

**Competing Models of Liaison Acquisition:**  
Evidence from Corpus and Experimental Data

Angelica Buerkin-Pontrelli  
*University of Pennsylvania*  
[abuerk@sas.upenn.edu](mailto:abuerk@sas.upenn.edu)

Jennifer Culbertson  
*University of Edinburgh*  
[jennifer.culbertson@ed.ac.uk](mailto:jennifer.culbertson@ed.ac.uk)

Géraldine Legendre  
*Johns Hopkins University*  
[legendre@jhu.edu](mailto:legendre@jhu.edu)

Thierry Nazzi  
*Université Paris Descartes*  
[thierry.nazzi@parisdescartes.fr](mailto:thierry.nazzi@parisdescartes.fr)

**Competing Models of Liaison Acquisition:**  
Evidence from Corpus and Experimental Data

## **Abstract:**

Given that nouns rarely appear in isolation in French, infants acquiring the language must often retrieve the underlying representation of vowel-initial lexical forms from liaison contexts which provide conflicting information about their initial phoneme. Given this ambiguity, how do learners represent these nouns in their lexicon, and how do these representations change as their knowledge of liaison and the lexicon become more adult-like? To explore this question, we analyze the types of errors children make, in both naturalistic and elicited speech, and how these are affected by input frequency. In doing so, we evaluate two major proposals for how children's early representations of liaison develop. The first model, couched in a constructionist framework, predicts relatively late mastery of liaison (age 5 or older) and heavy dependence on the contexts in which a particular noun appears in the input. The second model, takes an approach to liaison development which integrates it more closely with general phonological development, and predicts relatively early mastery (by age 3). The results of a corpus study reveal that by age 3 children are correctly producing liaison in the nominal domain and that their production errors are consistent with a phonological model of liaison acquisition. An elicitation task demonstrates that 3-year-olds' succeed at learning and correctly apply their knowledge of liaison to new nouns following brief exposure, though their productions continue to be influenced by nouns' input distributions. Taken together, our findings suggest that by age 3 children are well on their way to adult-like representations of liaison. A phonologically-based model, incorporating the effect of distributional context on early errors, provides a better overall fit to the data we present.\*

**Keywords:** language, language acquisition, French liaison, word segmentation, distributional cues, abstract knowledge

\*This study was conducted with the support of NSF grants BCS0446954 and BCS1251707 to GL as well as an REU to ABP while an undergraduate student at JHU. Special thanks Colin Wilson and Paul Smolenksy. We also thank Camillia Bouchon, Laurianne Cabrera, and Julien Marie who recorded our stimuli as well as the infants and parents who took part in our study.



## 1. INTRODUCTION

**1.1. LIAISON IN FRENCH.** French liaison is a relatively frequent (Boë & Tubach 1992) morpho-phonological process occurring at the boundary of two contiguous words. It is context dependent, causing an otherwise silent consonant (liaison consonant or LC) orthographically at the end of one word ( $W_1$ ) to surface in speech when a vowel-initial word ( $W_2$ ) follows. For example, in isolation or when followed by a consonant-initial word, *deux* ‘two’ is pronounced [dø]; however, when it is followed by a vowel-initial word, like *amis* ([ami] ‘friend’), a /z/ surfaces between the two words, yielding the string [dø.za.mi]. Liaison overwhelmingly resyllabifies the final consonant of  $W_1$  as the onset of  $W_2$  (Côté 2011). For example, in *deux amis*, the LC /z/ is syllabified as the onset of the  $W_2$  *amis*, resulting in [dø.za.mi]. Conventional analyses classify liaison as a means to ‘repair’ a hiatus (Tranel, 1996): insertion of an LC and resulting resyllabification avoids a realization of two contiguous vowels (as in \*[dø.ami]), and creates a universally-preferred consonant-initial syllable onset (Prince & Smolensky 1993/2004, McCarthy & Prince 1994).

Traditional approaches analyze the underlying position of the LC as the coda of  $W_1$ , following modern-day orthography (a reflection of historical pronunciation) e.g. *un* [œ̃] (see Côté 2011 for a thorough review of these theories). However, based on the observation that LCs are realized across intonational boundaries while their stable counterparts are not (e.g. *un robuste, mais petit, enfant* [pə.ti.tã.fã], but *une robuste, mais petite, étoile* [pə.tit || e.twal]), Morin (2005) suggests that LCs are underlying onsets of  $W_2$ . A third possibility, advocated by Tranel (1996) for example, is that LCs are “floating” consonants, inserted into the empty onset position of  $W_2$ . Finally, some researchers argue that the underlying representation of an LC depends on its context (Côté 2005, Durand & Lyche 2008).

The acquisition of liaison, regardless of its precise analysis, presents a puzzle for French learners. Here we focus on liaison in one particular domain—the noun phrase. French does not in general allow bare nouns, even in the case of generic (definite) subject NPs.<sup>1</sup> Instead, a determiner is required, which triggers mandatory liaison with vowel-initial nouns, as shown in (1). While most adjectives in French are post-nominal, a (small) set of relatively frequent pre-nominal adjectives also triggers mandatory liaison, as in (2).

(1) \*(Les) ours dorment pendant l'hiver. [le.zurs]<sup>2</sup>

‘Bears sleep during winter’

(2) Le petit ami de Claude. [pə.ti.ta.mi]

‘Claude’s boyfriend (or little friend)’

Given that nouns rarely appear in isolation in French, infants acquiring the language must retrieve the underlying representation of vowel-initial lexical forms like *ami* from contexts which provide conflicting information about their initial phoneme. For example, children may encounter a noun in many different liaison contexts, e.g. *les amis* [le.za.mi] ‘the friends’, *un ami* [œ.na.mi] ‘a friend’, *l’ami* [la.mi] ‘the friend’<sup>3</sup>, *petit ami* [pə.ti.ta.mi] ‘boyfriend’, and occasionally in a context where no liaison is triggered, e.g. *joli ami* [ʒo.li.ami] ‘pretty friend’. These examples potentially provide evidence for /z/, /n/, /l/, /t/, and /a/ as the initial segment of *ami*. Given this ambiguity, how do learners represent these nouns in their lexicon, and how do these representations change as their knowledge of liaison and the lexicon become more adult-like? To explore this question, we analyze the types of errors children make, in both naturalistic and elicited speech, and how these are affected by input frequency. In doing so, we evaluate two major proposals for how children’s early representations of liaison develop. These proposals contrast in the extent to which they posit early abstract knowledge. First, however we discuss in more detail the nature of the segmentation problem, which children must solve in order to acquire liaison productively.

**1.2 SEGMENTING LIAISON.** Liaison creates a misalignment of lexical and syllabic boundaries, which leads to segmentation challenges above and beyond what young learners already face. How French-speaking adults identify and segment liaison provides some insight into the sources of information which might be most useful in the acquisition process. In general, adults appear to primarily rely on lexical information, and top-down knowledge about liaison, rather than available acoustic cues. For example, LCs are acoustically distinct from fixed onset segments, exhibiting shorter VOTs and duration (Wauquier-Gravelines 1996, Gaskell et al. 2002).

Nevertheless, adults cannot reliably distinguish ambiguous familiar sequences like *son oeuf* [sɔ̃.nœf] ‘his egg’ and *son neuf* [sɔ̃.nœf] ‘his nine’ (Yersin-Besson & Grosjean 1996, Shoemaker & Birdsong 2008, Tremblay 2011, cf. Spinelli et al. 2003). This lack of sensitivity to acoustic cues has been found with nonce nouns as well (e.g. *ses onches* [se.zɔ̃ʃ] vs. *ses zonches* [se.zɔ̃ʃ], Babineau & Shi 2013.) Instead adults rely on lexical cues—for instance, whether a particular lexical item is known to be vowel-initial in non-liaison contexts or tends to exhibit an alternation involving a particular LC—to disambiguate LCs and stable consonants. Evidence for the use of lexical cues comes from observations that well-formed liaisons such as *petit avion* [pə.ti.ta.vjɔ̃] ‘small plane’ are more easily processed than ungrammatical cases such as *vrai avion* \*[vrɛ.ta.vjɔ̃] ‘real plane’ (Spinelli & Meunier 2005). Furthermore, knowledge of the frequency with which a given consonant occurs as an LC appears to play a role in processing nonce nouns embedded in ambiguous liaison contexts. In an implicit segmentation task, Babineau & Shi (2013) found that Canadian French speaking adults were less likely to treat /z/ and /n/, the two most frequent LCs, as fixed onsets of novel words compared to other potential LC consonants in onset position. Tremblay and Spinelli (2013) report that adults tend to interpret temporally ambiguous vowel-initial words in liaison contexts (e.g. *parfait*-/t/-*abri* ‘perfect shelter’ vs. *parfait*-/ø/-*tableau* ‘perfect painting’) as consonant initial if they began with a /t/ or an /n/ (equally likely to occur as word-initial and liaison consonants), but not a /z/ (more frequent as an LC). Taken together these studies suggest that top-down information outweighs bottom-up (acoustic) cues in liaison processing in adults.

As this top-down knowledge develops over the course of language acquisition, young children’s ability to segment words in liaison contexts is likewise expected to undergo considerable developmental changes. In very early stages, studies have shown that segmentation of vowel-initial words more generally—liaison aside—is known to pose a significant hurdle to learners. Studies conducted with American English-acquiring infants have shown that segmentation of vowel-initial words lags behind that of consonant-initial words, which emerges as early as 6 months old (e.g. Bortfeld, Morgan, Golinkoff, Rathbun 2005, Jusczyk & Aslin 1995). Using the Headturn Preference Procedure (HPP), Seidl & Johnson (2008) found that the earliest age at which children successfully segmented novel vowel-initial nouns was 11 months

(see also Mattys & Jusczyk 2001, Nazzi, Dilley et al. 2005). Even at this age, however, successful segmentation was limited to salient positions (e.g. sentence-initially as in *Eff runs a circus in Toronto*). These studies suggest the possibility that very young children may initially prefer to segment consonants as syllable onsets, potentially leading to mis-segmentation errors (following the Maximum Onset Principle, Prince & Smolensky 1993/2004).

Indeed, evidence from French-acquiring children supports an initial preference for segmenting LCs as the onset of the  $W_2$ . In a series of preferential-looking experiments, Babineau & Shi (2011) found evidence suggesting that 20-month-olds do not segment vowel-initial nonce words familiarized in multiple liaison contexts (e.g. *un onche* [œ̃.nɔ̃ʃ], *des onches* [de.zɔ̃ʃ]) as vowel-initial nonce words, but rather as consonant-initial nonce words including the LC. By 24-months of age, however, infants make use of the distributional cues in familiarization and switch from segmenting e.g. *zonches*, to segmenting vowel-initial e.g. *onches*. This suggests that by 24 months French-acquiring children can use language-specific distributional information to override an initial preference for segmenting CV syllables, and successfully retrieve new vowel-initial nouns from liaison contexts (for related work on English-acquiring children, see Johnson et al. 2003). However, the nature of children's early representations of vowel-initial nouns, and the age at which they can productively use knowledge of liaison to segment new nouns in an adult-like way remains unclear. In the next section we review two competing proposals which aim to answer these questions, and motivate the current investigation.

**2. TWO MODELS OF LIAISON ACQUISITION: CONSTRUCTIONIST VS. PHONOLOGICAL.** The two competing models of liaison acquisition we review here primarily focus on evidence for children's developing knowledge coming from their spontaneous productions. Both models incorporate the claim, discussed above, that very young children exhibit universal segmentation preferences. The differences lie in the predicted age at which liaison is expected to be fully productive, and what is claimed to drive production errors. The first model takes a constructionist approach, with distributional information in the input playing a major role and abstract knowledge of liaison developing relatively late (Chevrot & Fayol 2001, Chevrot et al. 2007, Dugua et al. 2009, Chevrot, et al. 2009, Chevrot, et al. 2013). The second model argues



that liaison develops in line with more general principles of phonological acquisition, with relatively early abstract knowledge (Wauquier-Gravelines & Braud 2005). We evaluate the distinct predictions of these two accounts using results from a new corpus study, focusing on two main questions:

- At what age do children master liaison production? Here we use a criterion, suggested by Chevrot et al. (2009), of 80% overall correct liaison production.
- Do the errors made in liaison production at various ages suggest a larger role for distributional information or principles of general phonological development?

**2.1. CONSTRUCTIONIST MODEL.** The constructionist approach to acquisition of liaison posits three main developmental stages. At Stage 1 children simply memorize unsegmented determiner + noun sequences they encounter. No systematic liaison errors are predicted to occur at this stage. During Stage 2, children extract multiple exemplars for each vowel-initial noun from the input, with a preference for assigning CV syllable structures to exemplars. For example, a child hearing *ami* ‘friend’ across multiple contexts, e.g. *les amis* [le.za.mi], *un ami* [œ.na.mi], *l’ami* [la.mi], *petit ami* [pə.ti.ta.mi], *joli ami* [ʒo.li.ami] will extract and store the set of exemplars /zami/, /nami/, /lami/, /tami/, and when no onset is available in the input, the vowel-initial /ami/. As children hear more and more determiner + noun sequences, they begin to link together sequences that contain overlapping parts, e.g. the singular indefinite determiner *un* followed by a noun (*un ami*, *un ballon*, etc.), to form schemas of the type *un* + X or *deux* + X (we refer to these collectively as *det* + X, but *det* does not correspond to an abstract determiner category in this model). The absence of a specified LC in the schema permits any of the (mis-)parsed exemplars to be inserted into the X-slot.

This model makes clear predictions about the types of errors children should make in liaison production at Stage 2, with three types in particular: insertions, omissions, and replacements. On the surface, insertions are characterized by the presence of an LC in a context which either does not require, or actually blocks liaison, e.g. *joli ami* [ʒo.li.a.mi] produced as

\*[zo.li.na.mi]. Replacement errors involve an inappropriate LC surfacing in a liaison context, e.g. *les arbres* ‘the trees’ [le.zarbr] produced as \*[le.narbr]. Omission errors occur when an LC is missing in an obligatory liaison context, e.g. *un arbre* ‘a tree’ [œ.narbr] produced as \*[œ.arbr]. Under the constructionist model, all three of these error types are the result of using the wrong exemplar in the X slot of a *det* + X schema. In addition to schema-driven errors, LC-initial exemplars may occur in isolation in the rare instance that a determiner is not required (e.g. *Escargot* \*[ne.skar.go], *que fais-tu?*), or if it has been omitted.

Importantly, Stage 2 schemas do not specify any phonological information about the LC that corresponds to a particular context. Rather, this model proposes that selection of a particular exemplar is determined by input frequency. If the strength of an exemplar (and therefore its probability of use) is based on context frequency, in principle each exemplar could be selected with equal probability. However, Dugua (2006, see also Dugua et al., 2009) suggests that many French nouns surface reliably more often in particular contexts. For example, a vowel-initial noun may be more frequently used in plural contexts, and thus it will tend to surface with a resyllabified [z] in onset position. Similarly, if a vowel-initial noun tends to surface in singular contexts, it is more likely to be preceded by an [n], the most frequent singular LC (for a more general discussion of LC frequencies see Ahmad 1993).

Some time later, children enter Stage 3, in which memorized schemas are generalized into constructions, or conventionalized patterns (Chevrot et al. 2009). This results in new schemas of the form *DET* + /LCX/, with phonetic information about the LCs in particular contexts specified. Because each Stage 3 schema is associated with a particular LC, use of that schema necessarily means use of that LC-initial exemplar (e.g. *les* + /zX/ allows only z-initial exemplars like [za.mi], [ze.skar.go]). As a consequence, use of these schemas with nouns beginning with consonants that also function as LCs—e.g. /n/-initial or /z/-initial nouns—is predicted to result in a new type of replacement error. In this type of error, LC-initial nouns are modified to fit the schema used. For example, use of the schema *les* + /zX/ with the noun *nuages* ‘clouds’ could result in \*[le.zy.aʒ], since the /n/ is erroneously analyzed as a LC and substituted with a /z/ to accommodate the schema.

Although individual variation in age of mastery is expected under this model, proponents have nevertheless suggested rough mappings of these stages to age ranges. For example Chevrot et al. (2009) suggest that children begin to enter Stage 2 at approximately 3;0, and Stage 3 at approximately 5;0.<sup>4</sup> A summary of the three stages of the constructionist model and error types predicted at each stage<sup>5</sup> is provided in Figure 1.

<INSERT FIGURE 1 ABOUT HERE>

Initial support for this model is provided in Dugua (2006), who analyzes liaison errors reported by the parent of a French-acquiring child, Sophie, between ages 2;1 and 6;4. The types of errors considered reflect the insertions, replacements, and omissions predicted by schemas present in Stages 2 and 3 of the constructionist model. Examples of each type are provided in (3). In terms of frequency, although only raw counts of errors (as opposed to error rates) are provided, insertions were most frequent (547), followed by replacements (370), and finally omissions (14).

(3) Example Stage 2/3 errors (Dugua 2006)

<INSERT EXAMPLE 3 ABOUT HERE>

As discussed above, the constructionist model predicts that children's liaison errors (particularly in Stage 2) should be affected by the frequency with which they hear nouns in particular liaison contexts. Some evidence for this can be seen in Sophie's most frequent errors; Dugua (2006) reports that Sophie's *ami* errors mostly involved /z/, while *anorak*, *arc-en-ciel*, *éléphant*, *escalier*, *ours*, and *avion* tended to surface with /n/. Dugua et al. (2009) report that according to adults' judgments and the French lexical database Lexique (New et al. 2001), *anorak*, *arc-en-ciel*, *avion*, *éléphant*, *escalier*, and *ours* are most likely to occur in singular contexts, but *ami* in plural. This subset of Sophie's frequent errors is thus plausibly attributed to the dominant context in which the relevant nouns appear in the input.

To further examine the role of input asymmetries in child liaison productions, Dugua et al. (2009) conducted an elicitation task with children aged 3-4 years. Children were shown a singular or plural picture of an object (e.g. *arbre* ‘a tree’) and asked to describe what they saw. Critically, the stimuli included eight vowel-initial nouns, half found more in the plural and half found more in the singular, according to adult judgments and Lexique. The results showed that nouns found more in the plural were more likely to be produced by children with initial [z] (e.g. *des arbres* [de.zar.br], *un arbre* \*[œ.zar.br]). By contrast nouns found more in the singular were more likely to be produced with initial [n] (e.g. *un avion* [œ.na.vjõ], *des avions* \*[de.na.vjõ]).

To summarize, Sophie’s errors (Dugua 2006) and the experimental data described above (Dugua et al. 2009) suggest that insertions, replacements, and omissions—driven by input frequency—are well-attested during liaison development. By contrast, data from additional corpora analyzed by Dugua (2006) suggest that at least in spontaneous speech, these error types may be somewhat marginal. In four additional corpora<sup>6</sup>, the average rates of replacements and omissions out of total liaison productions were 5.3% (68/1295) and 4.8% (62/1295) respectively, insertions rates were even lower at 0.8% (10/1295). Below we provide additional data from a corpus study, which corroborate the relatively marginal status of such errors. However, we first discuss an alternative model of liaison development, which seeks to explain it in the context of broader trends in phonological acquisition.

**2.2. A PHONOLOGICAL MODEL.** An alternative model of liaison development is proposed by Wauquier-Gravelines & Braud (2005). Generally speaking, the stages of this model are based on the evolution of syllabic templates and children advance relatively quickly in comparison with the predictions of the constructionist model. A schematic of this model including error types predicted at each developmental stage is shown in Figure 2.

<INSERT FIGURE 2 ABOUT HERE>

As in the constructionist model, this second model does not predict any systematic liaison errors in Stage 1. Rather, errors in this stage reflect more general phonological processes in early

speech such as those that preserve universally preferred binary minimal foot structure (Demuth 1994, Demuth & Johnson 2003).

At age 2;0, children enter Stage 2 in which they begin to parse the DETERMINER + NOUN sequences encountered in the input. Children at this stage are argued to maintain a strong preference to syllabify consonants in onset positions in their spontaneous productions (for comprehension see e.g. Babineau & Shi 2011, Johnson et al. 2003). Thus following the Maximum Onset Principle, V.CV syllables will be preferred over VC.V, leading to the misalignment of syllable and lexical boundaries, which is characteristic of liaison with resyllabification. Rather than proposing a set of exemplars that compete to fill X-slots in schemas based solely on their relative frequencies, this model proposes that children in this stage have an abstract awareness that a consonant occupies the onset position of  $W_2$ . Thus children are predicted to overwhelmingly produce forms with filled onset, and rarely if ever produce omission errors. Crucially however, at this stage the particular onset is not fully encoded and so children may rely on a number of distinct mechanisms to assign a particular segment to the position. While this segment may be the most frequent LC heard, it could also result from reduplication (e.g. *un avion* [œ.na.vjõ] produced as \*[œ.jo.jõ]), truncation (e.g. *un avion* [œ.vjõ]), or the insertion of a default or idiosyncratic consonant to fill the onset of  $W_2$ . A given noun is therefore expected to be produced in a variety of forms, e.g. *un ami* could be realized as [œ.la.mi], [œ.na.mi], [œ.ja.mi], [œ.za.mi], [œ.mi.mi], and [œ.mi].

By 3;0 to 3;5, children enter Stage 3 in which they use morphological bootstrapping to identify which LCs occur in which contexts. This is encouraged by exposure to allomorphs and derived forms of a particular  $W_1$  involved in liaison. For example, children encountering masculine and feminine allomorphs of determiners (e.g. *un(e)* as [œ] or [yn]) and prenominal adjectives (e.g. *petit(e)* as [pə.ti] or [pə.tit]) implicitly notice the presence of LCs. It remains unclear to what extent this process applies to  $W_1$ s such as *deux* or *les*. However, under this view, once children decipher the relationship between  $W_1$ s and the LC they trigger they are predicted to produce correct liaison sequences.

The bulk of the evidence in support of this model of liaison development comes from the longitudinal studies of two French-acquiring children, Clara (1;4-2;2) and Claire (1;10-1;11)

whose liaison productions were analyzed by Wauquier-Gravelines (2003). No vowel-initial nouns are attested among the earliest productions. As the children approach Stage 2 around 2;0, their productions illustrate truncation and reduplication processes predicted to be used in order to fill the onset position of  $W_2$ . Examples of LC insertions, replacements, and omissions are provided in (4). The last example illustrates the kind of idiosyncratic strategies children might use in order to fill  $W_2$  onsets: the lexical item *arc-en-ciel* ‘rainbow’ produced with two different onsets by Claire in the same recording session.

(4) Example Stage 2 errors (Wauquier-Gravelines 2003)

<INSERT EXAMPLE 4 ABOUT HERE>

Results from a picture-naming task conducted by Braud (1998) with 3-year-old children provide additional evidence for the prevalence of truncations and reduplications in child speech at the cusp of Stage 2. As shown in (5), examples involve reduplication of the initial syllable of  $W_2$ .

(5) Example reduplications (Braud 2003)

<INSERT EXAMPLE 5 ABOUT HERE>

**2.3. NEW CORPUS ANALYSIS AND MODEL COMPARISON.** The two alternative models discussed here differ in a number of crucial ways—in terms of the age at which mastery is predicted, the types of errors expected at a given age, and the role of input frequency in those errors. The corpus and experimental data used to support these two models provides some support for both accounts; all error types predicted by the constructionist account are found in Sophie’s data as well as the corpus analyzed by Wauquier-Gravelines & Braud (2005). In addition, evidence of input frequency effects was in Dugua et al. (2009), and in Sophie’s data. On the other hand, constructionist-type errors appear to be relatively rare in the other corpora investigated by Dugua (2006), and a number of other error types, which can be analyzed as heavily dependent on early

phonological constraints rather than input frequencies are reported by Wauquier-Gravelines & Braud (2005).

Below we further investigate age of mastery, error type, and influence of input frequency using additional corpus and experimental evidence. Section 3 examines these issues in light of a detailed quantitative analysis of the phonologically transcribed Lyon corpus (Demuth & Tremblay 2008) available in the CHILDES database (MacWhinney 2000). Our analysis has the advantage of providing a clearer picture of children's overall accuracy compared to the corpus results reported in the literature so far. Further, we will directly compare errors of the type predicted by *BOTH* accounts rather than focusing on a particular type. In section 4, we present a liaison production experiment, which focuses on age of mastery and the role of input frequency effects. To briefly preview our new corpus results, we find that (i) children master liaison earlier than expected according to constructionist model, (ii) the hallmark errors predicted by the constructionist model are rarely attested in child speech, (iii) errors predicted by the phonological account occur relatively often. The results of our production study, while supporting a role for input distributional asymmetries in children's errors, reveal that production of liaison in 3-year-olds is well on its way to matching the target adult grammar. The conclusion is supported by a simple computational model that combines input frequency and an abstract target-like grammar of liaison to predict performance in the task.

**3. STUDY 1: NEW EVIDENCE FROM THE LYON CORPUS.** This section evaluates the predictions of the two existing models of liaison acquisition discussed above in light of a detailed quantitative analysis of the liaison productions of three children (Théotime, Anaïs, and Marie) in the Lyon corpus. The corpus consists of approximately 119 hours of parent-child interactions recorded bi-monthly over a two-year period from age 1–3 years. These sessions occurred at the children's homes in the Lyon, France region. A total of 774 vowel-initial noun tokens in obligatory liaison and elision contexts were found across the three children. Though elision and liaison are separate linguistic phenomena, the models we compare do not make any theoretical distinction between the two processes in liaison acquisition. Following previous work, we have elected to include elisions in our data set for the purpose of the analyses we present here.<sup>7</sup>

Each vowel-initial production was coded as correct or incorrect, and errors were classified according to a set of 7 types. Recall that the errors predicted in Stage 2 of the constructionist model (ages 3;0-5;0), refer specifically to those which result in ungrammatical insertion of LC-initial or vowel-initial exemplars into early schemas. Here we refer to these error types separately as INSERTION, REPLACEMENT, and OMISSION, and collectively as LC ERRORS. This is intended to contrast with what we call PHONOLOGICAL ERRORS, namely reduplications and truncations, which result from a general constraint to fill onset positions of vowel-initial nouns at Stage 2 (ages 2;0-3;0) of the phonological model.

The remaining three error types are *OURS* ERRORS, GLIDE-INITIAL NOUN ERRORS, and OTHER errors. These are treated separately here. The noun *ours* ‘bear’ showed a strong effect of input context frequency in Dugua (2006) and Dugua et al. (2009), however it is not straightforward to diagnose errors with this word due to a related noun commonly used in child-directed speech: *nounours* ‘teddy bear’. While this word in fact arose initially from a liaison error (plus reduplication), as we discuss further below, it is now a distinct lexical item present in children’s input and child speech where it is subject to truncation. A second source of errors which we treat separately are glide-initial nouns (e.g. *oiseau* [wa.zo] ‘bird’) which present a particular problem for children that, as we argue below (section 3.3), is independent of liaison. Finally, we exclude 7 “other” tokens, which could not be categorized as either LC errors, phonological errors, or correct productions due to their ambiguity<sup>8</sup>.

**3.1. RESULTS.** Figure 3 illustrates the proportion (and raw frequency) of correct liaison productions as well as the proportion of errors of each type across all three children. The data are presented in 6-month age bins. At the two earliest age bins, children are producing relatively few vowel-initial nouns, and all errors are phonological in nature. This is expected if, at this age, children are producing unanalyzed determiner + noun combinations. From 2;0-2;6, raw token frequency increases sharply, and errors appear. Both token frequency and proportion of correct liaison productions increase from this point on to age 3;0. The children show some variation in overall accuracy rates, suggesting individual differences in development of liaison: Théotime and Marie showed relatively high proportions of correct liaison productions (83% and 73%



respectively), while Anaïs lags behind to some extent (61%).<sup>9</sup> By the last age bin analyzed (2;6–3;0) the children are on average producing correct liaison 86% of the time.

<INSERT FIGURE 3 ABOUT HERE>

As is clear from Figure 3, errors of the type predicted by the constructionist model—LC errors including insertions, replacements, and omissions—are overall quite rare, constituting 5.2% (40/774) of all responses across children in all age groups. Figure 4 illustrates the relative proportions of each error type more clearly. Importantly, LC errors also constitute only 24.7% (40/162) of all errors, and thus make up a small proportion relative to phonological errors. Regardless of age, this is unexpected under the constructionist model. Further, the errors appear at a particular age range, from 2;0–3;0, which is early relative to the age estimates described by the constructionist model. This range DOES accord well with Stage 2 (age 2;0–3;0) of the phonological model, where children are predicted to use various strategies to fill the onset position of  $W_2$ .

One possibility is that LC errors are in fact on the rise at the point where our data end. This is suggested by the slight increase in such errors shown in Figure 4. The constructionist account makes clear predictions that the proportion of LC errors should continue to increase relative to other production types from the age of 3;0 until liaison mastery. This is due to the insertion of consonant initial exemplars into schemas of the type DET + X. However, the slight increase shown in our data is not significant ( $\chi^2 = 1.55$ ,  $df = 1$ ,  $p = 0.21$ ), supporting the claims that LC errors remain comparatively low even after the age of 3;0. Further, data for one child, Marie, are available after age 3;0. Analysis of these data shows only 2/63 LC errors between the ages of 3;0–3;5, further corroborating the claim that errors of this type are rare overall, and appear between 2;0 and 3;0.

<INSERT FIGURE 4 ABOUT HERE>

While LC errors are relatively rare overall, almost all types predicted were nevertheless attested. In contrast to the corpus data reported by Dugua (2006), here omission errors, as in (6), were the most frequent LC error type, accounting for 14.2% (23/162) of children's errors. Replacement errors mainly came from a single child at age 2;6 (8/11 of those found), these are given in (7). Isolated errors of the type \*[na.mi], which according to the constructionist account result from storage of multiple, autonomous exemplars during Stage 1, occurred only once across the three children's productions. Such errors were quite common in the production task reported in Chevrot et al. (2009), where children (age 2;4-3;1) were asked to call puppets (e.g. *âne*, *viens ici!* 'donkey, come here!'). Interestingly, we found no LC replacement errors of the type \*[le.zy.aʒ] for *les nuages*, which were observed in Dugua (2006) and produced by over half of children tested in Chevrot et al. (2009). While this could be due to the early age of children in this corpus, the high accuracy rate at age 2;6-3;0 suggests this is unlikely (and we do not observe errors of this type later on for Marie). Below we discuss alternative explanations for this discrepancy between our data and these previous studies.

(6) Example omissions (all children, 2;3-2;10)

<INSERT EXAMPLE 6 ABOUT HERE>

(7) Example replacements (Marie, 2;6)

<INSERT EXAMPLE 7 ABOUT HERE>

In addition to the relative frequency of phonological errors, the diversity of strategies used to assign consonants to the onset of  $W_2$  provides support for the account of Wauquier-Gravelines & Braud (2005). Overall, 25% (11/44) of the vowel-initial noun types in the data showed some variation in their initial consonant, confirming that the strong preference to fill the onset position of the  $W_2$  at this stage can be satisfied in a number of different ways. For example, Marie produced the target form *assiette* 'plate' with four different onsets, not all of them possible LCs: [na.sje], [sa:.set], [ma.sje:t], [ta.sjet]. Thus while errors like those in (7) were analyzed here as replacements, errors involving the same lexical items clearly result from phonological processes which do not involve substituting one LC for another. For example, the productions in

(8), (9), and (10) represent the kinds of truncation, reduplication and idiosyncratic errors found in the corpus before age 3;0. The examples in (10) illustrate tokens in which Marie deletes the rhyme of the  $W_1$  and resyllabifies the leftover consonant as the onset of  $W_2$ , preserving the foot structure of the target and satisfying the preference for filled onsets.

(8) Example truncations (all children, 1;7-2;6)

<INSERT EXAMPLE 8 ABOUT HERE>

(9) Example reduplications (all children, 1;10-2;7)

<INSERT EXAMPLE 9 ABOUT HERE>

(10) Example idiosyncratic phonological processes (Marie, 2;0-2;7)<sup>10</sup>

<INSERT EXAMPLE 10 ABOUT HERE>

To summarize, the majority of errors these children make in liaison (and elision) production are best characterized as phonological errors rather than LC errors. The latter are relatively rare in this sample of spontaneous child speech, and appear to make up a subset of productions during a developmental phase in which liaison is not yet fully mastered but children have formed a strong constraint requiring that  $W_2$  onsets be filled. The particular consonant that satisfies this constraint may be an alternative LC, another segment present in the noun token, or some idiosyncratically chosen segment. That said, omission errors appear to be more frequent than predicted by either model. The strong constraint to fill onsets according to the phonological model leads to the clear prediction that omissions should be rare. In the constructionist model, at the time when children are nearing liaison mastery (approaching 80% correct), they should be using Stage 3 schemas containing LC information (e.g. des + /z/X) more than Stage 2 schemas (e.g. des + X) which allow some omission errors. In the following sections, we revisit two further issues which we believe might help to explain discrepancies between our results and those reported in Dugua (2006) and Dugua et al. (2009). In particular, we provide evidence from the Lyon corpus suggesting that errors with glide-initial words and the lexical item *nounours*,

which at first glance appear to be liaison-related, are best understood independently of it. First, however, we briefly discuss children's productions of liaison in the verbal domain.

**3.2. LIAISON MASTERY IN THE VERBAL DOMAIN.** While the focus of the current study is in the nominal domain, the most common context (Durand et al. 2011) for liaison production involves a preverbal subject clitic pronoun and a vowel-initial verb as in *ils arrivent* [il.za.riv] 'they arrive', *il est* [i.le] 'he is' (this qualifies as liaison in the register in which the final [l] of *il* is only pronounced before a vowel). Figure 5 illustrates the proportion (and raw frequency) of correct liaison productions as well as the proportion of errors of each type for vowel-initial verbs preceded by a liaison-triggering pronoun. Quite clearly, the children in the Lyon corpus show precocious mastery of liaison in this context; productions are nearly perfect (99.97% correct). The errors found (40/1149) consist of 36 omissions and 4 replacements, with more than half of the omissions produced by Théotime (29/40). No errors between a lexical subject and a following vowel-initial verb—a context which precludes liaison—were found (e.g. *le petit a réussi* [lə.pə.ti.a.re.y.si] → \*[lə.pə.ti.ta.re.y.si] 'the youngster succeeded'). This is in line with findings reported in Legendre et al. (2010), which provide evidence for successful comprehension of liaison-based subject-verb agreement by 2;6. We discuss the apparent asymmetry in verbal and nominal liaison acquisition further below.

<INSERT FIGURE 5 ABOUT HERE>

**3.3. A SEPARATE GLIDE PROBLEM?** One of the error types which we suggested above should be treated separately are those which involve a glide-initial noun. Though glide errors may arise from general phonological processes, peripheral to liaison per se, their surface forms invite misinterpretation as liaison errors. For instance, if the glide-initial noun *whisky* [wis.ki] is realized as [zis.ki], this could be either due to the use of the exemplar /ziski/ or more general phonological constraints. While few such nouns (or other word types for that matter) are attested in children's speech in the Lyon Corpus, the one which occurs with some frequency appears to cause consistent difficulty relative to other vowel-initial nouns. This is the word *oiseau* [wa.zo]

‘bird’. Out of all tokens of *oiseau*, 78.9% (71/90) involve deletion of the initial glide /w/, and before 2;5 only 3 productions retain it. Truncation and reduplication errors are frequent (49/90), and the realization [za.zo] is a particularly common error. Overall, the three children produced more *oiseau* errors than actual LC errors, with only 10/90 correct productions attested.<sup>11</sup> We discuss two possible explanations for children’s difficulty—the behavior of glides in liaison in the target grammar, and segmental problems with the glide /w/ itself, independent of liaison.

Glide-initial words in French exhibit variation with respect to triggering liaison (Côté 2011). For example, *oiseau* triggers liaison, but a semantically similar word like *oustiti* ‘marmoset’ does not. If lexically-based variability of this sort leads to difficulty in acquisition, one might expect children to commit errors of omission with glide-initial nouns. However, there were in fact very few omission errors with *oiseau* in liaison contexts (only 3/31 tokens), suggesting that some other factor may be at play.

In particular, it seems likely that children in the Lyon corpus do not have /w/ in their early segmental inventory. That this may be a general feature of phonological development in French is suggested by the fact that glides were almost entirely missing from early productions in the corpus data analyzed by Demuth & Johnson (2003). While no particular difficulty with /j/ is evident in the Lyon corpus (the children were able to produce *assiette* [a.sjet] and *avion* [a.vjõ] for example), there is systematic deletion of the /w/ in *oiseau*, and further analysis reveals a similar pattern of errors for other lexical items with /w/. The case of *voiture* [vwa.tyr] ‘car’ is illustrative. One child, Marie, systematically deletes the initial /vw/ cluster of *voiture*, and uses a number of different repair strategies: [ta.py], [pa.ty], [pa.py], [ba.tyr]. In the case of *oiseau*, we might then expect similar errors—in which the syllable structure is preserved, and the onset is filled idiosyncratically (Wauquier-Gravelines & Braud 2005, see also Demuth & Johnson 2003). Indeed, across the children, 17.8 % (16/90) of the consonants that appear in the onset are not possible LCs (e.g. [fa.zo], [ba.jo], [pa.so]). In light of this, *oiseau* errors like [za.zo] cannot be unambiguously analyzed as involving an underlying LC, but rather may simply be a consequence of children’s reduced productive segmental inventory. Hence, the pattern of appearance of these glide-initial nouns pattern strongly suggests that they are distinct from liaison errors, and should thus not be counted as such. Excluding these errors substantially decreases error rates, both in the

data in Dugua (2006), where *oiseau* accounts for 55 total errors, the second most error-triggering  $W_2$  after *âne* ‘donkey’, and in the Lyon Corpus, where removing this subset of errors reduces the total liaison error rate from 30% to 21%.

**3.4. INTERIM SUMMARY.** The results of the corpus analysis presented here provide some evidence consistent with both the constructionist and phonological models of liaison acquisition. Most importantly for the constructionist model, all types of LC errors predicted were in fact found in the corpus, including omissions (not expected under the phonological model). The trajectory of early phonological errors followed by a (brief) relative increase in LC errors suggests that as children’s phonological systems stabilize, they may produce errors more purely related to their developing knowledge of liaison. Nevertheless, taken together our results are better captured by the phonological model. First, LC errors were found to occur primarily at 2;0-3;0, about a year earlier than expected according to the stages laid out in Dugua (2006) and Chevrot et al. (2006). Second, phonologically-based errors were much more common than LC errors (15.7% versus 5.2%) suggesting that young children utilize a variety of phonological strategies in order to satisfy syllable structure constraints. Consonants which act as LCs may be implicated in these strategies, but more often they are not. Third, in terms of age of mastery, the Lyon corpus reveals that children begin to demonstrate adult-like liaison abilities well before the age of 5 or 6. Indeed, liaison accuracy reached 86% by 3 years of age (in line with error rates from other corpora reported in Dugua 2006).

Nevertheless, there remain interesting differences between the developmental picture painted by our corpus results and children’s behavior in production experiments reported by Dugua et al. (2009) and Chevrot et al. (2009). The latter reveal rather high error rates even at 3;0, and provide support for another important claim of the constructionist model—that differences in input frequency of a given noun in particular contexts might influence early liaison representations.

Several of the issues we have pointed out above may in part explain the high error rates found in these studies. For example, two of the experimental items used in Dugua et al. (2009)—*oiseau* [wa.zo] ‘bird’ and *yeux* [jø] ‘eyes’—may well have presented problems for children

independent of liaison since they are glide-initial. In addition to such stimuli, Dugua et al. (2009) and Chevrot et al. (2009) also use the lexical item *ours* [urs] ‘bear’, for which they observe a high rate of [nurs] errors. While they attribute such errors to mis-segmentation of liaison in the context of their task, another possibility is that children were substituting the alternative form *nounours* ‘teddy bear’. This related noun occurs commonly in child-(directed) speech (114 tokens of *nounours* relative to 125 *ours* in the Lyon corpus), and a truncated form [nurs] would be predicted by the phonological model (see also Allen & Hawkins 1978, 1980, Echols & Newport 1992, Echols 1993, Gerken 1994, Pater 1997, Kehoe 2000, for additional evidence of truncation errors in early child phonology). More generally, in light of the relative proportion of phonologically-driven errors found in the Lyon corpus it is critical to distinguish errors of this type from true LC errors in experimental studies.

The evidence provided thus far to suggest asymmetries in the frequency with which particular nouns occur in singular or plural contexts also deserves some further discussion here. Dugua et al. (2009) report a correlation between children’s errors and adults’ judgments of contextual frequency; more insertions of /n/ are found for nouns judged to occur more often in the singular, while more /z/ insertions are found for nouns judged to occur more in the plural. We were able to partially verify these asymmetries in CHILD-DIRECTED SPEECH from the Lyon corpus—for 6/8 nouns investigated by Dugua et al. (2009), we replicated the reported singular or plural-oriented asymmetry, for the remaining 2 (*arbre* ‘tree’, *escargot* ‘snail’) no asymmetry was found or the direction was reversed.<sup>12</sup> We take this to suggest that contextual asymmetry is at least somewhat variable, and thus a more controlled investigation of the role of input frequency is needed.

To summarize, we have so far reported corpus results which suggest relatively early mastery of liaison, discussed the potential mis-diagnosis of errors as liaison related in previous experiments, and pointed to the need for further investigation of input frequency effects. Motivated by these issues, we report the results of a production experiment with 3-year-old children below. Our experiment is designed to provide controlled input frequency asymmetries in order to investigate their effect on liaison production. Importantly, we test children’s use of

liaison with NOVEL NOUNS, thus providing additional evidence regarding the extent to which their early knowledge is productive.

#### 4. REPRESENTATIONS OF LIAISON AND THE ROLE OF EXPOSURE AT AGE 3

**4.1. STUDY DESCRIPTION.** This study extends our investigation of the two models of early liaison representations and how these differ in the extent to which they posit early abstract knowledge or frequency effects to account for development. To do so we examine knowledge of liaison in a relatively difficult task; we introduce 3-year-old French-acquiring children to novel vowel-initial nouns in a variety of contexts and ask whether they can learn and produce correct forms in new contexts—with and without liaison. Thus we are asking whether children by the age of 3 years have sufficiently abstract knowledge of liaison to enable them to generalize to novel nouns after only a few exposures. Importantly however, we manipulate the distribution of contexts in which children encounter these novel nouns to test the influence of input frequency on segmentation.

Each child learned two nonce nouns—one of which appeared mainly in /n/ liaison contexts (5/8) and only once in /z/, and the other which appeared mainly in /z/ contexts (5/8) and only once in /n/. Simulating typical input variation, each item also appeared in one /t/ and one non-liaison context. The 5:1:1:1 ratio of the familiarization provides an input asymmetry in favor of one particular exemplar while also providing sufficient distributional cues for a vowel-initial parse (Saffran et al. 1996, Gerken 2010). At test, children were prompted with phrases presenting different /z/, /n/, and zero-liaison contexts which they had to complete using the nonce nouns.

Based on the constructionist model we can hypothesize that at 3;0, the contexts in which nouns appear in the input are the main factor in determining what form is produced. This predicts that children in our experiment will overwhelmingly produce /n/-initial exemplars for new words encountered more in /n/ contexts e.g. \*[le.ni.vak] for *les ivaks*. Similarly, they will overwhelmingly produce /z/-initial exemplars for new words encountered more in /z/ contexts e.g. \*[mõ.zi.vak] for *mon ivak*. This should lead to more target-like utterances when the context (either singular or plural) would make use of the liaison consonant encountered more often in the input with a particular nonce noun. By contrast, the phonological model states that, though phonological development may still be underway, relatively abstract knowledge of liaison should



be in place by 3;0. The strong prediction from this hypothesis is that children will produce the correct exemplar in each context, regardless of the asymmetries in the input.

**4.2. PARTICIPANTS.** Participants were 24 monolingual French-learning children aged 3;0-3;8 (M=3;2). An additional 12 children were run but excluded due to bilingualism or knowledge of non-hexagonal French (3), experimenter error (1), caregiver providing noun forms in isolation (3), fussiness or failure to complete the task (4), or caregiver-reported language delay (1). In addition, 24 adult monolingual French-speaking controls (18-45 years old) participated in the study. Adult controls were recruited through the RISC (Relais d'Information sur les Sciences de la Cognition) network and received €10 for their participation.

**4.3. MATERIALS.** Two vowel-initial nonce nouns were used as test stimuli: *oupel* [upel] and *ivak* [ivak]. The nonce noun *ivak* originates from Chevrot et al. (2009). We chose the form *oupel* to provide sufficient contrast in terms of vowel height and rounding. Both nonce nouns avoid word-medial /z/, /n/, and /t/ to prevent ambiguous cases of reduplication (as in [za.zo] for *oiseau* discussed above). Importantly, neither nonce word is likely to be mistaken for a known word; according to Lexique (New et al. 2001) *ivak* has no phonological neighbors, while *oupel* has 2 verb neighbors *appelle* ‘call’ and *épèle* ‘spell out a word’, the latter not likely known by 3-year-olds.

For familiarization, the two nonce nouns were embedded in noun phrases which included a determiner or a determiner plus an adjective. The word immediately preceding the noun, the  $W_1$ , could trigger liaison or not. The set of  $W_1$ s included: three /n/-liaison-inducing prenominal elements (*un* ‘one’, *mon* ‘my-sg’, *bon* ‘good-sg’), three /z/-liaison-inducing prenominal elements (*les* ‘the-pl’, *mes* ‘my-pl’, *bons* ‘good-pl’), one /t/-liaison adjective (*petit* ‘small’), and one adjective that did not trigger liaison (*vrai* ‘real’). For testing, the noun phrases included one /n/-liaison-triggering determiner (*un* ‘one’), one /z/-liaison-triggering determiner (*deux* ‘two’) and one prenominal adjective which does not trigger liaison (*joli* ‘pretty’).<sup>13</sup>

Familiarization trials were presented using videos of a cartoon female who presented objects to the children (Figure 6). Accompanying audio stimuli were recorded by a female native

French speaker using Cool Edit Pro in an IAC acoustic chamber at 4410 Hz sampling frequency with a bit rate of 16 bits. Audio stimuli for the testing phase were recorded by a male native French speaker under the same acoustic conditions.

**4.4. PROCEDURE.** The experiment consisted of three phases: training with consonant-initial nouns, familiarization with vowel-initial nonce nouns, and testing with those nonce nouns.

**TRAINING.** During the training phase children were familiarized with the procedure that would be used for the following two phases. Each child was exposed to up to 4 familiar objects whose labels begin with consonants that are not typical LCs: *ballon* ‘ball’, *cheval* ‘horse’, *poisson* ‘fish’, *vélo* ‘bicycle’. Each object moved across the screen accompanied by pre-recorded sentences spoken by a female voice. These sentences were always of the form:

- (11) *C’est un [noun]. Voici le vrai [noun].*  
 ‘This is a [noun]. Here is the real [noun].’

After this initial presentation, the object reappeared in the center of the screen accompanied by the following pre-recorded prompt, designed to elicit the noun:

- (12) *Maintenant à toi, dis-moi sur cette image il y a un joli...*  
 ‘Now it’s your turn, tell me, in this image there is a pretty...’

If the child did not respond within ten seconds a second prompt followed:

- (13) *Dis-moi, sur cette image il y a un joli...*  
 ‘Tell me, in this image there is a pretty...’

If again, there was no response the third and last prompt began another ten seconds later:

(14) Alors, qu'est-ce qu'il y a sur cette image?

'So tell me, what is in this image?'

Adult controls completed a shortened version of this training phase in which they received a single practice trial with *ballon* 'ball'. After children completed at least two correct practice trials with known consonant-initial nouns, they completed a final round of practice using consonant-initial nonce nouns (bipote [bipot], mefype [mefyp]).

FAMILIARIZATION. After training, participants completed up to two rounds of familiarization per nonce noun. In each round, eight carrier sentences similar to (11) above were presented. Crucially, the carrier sentences themselves consisted of DET N or DET ADJ N according to the distributions described in section 4.1 above. One round presented the first nonce noun in the singular-biased condition (5 instances in an /n/ liaison-triggering context, and 1 in each of /z/, /t/, and non-liaison contexts), the other presented the second nonce noun in the plural-biased condition (5 /z/, 1/n/, 1 /t/, and 1 non-liaison context). The conditions for each noun were counterbalanced: half of the participants heard *ivak* in the singular-biased condition, and *oupel* in the plural-biased condition, and half heard the opposite. The condition presented first was also counterbalanced across participants.

TESTING. The testing phase immediately followed familiarization, and was similar to what children were asked to do during training (12-14). The testing procedure used—designed to elicit the noun—was intended to limit productions to those which are most relevant for investigating knowledge of liaison. It was based on the observation that children in the Lyon Corpus spontaneously provided missing liaison in contexts where they were completing a phrase uttered by a parent or caregiver. At the start of each trial, one of the nonce objects appeared in the center accompanied by a pre-recorded, male voice which prompted the participant to complete a short phrase. The last word of the incomplete phrase triggered an /n/ liaison, a /z/ liaison, or no liaison. There were two trials for each of these three contexts, resulting in 6 trials per child. If the child failed to complete the first prompt, a second prompt played 10 seconds later, and a third another

10 seconds later if necessary. In the event that the child failed to answer after the three prompts, the experimenter repeated the prompt and/or asked the child to describe what she saw.<sup>14</sup> Examples of the prompts are provided in (15), where the three possible contexts are in curly brackets. For 6 out of 24 children only a single round of familiarization was necessary to elicit responses at test. In the event that the child did not answer all three test prompts the first time around, the familiarization was repeated before the test prompts were played again. This was the case in 18 of the 24 children.

<INSERT FIGURE 6 ABOUT HERE>

(15) a. *O regarde! Maintenant à toi! Dis-moi! Sur cette image il y a {un, deux, un joli}...*

‘Oh look! Now your turn! Tell me! In this image there is {a, two, a pretty}...

b. *Dis-moi, sur cette image il y a {un, deux, un joli}...*

‘Tell me, in this image there is {a, two, a pretty}...

c. *Alors, qu’est-ce qu’il y a sur cette image?*

‘So tell me, what is in the image?’

**4.5. CODING.** The experiment procedure was designed to elicit productions of the two nonce nouns during the testing phase. However, children often produced these forms during the familiarization phase, or before a test prompt, and in some cases directed them at their caretaker rather than the experimenter. In order to maximize the amount of data for each child, we therefore considered a subset of these somewhat more spontaneous responses in our analyses. In particular, we included any responses that were made *after* the child had been exposed to the full distribution of contexts for each nonce form (i.e. 5 instances of a nonce noun in its dominant context, and 1 in each of three other contexts). This included the testing phase and the second repetition of the familiarization phase (for those children who heard it twice). Importantly, we discarded responses made during familiarization which were exact repetitions of prompts. For

example, if the child heard a prompt containing *vrai ivak* [vrɛ.i.vak] and proceeded to say [vrɛ.i.vak], this was considered a repetition. However, if in this instance the child produced *un ivak*, *deux ivak*, *joli ivak*, etc., this did not constitute a repetition.

Adult participants generally answered the pre-recorded prompts upon request. When an adult participant provided an incorrect isolated noun completion to the pre-recorded prompt (e.g. [ivak] to complete the prompt *sur cette image il y a deux...* ‘in this image there are two...’), the experimenter would repeat the prompt a second time. For both adults and children, responses included both isolated nouns (e.g. [zi.vak]) and phrasal productions (e.g. [dø.zi.vak]), either in response to a prompt or spontaneously. Both types were used in the analyses, however we discuss distinctions between the two types of responses below. The intended number (singular or plural) of each utterance was either determined by context in a response phrase (e.g. a response *un ivak* was coded as singular, and *deux ivaks* as plural), or by the context in the preceding phrase (e.g. the prompt *sur cette image il y a deux...* with a noun only response coded as plural). Noun only responses for which the context number could not be determined were discarded (3% of child responses, 0.5% of adult responses).

**4.6. RESULTS: OVERALL ACCURACY AND ERROR TYPES.** Across all contexts and input conditions, the mean accuracy level for adults in this task was 75% (SE = 0.04), and for children 57% (SE = 0.03). Generally then, adults did not perform at ceiling, however their responses were nevertheless more accurate than children’s. Recall that the contexts at test involved either a liaison-triggering  $W_1$  which could be singular or plural (*un*, *deux*) or a non-liaison-triggering  $W_1$  which was necessarily singular (*joli*). Figure 7 shows response types for non-liaison triggering contexts, including correct responses and insertions (use of a spurious liaison consonant in a non-liaison-triggering context). Figure 8 shows performance in liaison-triggering contexts, for each context number by input condition combination. Response types here include correct responses, replacements (use of the wrong liaison consonant), and omissions (missing liaison consonant where one is required). Importantly, the difference between children and adults in non-liaison-triggering contexts appears to be much more dramatic than in liaison-triggering contexts.

To confirm this, accuracy in singular trials (where both liaison- and non-liaison-triggering contexts are present) was modeled using mixed-effects logistic regression with age (child or adult), context type (liaison-triggering or not), and input condition (singular or plural bias) as fixed effects, and participant as a random effect.<sup>15</sup> A significant main effect of age ( $\beta = -0.80 \pm 0.23$ ,  $p < .001$ ) and a significant interaction between age and context type ( $\beta = 0.95 \pm 0.20$ ,  $p < .001$ ) were found, confirming that children produce fewer correct responses overall, and that the difference between the age groups is greater in non-liaison-triggering contexts. What drives the latter effect is suggested by the error types shown in Figure 7 and Figure 8; children's errors are equally likely to involve replacement, omission, or insertion (27 of each type in fact). By contrast, adults' errors almost exclusively consist of omissions (45 out of a total of 50 errors). Adults are thus almost at ceiling in contexts where no liaison is in fact required, while children clearly make insertion errors in these context. We return to this below, after first evaluating the effect of input distribution on accuracy.

<INSERT FIGURE 7 ABOUT HERE>

The constructionist hypothesis predicts that liaison accuracy should be affected by the contexts in which a noun is found in the input. Here, this means that accuracy in singular /n/ liaison-triggering contexts should be higher for singular biased nouns since the latter were heard with /n/ in the input much more frequently. Similarly, accuracy in plural /z/ liaison-triggering contexts should be higher for plural biased nouns. The opposite should also hold: accuracy should be relatively low in singular /n/ liaison-triggering contexts for plural biased nouns as well as plural /z/ liaison-triggering contexts for singular biased nouns. These predictions appear to be borne out numerically for both children and adults; as shown in Figure 8, there is a trend in the right direction for both singular and plural contexts. To assess this effect statistically, a mixed effects model of accuracy in liaison-triggering contexts (trials with *un* and *deux*) was run with age, context number, and bias condition as fixed effects, and participant as a random effect. A significant interaction between context number and bias condition ( $\beta = 0.37 \pm 0.14$ ,  $p = .008$ ) was found, confirming that for both age groups, accuracy in a particular context number changes

depending on the bias condition of the noun. Accuracy decreases in liaison-triggering contexts that do not match the most frequent context heard in the input. In addition, a significant interaction between context number and age was found ( $\beta = 0.36 \pm 0.14$ ,  $p = .01$ ), reflecting children's relatively poor performance in plural contexts, regardless of bias condition. Two additional models were run to determine whether the effect of bias condition was driven primarily by singular or plural context trials. This analysis revealed a significant effect of bias condition only for plural trials ( $\beta = -0.49 \pm 0.22$ ,  $p = .03$ ).<sup>16</sup>

<INSERT FIGURE 8 ABOUT HERE>

A related prediction made by the constructionist model is that nouns heard more often in singular contexts should be more likely to trigger errors involving mis-use of /n/, while those heard more often in plural contexts would trigger /z/ errors. As Figure 9 shows, children indeed showed a tendency to overuse initial /n/ for nonce nouns they heard more in singular contexts (e.g. *deux* [ni.vak]), and to overuse /z/ for nouns they heard more in the plural (e.g. *un* [zi.vak]). Such errors were made both in environments triggering liaison—replacement errors—and those not triggering liaison—insertion errors (e.g. *joli* [ni.vak])—suggesting that segmentation of the nonce words was problematic for children. A single dependent variable was created to code whether a particular /n/ or /z/ error involved using the consonant heard most in the input. A mixed effects model predicting this aggregate variable from bias condition (singular or plural biased) revealed no significant main effect of bias condition ( $\beta = -0.57 \pm 0.62$ ,  $p = .36$ ). However the intercept term was significant, indicating that errors using the liaison consonant heard most in the input were indeed more likely than those which used the alternative consonant ( $\alpha = 2.75 \pm 1.20$ ,  $p = .02$ ). As we have seen, adults overwhelmingly made errors of omission, and thus very few such /n/ or /z/ liaison errors were made (4 total, though all in the direction predicted by the input distribution).

The results of these two analyses confirm that at age 3;0, children's productions are sensitive to input frequency. This is shown both in their accuracy rates (higher in contexts which follow the input bias, particularly in the plural), and in the types of errors they make (involving the LC

heard most frequently in the input). Interestingly, the former effect was shown in the adult data as well (although it is numerically weaker). In order to better quantify the RELATIVE contributions of input frequency and correct target-like representation in explaining the child and adult data, below we discuss a simple probabilistic model of liaison production.

<INSERT FIGURE 9 ABOUT HERE>

However, before we turn to our proposed model of liaison production, we discuss the somewhat surprising performance of our adult controls. To summarize: in contexts which do *not* trigger liaison (e.g. ...*il y a un joli*... ‘...there is a pretty...’) adults are near ceiling. Where they appear to have trouble in this task is in contexts which DO trigger liaison. In these contexts, adults overwhelmingly commit errors of omission. Thus while the accuracy level of adults is similar to that of children in liaison-triggering contexts, the underlying source of the errors is at least partially distinct. Children in our experiment exhibit clear evidence of a strong constraint to fill onsets, potentially indicating trouble with segmentation or strong constraints on producing C-initial syllables, by contrasts adults regularly produce vowel-initial tokens—even in liaison-triggering contexts, suggesting they have correctly segmented the nonce nouns.

What this suggests is that adults are to some extent unwilling to use liaison with the novel forms they were taught. This is likely related to the nature of the prompts at test<sup>17</sup>; our task involved sentence (or phrase) completion, which successfully limits the class of responses participants can give, but necessarily involves a prosodic break following the last word of the prompt. Although the desired response involved liaison completion where appropriate, (e.g. prompt: *sur cette image il y a un*... response: [ni.vak]), the prosodic break may have encouraged the use of the segmented citation form. The fact that adult and child participants sometimes produced a full phrase rather than the noun alone in their completions allows us to evaluate this possible explanation concretely.

<INSERT FIGURE 10 ABOUT HERE>



In fact as Figure 10 illustrates, for liaison-triggering prompts, adults were more likely to use (correct) liaison when they provided a full phrase compared to noun only completions ( $\beta = 1.41 \pm 0.51$ ,  $p = .006$ ). This was true of children as well, who were much more likely to produce the correct liaison form in a full phrase—in fact their accuracy reaches adult-like levels in this context. This suggests that the prosodic break introduced for the completion task encouraged more use of the vowel-initial citation form, particularly for adults, but to some extent for children as well. While it remains true that adults and children exhibit distinct patterns of errors in both phrase and noun only completions, the clear jump in accuracy for phrasal responses suggests that the results of this study may still present a somewhat conservative picture of liaison productivity. Importantly, the effect of the input distribution on accuracy appears to be found independent of response type; a model of children's accuracy including context number, bias condition, and response type (noun only or phrase) as fixed effects revealed a significant main effect of response type ( $\beta = -0.62 \pm 0.22$ ,  $p = .005$ ), and a significant interaction between context number and bias condition ( $\beta = 0.49 \pm 0.21$ ,  $p = .03$ ), but no other main effects or interactions. The amount of data we have to assess any difference between response types is, however, small (recall that the study was designed to elicit 6 responses from each child), as many child do not provide both response types.

**4.7. QUANTIFYING THE CONTRIBUTION OF INPUT FREQUENCY.** The results reported above suggest that input frequency influences children's liaison productions for newly learned words. However, in order to more precisely quantify the relative contributions of input frequency and target-like knowledge, we constructed two simplified models of liaison production. Figure 11 illustrates both models schematically. For any given target, each model assigns some probability to that target and to each other form consistent with the input.<sup>18</sup> The input frequency model assigns those probabilities according to their frequency in the input. Thus, for example, if a nonce word was plural biased in the input for a particular child, then the probability of the /z/-initial exemplar is 5/8 or 0.625. The probability of any other response is 1/6 or 0.125. By contrast, the target representation model assigns probability 1 to forms which match the target grammar, regardless of the input, and assigns probability 0 to any non-target-like forms.

In order to determine the contribution of each model in explaining adult and child productions, we ran a general-purpose optimization algorithm to find the weighted mixture of the two models which assigns the highest likelihood to each set of data. As shown in Figure 11, the best-fitting mixture for the child productions assigned a weight of 42% to the target representation model, and 58% to the input frequency model. The best-fitting mixture for the adults essentially reversed these two, giving a weight of 64% to the target representation grammar and 36% to the input frequency grammar.

As discussed above, there is reason to believe that the prosodic break introduced by our test prompts may have encouraged participants to omit liaison (and produce the citation form). However, the model as stated has no way of accounting for these types of responses other than to assign them a small probability using the input frequency model—they are always assigned zero probability according to the target representation model. However, if the latter is supposed to reflect adult-like behavior in some sense, then this part of the model should have a way of choosing to omit liaison in noun only responses. In order to capture this, we incorporated a probability of 0.4 assigned by the target representation model to noun only responses, estimated from the rate at which adults in our task omitted liaison in such responses. For the child data, this resulted in a mixture weight of 53% for the target representation model and 47% for the input frequency model. For adults, the weights were 86% and 14% respectively.

<INSERT FIGURE 11 ABOUT HERE>

**4.8. DISCUSSION.** Study 2 was designed to investigate the extent to which 3-year-old French-learning children are able to correctly produce novel V-initial nouns in liaison and non-liaison contexts after only a short exposure period. Recall that in the experiment reported here, familiarization consisted of a total of 8 tokens of each novel word (or 16 if a child was given a second round of familiarization). Only 1/8 of these tokens provided the segmented vowel-initial form. From this type of token, plus evidence for segmentation from the variation of liaison consonants, participants had to learn the correspondence between an unfamiliar object and its new label, and correctly infer the underlying form of the label. Our results suggest that at 3-years

of age, French-learning children can to some extent successfully segment and use liaison after this very brief experience with a new noun. Indeed, children's accuracy rates for responses including a full phrase (e.g. *deux ivak* as opposed *ivak* alone) approach those of adults. Nevertheless, the major source of children's errors appears to be segmentation of the vowel-initial noun; the analysis of errors provided above shows that children clearly differ from adults in their ability to provide vowel-initial productions in non-liaison-triggering contexts. In addition, of their error types, approximately 2/3 involved replacements or insertions. Thus, there is evidence of a strong constraint to fill the onset in productions, which is influenced by the particular distribution of onsets a noun appears with in the input. A simple model quantifying the relative contributions of input frequency (of liaison consonants with each nonce noun) and target-like representation, resulted in a mixture favoring the former over the latter. These relative contributions swapped when the target model incorporated the possibility—apparently used by adults in our task—of omitting liaison following a prosodic break. Interestingly, several children produced forms reminiscent of the types of idiosyncratic onset-filling strategies found in the corpus study reported here. For example, one child consistently used the form *loupel* for *oupel*, using an onset which is not a possible nominal liaison consonant, much less a possible segmentation of an input form. This suggests that, to some extent, the same liaison-independent features of phonological development may be at play in the experiment.

**5. GENERAL DISCUSSION.** In this paper, we have presented two new sources of evidence—a corpus study of children's spontaneous speech and an elicitation experiment with novel nouns—which shed light on French-acquiring children's early representations of liaison. We framed our discussion throughout in terms of two competing models of liaison development. The first model is couched in a constructionist framework and is exemplified in the work of Dugua (2006), Chevrot et al. (2009), Dugua et al. (2009), and Chevrot et al. (2013). This model predicts relatively late mastery of liaison (age 5 or older), and heavy dependence on the contexts in which a particular noun appears in the input. The second model, put forward by Wauquier-Gravelines & Braud (2005), takes an approach to liaison development which integrates it more closely with general phonological development, and predicts relatively early mastery (by age 3).

The corpus study reported here analyzed the vowel-initial noun productions of three children in the Lyon corpus from onset of first words to age 3 years. We focused both on age of mastery, and on the types of errors predicted by both models of liaison acquisition. The results reveal that by age 3 the children are correctly producing liaison in the nominal domain over 85% of the time. We also briefly presented data on liaison in the verbal domain (between a subject clitic and a following vowel-initial verb), which reveal near perfect performance by age 2;5. In terms of error types, those predicted by a constructionist model of liaison acquisition—insertions (use of a liaison consonant in non-liaison-triggering contexts), replacements (use of the wrong liaison consonant), and omissions (missing liaison consonants)—were relatively rare. By contrast, errors of the type predicted by a phonological model—idiosyncratically filled onsets, and reduplication or truncation errors—were more frequent.

Our corpus results highlight the importance of factoring phonological development into the analysis of apparent liaison errors. For example, our data suggest that glide-initial nouns present problems for children that are independent of liaison. Further, patterns of reduplication and truncation, which can be seen in liaison contexts, are also clearly found outside of them. For example, analysis of one of the children in the Lyon corpus, Théotime, revealed that in definite determiner + consonant-initial noun contexts (e.g. *le livre* ‘the book’), such errors occurred in comparable frequencies and patterned similarly to vowel-initial nouns. Overall, truncation errors were found at a rate of 18.4% (113/615) comparable to the 25.2 % (56/222) vowel-initial truncation rate across the three children. Reduplications occurred at a rate of 2.9% (18/615) for consonant initial forms, similar to the 6.3% (14/222) found for vowel initial nouns in the corpus. While the frequencies across these two contexts are not identical, it is clear that base-rates must be taken into account in order to determine the extent to which errors reflect general processes of phonological acquisition or liaison-specific processes.

The elicitation task reported here investigated 3-year-olds’ ability to apply their knowledge of liaison to new nouns following brief exposure. The design of the task was motivated by some potential limitations of previous liaison production experiments. For example, previous research has tested known words, including several which we have argued to be independently problematic (e.g. glide-initial nouns and *ours* ‘bear’). Our experiment introduced children to two

nonce nouns in both liaison and non-liaison-triggering contexts. Crucially, each noun appeared in an asymmetrical distribution of contexts. One noun appeared in singular contexts triggering /n/ liaison most often, and the other appeared most often in plural contexts triggering /z/ liaison. This allowed us to test whether children's subsequent liaison productions were influenced by input frequency as predicted by the constructionist model. The results suggest that early stages of learning are indeed influenced by frequency; children's insertion and replacement errors reflected the input distribution. In addition, while adult controls overwhelmingly made omission errors, children's errors clearly reflected a strong constraint to fill onsets in their productions. This resulted in both replacement and insertion errors of the type predicted by the constructionist account. Nevertheless, as our computational model confirms, children at this age are also making significant use of target-like knowledge in their productions of novel vowel-initial nouns.

Taken together, our findings suggest that by age 3 children are well on their way to adult-like representations of liaison. The relatively low frequency of errors clearly related to liaison (e.g. replacement, insertion, and omission errors), and early signs of mastery observed in our corpus study support this claim. While these types of errors—dictated to some extent by input frequency—were made by children in our elicitation task, we would argue that this is likely a feature of very early word learning. The constructionist model of liaison therefore largely underestimates children's progress by age 3. A phonologically-based model, incorporating the effect of distributional context on early errors, provides a better overall fit to the data we have presented. More generally, we believe our results support the need for explicit formal models of the acquisition of complex processes like liaison, tested on multiple data sources, incorporating insights from usage-based and generative frameworks. This approach has the potential to further one of the major research goals in the field, namely understanding the trade-off between distributional cues and abstract knowledge in early language development. While here we have focused on production data, the models of development under discussion make clear predictions for perception or comprehension as well. In particular, both models make specific claims about developmental changes in the underlying representations of liaison. These are used to explain patterns of production errors, but should also be reflected in children's sensitivity to or perception of forms they hear. Integrating such additional sources of data is a clear step for

future research.

## Appendix.

Frequency distribution, W1 exemplars, and example sentences for familiarization phase of elicitation experiment.

Familiarization	LC Distribution	Word 1s	Example familiarization
<b>/n/ condition</b>	/n/ (x5) /ø/, /t/, /z/ (x1)		1. <i>C'est un ivak.</i> [œ.ni.vak] This is an ivak.
			2. <i>Quel bon ivak!</i> [bo.ni.vak] What a good ivak.
		<i>un</i> [œ]	3. <i>Tu vois les ivaks?</i> [le.zi.vak] Do you see the ivak ?
		<i>mon</i> [mõ]	4. <i>Regarde ce que fait mon ivak.</i> [mo.ni.vak] Watch what my ivak is doing.
		<i>bon</i> [bõ]	5. <i>Il est où le petit ivak?</i> [pø.ti.ti.vak] Where is the little ivak ?
		<i>petit</i> [pøti]	6. <i>Regarde bien le bon ivak.</i> [bo.ni.vak] Carefully watch the good ivak
		<i>les</i> [le]	7. <i>Il y a un vrai ivak ici.</i> [vrø.ivak] There is a real ivak ici.
		<i>vrai</i> [vrø]	8. <i>Voici mon ivak.</i> [mo.ni.vak]
<b>/z/ condition</b>	/z/ (x5) /ø/, /t/, /n/ (x1)	<i>les</i> [le]	1. <i>Ce sont des ivaks.</i> [de.zi.vak] There are ivaks.
		<i>mes</i> [me]	2. <i>Tu vois les ivaks?</i> [le.zi.vak] Do you see the ivaks ?
		<i>des</i> [de]	3. <i>Quel bon ivak!</i> [bo.ni.vak] What a good ivak !
		<i>petit</i> [pøti]	4. <i>Regarde ce que font mes ivaks.</i> [me.zi.vak] Look at what my ivaks are doing.
		<i>bon</i> [bõ]	
		<i>vrai</i> [vrø]	

---

5. *Il est où le petit ivak?* [pø.ti.ti.vak]

Where is the little ivak ?

6. *Regarde bien les ivaks.* [**le.zi.vak**]

Watch carefully the ivaks.

7. *Il y a un vrai ivak ici.* [vrɛ.ivak]

There is a real ivak here.

8. *Voici mes ivaks.* [**me.zi.vak**]

Here are my ivaks.

---



## References

- Ahmad, M. 1993. *Vingt heures de français parlé...Aspects phonétiques de la liaison*. Thèse de doctorat, Université Stendhal, Grenoble.
- Allen, G., and S Hawkins. 1978. The development of phonological rhythm. In *Syllables and segments*, ed. by Allan Bell and Joan B. Hooper, 173-185. North Holland: Amsterdam.
- Allen, G., and S Hawkins. 1980. Phonological rhythm: Definition and development. In *Child Phonology*, ed. by Grace H. Yeni-Komshian, James F. Kavanagh, and Charles A. Ferguson, 1, 227-256.
- Babineau, M. and R. Shi. 2011. Processing of French liaisons in toddlers. In Danis, N., Mesh, K., and Sung, H., editors, *Proceedings of the 35th Annual Boston University Conference on Language Development*, pages 25–37. Cascadilla Press, Somerville, MA.
- Babineau, M., and R. Shi. 2013. Acoustical cues versus top-down bias in infants' parsing. In *Proceedings of Meetings on Acoustics* (Vol. 19, No. 1, p. 060073). Acoustical Society of America.
- Basset, B. 2000. La liaison à 3, 7 et 11 ans: description et acquisition. Grenoble, France: Université de Grenoble Masters Thesis.
- Battye, A.; M.-A. Hintze; and P. Rowlett. 2000. The French language today: a linguistic introduction. Routledge, New York.
- Bates, D. 2010. lme4: Mixed-effects modeling with R. <http://lme4.r-forge.r-project.org/book/>
- Boë & Turbach. 1992. De A à Zut: dictionnaire phonétique du français parlé. Grenoble: Ellug.
- Bortfeld, H.; J. Morgan, R. Golinkoff; and K. Rathbun. 2005. Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science*, 16(4). 298-304.
- Braud, V. 1998. Acquisition de l'information phonologique: exemple de la liaison. Nantes, France: Université de Nantes Masters Thesis.
- Braud, V. 2003. Acquisition de la prosodie chez des enfants francophones: les phénomènes de Troncations. Nantes, France: Université de Nantes Doctoral dissertation.
- Chabanal, D. 2003. Un aspect de l'acquisition du français oral: la variation socio-phonétique chez l'enfant francophone. Montpellier, France: Université de Montpellier 3 Doctoral

dissertation.

- Chevrot, J.P., Chabanal, D. and Dugua, C. 2007. Pour un modèle de l'acquisition des liaisons basé sur l'usage: trois études de cas. *Journal of French Language Studies* 17(01). 103-128.
- Chevrot, J-P.; C. Dugua; and M. Fayol. 2009. Liaison acquisition, segmentation and construction in French: a usage based account. *Journal of Child Language*. 36: 557-596.
- Chevrot, J.-P., Dugua, C., Harnois-Delpiano, M., Siccardi, A., and Spinelli, E. (2013). Liaison acquisition: debates, critical issues, future research. *Language Sciences*, 39:83 – 94.
- Chevrot, J.P. and Fayol, M., 2001. Acquisition of French liaison and related child errors. In *Proceedings of the 8th Conference of the International Association for the Study of Child Language*. 760-774.
- Côté M.-H. 2005. Le statut lexical des consonnes de liaison, *Langages* 158, 66-78.
- Côté, M.-H. 2011. French liaison. *Companion to phonology* (ed.) by Marc van Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice, 2685-2710. Malden: Wiley-Blackwell.
- Demuth, K., 1994. On the underspecification of functional categories in early grammars. *Syntactic theory and first language acquisition: Cross-linguistic perspectives* 1.119-135.
- Demuth K., and M. Johnson. 2003. Truncation to subminimal words in Early French. *The Canadian Journal of Linguistics/ La Revue Canadienne de Linguistique*. 211 241.
- Demuth, K. and A. Tremblay. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, 34, 1-29.
- Dugua, C. (2006). *Liaison, segmentation lexicale et schémas syntaxiques entre 2 et 6 ans. Un modèle développemental basé sur l'usage*. Thèse de doctorat. Grenoble: Université Stendhal Grenoble 3.
- Dugua, C.; E. Spinelli; J.P. Chevrot; and M. Fayol. 2009. Usage-based account of the acquisition of Liaison: Evidence from sensitivity to the singular/plural orientation of nouns. *Journal of Experimental Child Psychology*. 102: 342-350.
- Durand, J., Laks, B., Calderone, B., & Tchobanov, A. 2011. Que savons-nous de la liaison

- aujourd'hui?. *Langue française* 1. 103-135.
- Durand, J. and C Lyche. 2008. French liaison in the light of corpus data. *Journal of French Language Studies*, 18(01):33–66.
- Echols, C., and E. Newport. 1992. The role of stress and position in determining first words. *Language Acquisition*, 2. 189-220.
- Echols, C. 1993. A perceptually based model of children's earliest productions. *Cognition*, 46. 245-296.
- Gaskell, M. G., Spinelli, E., & Meunier, F. 2002. Perception of resyllabification in French. *Memory & cognition*, 30(5), 798-810.
- Gerken, L. 1994. A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language* 21, 565-584.
- Gerken, L. 2010. Infants use rational decision criteria for choosing among models of their input. *Cognition*, 115(2). 362-366.
- Jusczyk, P. W. and R. N. Aslin. 1995. Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, 29, (1), 1-23.
- Johnson, E. K., and P.W. Jusczyk. 2001. Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44, 548- 567.
- Johnson, E. K.; P.W. Jusczyk; A. Cutler; and D. Norris. 2003. Lexical viability constraints on speech segmentation by infants without a lexicon. *Cognitive Psychology*, 46, 65-97.
- Kehoe, M. 2000. Truncation Without Shape Constraints: The Latter Stages of Prosodic Acquisition. *Language Acquisition* Vol. 8, Iss. 1.
- Legendre, G., I. Barrière; L. Goyet; and T. Nazzi. 2010. Comprehension of infrequent subject-verb agreement forms: Evidence from French-learning children. *Child Development*. 81:6, 1859–1875.
- Legendre, G.; J. Culbertson; I. Barrière; T. Nazzi; and L. Goyet. 2010. Experimental and empirical evidence for the status and acquisition of subject clitics and agreement marking in adult and child Spoken French. In Torrens, V., Escobar, L., Gavarro, A., J.G. Mangado (eds.) *Movement and Clitics*. Newcastle: Cambridge Scholars Publishing.
- Lehmann, E. L. 1986. *Testing Statistical Hypotheses*. Wiley, New York, NY, 2nd edition.

- MacWhinney, B. 2000. *The CHILDES Project: Tools for analyzing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mattys, S. L., and P.W. Jusczyk. 2001. Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, 27 (3), 644-655.
- McCarthy, J. J., and A. Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. In M. González (Ed.), *Proceedings of the North East Linguistics Society* 24 (333-379). Amherst, MA: GLSA.
- Morin, Y.C. 2003. Remarks on pronominal Liaison consonants in French. In Stefan Ploch (ed.) *Living on the Edge: 28 Papers in Honour of Jonathan Kaye* pp 385-400. Berlin: Mouton de Gruyter.
- Morin, Y.C. 2005. La Liaison relève-t-elle d'une tendance à éviter les hiatus? Réflexions sur son évolution historique. *Langages* 158. 8-23.
- Morin, Y.C., and J. Kaye. 1982. The Syntactic base for French Liaison. *Journal of Linguistics* 18. 291-330.
- Nazzi, T., Dilley, L.C., Jusczyk, A.M., Shattuck-Hufnagel, S. and Jusczyk, P.W., 2005. English learning infants' segmentation of verbs from fluent speech. *Language and Speech* 48(3). 279-298.
- Nazzi, T.; G Iakimova; J. Bertoncini; S. Fredonie; and C. Alcantara. 2006. Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54 (3), 283-299.
- New, B., C. Pallier; L. Ferrand; and R. Matos. 2001. Une base de données lexicales du français contemporain sur internet: LEXIQUE. *Année Psychologique*, 101:447-462.  
<http://www.lexique.org>.
- Norris, D.; J.M. McQueen; A. Cutler; and S. Butterfield. 1997. The possible word-constraint in the segmentation of continuous speech. *Cognitive Psychology* 34 pp191-243
- Pater, Joe. 1997. Minimal Violation and Phonological Development. *Language Acquisition* 6. 201-253.
- Prince, A., and P. Smolensky. 1993/2004. Optimality Theory: Constraint interaction in

- generative grammar. John Wiley and Sons.
- Projet Phonologie du Français Contemporain: <[www.projet-pfc.net](http://www.projet-pfc.net)>, Code du locuteur: 75clb1 (09.03.2012).
- Saffran, J. R.; E.L. Newport; and R.N. Aslin. 1996. Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4):606–621.
- Sampson, R. 2001. Liaison, nasal vowels and productivity. *Journal of French language studies*, 11(02):241–258.
- Schane, S. A. 1965. *The phonological and morphological structure of French*. PH. D., MIT, Cambridge, Massachusetts.
- Schane, S. A. 1967. L'élision et la liaison en français. *Langages*, 8, 37-59.
- Seidl, A., and E. K. Johnson. 2008. Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *Journal of Child Language*, 35, 1-24.
- Shoemaker, E. and Birdsong, D., 2008. La résolution de la liaison en français par des locuteurs natifs et non-natifs. *Acquisition et interaction en langue étrangère* 27. 43-62.
- Spinelli, E., McQueen, J.M. and Cutler, A., 2003. Processing resyllabified words in French. *Journal of Memory and Language* 48(2). 233-254.
- Spinelli, E. and F. Meunier. 2005. Le traitement cognitif de la liaison dans la reconnaissance de la parole enchaînée. *Langages* 158. 79-88.
- Stoeckel, V., & Siccardi, A. 2004. Liaison et segmentation des mots chez un enfant québécois de 44 mois-Exploitation d'un corpus dense. Grenoble, France: Université Stendhal Masters Thesis.
- Tranel, B. 1996. Exceptionality in optimality theory and final consonants in French. In Zagona, K. (Ed.), *Grammatical theory and Romance languages - Selected papers from the 25th linguistic symposium on Romance languages (LSRL XXV)* (275-291). Amsterdam/Philadelphia: John Benjamins Publishing Company
- Tranel, B. 2000. Aspects de la phonologie du français et la théorie de l'optimalité. *Langue française*, 126, 39-72.
- Tremblay, A., 2011. Learning to parse liaison-initial words: An eye-tracking study. Bilingualism: *Language and Cognition* 14(03). 257-279.

- Tremblay, A., & Spinelli, E. (2013). Segmenting liaison-initial words: the role of predictive dependencies. *Language and Cognitive Processes*, 28(8), 1093-1113.
- Wauquier-Gravelines, S. 1996. *Organisation phonologique et traitement de la parole continue*. Dissertation. Atelier national de reproduction des thèses.
- Wauquier-Gravelines, S., 2003. Du réalisme des formalisations phonologiques contemporaines: que nous apprennent les données d'acquisition?. *Langages*. 9-34.
- Wauquier-Gravelines, S. and Braud, V., 2005. Proto-déterminant et acquisition de la liaison obligatoire en français. *Langages* 2. 53-65.
- Yersin-Besson, C., and F. Grosjean. 1996. L'effet de l'enchaînement sur la reconnaissance des mots dans la parole continue. *L'Année psychologique*, 96(1), 9-30.

## Notes

<sup>1</sup> Bare nouns are typically restricted to vocative use in utterance-initial position in the adult French grammar (e.g. *Escargot, que fais-tu?* ‘Snail, what are you doing?’).

<sup>2</sup> We will represent [ʁ] and [ʀ] as [r] throughout for simplicity.

<sup>3</sup> This example illustrates *elision*, deletion of the vowel in forms like *le* and *la* to accommodate a vowel initial following word. While both elision and liaison result in resyllabification, they are typically treated as distinct phenomena (i.e. Tranel 1996, Durand et al. 2011).

<sup>4</sup> Alternative age ranges have been suggested elsewhere. For example, Chevrot et al., 2013 suggest only that children progress through these stages between the ages of 2 and 6, with liaison mastery sometime between 4-6. As an anonymous reviewer points out, estimates of the age of mastery may depend on the source of evidence. For example, Dugua (2006) and Chevrot et al. (2006) report corpus data where 4;2–5;0 is the earliest range at which 80% accuracy is achieved. Chevrot et al. (2009) report at least one elicitation task where LC errors are reported for the majority of children tested up to 5;2.

<sup>5</sup> Chevrot et al. (2013) state that at each stage, errors from subsequent stages are predicted to be possible. However, this severely restricts the predictions generated by the constructionist model. For example, contextually determined replacement errors would be predicted to only occur at Stage 3, but all other error types would be predicted at either stage. For simplicity, we evaluate the model instead by error types at each stage under the assumption that, at the very least, the model should predict fewer Stage 2 errors at Stage 3.

<sup>6</sup> The corpora include: Basset (2000), Stoeckel & Siccardi (2004), Chabanal (2003, child 1 and 2), and Dugua (2006). Four further corpora were analyzed, however we do not discuss these here as the children’s ages exceed 5 years. For a more detailed description of all the corpora see Dugua (2006, pp 111-114).

<sup>7</sup> Elision and liaison do not differ on theoretic grounds in the constructionist and phonological models: /l/ may be the onset of an exemplar in the constructionist model as well as an onset inserted into the syllabic templates in the phonological model. Though the overall accuracy rate in elision (91%, 60/66) exceeded liaison performance (74%, 454/616) adding elisions to the

liaison productions does not have a significant impact on the pattern of errors in our data given that they are much less frequent.

<sup>8</sup> These errors included forms such as [le.le.le.fã] which could have either represented a phonological error (reduplication of the initial syllable), an LC error with *le* appearing twice as the determiner (as in stuttering speech), or as an autocorrection where *le* is corrected to *l'* online by the child.

<sup>9</sup> Théotime had 13 LC errors, 52 phonological errors, and 314 correct productions. Marie had 17 LC errors, 30 phonological errors, and 128 correct productions. Anaïs had 10 LC errors, 40 phonological errors and 78 correct liaisons.

<sup>10</sup> We thank an anonymous reviewer for pointing out that one could also interpret these productions as a result of the mis-segmentation of the sequence /la. sjet/ (with the determiner *la* and the noun *siette*). The productions could then have arisen from the application of the paradigm *ma*+X, *ta*+X, *sa*+X. However, it is unclear how such a scenario would result from the child hearing other determiners such as *mon* or *ton* preceding the nouns.

<sup>11</sup> Marie produced 35 *oiseau* errors (compared to 17 LC and 30 phonological errors). Théotime produced 26 *oiseau* errors (compared to 13 LC and 52 phonological errors). Anaïs produced 10 *oiseau* errors (compared to 10 LC and 40 phonological errors).

<sup>12</sup> The total occurrences of these nouns were as follows: *oeil*: 242 (41sg/201pl), *oeuf*: 113 (37sg/76pl), *arbre*: 207 (45sg/49pl), *escargot*: 25 (19sg/6pl), *éléphant*: 128 (97sg/31pl), *avion*: 60 (47sg/13pl), *ordinateur*: 2 (2s/0pl), *ours*: 125 (107sg/18pl).

<sup>13</sup> Note that the singular determiner *un* occurs in both the familiarization and testing for the singular condition. While this introduces a potential advantage in singular testing contexts for this condition, there are relatively few candidate  $W_1$ 's for /n/ liaison, even in the adult grammar. Many /n/-final words no longer trigger liaison reliably, and of those, few are highly likely to be known by children and contextually appropriate (Sampson 2001, Durand & Lyche 2008). This is the reason why we used *un* in familiarization. Importantly though, it only appeared once (the same number of instances /z/-liaison in that condition, see Appendix for details) to minimize the likelihood that it would influence the results unduly. Moreover, in the results section, we will



report data that fails to show an advantage of singular-biased forms occurring in a singular context as compared to plural-biased forms occurring in a plural context.

<sup>14</sup>On occasion children responded to prompts provided by caregivers. In anticipation of this, before the experiment began, caregivers were instructed *not* to use the names of the nonce nouns in any speech to their child. Therefore additional prompts did not provide additional evidence about these forms, but instead were expressions such as *Comment ça s'appelle?*, 'What's that called?' *C'est quoi ça?* 'What is that?' *Il y a un/deux...* 'There is one/two...').

<sup>15</sup> All models reported here were run with the lme4 package in R (Bates 2010), and include random intercept terms for subject (due to convergence failures, random slopes were not included). All fixed effects were sum coded, reference levels were as follows: child (age), singular input condition, singular context number, liaison triggering context type. Trial number is not included in reported models, as according to standard Likelihood Ratio tests (Lehmann 1986) it did not contribute significantly to explaining the data in any case (all  $p > 0.05$ ). In addition, all analyses were also modelled using ANOVA, all effects remained the same.

<sup>16</sup> Here we can confirm that the single exposure to *un* in the singular-biased familiarization trials does not appear to have affected performance. Though children's accuracy was highest with singular responses, there is no particular advantage for *un* test trials with a singular-biased noun compared to *deux* test trials with a plural-bias noun (where no *deux* was present during familiarization). This is confirmed by model predicting accuracy on bias-supported trials (*un* following the singular-biased familiarization and *deux* following a plural-biased familiarization) from age and context number. This model failed to find a reliable difference between singular and plural contexts ( $\beta = 0.02 \pm 0.20$ ,  $p = .91$ ), and no interaction with age was found ( $\beta = 0.31 \pm 0.21$ ,  $p = .13$ ).

<sup>17</sup> Another possible contributor is a general restriction on the productivity of liaison in the adult French grammar. This comes from evidence that vowel-initial loan words tend not to trigger liaison in French (e.g. *le underground*, *le upgrade*), and orthographically h-initial loan words which are phonologically vowel-initial in French are treated as *h-aspiré*—as opposed to *h-muet*—and thus block liaison (e.g. *le hamburger*, *la hype*, Battye et al. 2000).

---

<sup>18</sup> For simplicity, we have removed productions which do not correspond to a potential target form according to the input. For example, forms like *loupel* neither match the underlying vowel-initial target, nor the target with a realized liaison consonant /z/, /n/ or /t/. Both models assign zero probability to such forms. Including these forms and incorporating a noise probability mass of 0.001 for productions otherwise assigned probability zero by each model gives similar results (43% target representation for children, 64% for adults).

## Tables

### (3) Example Stage 2/3 errors (Dugua 2006)

- |   |   |   |
|---|---|---|
| <p>a. <i>Insertion</i> (Stage 2)</p> <p>[le.fi.õ]→[ le.fi.zõ]</p> <p>‘les filles ont’</p> | <p>b. <i>Replacement</i> (Stage 2)</p> <p>[de.zarbr]→[de.narbr]</p> <p>‘des arbres’</p> | <p>c. <i>Omission</i> (Stage 2)</p> <p>[œ.narbr]→[œ.arbr]</p> <p>‘un arbre’</p> |
| <p>d. <i>Replacement</i> (Stage 3)</p> <p>[le.nyaʒ]→[le.zyaʒ]</p> <p>‘les nuages’</p>     | <p>e. <i>Insertion</i> (Stage 3)</p> <p>[a.mi]→[za.mi]</p> <p>‘un ami’</p>              |   |

## (4) Example Stage 2 errors (Wauquier-Gravelines 2003)

- |   |   |  |
|---|---|--|
| <p>a. <i>Reduplication</i></p> <p>[œ̃.na.ro.zwar]→[ro.ro.zwar]</p> <p>‘un arrosoir’</p> | <p>b. <i>Insertion</i></p> <p>[lan]→[lə.tan]</p> <p>‘l’âne’</p>   | <p>c. <i>Omission</i></p> <p>[dø.ze.le.fã]→[dø.eã]</p> <p>‘deux éléphants’</p> |
| <p>d. <i>Replacement</i></p> <p>[le.zan]→[le.tan]</p> <p>‘les ânes’</p>                 | <p>e. <i>Idiosyncratic error</i></p> <p>[lar.kã.sjɛl]→[lə.pa.ka.sjɛl]</p> <p>[lar.kã.sjɛl]→[lə.ma.ka.sjɛl]</p> <p>‘l’arc-en-ciel’</p> |  |

## (5) Example reduplications (Braud 2003)

- a. [œ.nurs]→[ε.nu.nurs]    b. [œ.na.ro.zwar]→[ε.ro.rɔ.zwar]    c. [œ.na.kɔr.de.ð]→[ε.ra.ra.jɔ]  
     ‘un ours’                      ‘un arrosoir’                      ‘un accordéon’
- d. [œ.na.spi.ra.tœr]→[ε.ra.ra.sta.tœ]                      e. [œ.nɛ.lɛ.kɔp.tɛr]→[ε.ni.ni.kɔp.tɛr]  
     ‘un aspirateur’                      ‘un hélicoptère’

(6) Example omissions (all children, 2;3-2;10)

- |    |                                   |    |                          |
|----|-----------------------------------|----|--------------------------|
| a. | [mõ.na.ni.ver.sɛr]→[mõ.a.ver.sɛr] | b. | [sõ.na.sjɛt]→[sõ.a.sjɛt] |
|    | ‘mon anniversaire’                |    | ‘son assiette’           |
| c. | [de.zi.stwar]→[də.i.stwar]        | d. | [le.zã.fã]→[le.ã.fã]     |
|    | ‘des histoires’                   |    | ‘les enfants’            |

(7) Example replacements (Marie, 2;6)

- a. [dø.zar.ʒã]→[dø.nar.ʒã]    b. [me.zar.ʒã]→[me.nar.ʒã]    c. [de.za.sjet]→[de.na.sje]  
‘deux argents’                      ‘mes argents’                      ‘des assiettes’

(8) Example truncations (all children, 1;7-2;6)

- |    |             |    |                 |    |                |    |                |
|----|-------------|----|-----------------|----|----------------|----|----------------|
| a. | [a.bɛ]→[vɛ] | b. | [a.sjɛt]→[sjɛt] | c. | [a.vjõ]→[vjõ:] | d. | [e.le.fã]→[fã] |
|    | ‘abeille’   |    | ‘assiette’      |    | ‘avion’        |    | ‘éléphant’     |



(9) Example reduplications (all children, 1;10-2;7)

- |    |                             |    |                       |
|----|-----------------------------|----|-----------------------|
| a. | [œ.ne.le.fã]→[a.te.te.fõ]   | b. | [œ.es.kar.go]→[økako] |
|    | ‘un éléphant’               |    | ‘un escargot’         |
| c. | [i.po.po.tam]→[fõ.pɔ.tɔ.ta] | d. | [a.ni.mal]→[ja.me.ma] |
|    | ‘hippopotame’               |    | ‘animal’              |

(10) Example idiosyncratic phonological processes (Marie, 2;0-2;7) [note 19]

- |    |                         |    |                        |    |                      |
|----|-------------------------|----|------------------------|----|----------------------|
| a. | [la.sjɛt]→[la.be.fɛt]   | b. | [de.zorej]→[du.lɛj]    | c. | [də.lar.zã]→[dar.zã] |
|    | ‘l’assiette’            |    | ‘des oreilles’         |    | ‘de l’argent’        |
| d. | [mõ.na.sjɛt]→[ma.sjɛ:t] | e. | [tõ.na.sjɛt]→[ta.sjɛt] |    |                      |
|    | ‘mon assiette’          |    | ‘ ton assiette’        |    |                      |

## Figures

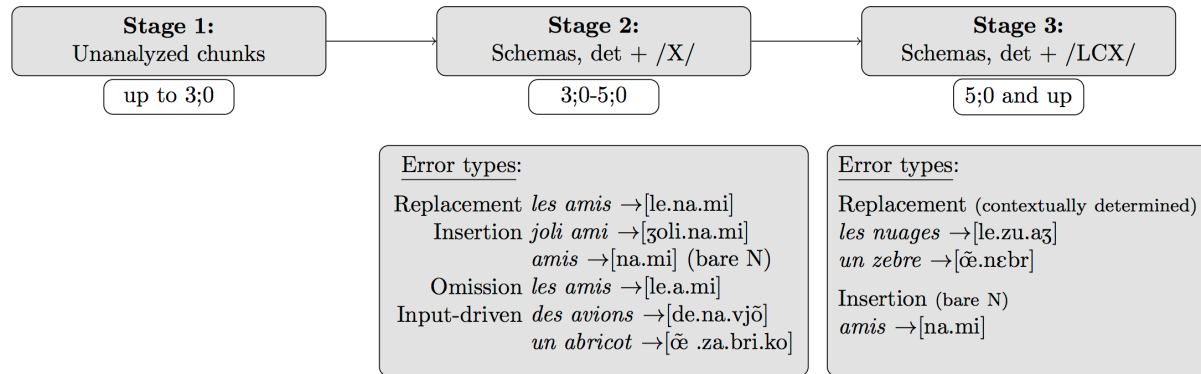


Figure 1. Predictions of the constructionist model of liaison development (Chevrot et al. 2009).

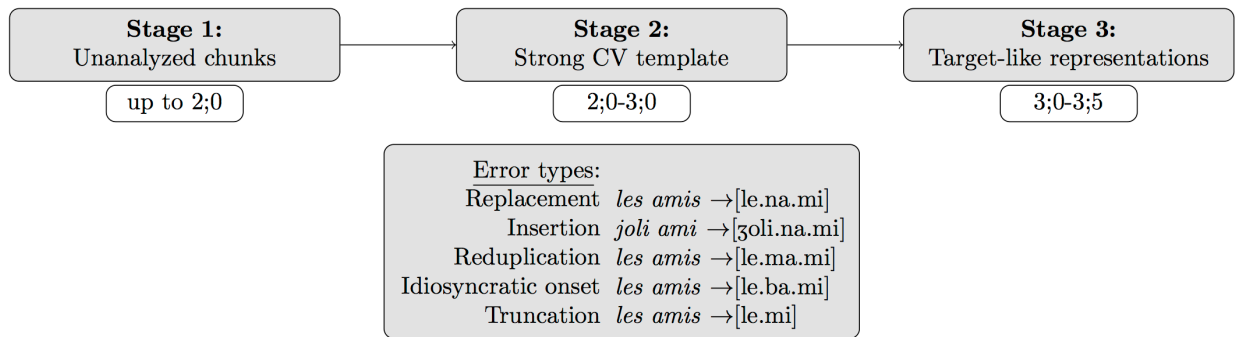


Figure 2. Age and predicted error types at each stage according to the phonological model of liaison development (based on Wauquier-Gravelines & Braud 2005)

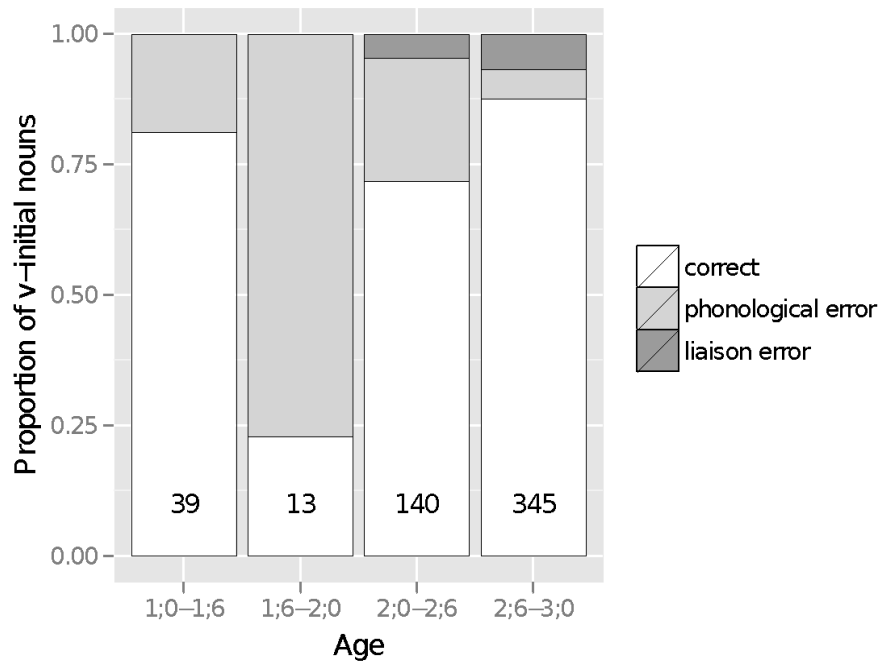


Figure 3. Proportion of each response type for vowel-initial noun targets in Lyon Corpus by binned age. Raw frequency of correct targets is also provided numerically for each age bin.

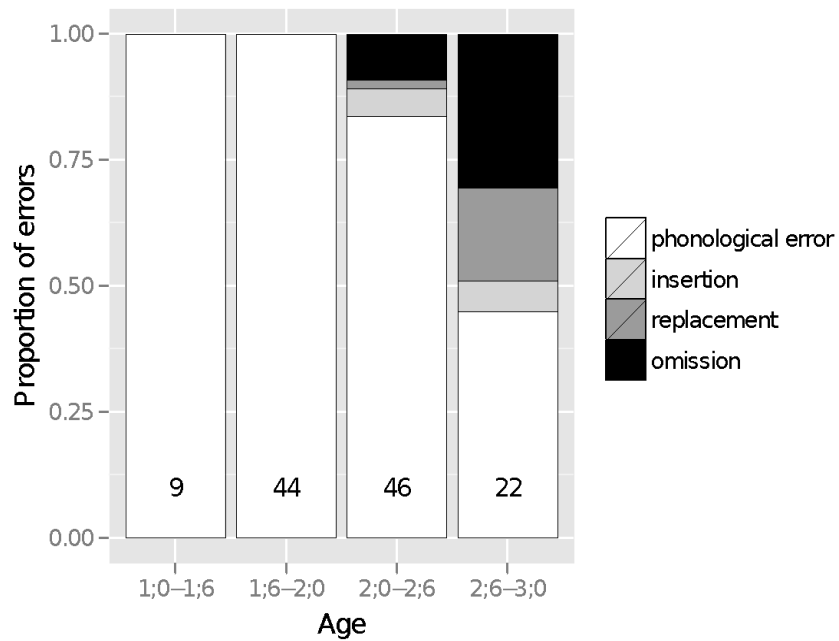


Figure 4. Proportion of each error type for vowel-initial nouns in the Lyon Corpus, binned by age. Phonological errors are shown in white, the three types of LC errors in gray. Raw frequency of phonological errors is also provided numerically for each age bin.

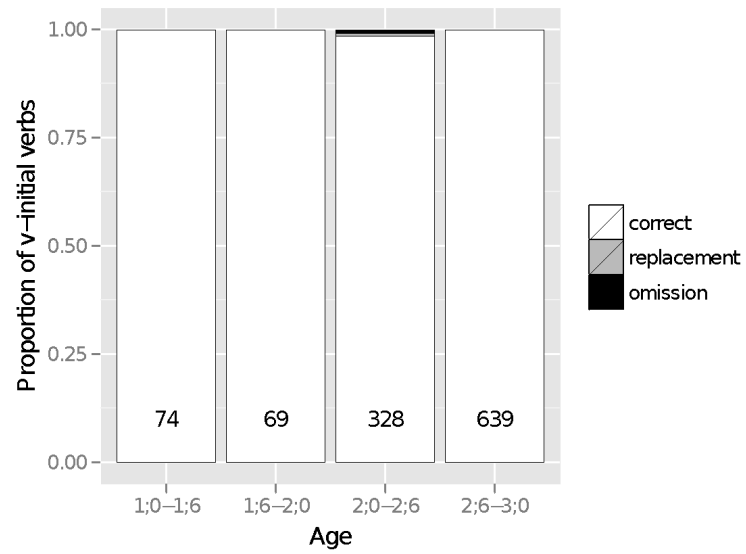


Figure 5. Proportion of each response type for vowel-initial verb targets in Lyon Corpus by binned age. Raw frequency of correct targets is also provided numerically for each age bin.

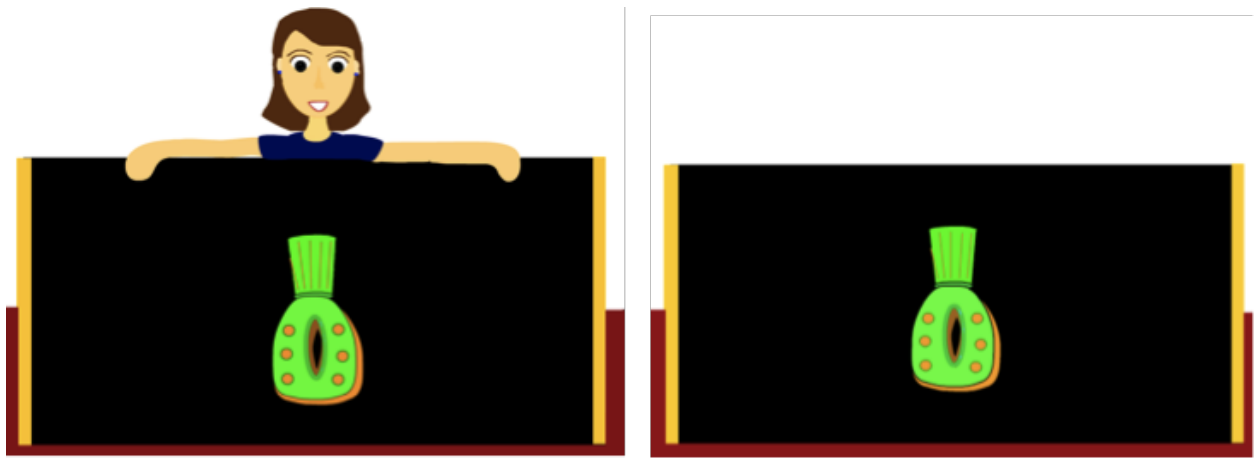


Figure 6. Example video used for training and familiarization(left) and test (right).



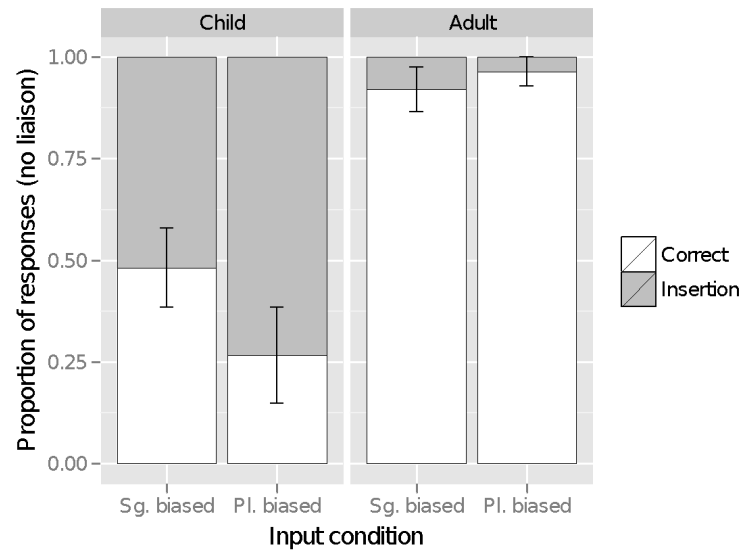


Figure 7. Accuracy by age in contexts not triggering liaison (e.g. *joli*), singular only. Error bars on correct proportions represent standard error of the mean.

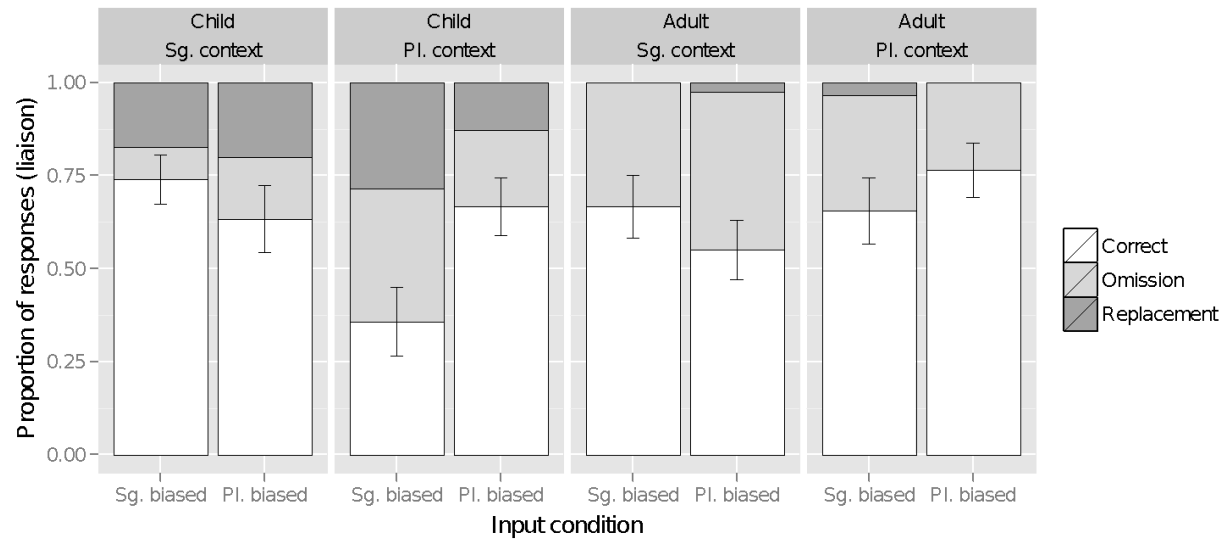


Figure 8. Accuracy by age in contexts triggering liaison (e.g. *un, deux*). Error bars on correct proportions represent standard error of the mean.

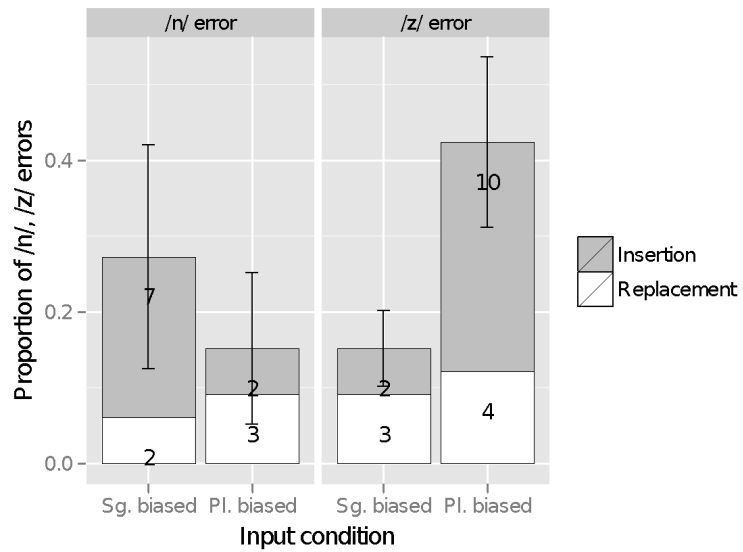


Figure 9. Influence of bias condition on children's /n/ and /z/ LC errors (replacements and insertions). Error bars represent standard error of the mean for total proportion of errors.

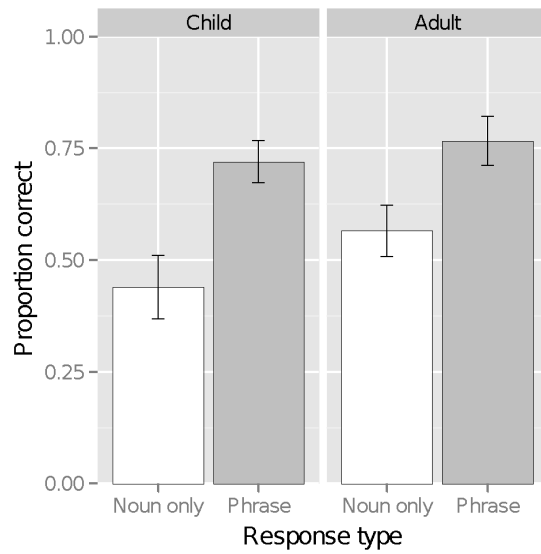


Figure 10. Accuracy in liaison-triggering contexts for noun only compared to phrasal responses by age. Error bars represent standard error of the mean.

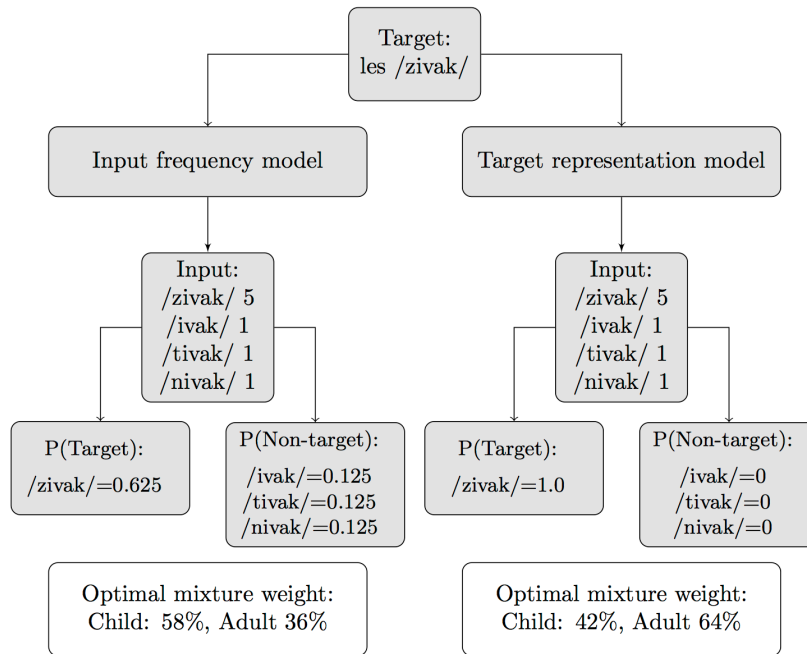


Figure 11. Two models of liaison production and optimized mixture for children and adults.