide no evidence against the STH because they are 1-to-1 mappings. However, there is a many-to-1 mapping for the concept of FRIEND. More specifically, there is a 3-to-1 mapping so it consists of 2 counts against the STH. This corresponds to the bolded links in figure 1. Because any one of the links in the many-to-1 mapping could be part of a 1-to-1 mapping if the others were not also present, only x - 1 of the links in a many-to-1 mapping count as evidence against the STH.

Now that we have a precise quantification of evidence against the STH, we just need to define "enough" numerically in order to answer the question of when a child has enough evidence to abandon the STH. We define "enough" via the Tolerance Principle (TP, Yang, 2016), such that we abandon the STH when we have more than $\frac{N}{ln(N)}$ exceptions. N is the total number of lexical entries, or phonological word to concept mappings. The number of exceptions is simply the quantification of evidence against the STH that we described in the previous paragraph.

Putting all this together, in figure 1, we have 10 links total so N=10 and the TP threshold is $\frac{N}{ln(N)} = \frac{10}{ln(10)} = 4.34$. Therefore, if we have more than 4 exceptions, we exceed the threshold. The exceptions in figure 1 are bolded. Because we only have 2, we do not have enough evidence to abandon the STH in this case. Crucially, the larger N is, the the smaller the proportion of N that is tolerable as exceptions. In other words, a smaller percentage of the data needs to consist of exceptions in order to exceed the threshold when N is large.

4. Simulation of the Acquisition of Liaison

We were precise in our formulation of the STH and at what point the learner must abandon it in the above section. This allows us test how many words a French-learning child needs in their vocabulary in order to accumulate enough exceptions against the STH. We validate our model if it predicts this number of words to be approximately the vocabulary size that we would expect a child to have at the age of transition between stage 1 and stage 2 of liaison acquisition.

There are two types of vocabularies that we could be referring to: expressive and receptive vocabulary (Fenson et al., 1994). Expressive vocabulary refers to the set of words that a child can produce while receptive vocabulary refers to the set of words that they can comprehend. Expressive vocabulary is much easier to quantify with access to corpus data or production experiments. It is also easier to assess whether the meaning a child has for a word in their expressive vocabulary is accurate. If the child uses a word in the right contexts, then we can be fairly confident that the meaning associated with that word is accurate.

However, children's expressive vocabularies lag behind their receptive vocabularies, especially early in their lexical development (Benedict, 1979). Bergelson and Swingley (2012) demonstrated that preverbal infants can recognize common food- and body-related words at 6 months. When presented with two object images, the infants in this study looked more at the named object than the distractor.

Success in this task is consistent with what it means to know a word in the context of our model. Therefore our model's lexicon represents children's receptive vocabularies rather than their expressive vocabularies. This is appropriate for a model of phonological representation learning because of evidence that children's lexicons become more phonologically specified based on their receptive vocabularies. While 6-month-old infants' comprehension of common words is not impeded by mispronunciations, 11- to 14-month-olds' comprehension is (Bergelson and Swingley, 2018). This suggests that the words in their receptive vocabularies become more phonologically specified before children start producing these words.

We simulate the acquisition of a receptive vocabulary by sampling words from French child-directed speech data. This data consisted of 15 French CHILDES (MacWhinney, 2014) corpora of data from 336 children. The transcribed data was part of speech (POS)-tagged and lemmatized via TreeTagger using the French parameter file (Schmid et al., 2007). We compiled all bigrams where the two words of the same sentence were not separated by punctuation. We cleaned the resulting data by excluding any bigrams where either word had an "unknown" lemma or a POS tag that was different from that word's majority tag, for example. The word frequencies and whether a liaison form would be realized for any given bigram was calculated over this resulting data.

We sampled words with replacement based on their frequency. Frequency is a relatively strong predictor of the age at which children will learn a word (Braginsky et al., 2016; Swingley and Humphrey, 2018; Braginsky et al., 2019). Only the top 1000 words were eligible to be sampled. This allowed us to emulate the similarity in children's vocabularies as they grow in size. Early vocabularies are highly variable (Fenson et al., 1994), but they begin overlapping more with one another as they get larger (Richter, 2021b).

We continued sampling words until we had a lexicon with M lemmas. Especially as M increased, we typically sampled more than M words because some of the sampled words were inflected forms of lemmas that we already had. Therefore a sampled lexicon of size M lemmas has W words, where $W \geq M$. We then checked whether any liaison phonological forms resulted from bigrams of any of the W words in our lexicon. For example, if both "un" and "ami" were in the lexicon's W words, then the phonological form /nami/ would be mapped to the lemma "ami." If "petit" were also in this lexicon, then /tami/ would also be mapped to the lemma "ami," resulting in an STH exception. Note that adding a mapping for /tami/ does not increase our M, because "ami" is already a represented lemma. It increments our N though. Therefore our N, over which we calculate our TP threshold, is equal to M + the number of exceptions.

N is not easy to compare with receptive vocabulary estimates in the literature, but M is. We therefore calculate and report the proportion of 100 lexicons of size M for which the number of exceptions exceeded the TP threshold. This proportion also helps us determine the likelihood of a child having abandoned the STH with

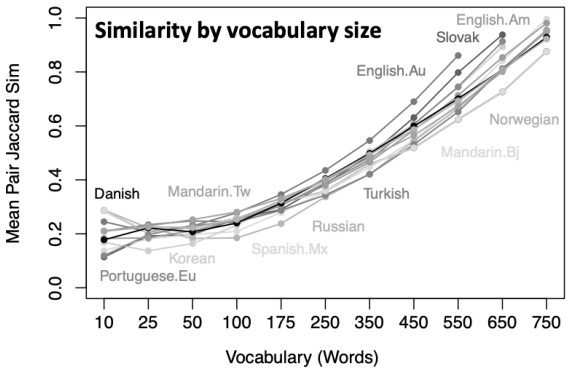

Figure 2a: Reference vocabulary Jaccard similarities (Richter, 2021b).

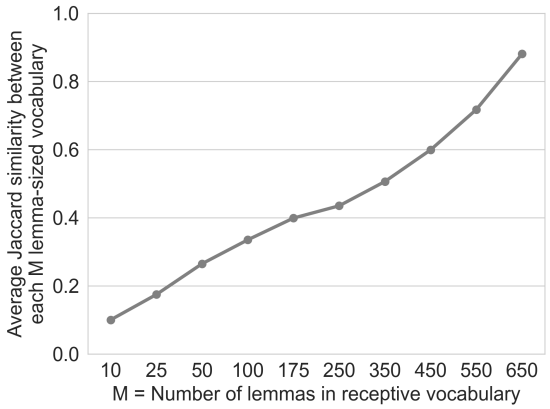

Figure 2b: Average Jaccard similarities of our simulated French vocabularies

a receptive vocabulary of that size. We can thereby predict the age at which we would expect most French-learning children to have transitioned between stage 1 and stage 2 of liaison acquisition.

5. Results

We built 100 vocabularies each of size 5 to 500 lemmas. We first verified that the lexical variability of these 100 vocabularies is comparable to that across children's vocabularies as estimated via the MacArthur-Bates Communicative Development Inventory (CDI) (Fenson et al., 2007). Figure 2a shows the average Jaccard similarities calculated over thousands of children's vocabularies of different sizes for each of 12 language varieties (Richter, 2021b). Figure 2b shows the average Jaccard similarities of our simulated vocabularies. These averages are quite comparable. Notably, the average Jaccard similarity of our simulated vocab-

ularies increases more quickly between vocabulary sizes of 25 and 250. Note that the top 1000 most frequent words only represented 701 lemmas, so we could not include a point of comparison for M=750 in 2b.

Having verified that the vocabularies built by our simulation are approximately as varied as those of actual children, we now turn to the results of the simulation. Figure 3 demonstrates that once a French-learning child's receptive vocabulary has between 350 and 400 lemmas, they have no choice but to abandon the STH.

At 20 months old, French-learning children have expressive vocabularies of about 125-150 words as estimated by CDI surveys adapted for French. At 24 months old, their expressive vocabularies have about 250-275 words (Bouchard et al., 2009; Trudeau and Sutton, 2011). We still expect children's receptive vocabularies to be larger than their expressive vocabularies at this age. Therefore, an expressive vocabulary of 250-275 words corresponding to a receptive vocabulary of 350-400 lemmas seems reasonable. We accordingly conclude that our model successfully predicted that French-learning children will enter stage 2 of liaison acquisition at 24 months.

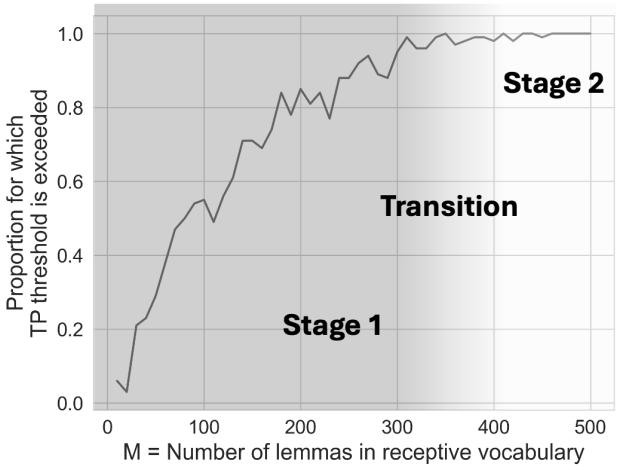

Figure 3: The number of exceptions exceed the TP thresholds in almost all vocabularies of more than 350 lemmas, so the transition likely occurs around then.

6. Discussion

Our model succeeds in predicting *when* children will enter the next stage of liaison acquisition by explaining *why* they must leave the first stage. Our proposal for why is that children start with surface-true URs that they abandon with enough evidence against them. We defined precisely what constitutes enough evidence

via a lexical acquisition model and the Tolerance Principle. Then we simulated the acquisition of receptive vocabularies of different sizes to predict the age at which children will transition between stage 1 and stage 2 of liaison acquisition. Our model prediction is corroborated by the approximate vocabulary size of 24-month-olds, as estimated by French-adapted CDI surveys.

The major advantage of our model over the constructionist and phonological models of liaison acquisition is that it can explain how children's input causes their linguistic representations to change. This means that our model can be validated via a lexical acquisition simulation, unlike the phonological and constructionist models. We also ensured that the STH and its consequent abandonment is consistent with the experimental results that we outlined in section 2.1.

We demonstrate with this model that abstract URs can be discovered from alternations in the input data, with minimal assumptions of morphological knowledge. This means that URs do not need to depend on innate knowledge of Universal Grammar, contra the OT learning account and an argument that some linguists make against URs altogether (Hyman, 2018). Rather, the STH posits that children's early URs are all represented in the input data. It is only when these URs fail, that a child has no choice but to embrace abstract URs.

To further explore how children might embrace abstract URs, future work includes specifying the URs French-learning children embrace in order to be able to learn a phonological rule during stage 3. Validation of these URs would come from success in modeling the timeline of learning this rule. The specification of the stage 2 URs would also allow us to infer a path to these URs from the STH URs. We believe that all children, regardless of what language they are learning, start with the STH. Therefore, an explanation for how French-learning children arrive at their stage 2 URs from the STH is likely to lead to a theory for how children cross-linguistically learn abstract URs not represented in their input data.

References

Babineau, Mireille, Emond, Emeryse, and Shi, Rushen (2023). When language-general and language-specific processes are in conflict: The case of sub-syllabic word segmentation in toddlers. *Infancy*, 28(2):301–321.

Babineau, Mireille, Legrand, Camille, and Shi, Rushen (2021). Variable forms in French-learning toddlers' lexical representations. *Developmental Psychology*, 57(4):457.

Babineau, Mireille and Shi, Rushen (2014). Distributional cues and the onset bias in early word segmentation. *Developmental Psychology*, 50(12):2666.

Babineau, Mireille, Shi, Rushen, and Achim, André (2017). Contextual factors in lexical processing: The case of French Liaison. *Language, Cognition and Neuroscience*, 32(4):457–470.

Benedict, Helen (1979). Early lexical development: Comprehension and production. *Journal of child language*, 6(2):183–200.

Bergelson, Elika and Swingley, Daniel (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9):3253–3258.

Bergelson, Elika and Swingley, Daniel (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child development*, 89(5):1567–1576.

Bouchard, Caroline, Trudeau, Natacha, Sutton, Ann, Boudreault, Marie-Claude, and Deneault, Joane (2009). Gender differences in language development in French Canadian children between 8 and 30 months of age. *Applied psycholinguistics*, 30(4):685–707.

Braginsky, Mika, Yurovsky, Daniel, Marchman, Virginia A, and Frank, Mike (2016). From uh-oh to tomorrow: Predicting age of acquisition for early words across languages. In *CogSci*, volume 6.

Braginsky, Mika, Yurovsky, Daniel, Marchman, Virginia A, and Frank, Michael C (2019). Consistency and variability in children's word learning across languages. *Open Mind*, 3:52–67.

Chevrot, Jean-Pierre, Dugua, Céline, and Fayol, Michel (2009). Liaison acquisition, word segmentation and construction in French: a usage-based account. *Journal of child language*, 36(3):557–596.

Chevrot, Jean-Pierre, Dugua, Céline, Harnois-Delpiano, Mylène, Siccardi, Anne, and Spinelli, Elsa (2013). Liaison acquisition: debates, critical issues, future research. *Language Sciences*, 39:83–94.

Cŏté, Marie-Hélène (2011). French liaison. *The Blackwell companion to phonology*, pages 1–26.

Durand, Jacques and Lyche, Chantal (2008). French liaison in the light of corpus data. *Journal of French language studies*, 18(1):33–66.

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

Fenson, Larry, Dale, Philip S, Reznick, J Steven, Bates, Elizabeth, Thal, Donna J, Pethick, Stephen J, Tomasello, Michael, Mervis, Carolyn B, and Stiles, Joan (1994). Variability in early communicative development. *Monographs of the society for research in child development*, pages i–185.

Fenson, Larry et al. (2007). Macarthur-Bates communicative development inventories.

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

Hale, Kenneth (1973). Deep-surface canonical disparities in relation to analysis and change: An Australian example. *Current trends in linguistics*, 11(19731):401–458.

Hyman, Larry M (2018). Why underlying representations? *Journal of Linguistics*, 54(3):591–610.

Kim, Yun Jung and Sundara, Megha (2021). 6–month–olds are sensitive to English morphology. *Developmental science*, 24(4):e13089.

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

Kiparsky, Paul (2012). *Explanation in phonology*, volume 4. Walter de Gruyter.

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

Markman, Ellen M and Wachtel, Gwyn F (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, 20(2):121–157.

Markman, Ellen M, Wasow, Judith L, and Hansen, Mikkel B (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive psychology*, 47(3):241–275.

Marquis, Alexandra and Shi, Rushen (2012). Initial morphological learning in preverbal infants. *Cognition*, 122(1):61–66.

Meinschaefer, Judith, Bonifer, Sven, and Frisch, Christine (2015). Variable and invariable liaison in a corpus of spoken French. *Journal of French Language Studies*, 25(3):367–396.

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

Peters, Ann M and Zaidel, Eran (1980). The acquisition of homonymy. *Cognition*, 8(2):187–207.

Prince, Alan (2004). Optimality Theory: Constraint interaction in generative grammar. *University, New Brunswick, and University of Colorado*.

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

Richter, Caitlin (2021b). Consequences of lexical variability in toddlers' vocabularies.

Ringe, Don and Eska, Joseph F (2013). *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge University Press.

Schmid, Helmut, Baroni, Marco, Zanchetta, Erika, and Stein, Achim (2007). Il sistema 'tree-tagger arricchito'–the enriched treetagger system. *IA Contributi Scientifici*, 4(2):22–23.

Shi, Rushen and Cyr, Marilyn (2008). Processing of morphological variations in toddlers. In *BUCLD*, volume 34, pages 363–374.

Smolensky, Paul and Goldrick, Matthew (2016). Gradient symbolic representations in grammar: The case of French liaison. *Rutgers Optimality Archive*, 1552:1–37.

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

Swingley, Daniel and Aslin, Richard N (2007). Lexical competition in young children's word learning. *Cognitive psychology*, 54(2):99–132.

Swingley, Daniel and Humphrey, Colman (2018). Quantitative linguistic predictors of infants' learning of specific English words. *Child development*, 89(4):1247–1267.

Tesar, Bruce and Smolensky, Paul (1998). Learnability in optimality theory. *Linguistic inquiry*, 29(2):229–268.

Tessier, Anne-Michelle, Jesney, Karen, Vesik, Kaili, Lo, Roger, and Bouchard, Marie-Eve (2022). The Productive Status of Laurentian French Liaison: Variation across Words and Grammar. In *Proceedings of the Annual Meetings on Phonology*.

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

Wauquier, Sophie (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, Sophie and Braud, Virginie (2005). Proto-déterminant et acquisition de la liaison obligatoire en français. *Langages*, 158(2):53–65.

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