Gonzalez-Gomez, N and Nazzi, T

Constraints on statistical computations at 10 months of age: the use of phonological features.

Gonzalez-Gomez, N and Nazzi, T (2014) Constraints on statistical computations at 10 months of age: the use of phonological features.. *Developmental Science*, 18 (6). pp. 864-876.

doi: 10.1111/desc.12279

Constraints on statistical computations at 10 months of age:

The use of phonological features

Nayeli Gonzalez-Gomez[ab] & Thierry Nazzi [bc]

[a] Oxford Brookes University, Department of Psychology, Oxford, England.

[b] Université Paris Descartes, Sorbonne Paris Cité, Paris, France

[c] CNRS (Laboratoire Psychologie de la Perception, UMR 8158), Paris, France

Address for correspondence: Nayeli Gonzalez-Gomez, Oxford Brookes University, Psychology Department, Gipsy Lane, Oxford, OX3 0BP, UK.

Email: ngonzalez-gomez@brookes.ac.uk; tel: +44 1 865 48 36 76

Abstract

Recently, several studies have argued that infants capitalize on the statistical properties of natural languages to acquire the linguistic structure of their native language, but the kinds of constraints which apply to statistical computations remain largely unknown. Here we explored French-learning infants' perceptual preference for labial-coronal (LC) words over coronal-labial words (CL) words (e.g., preferring *bat* over *tab*), to determine whether this phonotactic preference is based on the acquisition of the statistical properties of the input based on a single phonological feature (i.e., place of articulation), multiple features (i.e., place and manner of articulation), or individual consonant pairs. Results from four experiments revealed that infants had a labial-coronal bias for nasal sequences (Exp. 1) and for all plosive sequences (Exp. 2 & 4) but a coronal-labial bias for all fricative sequences (Exp. 3 & 4), independently of the frequencies of individual consonant pairs. These results establish for the first time that constellations of multiple phonological features, defining broad consonant classes, constrain the early acquisition of phonotactic regularities of the native language.

Key words: statistical learning, phonotactics, phonological features

Introduction

There is ample evidence that humans possess powerful statistical learning capacities. The ability to compute distributional regularities in visual or auditory input has been found in infants from birth (Gomez & Gerken, 1999; Kirkham, Slemmer, & Johnson, 2002; Mersad & Nazzi, 2012; Saffran, Aslin, & Newport, 1996; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009), in adults (Cleeremans, 1993; Saffran, Newport, & Aslin, 1996; see review by Romberg & Saffran, 2010) and to a certain degree, even in non-human primates (Fitch & Hauser, 2004; Greenfield, 1991; Hauser, Newport, & Aslin, 2001;Savage-Rumbaugh et al., 1993). This capacity is assumed to be very useful for language acquisition, facilitating the discovery of linguistic structure, including vowel and consonant categories (see Werker, Yeung, & Yoshida, 2012 for review), phonotactic regularities (i.e., where in a word vowels and consonants can occur; Chambers, Onishi, & Fisher, 2003; Cristià & Seidl, 2008; Onishi, Chambers, & Fisher, 2002; Seidl & Buckley, 2005;), word forms (Johnson & Tyler, 2010; Mersad & Nazzi, 2012; Saffran et al., 1996), and rudimentary syntax (Gomez & Gerken, 1999; Saffran & Wilson, 2003). The goal of the present study is to explore constraints on statistically-based acquisition, and more precisely, the units over which phonotactic acquisition operates.

Although this study addresses specific questions in phonological development, exploring constrains on statistical computations is an issue that bears more generally on acquisition in other cognitive domains. Various studies have shown that statistical learning is not a specific linguistic mechanism, but rather a domain general mechanism designed to detect structural regularities in the environment. For example, infants have been found to be able to learn regularities in non-linguistic tone sequences (Saffran, Johnson, Aslin, & Newport, 1999), to detect transitional probabilities of visual stimuli sequences (Kirkham et al., 2002) and to be sensitive to high-order statistical structure of visual scenes (Fiser & Aslin, 2002). Furthermore, there is evidence showing that infants use their knowledge about visual features that co-occur for object individuation, recognition and categorization (Wu, Gopnik, Richardson, & Kirkham, 2011). Exploring the constraints which apply to statistical learning in language acquisition is thus a very important issue with potential implications in other domains.

A few studies have explored the units used to compute statistical patterns when learning sound patterns. One study explored adult learning of non-adjacent regularities between either syllables or phonemic segments (consonants or vowels) from an artificial grammar (Newport & Aslin, 2004). Learners did not readily acquire regularities between non-adjacent syllables (e.g., badite where “ba” predicts “te”), but did so between non-adjacent phonemic segments involving either consonants or vowels (e.g. “pigute” where p predicts g and t; see Bonatti, Peña, Nespor, & Mehler, 2005, for evidence of easier learning of consonantal over vocalic regularities). This parallels natural languages, which use non-adjacent dependencies between phonemic segments (e.g. vowel harmony, Turkish, for example, presents front–back harmony, according to which words cannot contain both front vowels, such as /i/ or /e/, and back vowels, like /o/ and /u/) much more than between syllables.

Similarly, infants are able to extract generalizations over sets of sounds defined by a shared phonological feature. For example, 9-month-olds can form generalizations across different segments on the basis of place of articulation (e.g. labial consonant + round vowel versus coronal consonant + front vowel, for example /bu/ versus /de/; Seidl & Buckley, 2005). Furthermore, 7-month-olds can learn (and generalize) different constraints on consonant categories, but only when those categories are defined by a single phonological feature (e.g., not a continuant, continuants being sounds in which the closure of the vocal tract is incomplete, allowing the continuous passage of airflow, such as /r/ or /s/; Cristià & Seidl, 2008). However, 4-month-olds do not seem constrained in the same way, suggesting that the acquisition of phonological features starts between 4 and 7 months (Cristià, Seidl & Gerken, 2011). Finally, 9-month-olds can learn constraints between non-adjacent consonants, but only when the consonants share a phonological feature (e.g., $\mathbf{C_1}V_1C_2\mathbf{C_3}V_2C_4$, in which consonants C_1 and C_3 were voiced, that is produced with vocal cord vibration, such as /b/ or /d/; Saffran & Thiessen, 2003). These findings show that infants can compute statistical computations over sets of sounds sharing a phonological feature in controlled laboratory experiments.

In summary, both adult and infant studies on statistical learning in the laboratory have shown that regularities can be extracted over non-adjacent vowel or consonant segments (Newport & Aslin 2004; Saffran & Thiessen, 2003), as well as over groups of sounds defined by phonological features (Cristià et al., 2011; Cristià & Seidl, 2008; Maye, Weiss, & Aslin, 2008; Saffran & Thiessen, 2003; Seidl & Buckley, 2005;). However, many questions remain, including the nature of the units that enter into statistical computations. The present study explored this question, examining how infants learn from the naturally occurring and complex language input perceived outside of the laboratory.

Specifically, we identified a phonological dependency that is not homogenously distributed in the lexicon of French, and for which calculations based on phonological features would pattern differently than calculations based on individual consonants. We first present a corpus-analysis of this phonological dependency (the *Labial-Coronal bias*), and then present experiments examining French-learning infants' preferences for sound patterns that follow or do not follow statistical patterns defined by various units (i.e., individual consonants, groups of consonants defined by a single feature, or groups of consonants defined by two phonological features).

A corpus analysis of the Labial-Coronal (LC) bias

The Labial-Coronal bias corresponds to the prevalence of sequences starting with a labial consonant (produced with one or both lips, e.g., /b/, /p/) followed by a coronal consonant (articulated with the front part of the tongue, e.g., /t/, /d/) such as “bat” (LC pattern) over the opposite pattern: sequences starting with a coronal consonant followed by a labial one, as in the word “tab” (CL pattern). This bias was first found in early word production; English-, German-, Dutch-, French- or Czech-learning infants in the 50-word stage tend to produce more LC than CL sequences (Ingram, 1974; MacNeilage & Davis, 2000; MacNeilage, Davis, Kinney, & Matyear, 1999). Additionally, the prevalence of LC over CL sequences has also been found across languages in typological studies. This tendency was found in 16 out of 18 languages examined (e.g., French, English and Spanish), Japanese and Swahili being the exception (MacNeilage et al., 1999; Vallée, Rousset, & Boë, 2001). More recently, this bias was also found at the perceptual level. French-learning infants start preferring LC over CL sequences (e.g. *bat* over *tab*) in both bisyllabic Consonant-Vowel-Consonant-Vowel (CVCV) words and monosyllabic Consonant-Vowel-Consonant (CVC) words between 7 and 10 months of age (Gonzalez-Gomez & Nazzi, 2012; Nazzi, Bijeljac-Babic, & Bertoncini, 2009). Importantly though, 13-month-old Japanese-learning infants show a preference for the opposite pattern, that is CL sequences, which is the more frequent pattern in the Japanese lexicon (Gonzalez-Gomez, Hayashi, Tsuji, Mazuka, & Nazzi, 2014). Taken together, these results indicate that exposure to linguistic input is a key factor in the emergence of the perceptual LC bias, establishing that an opposite CL bias can emerge if supported by the input.

With respect to French, the LC preference found in French-learning infants (Gonzalez-Gomez & Nazzi, 2012; Nazzi et al., 2009) is assumed to be related to the fact that LC words are overall more frequent than CL words (e.g. there are more words like *bat* than *tab*) in French, as established by Gonzalez-Gomez and Nazzi (2012) using Lexique 3 (New, Pallier, Ferrand, & Matos, 2001), a corpus of French texts containing 31 million words. However, the LC bias is interesting to further explore because it appears not to be homogenously distributed in the lexicon. A previous analysis of the French lexicon had reported that out of the four possible pairs of plosive consonants (produced by stopping airflow through the mouth, i.e., labials /p/ and /b/ versus coronals /t/ and /d/), three pairs (/p-t/, /p-d/, /b-t/) had an LC bias while the fourth pair (/b-d/) had the opposite pattern, that is, more /d-b/ than /b-d/ sequences (Sato, Vallee, Schwartz, & Rousset, 2007). Based on this finding, we first decided to conduct more detailed analyses of the French lexicon using Lexique 3 given that it is the biggest available corpus of the French lexicon. Second, additional frequency analyses were conducted on a smaller child-directed corpus (Ngon et al., 2013), to verify that child-directed speech contains the same biases (or at least tendencies) as those found in the adult corpus. The child-directed corpus contains over 285,000 word tokens (Ngon et al., 2013). The corpus is a compilation of several French corpora from the CHILDES database (MacWhinney, 2000), consisting of parent–infant speech dialogues addressed to infants from French-speaking families who were at most 24 months of age (Bassano & Maillochon, 1994; De Cat & Plunkett, 2002; Demuth & Tremblay, 2008; Hamann et al., 2003; Hunkeler, 2005; MacWhinney, 1995; Morgenstern, 2006; Suppes, Smith, & Leveille, 1973).

We first conducted a token frequency analysis in *Lexique 3* of all possible word-initial Consonant-Vowel-Consonant (CVC) sequences containing a labial and a coronal consonant (see Table 1 for data on the consonant paris used in Exp. 1-4, and Appendix for data on all pairs). These analyses revealed that out of 40 possible consonant pairs between 5 labials (/p/, /b/, /f/, /v/, /m/) and 8 coronals (/t/, /d/, /s/, /ʃ/, /z/, /ʒ/, /n/, /l/), 14 pairs presented a significant LC bias on both corpora and 4 pairs a significant reversed CL frequency bias (7 pairs do not have any occurrences for CL sequences; 6 pairs showed a significant opposite bias in both corpora; 5 pairs showed an LC tendency and 2 a CL one but it was not significant in one of the corpus; and 2 pairs showed a non-significant opposite tendency). These analyses confirm that the LC bias is the most common pattern, but is not uniformly distributed across all pairs of consonants. Similar analyses were conducted on the *CHILDES* database, and revealed the same pattern of biases (except for /b-t/, this reverse pattern being mainly caused by the high frequency of occurrence of the word /tɔ̃be/, meaning to fall, that accounted for 442 of the 627 occurrences of all /b-t/ tokens), although some of the biases that were significant in the adult database failed to reach significance in the infant database, possibly due to the small numbers of tokens in the infant database for some of the consonant pairs.

Table 1. Word-initial LC to CL ratios in French words for the consonant pairs used in Experiments 1-4 according to *Lexique 3* (left) and *CHILDES* (right). Ratios above 1 indicate an LC bias, ratios below 1 indicate a CL bias (marked in **bold**).

	Pair Bias	Pair	Ratio Lexique 3	Ratio CHILDES
Plosives	LC	p-t	6.29***	9.28***
		b-t	1.08*	**0.44*****
	CL	p-d	**0.52*****	**0.89** ns
		b-d	**0.29*****	**0.30*****
Fricatives	LC	f-s	1.64***	1.57***
		v-ʒ	1.46***	1.50 ns
	CL	v-s	**0.10*****	**0.13*****
		f-ʃ	**0.38*****	**0.86** ns
		v-ʃ	**0.13*****	**0.74****
Nasals	LC	m-n	9.18***	16.75***

Chi-square test of goodness-of-fit: ns: $p > 0.10$; †: $p \leq 0.09$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$.

This variability in biases across pairs could affect acquisition differently depending on the level at which the previously reported LC bias is acquired. One possibility, based on the pair analysis, is that infants compute statistics for each pair of consonants, and learn a different (LC or CL) bias for each pair. A second possibility is that infants learn a single-feature-based LC bias for their native language, i.e., that they learn that LC sequences are overall the more frequent, ignoring statistics for individual pairs of consonants. Since previous studies have not controlled this factor, it is unclear which hypothesis is supported by the LC biases found in early perception.

A third hypothesis was considered in light of the work discussed above showing that infants are sensitive to phonological features (Cristià et al., 2011; Cristià & Seidl, 2008; Maye et al., 2008; Saffran & Thiessen, 2003; Seidl & Buckley, 2005). We conducted two additional analyses in *Lexique3* and *CHILDES* to determine whether the consideration of multiple phonological features (in addition to just labial versus coronal: place of articulation) would reveal patterns of variation predicting infant behavior. In the first analysis, we grouped consonants by voicing (whether or not sounds involve the vibration of the vocal cords; e.g., voiceless sounds: /p/, /t/, /f/;   voiced sounds: /b/, /d/, /n/). The results (Table 2, top panel) showed the existence of an LC bias for three of the four possible combinations, while the voiceless-voiced pairs showed a CL bias in both databases. However, given that the main differences were only found in the mixed pairs (Voiced-Voiceless and Voiceless-Voiced) and that testing these pairs was problematic in terms of experimental design (since we cannot include the same phonemes in the LC and CL lists), this feature was not further investigated.

Table 2. Word-initial LC to CL ratios by voicing (top) and by manner of articulation (bottom) in French words according to *Lexique 3* (left) and *CHILDES* (right). Ratios above 1 indicate an LC bias, ratios below 1 indicate a CL bias (marked in **bold**).

	Lexique 3			CHILDES		
1st\2nd Cons	Voiced	Voiceless		Voiced	Voiceless	
Voiced	1.38 ns	3.28***		4.07***	16.53***	
Voiceless	**0.85*****	2.94***		**0.72*****	4.00***	
1st\2nd Cons	Plosive	Nasal	Fricative	Plosive	Nasal	Fricative
Plosive	1.4***	**0.48*****	1.68***	1.74***	1.41***	11.64***
Nasal	24.62***	9.18***	3.00***	193.42***	16.75***	14.79***
Fricative	**0.78*****	1.05†	**0.62*****	**0.94**ns	2.34***	**0.90**†

Chi-square test of goodness-of-fit: ns: $p > 0.10$; †: $p \leq 0.09$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$.

In the second analysis, we grouped consonants by manner of articulation, examining plosive (produced by stopping airflow through the mouth, i.e., labials /p/ and /b/ versus coronals /t/ and /d/), nasal (produced a lowered velum in the mouth, allowing air to come out through the nose, i.e., labial /m/ versus coronal /n/) and fricative (produced by forcing air through a narrow channel made by placing two articulators close together, i.e., labials /f/ and /v/ versus coronals /s/ and /z/) consonants separately. Results are presented in Table 2 (bottom panel). In Lexique 3, for the heterogeneous pairs (accounting for 54% of all sequences), this analysis showed that 4 combinations have a LC bias (plosive-fricative, nasal-plosive, nasal-fricative and fricative-nasal) and 2 a CL bias (fricative-plosive and plosive-nasal). For homogenous sequences, that is sequences having the same manner of articulation in both consonants (i.e. plosive-plosive, fricative-fricative and nasal-nasal; accounting for 28% of all sequences[1]), this analysis revealed that the LC bias is present for sequences of plosive and nasal consonants. For fricative sequences there was a CL bias. When taking manner of articulation into account, results reveal clear variations in the prevalence of LC versus CL input biases in Lexique 3. The results on the CHILDES database reveal a similar pattern of biases (except for plosive-nasal sequences), although again some of the biases are not significant again probably due to lack of power related to the smaller size of the database.

These analyses therefore suggest a third possibility, namely that infants compute statistics considering groups of consonants defined by both place (LC or CL) and manner of articulation (plosive, fricative, or nasal). In order to evaluate this possibility together with the two previously discussed hypotheses (i.e., a global bias for LC over CL sequences, or a bias that varies depending on individual consonant pairs), our study focused on homogeneous

[1] The remaining 18% includes the pairs with the liquid consonant /l/, which have not been included in this analysis given that /l/ does not belong to any of the three categories analyzed here and does not have a labial equivalent.

sequences. Our predictions were that infants would learn LC biases for both plosive and nasal sequences, while they would learn a CL bias for fricative sequences.

Overview of the experimental hypotheses and design

To sum up the above corpus analysis, the French lexicon shows an overall LC bias. However, we identified 4 consonantal pairs showing a significant opposite CL pattern in both of our corpora analyses (+ one only in lexique, b-t, mainly due to the word /tɔ̃be/). In addition, the analysis grouping consonants by manner of articulation revealed that the LC bias is clearly present for homogeneous sequences of plosive and nasal consonant sequences, but not for homogeneous fricative sequences. This complex structure of the French lexicon thus allowed us to distinguish three hypotheses about infants' preference for LC words (see also Table 3):

Single-feature hypothesis: The LC bias is learned at the level of a single phonological feature (place of articulation). This computation requires only one global statistical analysis contrasting LC *versus* CL sequences, and should lead to the acquisition of an overall LC bias.

Multiple-feature hypothesis: The LC bias is learned at the level of multiple phonological features (place and manner of articulation). This possibility requires 9 different analyses combining the three different manners of articulation (plosive, fricative and nasal). In the case of the homogeneous sequences investigated here, infants would learn LC biases for plosive and nasal sequences and a CL bias for fricative sequences.

Phoneme-based hypothesis: Infants learn those biases at the level of individual consonant pairs, features playing no role in these computations. This possibility requires the computation of 40 statistical analyses, one for each pair, predicting an LC bias for 27 pairs and a CL bias for 10 pairs (the remaining pairs showing not significant input patterns).

Table 3. Summary of the experimental hypotheses and predictions.

	Computations based on	Number of computations required	Predicted bias/biases (based on Lexique)
Single-feature hypothesis	A group of consonants sharing a single phonological feature (in this case place of articulation)	1	1 LC global bias
Multiple-feature hypothesis	Groups of consonants sharing characteristics of several common features (in this case place and manner of articulation)	9	6 LC biases 3 CL biases
Phoneme-based hypothesis	Individual consonant pairs	40	27 LC biases 10 CL biases

To explore these three hypotheses about the level of generalization at which the LC bias in French is learned, four experiments were conducted. All experiments presented French-learning infants from Paris, France with lists of LC versus CL words, measuring infants' preference for each list.

Experiment 1: Nasal sequences

Infants were presented with stimuli made of the only existing pair of nasal consonants in French, /m/ and /n/, which has a statistical LC bias in French (c.f. Table 1). Although all three hypotheses predict an LC bias, this experiment was conducted to provide a comprehensive picture of the acquisition in French of LC/(CL) biases in homogeneous sequences across all manners of articulation.

Materials and Methods

Participants. Sixteen 10-month-old infants from French-speaking families were tested (mean age = 10 months 12 days; range: 10 months 1 day – 21 days; 9 girls, 7 boys). The data of two additional infants were not included in the analyses due to fussiness/crying.

Stimuli. Twelve monosyllabic C_1VC_2 items were selected, combining the labial consonant /m/ and the coronal consonant /n/, which occur in the French lexicon more often as LC than CL (c.f. Table 1): 6 items with a labial-coronal (LC) structure (mVn: /**mɔn**/, /**mon**/, /**mun**/, /man/, /**myn**/, /**møn**/) and 6 items with a coronal-labial (CL) structure (nVm: /nɔm/, /**nom**/, /**num**/, /**nam**/, /**nym**/, /**nøm**/).

Vowels in all sub-experiments were chosen to obtain balanced adjacent dependencies between the LC and CL lists for the C_1V_1, V_1C_2 and $C_1V_1C_2$ phoneme sequences according to Lexique 3. Due to this constraint, we had to use both low frequency French words and pseudowords legal in French (marked in **bold** in the stimuli section). In each sub-experiment, the consonants and vowels used in the LC and CL structures were identical.

The stimuli were recorded in a sound-attenuated booth by a French female native speaker who was naive to the purpose of the study. Two tokens of each item were selected. Two LC lists were created, one containing the first token of each LC item and the other the second token. Within each list, all 6 items were repeated twice (leading to a total of 12 items), and were arranged in semi-random order. Two CL lists were constructed in the same way. The duration of all the lists was 18.00 s.

Procedure and Apparatus. The experiment was conducted inside a soundproof booth. The booth had a red light and a loudspeaker (SONY xs-F1722) mounted at eye level on each of the side panels and a green light mounted on the center panel. A response box (connected to Dell Optiplex computer) and a TV screen (connected to a camera inside the booth) were located outside the booth. The observer, who looked at the video of the infant on the TV screen, pressed the buttons of the response box according to the direction the infant's head, thus starting and stopping the flashing of the lights and the presentation of the sounds, and recording the looking times. The observer and the infant's caregiver wore earplugs and listened to masking music over tight-fitting closed headphones, which prevented them from hearing the stimuli presented. Information about the duration of the head-turn was stored on the computer.

The classic version of the Head-turn Preference Procedure (HPP) was used (Jusczyk, Cutler, & Redanz, 1993). Each infant was held on a caregiver’s lap in the center of the test booth. Each trial began with the green light on the center panel blinking until the infant had oriented to it. Then, the red light on one of the side panels began to flash. When the infant turned in that direction, the stimulus for that trial began to play. The stimuli were delivered by the loudspeakers via an audio amplifier (Marantz PM4000). Each stimulus was played to completion or stopped after the infant failed to maintain the head-turn for 2 consecutive seconds. If the infant turned away from the target by 30° in any direction for less than 2s and then turned back again, the trial continued but the time spent looking away (when the experimenter released the buttons of the response box) was automatically subtracted from the orientation time by the program. Thus, the maximum orientation time for a given trial was the duration of the entire speech sample. If a trial lasted less than 1.5 s, the trial was repeated and the original orientation time was discarded.

Each session began with 2 musical trials, one on each side to give infants an opportunity to practice one head-turn to each side. The test phase consisted of 8 trials divided in 2 blocks (in each of which the two lists of each structure were presented). Order of the different lists within each block was randomized.

Results

Mean orientation times to the LC and CL lists were calculated for each infant. The means for the group (M_{LC} = 8.75 s, *SD* = 1.80 s; M_{CL} = 7.33 s, *SD* = 1.70) are presented in Figure 1 (left panel). A t-test revealed that the difference between the LC and CL trials was significant, $t(15) = 2.43$, $p = .02$, $d = .81$, infants having longer orientation times for the LC lists. This pattern was present in 13 of the 16 infants tested (binomial test $p = .01$).

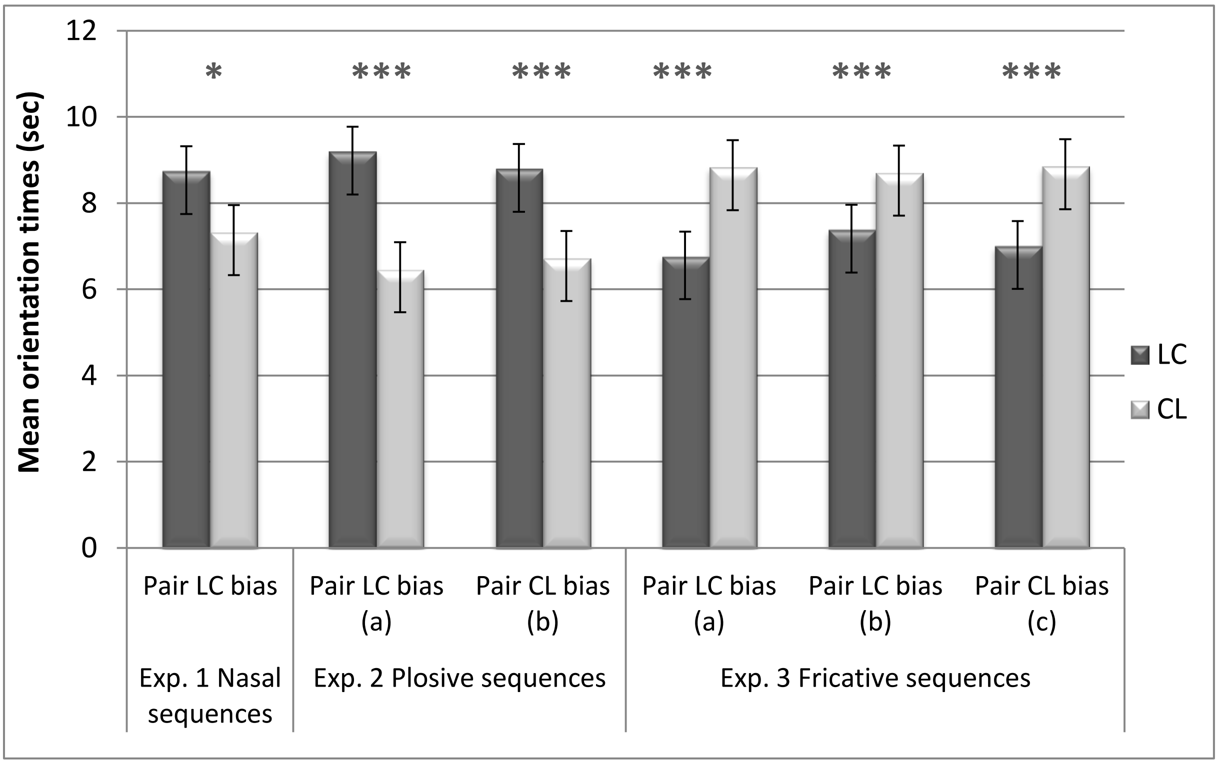

Figure 1. Mean orientation times (and SEs) to the LC and CL stimuli. Left panel: plosive sequences (Exp. 1); middle panel: fricative sequences (Exp. 2); right panel: nasal sequences (Exp. 3). Letters a-c denote each subexperiment. *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$.

Experiment 1 shows that French-learning 10-month-old infants prefer nasal LC sequences over nasal CL sequences. These results establish the existence of an LC bias for nasal consonant sequences, importantly extending the results of previous studies, which had shown a perceptual LC bias from plosive consonant sequences (Nazzi et al., 2009; Gonzalez-Gomez & Nazzi, 2012), to sequences from another manner of articulation. However, given that there is also an LC bias for nasal consonant sequences in the French lexicon, all three of the different acquisition hypotheses we presented above can account for this bias. To disentangle these hypotheses, two more experiments were conducted, one using plosive sequences (Exp. 2), one using fricative sequences (Exp. 3).

Experiment 2: Plosive sequences

Infants were presented with lists of LC versus CL words containing only plosive consonants. Two different sub-experiments were conducted: one used words with a statistical LC bias in the French lexicon (Experiment 2a) while the other used words with a statistical CL bias (Experiment 2b; c.f. Table 1). Both the single-feature and the multiple-feature hypotheses predicted an LC bias for both experiments. However, the phoneme-based hypothesis predicted the existence of two opposite biases: an LC bias for Experiment 2a and a CL bias for Experiment 2b.

Materials and Methods

Participants. Two different groups of sixteen 10-month-old infants from French-speaking families were tested (mean age = 10 months 17 days; range: 10 months 1 day – 26 days; 13 girls, 19 boys). The data of five additional infants were not included in the analyses due to fussiness/crying.

Stimuli.

Experiment 2a: LC bias. Twelve monosyllabic C_1VC_2 items were selected, combining the labial consonant /p/ and the coronal consonant /t/, which occur in the French lexicon more often as LC than CL (c.f. Table 1): 6 LC items (pVt: /pɔt/, /pat/, **/put/**, /**pɔt**/, p t/, /**pot**/) and 6 CL items (/**tɔp**/, /tap/, /**tup**/, /tɔp/, / **p**/, /top/).

Experiment 2b: CL bias**.** Twelve monosyllabic C_1VC_2 items were selected, combining the labial consonant /b/ and the coronal consonant /d/, which occur in the French lexicon more often as CL than LC (c.f. Table 1): 6 LC items (/bɔd/, /bad/, /bud/, /**bɔd**/, / **d**/, /**bod**/) and 6 CL items (/**dɔb**/, /dab/, /**dub**/, /**dɔb**/, / **b**/, /dob/).

The steps in stimuli preparation were the same as in Experiment1.

Procedure and Apparatus. Same as in Experiment 1.

Results

Mean orientation times to the LC and CL lists were calculated for each infant. Group averages for Experiments 2a and 2b are presented in Figure 1 (middle panel). The means for the group in Experiment 2a were M_{LC} = 9.20 s (*SD* = 2.86) and M_{CL} = 6.47 s (SD = 2.93). This pattern was present in 13 of the 16 infants tested (binomial test p = .01). The means for the group in Experiment 2b were M_{LC} = 8.80 s (*SD* = 2.96) and M_{CL} = 6.73 s (*SD* = 2.19). This pattern was present in 13 of the 16 infants tested (binomial test $p = .01$). A two-way ANOVA with the between-subject factor of Experiment (2a versus 2b) and the within-subject factor of lexical structure (LC versus CL) was conducted. The effect of lexical structure was significant ($F(1,30) = 18.89$, $p < .001$, $\eta p^2 = .39$), infants having longer orientation times for the LC lists. Neither the effect of experiment ($F(1,30) = .008$, $p = .93$) nor the interaction between experiment and lexical structure ($F(1,30) = .36$, $p = .56$) reached significance, indicating that the effect did not change across sub-experiments.

The results of Experiment 2 show that French-learning 10-month-olds have a clear preference for LC plosive sequences, even for the pair with a CL statistical bias in the lexicon. These findings are not predicted by the phoneme-based hypothesis, according to which infants learn phonotactic biases at the level of each consonant pair. However, they are compatible with both the single-feature and multiple-feature hypotheses. To further explore this question, and more importantly to rule out one of these remaining hypotheses, Experiment 3 was conducted using sequences of fricative consonants.

Experiment 3: Fricative sequences

Infants were presented with lists of LC versus CL words, using only fricative consonants. As in Experiment 2, two different sub-experiments were conducted: one used words with a statistical LC bias in the French lexicon (Experiment 3a) while the other used words with a statistical CL bias (Experiment 3c; c.f. Table 1). These pairs were selected because those pairs were the ones having the most clear biases in the lexicon according to lexique 3. However, given that both pairs contained the phoneme /f/, an additional pair of phonemes having an LC bias was selected (Experiment 3b) to rule out the possibility of a preference for f-final sequences if CL biases were obtained in both Exp. 3a and 3c. Fricative sequences are a crucial case to explore given that they do not follow the same statistical pattern as plosive and nasal sequences in French: there are more CL than LC fricative sequences. Therefore, while the single-feature hypothesis also predicts an LC bias overall for fricative sequences, the multiple-feature hypothesis predicts a CL bias for this class of consonants. Lastly, the phoneme-based hypothesis predicts the existence of two opposite biases: an LC bias for Experiments 3a and 3b and a CL bias for Experiment 3c.

Materials and Methods

Participants. Three different groups of sixteen 10-month-old infants from French-speaking families were tested (mean age = 10 months 12 days; range: 10 months 1 day – 27 days; 20 girls, 28 boys). The data of six additional infants were not included in the analyses due to fussiness/crying.

Stimuli.

Experiment 3a: LC bias. Twelve monosyllabic C_1VC_2 items were selected, combining the labial consonant /f/ and the coronal consonant /ʃ/, which occur in the French lexicon more often as LC than CL (c.f. Table 1): 6 LC items (/**fɔʃ**/, /fiʃ/, / **f ʃ**/, /**fuʃ**/, /**fyʃ**/, /**føʃ**/) and 6 CL items (/**ʃɔf**/, /ʃif/, /**ʃ f**/, /**ʃuf**/, /**ʃyf**/, /**ʃøf**/).

Experiment 3b: LC bias. Twelve monosyllabic C_1VC_2 items were selected, combining the labial consonant /v/ and the coronal consonant /ʒ/, which occur in the French lexicon more often as CL than LC (c.f. Table 1): 6 LC items (/vɔʒ/, /vœʒ/, /voʒ /, /vəʒ /, /vyʒ/, /vɛʒ/) and 6 CL items (/ʒɔv /, /ʒœv/, /ʒov/, /ʒəv/, /ʒyv/, /ʒɛv/).

Experiment 3c: CL bias. Twelve monosyllabic C_1VC_2 items were selected, combining the labial consonant /f/ and the coronal consonant /s/, which occur in the French lexicon more often as CL than LC (c.f. Table 1): 6 LC items (/fɔs/, /fis/, /**f** /, /**fus**/, /**fys**/, /**føs**/) and 6 CL items (/**sɔf**/, /**sif**/, /**s** /, /**suf**/, /**syf**/, /**søf**/).

The steps in stimuli preparation were the same as in Experiment1.

Procedure and Apparatus. Same as in Experiment 1.

Results

Mean orientation times to the LC and CL lists were calculated for each infant. Group averages for Experiments 3a, 3b and 3c are presented in Figure 1 (right panel). The means for the group in Experiment 3a were M_{LC} = 6.77 s (*SD* = 2.84) and M_{CL} = 8.84 s (*SD* = 3.75). This pattern was present in 13 of the 16 infants tested (binomial test p = .01). The means for the group in Experiment 3b were M_{LC} = 7.34 s (*SD* = 1.90) and M_{CL} = 8.71 s (SD = 2.61). This pattern was present in 13 of the 16 infants tested (binomial test p = .01). The means for the group in Experiment 3c were M_{LC} = 7.01 s (*SD* = 1.95) and M_{CL} = 8.86 s (*SD* = 2.47). This pattern was present in 12 of the 16 infants tested (binomial test p = .04). A three-way

ANOVA with Experiment (3a, 3b and 3c) and lexical structure (LC versus CL) was conducted. The effect of lexical structure was significant ($F(1,45) = 19.55$, $p \leq .001$, $\eta p^2 = .31$), infants having longer orientation times for the CL lists. Neither the effect of Experiment ($F(2,45) = .05$, $p = .95$ nor the interaction between experiment and lexical structure ($F(2,45) = .31$, $p = .73$) reached significance, indicating that the effect did not change across experiments.

The results of Experiment 3 show a CL bias for the three pairs of fricative sequences, supporting the multiple-features hypothesis, which states that the LC bias is learned by consonant groups that share multiple phonological features (i.e., fricatives having an LC structure). Indeed, the single-feature hypothesis predicted an LC bias in Experiments 3a 3b and 3c, while the phoneme-based hypothesis predicted an LC bias in Experiments 3a and 3b and a CL bias in Experiment 3c. These new results allow us to discard the single-feature hypothesis that could have accounted for the results of Experiments 1 and 2. Before further discussing implications of these results, the multiple-features hypothesis was given a direct test in the next experiment. There we did not present infants with stimuli having either an LC or CL bias based on individual pairs of consonants (as in Experiments 1-3). Rather we used a set of mixed stimuli, which had either an LC or CL bias when calculating statistics based on manner of articulation, but not when calculating statistics across individual items.

Experiment 4: Biases beyond individual pairs

Infants were presented with lists of LC versus CL words, combining either four plosive consonants or four fricative consonants. For Experiment 4a, we used the only labial (/p/ and /b/) and coronal (/t/ and /d/) plosive consonants in French. For Experiment 4b, we used the only labial fricative consonants in French (/f/ and /v/), and selected two coronal fricatives having the higher frequency of occurrence (/s/ and /ʃ/). The combination of the two labial and the two coronal consonants resulted in a total of 4 different pairs of consonants for each structure (LC or CL) in each subexperiment. Among these pairs of consonants, some pairs showed a statistical LC bias (Exp. 4a: p-t and b-t; Exp. 4b: f-s) and some pairs showed a statistical CL bias (Exp. 4a: d-b and p-d; Exp. 4b: ʃ-f, ʃ-v and s-v; c.f. Table 1). Hypothesis 2 (multiple-feature), which accounts for our findings so far, predicts an LC bias for plosives (Experiment 4a) but a CL bias for fricatives (Experiment 4b), in spite of the fact that some individual consonant pairs within each sub-experiment present the opposite statistical bias.

Materials and Methods

Participants. Two different groups of sixteen 10-month-old infants from French-speaking families were tested (mean age = 10 months 12 days; range: 10 months 2 day – 26 days; 18 girls, 14 boys). The data of three additional infants were not included in the analyses due to fussiness/crying.

Stimuli.

Experiment 4a: Plosives (from Gonzalez-Gomez & Nazzi, 2012). Twenty-four monosyllabic C_1VC_2 items were selected, combining labial consonants /p/ and /b/, and coronal consonants /t/ and /d/: 12 items with a labial-coronal (LC) structure (3 bVd: /bɔd/, /**byd**/, /**bad**/; 3 pVt: /pɔt , p t/, /**pot**/; 3 bVt: /bɔt/, /byt/, /bat/; and 3 pVd: /pad/, /**pod**/, /**p** /) and 12 items with a coronal-labial (CL) structure (3 dVb: /**d** /, /dob/, /**dab**/; 3 tVp: /t p , tap , top ; t b: t b , tɔb , **tab**/; and 3 dVp: /**dap**/, / **p**/, /dɔp/).

Experiment 4b: Fricatives. Twenty-four monosyllabic C_1VC_2 items were selected, combining labial consonants /f/ and /v/, and coronal consonants /ʃ/ and /s/: 12 items with a labial-coronal (LC) structure (3 fVs: /fɔs/, /fos/, / **s**/; 3 fVʃ: /**fyʃ**/, /faʃ/, /**f ʃ** ; v s: v s/, /**v** /*, /**vos**/*; and 3 vVʃ: /vaʃ/, /vyʃ/, /**vɔʃ**/*) and 12 items with a coronal-labial (CL) structure

(3 ʃVf: /**ʃ f**/, /ʃof/, /**ʃaf**/; 3 ʃVv: /**ʃ v**/, /**ʃav**/, /**ʃyv**/; 3 sVf: /**sɔf**/, /sof/, / **f**/; and 3 sVv: /**syv**/, /**s** /, /**sɔv**/).

In both sub-experiments, items in both lists (LC and CL) were made up of exactly the same consonants, and the vowels were almost completely balanced across lists. Otherwise, the steps in stimuli preparation were the same as in Experiment1.

Procedure and Apparatus. Same as in Experiment 1.

Results

Mean orientation times to the LC and CL lists were calculated for each infant. Group averages for Experiments 4a and 4b are presented in Figure 2. The means for the group in Experiment 4a were M_{LC} = 10.41 s (*SD* = 2.80) and M_{CL} = 7.64 s (*SD* = 2.00). This pattern was present in 15 of the 16 infants tested (binomial test $p < .001$). The means for the group in Experiment 4b were M_{LC} = 7.24 s (*SD* = 1.72) and M_{CL} = 8.88 s (SD = 2.38). This pattern was present in 12 of the 16 infants tested (binomial test p = .04). A two-way ANOVA with the between-subject factor of Experiment (4a versus 4b) and the within-subject factor of lexical structure (LC versus CL) was conducted. Neither the effect of lexical structure ($F(1,30) = 2.01$, $p = .17$) nor the effect of experiment reached significance ($F(1,30) = 1.89$, $p = .18$). Importantly though, the interaction between experiment and lexical structure was significant ($F(1,30) = 30.93$, $p \leq .001$, $\eta p^2 = .51$), showing that the effect of lexical structure changed across experiments, due to the fact that infants had longer orientation times for the LC sequences in Experiment 4a but longer orientation times for the CL sequences in Experiment 4b. Planned comparisons confirmed that the lexical structure effect was significant in both experiments (Experiment 4a: ($F(1, 30) = 24.36$, $p < .001$, $d = 1.14$; Experiment 4b: $F(1, 30) = 8.58$, $p = .006$, $d = .79$).

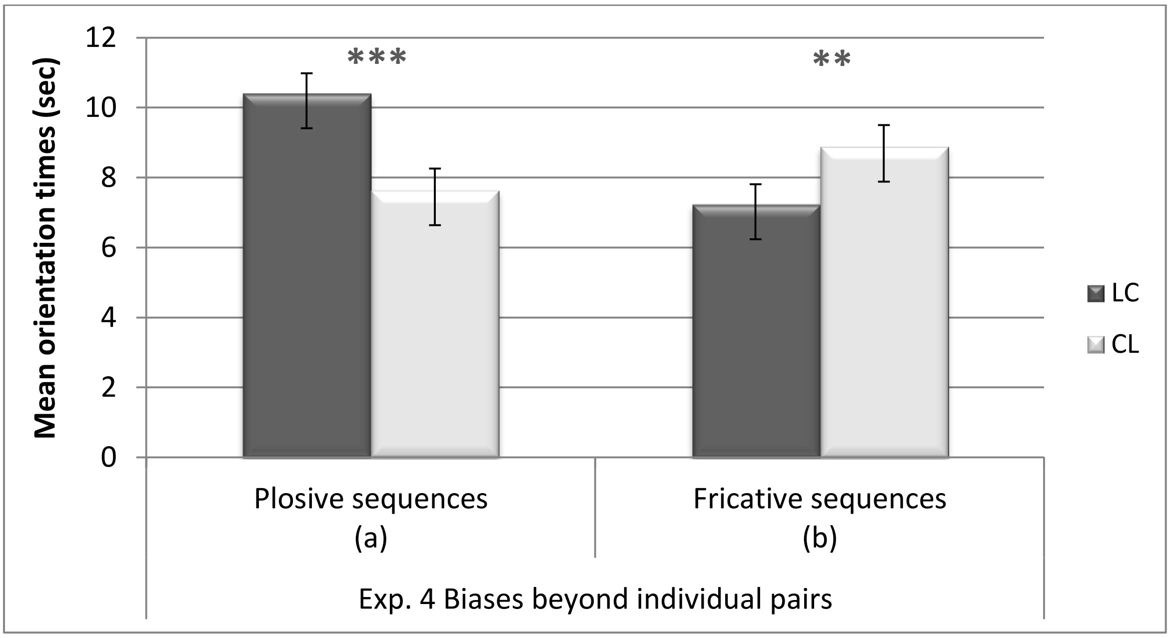

Figure 2. Mean orientation times (and SEs) to the LC and CL stimuli in Experiment 4. Left panel: plosive sequences; right panel: fricative sequences. *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$.

Taken together, the results of Experiments 1-4 show an LC bias for all plosive and nasal sequences, and a CL bias for all fricative sequences. Interestingly, nasals pattern like plosives and both form the class of non-continuants, which might suggest that non-continuants show an LC bias, while fricatives, which are continuants (i.e., consonants produced with the vocal tract only partly closed, allowing the airflow to pass through and the sound to be prolonged), show a CL bias. Independently, these results support the multiple-feature hypothesis, which states that the LC bias is learned by computing statistical analyses on consonant groups sharing particular configurations of multiple phonological features, rather than at the phoneme-based level or at the level of a single phonological feature (c.f. Table 3).

General Discussion

The goal of the present study was to explore constraints on statistical learning, and more precisely, the perceptual level at which phonotactic acquisition operates. The LC bias was used to explore this question, given that this bias is not uniformly present in the French lexicon. Three different hypotheses were evaluated. According to the “single-feature” hypothesis, the LC bias is learned at the level of a single feature (i.e., place of articulation), as infants make only one global statistical analysis contrasting LC *versus* CL sequences (e.g., sequences like *bat* versus *tab*), which should result in the acquisition of an overall LC bias. According to the “multiple-feature” h pothesis, these computations are made at a more specific level, grouping consonant sequences by both place and manner of articulation. This possibility requires the computation of 9 separate analyses. In the case of the homogeneous sequences tested in the present study, this second hypothesis should result in the acquisition of LC biases for plosive and nasal sequences, and a CL bias for fricative sequences (see more on this below). Third, the “phoneme-based” h pothesis predicts that infants learn LC or CL biases at the level of individual consonant pairs. This possibility requires the computation of 40 statistical analyses, one for each consonant pair, predicting the acquisition, based on the Lexique 3 analyses, of 27 LC biases and 10 CL biases (3 cases being statistically non-significant). Importantly though, exploring whether statistical learning is constrained by single features, features combinations or full exemplars is a very important issue with potential implications in other domains (e.g., learning of visual categories).

Four experiments were conducted, focusing on sequences homogeneous in terms of manner of articulation that differed in their statistical biases at the feature level: LC bias for plosive and nasal sequences, and CL bias for fricative sequences. The only possible pair of nasal sequences (LC bias) was tested in Exp. 1. For both plosive and fricative sequences, we first tested two pairs, one with an LC and one with a CL statistical bias (Exp. 2a-b & 3a,c), an extra pair of fricatives with an LC bias was also tested to rule out a possible positional interpretation of the results (Exp 3b), and then a mix of pairs with various individual biases (Exp. 4a-b). Our results support the predictions of the multiple-feature hypothesis of the acquisition of an LC bias for all plosive and nasal sequences, and of a CL bias for all fricative sequences. The CL bias for fricative sequences does not fit the overall LC bias predicted by the single-feature hypothesis. The fact that the biases observed for the plosive and fricative sequences were determined by the biases at the level of these classes of consonants rather than those at the level of each individual pair contradicts the predictions of the phoneme-based hypothesis.

The present results thus establish that infants' preference for LC/CL sequences is neither based on a single feature statistical analysis (place of articulation), nor based on an analysis of individual pairs of consonants. Rather, our findings support the proposal that the LC/CL biases are determined by both place and manner of articulation. These results suggest that infants use multiple phonological features to group consonants together and track statistics on these consonant groups. Infants must have inferred from natural language input in French that LC sequences are more common when presented within homogeneous plosive and nasal sequences, but CL sequences are more common when presented within homogeneous fricative sequences.

Furthermore, these results confirm that the LC bias is an effect resulting from exposure to the linguistic input. This is in line with recent results obtained with Japanese-learning infants showing that CL biases can also be found in perception if supported by the statistical properties of the input (see Gonzalez-Gomez et al., 2014 for further discussion on the origins of the LC bias). However, further studies are needed in Japanese and in other languages to investigate whether or not the LC bias is also learned based on groups of consonants sharing several phonological features.

The use of phonological features to acquire phonotactic regularities

Linguistic descriptions have shown that phonological and phonotactic regularities are often governed by natural feature classes (e.g., Kuo, 2009). Our results showing that 10-month-olds learn phonotactic patterns by computing constellations of multiple phonological features are in line with studies showing that infants use features to find phonotactic regularities in artificial language experiments (Cristià et al., 2011; Cristià & Seidl, 2008; Maye et al., 2008; Saffran & Thiessen, 2003; Seidl & Buckley, 2005). However, our study is the first to show feature-guided phonotactic learning from the complex natural input of the native language, which occurs before infants come to the laboratory.

While our results conclusively show that statistical mechanisms in phonotactic acquisition are constrained by phonological features, several questions about this mechanism require further research. For example, we tested infants' preferences for homogenous LC versus CL sequences (plosive-plosive, nasal-nasal, or fricative-fricative sequences). Nevertheless, our feature-based hypothesis defines 9 different possible combinations, 3 homogeneous and 6 heterogeneous ones. How could a statistical mechanism based on multiple features allow the acquisition of all of these combinations? One possibility is that the learning mechanism identifies the manner of articulation of the first and second consonant in each sequence, and then computes biases for all of the 9 possible combinations. A second possibility is that this mechanism only takes into account the manner of articulation of the first consonant, a proposal motivated by findings regarding the importance of word-initial positions found in some studies (e.g., Swingley, 2005), though not others (e.g., Nazzi & Bertoncini, 2009). In addition, further research will be needed to explore the generality of this finding for the acquisition of other kinds of phonotactic patterns.

Another important issue refers to the developmental trajectory of this mechanism. Are there changes along development on the kinds of constraints applying to statistical computations? Unfortunately, this question cannot be answered using the LC bias, since the youngest infants showing this bias are 10 months of age (the age tested in this study) and younger infants have failed to show such a preference (6 and 7 month-olds; Gonzalez-Gomez & Nazzi, 2012; Nazzi et al., 2009). However, studies by Cristià and colleagues can shed some light on this issue (Cristià et al., 2011; Cristià & Seidl, 2008). These studies found that 4-month-olds were able to learn (and generalize) a constraint (i.e., consonant restricted to word-initial position) applying to consonant categories whether or not those consonant categories were defined by a phonological feature. However, 7-month-olds were only able to learn (and generalize) those constraints when the consonant categories were defined by a phonological feature. Taken together, these results suggest that infants' sensitivit   changes across development and that phonological features emerge over the course of development (Mielke, 2008). However, further research is needed to explore this issue.

Conclusion

The present research is a first attempt at exploring the level at which phonotactic acquisition operates when learning one's native language outside the laboratory (rather than an artificial language in the laboratory). It provides the first piece of evidence that the acquisition of a phonotactic property of the native language, the LC bias, is made based on groups of consonants sharing characteristics of several phonological features (i.e. place and manner of articulation). These results indicate that constellations of multiple phonological features constrain the acquisition of phonotactic regularities. Furthermore, they suggest that phonological features play an important role in the acquisition of native phonotactic regularities.

Acknowledgments

This study was conducted with the support of a CONACYT grant to NGG, and ANR-13-BSH2-0004 and LABEX EFL (ANR-10-LABX-0083) grants to TN. We thank Henny Yeung for helpful comments on a previous of this manuscript. Special thanks to the infants and their parents for their kindness and cooperation.

References

Bassano, D., & Maillochon, I. (1994). Early grammatical and prosodic marking of utterance modality in French: a longitudinal case study. *Journal of Child Language*, 21, 649–675.

Bonatti, L.L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations. *Psychological Science*, 16, 451-459.

Chambers, K.E., Onishi, K.H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experiences. *Cognition*, 87, B69–B77.

Cleeremans, A. (1993). Mechanisms of implicit learning: Connectionist models of sequence learning. *Cambridge, MA:* MIT Press.

Cristià, A., & Seidl, A. (2008). Is infants' learning of sound patterns constrained b phonological features? *Language Learning and Development,* 43, 203–227.

Cristià, A., Seidl, A., & Gerken, L.A. (2011). Learning classes of sounds in infancy. *U. Penn Working Papers in Linguistics*, 17, 68-76.

De Cat, C., & Plunkett, B. (2002). Qu'est ce qu'i (l) dit, celui-la? Notes méthodologiques sur la transcription d'un corpus francophone. In C.D. Pusch & W. Raible (Eds.), *Romanistische Korpuslinguistik: Korpora und gesprochene Sprache* (=ScriptOralia; 126).Tubingen: Narr, CD-rom.

Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, 35, 99–127.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822–15826.

Fitch, W.T., & Hauser, M.D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science,* 303, 377–380.

Gomez, R.L., & Gerken, L. (1999). Artificial language learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109–135.

Gonzalez-Gomez, N., & Nazzi, T. (2012). Acquisition of non-adjacent phonological dependencies in the native language during the first year of life. *Infancy*, 17, 498-524.

Gonzalez-Gomez, N., Hayashi, A., Tsuji, S., Mazuka, R., & Nazzi, T. (2014). The role of the input on the development of the LC bias: A crosslinguistic comparison. *Cognition*, 132, 301-311.

Greenfield, P.M. (1991). Language, tools, and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14, 531–595.

Hamann, C., Ohayon, S., Dube, S., Frauenfelder, U., Rizzi, L., Starke, M., & Zesiger, P. (2003). Aspects of grammatical development in young French children with SLI. *Developmental Science*, 6, 151–158.

Hauser, M.D., Newport, E.L., & Aslin, R.N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, 78, B53–B64.

Hunkeler, H. (2005). Aspects of the evolution of the early lexicon in mother–child interactions: case study of two dizygotic twin children between 15 and 26 months. Unpublished ms. University of Rouen.

Ingram, D. (1974). Fronting in child phonology. *Journal of Child Language*, i, 233-241.

Johnson, E.K., & Tyler, M. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science,* 13, 339-345.

Jusczyk, P.W., Cutler, A., & Redanz, N.J. (1993). Preference for the predominant stress patterns of English words. *Child Development*, 64, 675-687.

Kirkham, N.Z., Slemmer, J.A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35–B42.

Kuo, L.J. (2009). The role of natural class features in the acquisition of phonotactic regularities. *Journal of Psycholinguistic Research*, 38, 129-150.

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*, 527-531.

MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (1999). Origin of serial-output complexity in speech. *Psychological Science*, *10*, 459-460.

MacWhinney, B. (1995). *The CHILDES project: Tools for analysing talk* (2nd edn.). Hillsdale, NJ: L. Erlbaum.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analysing talk* (3rd edn.). Mahwah, NJ: Lawrence Erlbaum Associates.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental science*, 11, 122-134.

Mersad, K., & Nazzi, T. (2012). When Mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning & Development*, 8, 303-315.

Mielke, J. (2008). *The emergence of distinctive features*. New York: Oxford University Press.

Morgenstern, A. (2006). *Un JE en construction. Ontogenese de l'auto-designation chez l'enfant*. Ophrys: Bibliotheque de Faits de langues.

Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: New evidence from consonants in coda positions. *Language and Speech*, *52*, 463-480.

Nazzi, T., Bijeljac-Babic, R., & Bertoncini, J. (2009). Early emergence of a perceptual LC bias. *Journal of the Acoustical Society of America*, 126, 1440-1446.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique*, 101, 447-462.

Newport, E.L., & Aslin, R.N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162.

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non) words, (non) words, (non) words: evidence for a protolexicon during the first year of life. *Developmental Science,* 16, 24-34.

Onishi, K.H., Chambers, K.E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83, B13-B23.

Romberg A.R., & Saffran J.R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews-Cognitive Science,* 1, 906–914.

Saffran J.R., & Wilson, D.P. (2003). From syllables to syntax: Multi-level statistical learning by 12-month-old infants. *Infancy*, 4, 273–284.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27-52.

Saffran, J.R., & Thiessen, E.D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39, 484–494.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-olds. *Science*, 274, 1926–1928.

Saffran, J.R., Newport, E.L., & Aslin, R.N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.

Sato, M., Vallee, N., Schwartz, J.L., & Rousset, I. (2007). A perceptual correlate of the labial–coronal effect. *Journal of Speech Language and Hearing Research*, 50, 1466–1480.

Savage-Rumbaugh, E.S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, 58, 1–221.

Seidl, A., & Buckely, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development*, 1, 289–316.

Suppes, P., Smith, R., & Leveille, M. (197). The French s ntax of a child’s noun phrases. *Archives de Psychologie*, 42, 207–269.

Swingley, D. (2005). 11-month-olds’ knowledge of how familiar words sound.*Developmental science*, *8*(5), 432-443.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10:21.

Vallée, N., Rousset, I., & Boë, L. J. (2001). Des lexiques aux syllabes des langues du monde. Typologies, tendances et organisations structurelles. *Linx. Revue des linguistes de l’université Paris X Nanterre*, 45, 37-50.

Werker, J.F., Yeung, H.H., & Yoshida, K.A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, 21, 221-226.

Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. Developmental Psychology, 47, 1220.

Appendix

Word-initial LC and CL frequencies and ratios for each consonant pair in French words according to *Lexique 3* (left) and *CHILDES* (right). *Lexique* frequencies correspond to the number of occurrences of a given pattern per word million. Ratios above 1 indicate an LC bias, ratios below 1 indicate a CL bias (marked in **bold**).

Pair	Lexique 3			CHILDES		
	LC Frequency	CL Frequency	Ratio	LC Frequency	CL Frequency	Ratio
p-t	3654	581	6.29***	1484	160	9.28***
p-d	764	1481	**0.52*** **	74	83	**0.89** ns
p-s	4273	1661	2.57***	957	261	3.67***
p-ʃ	470	294	1.6***	84	136	**0.62*** **
p-z	709	1	611.35***	167	0	---
p-ʒ	190	92	2.07***	142	4	35.50***
p-n	830	47	17.71***	65	1	65.00***
p-l	1064	272	3.92***	344	302	1.14†
b-t	1330	1230	1.08*	279	627	**0.44*** **
b-d	291	1009	**0.29*** **	68	224	**0.30*** **
b-s	510	1257	**0.41*** **	91	51	1.78***
b-ʃ	470	436	1.08 ns	278	112	2.48***
b-z	576	18	32.64***	290	23	12.61***
b-ʒ	302	274	1.1 ns	225	78	2.88***
b-n	451	210	2.15***	401	13	30.85***
b-l	927	655	1.41***	556	8	69.50***
f-t	1183	38	31.18***	227	0	---
f-d	304	930	**0.33*** **	71	99	**0.72***
f-s	1353	826	1.64***	280	178	1.57***
f-ʃ	139	369	**0.38*** **	25	29	**0.86** ns
f-z	210	5	40.84***	50	0	---
f-ʒ	74	37	2.01***	1	0	---
f-n	962	266	3.62***	384	22	17.45***
f-l	2050	12	167.35***	198	0	---
v-t	1160	11	107.32***	215	0	---
v-d	492	3477	**0.14*** **	110	93	1.18ns
v-s	334	3266	**0.10*** **	22	173	**0.13*** **
v-ʃ	84	673	**0.13*** **	160	216	**0.74*** *
v-z	979	0	---	36	0	---
v-ʒ	95	65	1.46***	3	2	1.50 ns
v-n	1735	892	1.95***	83	92	**0.90** ns
v-l	3315	1273	2.6***	246	666	**0.37*** **
m-t	4580	253	18.1***	1977	92	21.49***
m-d	1742	2397	**0.73*** **	731	239	3.06***
m-s	1347	836	1.61***	211	4	52.75***
m-ʃ	490	467	1.05 ns	220	58	3.79***
m-z	1184	11	107.63***	305	49	6.22***
m-ʒ	453	1261	**0.36*** **	950	84	11.31***
m-n	1648	180	9.18***	201	12	16.75***
m-l	2599	500	5.2***	422	104	4.06***

Chi-square test of goodness-of-fit: ns: p > 0.10; †: p ≤ 0.09; *: p ≤ 0.05; **: p ≤ 0.01; ***: p ≤ 0.001.