Functional load and frequency as predictors of consonant emergence across five languages

Margaret CYCHOSZ*
University of California, Berkeley
Department of Linguistics
1203 Dwinelle Hall
Berkeley, CA 94720, United States

*Corresponding author: Margaret Cychosz, 1203 Dwinelle Hall, Department of Linguistics, University of California, Berkeley, mcychosz@berkeley.edu

Abstract: Recent cross-linguistic inquires suggest that consonants do not emerge in a universal, preordained order in child speech. The order of emergence is instead heavily influenced by the ambient language. An influential hypothesis in this realm predicts that phones with a high functional load, those that contrast many words in the child's environment, may predict the order of emergence as children selectively converge upon meaningful contrasts. Current statistical modeling techniques and large databases of naturalistic child language now permit an expansive cross-linguistic test of this hypothesis. Crucially, we can control for additional factors known to influence the order of speech sound development such as phone frequency and articulatory complexity. Results demonstrate that the role of functional load is language-dependent: functional load predicts the order of consonant emergence in three of five tested languages. Additional models fit to bootstrapped corpus data suggest that frequency is often the stronger predictor of emergence.

Keywords: speech; child development; input statistics; phonology; corpus

Introduction

In the first years of life, children advance rapidly from producing their first babbling patterns to eventually articulating entire words in their native language. For decades, it was assumed that the order that consonant phonemes reliably appeared in a child's speech, or CONSONANT EMERGENCE, was universal and that children followed an implicational hierarchy of sound production (Jakobson, 1941/1968). For example, a child's production of the fricative [s] implied that the child was already producing the plosive [t]. This order of emergence was strict, without room for environmental effects like phoneme frequency or language phonotactics. Children would follow the order regardless of the language of exposure.

Such a proposal has been largely disproven (Edwards & Beckman, 2008; Vihman, 1993). Now we understand that even very early speech patterns such as babbling are heavily influenced by the native language (de Boysson-Bardies & Vihman, 1991). Still, the cross-linguistic similarities in speech sound emergence are uncanny and something must account for them. Non-cognitive factors, such as how many coordinated oral movements are required to produce a certain sound, have some explanatory power. However, many factors simultaneously predict when a child will first start producing a sound of their native language: articulatory difficulty (Green et al., 2000), frequency in the input (Edwards et al., 2015), and word structure (Vihman, 2014).

All of these factors contribute to consonant emergence. Yet the relative contribution of each has eluded speech development researchers. This is due in part to the (until quite recently) paucity of naturalistic child language data. But the question of what factors predict consonant emergence has also remained unanswered due to the logic that each factor's role for speech development likely varies by language and interacts in unique ways depending upon the linguistic system. For example, with a lingual constriction towards the front, and at times back, of the mouth, the lateral approximant [l] has a complex articulation. It is, accordingly, late to emerge in child English (Lin & Demuth, 2015). However, children acquiring Quiché Mayan begin to use lateral approximants fairly early in development, as soon as 1;7, perhaps because [l] is highly frequent in the ambient language (Pye et al., 1987). If solely the articulatory complexity of a phoneme predicted its emergence in child speech, we should not see such differences between languages.

Beyond frequency and articulatory complexity, another predictor of when children start using a sound may be functional load (FL). Though it is a natural correlate with frequency, FL, unlike frequency, encompasses semantic contrast and the structure of the mental lexicon. This may be critical for development since phonetic categories derived from the lexicon, over those inferred purely from input distributions, have resulted in more robust category acquisition, at least in models of infant learners (Feldman et al., 2013).

Formally, FL has been defined as the entropy reduction a system undergoes due to minimal pair loss (Hockett, 1955). In implementation, FL can measure the effect of removing a phoneme from a language by quantifying how many minimal word pairs that phoneme distinguishes (e.g. pat~bat).

The relevance of FL for language harkens back to the early 20th century (Mathesius, 1929; Trubetskoy, 1939). Recently, interest has resurfaced as FL may explain sound mergers in language change (Oh et al., 2013; Surendran & Levow, 2004; Surendran & Niyogi, 2003; 2006; Wedel, Jackson, & Kaplan, 2013; Wedel, Kaplan, & Jackson, 2013). This FUNCTIONAL LOAD HYPOTHESIS has also been influential in phoneme emergence. There, the FL Hypothesis predicts that phones that contrast many words in the child's linguistic environment will emerge first as children selectively converge upon the most meaningful contrasts (Dietrich et al., 2007). The relevance of FL or "maximal opposition" for consonant emergence was suggested as early as Pye et al. (1987: 187), though the authors stressed that it lacked formalization. Elsewhere, the FL's potential as a metric in development is often suggested, but not implemented (So & Dodd, 1995) or conflated with frequency (Amayreh, 2003).

Two studies have explicitly tested the FL Hypothesis for consonant emergence. Like the relationship between the emergence and articulatory complexity of the lateral approximant [l], the role of FL may differ by language. Van Severen et al. (2013) found that FL was more predictive of consonant emergence than frequency in Dutch. However, since the authors only analyzed Dutch, it is difficult to generalize from the results. Stokes & Surendran (2005) compared the role of frequency and FL in Cantonese and English and found that FL was a predictor of speech sound emergence for English, but not Cantonese. They attributed this to the high FL of tone in Cantonese.

These studies calculated FL differently and made different methodological choices, such as the decision to calculate FL over child-directed versus adult-directed speech corpora. Consequently, it is not clear if the different conclusions they came to regarding the role of FL and frequency for consonant emergence are due to the languages studied or methodological choices. A standard FL calculation, measured over many languages, can address what parameters of the surrounding environment predict speech emergence and the universality of these parameters across different children and languages.

In this work we ask: is the contribution of FL and frequency the same for all children? Or does it depend on the language of exposure? Furthermore, though the predictive role of FL for the age of consonant emergence is intuitive, why a child would prioritize the use of highly contrastive phones is less straightforward. It could be that learners of distinct languages employ distinct distributional learning mechanisms – mechanisms that take advantage of a language's structural idiosyncrasies such as the phonotactic structure or phoneme inventory. Here this question is addressed by computing FL and phone frequency over the child-directed speech (CDS) of five typologically diverse languages that vary by consonant inventory size and structure as well as word composition. This cross-linguistic comparison evaluates if and how frequency and FL vary by the language of exposure.

Methods

Corpora Preprocessing

FL and frequency calculations were made over naturalistic, monolingual corpora of American English, Japanese, Shenzhen Mandarin, Peninsular Spanish, and Turkish available in CHILDES (MacWhinney, 2000) (Table 2). Language selection criterion was based on: 1) if large-scale, naturalistic, monolingual corpora were available and 2) if the corpus was already phonologically transcribed/if the language was relatively orthographically transparent to permit algorithmic grapheme-to-phoneme conversion. Only CDS directed towards the child from 1;0-3;0 and from the target child's mother, father, grandparents, and adult interlocutors was included. Sibling utterances were excluded since age and presence was corpus-dependent. To further increase corpus generalizability, the following were also removed: all proper nouns, with the exception of familial terms (e.g. "Mama"), child- and family-specific forms, second language items, and investigator speech.[i] Details of corpora size and composition are listed in Table 1. Discrepancies in corpus size are counteracted in a bootstrap procedure in the Results section. Code to replicate corpus cleaning procedures, FL/frequency measurements, and statistical model fitting are available in the OSF project page associated with this article (Author, 2018).

Table 1: Description of corpora used to compute functional load and frequency

LANGUAGE	CORPUS	SIZE (TOKENS)	SIZE (TYPES)	PARTICIPANTS	REFERENCE
Amer. English	Bernstein-Ratner; Brent-Ratner	32,993	1,321	(*N*=9) children interacting with mothers	Bernstein-Ratner (1987); Brent & Cartwright (1996)
Japanese	MiiPro	235,705	10,412	(*N*=4) children interacting with mothers and occasionally father (for 2 children)	Miyata (2012)
Shenzhen Mandarin	Tong	72,908	2,200	(*N*=1) child interacting with mother, father, grandmother, & occasionally grandfather	Deng & Yip (2017)
Peninsular Spanish	Aguirre	44,440	2,304	(*N*=1) child interacting with mother, father, & occasionally family friend	Aguirre (2000)
Turkish	Aksu; Altinkamis	10,977	2,216	(*N*=9) children interacting with mother, father, and occasionally grandmother	Slobin (1982); Türkay (2005)

The Brent-Ratner corpus for American English was already transcribed phonologically. The MiiPro Japanese corpus was transcribed in Hepburn Romanization which is orthographically transparent (Miyata, p.c.). However, the Spanish, Mandarin (Pinyin transcription), and Turkish corpora were only orthographically annotated. Consequently, a grapheme-to-phoneme conversion, utilizing the Montreal Forced Aligner (McAuliffe et al., 2017), was conducted for all three. Lexical items found in the Spanish, Mandarin, and Turkish corpora that did not have corresponding representation in the dictionary were discarded. These unknown words made up 0.39%, 0.86%, and 10.71% of the Spanish, Mandarin, and Turkish corpora, respectively. Although the crucial role of tone for contrast in Mandarin is acknowledged, we follow the methodology of Stokes & Surendran (2005) and measure the relationship of segments and FL. Consequently, tone was removed from the Mandarin corpus.

Formalizing Functional Load

In its basic form, FL has been defined as the system entropy reduction resulting from minimal pair loss (Hockett, 1955). However, for the present work, FL is instead formalized as information loss following Surendran & Niyogi (2006):

$$FL_U(a) = \frac{H(L_U) - H(L_U^{-a})}{H(L_U)}$$

where a is the linguistic unit (phone or word), H is entropy, and L_U is the language system (i.e. lexicon). This information-theoretic approach assumes that any given language system is composed of discrete units (phones, words) (Shannon, 1948). This calculation was made at the word level, but phoneme level is another alternative (see Surendran & Niyogi, 2003; 2006). A common approach to FL calculation (of phones) is a weighted sum FL of the phones with which the phone could merge or a binary contrast between two phones. In previous instantiations of phoneme FL, calculated to predict language diachrony for example, the merger of a phone with these "neighbors" makes sense. Sounds rarely disappear completely from a language inventory – they merge with or split from acoustic and articulatory neighbors (e.g. /k/ > /tʃ/ before front vowels). However, there is no evidence to suggest that speech emerges in this manner in ontogeny, with neighboring phonemes diverging from one another once they reach a particular frequency or FL threshold in the lexicon (though this is an interesting hypothesis). Consequently, here FL is calculated solely over individual phones, representing the importance of the single phoneme for the child, and not in relation to a subset of other phones in the language system.

Phone frequency was measured as the number of occurrences in the corpus divided by the number of total phones in the corpus. FL and frequency were calculated over type frequencies. Van Severen et al. (2013) found that FL and frequency calculated on word types correlated more strongly with consonant emergence than calculations over word tokens. This conclusion supports other work arguing that when the lexicon is used to bootstrap into phonological categories, children employ word types, not tokens (Edwards et al., 2015).

FL was measured over the entire consonant inventory of each language (Table 2) except Japanese geminate stops such as /pp/[ii] and Japanese /ʃ/ because there was not developmental data on these segments. Spanish /s/ was also excluded because the Spanish CDS corpus is a *ceceo* dialect (orthographic 'z' and 'c' are realized as [θ] instead of [s]) and the developmental data report on non-*ceceo* dialects (Lipski, 1994). Finally, nasals like /n/ and /m/ are excluded in all languages because they are ubiquitous from early babbling. They likely emerge too early to be lexically meaningful and are not good candidates to test the FL Hypothesis. The decision to make these exclusions was made prior to any data analysis.

Table 2: Consonants measured

LANGUAGE	STOPS	AFFRICATES	FRICATIVES	LIQUIDS/GLIDES
English	p, t, k, b, d, g	tʃ, dʒ	f, v, θ, ð, s, z, ʃ, ʒ	l, ɹ, w, j
Japanese	p, t, k, b, d, g	ts, tʃ	s, z, h	r, w, j
Mandarin	p, t, k, p^h, t^h, k^h	ts, ts^h, tʂ, tʂ^h, tɕ, tɕ^h	s, f, ʂ, ɕ, x	ɻ̩, l
Spanish	p, t, k, b, d, g	tʃ	f, x	l, ɾ, r, j̞
Turkish	p, t, k, b, d, g	tʃ, dʒ	s, f, v, ʃ, ʒ, ɣ, h	l, r, j

There is disagreement concerning best practices for FL calculation, even within child language development. Stokes & Surendran justify the choice to calculate FL only in word-initial position since "children pay attention to the onsets of words" (2005: 581). However, this is not universal. In early word production, French children actually tend to omit word-initial segments, likely due to exclusive word-final stress in French (Vihman, 2014). Consequently, here FL is calculated over all segments. Elsewhere, FL calculations are limited to the lemma (Wedel, Jackson, & Kaplan, 2013). But since Turkish, a highly agglutinating language, is included in this analysis, all inflected and derived forms in the remaining languages are also included. Unfortunately, including this greatly increases the occurrence of English /s/ and /z/ – the plural allomorphs. It was decided to exclude /s/ and /z/ from the English analysis. Future work will compare the effects of frequency and FL in morphologically decomposed corpora and will include English /s/ and /z/.

Developmental data

Age of consonant emergence (AoE) was determined from previous peer-reviewed works of developmental phonology. Studies employ distinct metrics to qualify a consonant as "emerged" in a child's phonological repertoire: if the sound is produced correctly two times in a picture naming task (Dinnsen et al., 1990), or if 75% of the children produced the sound correctly in word-initial and final position (Prather et al., 1975), or 90% of participants articulated the sound correctly, in the proper position, in at least two words (So & Dodd, 1995). Data collection methodologies – naturalistic, elicited, etc. – also differ. Table 3 lists studies referenced and the metrics employed.[iii]

Table 3: Metric for consonant emergence

LANGUAGE	METRIC FOR EMERGENCE	METHOD	REFERENCE
English	75% of children (*N*=147) used consonant in initial and final position	picture-prompted word elicitation	Prather et al. (1975)
Japanese	mean age of first appearance across (*N*=10) children	picture-prompted word elicitation	Nakanishi (1982); Ota (2015)
Mandarin	90% of children in the given age group (~*N*=20) produced sound one time (*N*=129 children total)	picture-prompted word elicitation	Hua & Dodd (2000)
Spanish	occurred at least two times during elicitation from *N*=16 children	picture-prompted word elicitation	Cataño et al. (2009) (metanalysis)
Turkish	produced in all seven possible word positions by *N*=22 children	word elicitation during structured play session	Topbaş (1997)

Given discrepancies between studies, the metric for consonant emergence is not standard. For example, consonants appear to emerge much later in the English data but this is due to the more stringent emergence criterion that Prather et al. (1975) used. As a result, AoE is not directly comparable between languages but it was consistent within each language. When a range of ages was specified for a phone (e.g. for /d/ at 18-22 months), mean month is AoE. The result is that AoE is not a continuous variable but instead reflects discrete, chronologically ordered developmental stages.

Though age and order of emergence can be unstable measures in language development – individual children may deviate from language-specific norms (Lahey et al., 1992) – we are attempting to replicate the methodologies of the two previous papers on this topic (Stokes & Surendran, 2005; Van Severen et al., 2013). Both of these works used AoE as the outcome measure. Furthermore, we stress that for each language studied, emergence was determined by measuring across *multiple* children and should be relatively robust to individual differences.

Results

Single sample correlations

Figure 1 maps the relationship of normalized FL to age of emergence (AoE). (Though it is the outcome variable, AoE is listed on the x-axis throughout the Results to reflect a chronological timeline.) To compare FL measurements between languages, FL was normalized by the sum of all FL calculations within each language: $FL(a) / \sum Fx_i$. FL was negatively correlated with AoE for each language, but the strength of the relationship varied widely: English (Pearson r = -.50), Japanese (r = -.29), Mandarin (r = -.04), Spanish (r = -.13), and Turkish (r = -.58). These correlations suggest that in some languages, children tend to acquire phones with higher FL first. However, there is almost no relationship between FL and AoE in Mandarin.

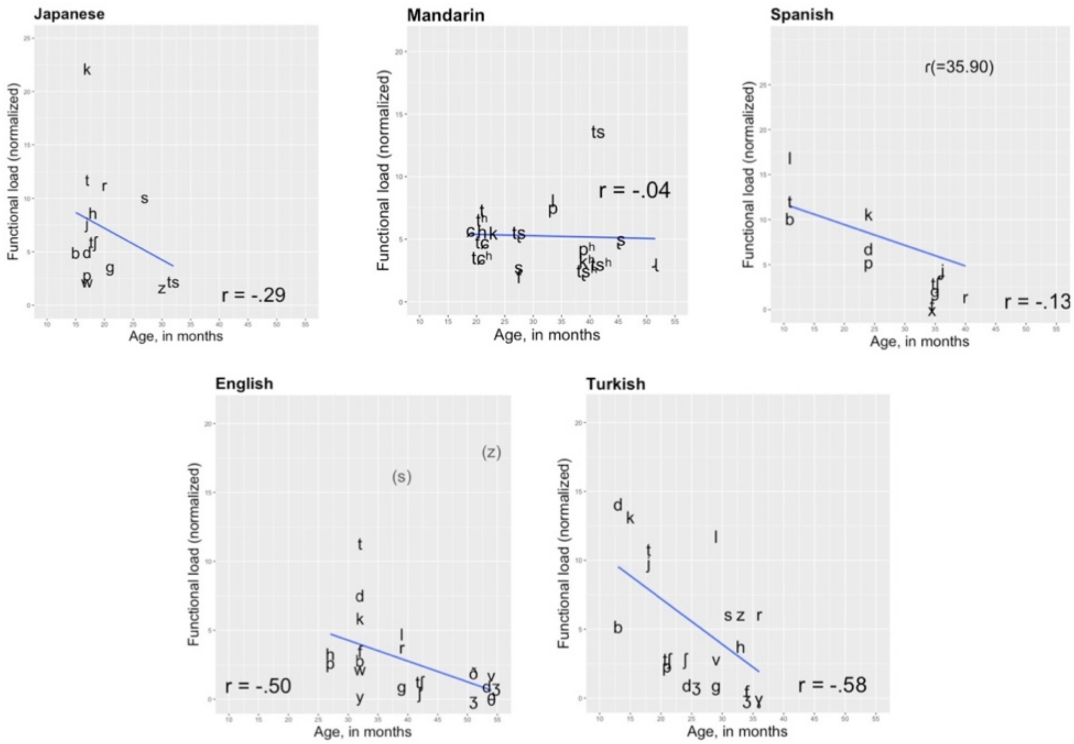

Figure 1: Functional load by age of emergence:
single sample (not bootstrapped)

Comparing frequency to FL for each language, results show that frequency is similarly negatively correlated with AoE for each language: English (*r* = -.39), Japanese (*r* = -.20), Mandarin (*r* = -.47), Spanish (*r* = -.58), and Turkish (*r* = -.29). FL correlated more strongly than frequency with AoE for three out of the five languages. This suggests that FL plays a role in consonant emergence for some languages. However, single-sample correlations are wholly inadequate to address the research questions. For one thing, FL and frequency are highly correlated: a phone that contrasts many words in a language is by necessity going to be more frequent. To further investigate these correlational trends, control for factors that are interwoven with FL, and avoid corpus sampling biases, we scaled up the analysis by simulating consonant emergence via a bootstrapping with replacement procedure.

Bootstrapping models

A bootstrapping with replacement procedure was employed over each of the CDS type-frequency language corpora in one hundred samples of 750 words.[iv] FL was then normalized over the sum of phone FL measurements within each sample.

Stepwise cumulative link models were fit to the bootstrapped data for each language using the clm function in the R package ordinal (Christensen, 2015). These models are similar in implementation and interpretation to other regression models fit via maximum likelihood estimation but are instead specified for ordinal response variables. Forwards stepwise model fitting was evaluated through log-likelihood tests, AIC comparison, and parameter significance within model summary.

Multivariate linear and even logistic models cannot perform as well as clms on this dataset. Linear models predict values outside of the realistic range of the response variable meaning that the model would predict a relationship both before the age children actually begin producing consonants and a relationship after all the native consonants have emerged in children's speech. Furthermore, AoE is not a continuous developmental period. It consists of discrete, chronologically ordered developmental stages. This is because diary reports on phonological development are not continuous, but may, for example, give a report of the child's phonological repertoire every six or eight weeks. Because this reporting 1) differed between languages and 2) could be inconsistent within studies (i.e. noting the child's productions irregularly every 4-6 weeks), the variable AoE was further binned into standardized developmental periods of 3 months (e.g. year;month: 0;11-1;1, 1;2-1;5). Binning into developmental periods standardizes the measurement for all studies referenced on all languages tested. Binning also more closely approximates the discrete measures of phonological development employed in the studies referenced here.

Still, while binning into 3 month time frame permits some flexibility in AoE, the fact remains that not all English-learning children will, for example, start reliably producing /k/ between 32 and 34 months. Consequently, some developmental variability was modelled into the binned AoE metric for each language by randomly sampling the 100 AoE measurements around the mean AoE for each phone with σ =1.

The choice to fit multiple models, one for each language, instead of a large model with a factor for language was made for several reasons. First, the theoretical interest here is understanding how these predictors of consonant emergence differ by language. However, we already know that a child's speech development varies as a function of the language to which they are exposed (Edwards & Beckman, 2008). More specifically, previous research suggests that the role of FL and frequency for speech emergence will vary by language (Stokes & Surendran, 2005; Van Severen et al., 2013). Therefore, our interest is not in one large, explanatory model of frequency and FL because we acknowledge that these differences exist.

Furthermore, fitting separate models by language aids in model parameter interpretability. While one can interpret an interaction of frequency with a language, it is much less clear how to interpret a three-variable interaction of frequency, FL, and a language (i.e. the role of frequency on AoE is language-dependent but also FL-dependent?).

Finally, and perhaps most crucially, researchers are increasingly acknowledging the inflated risk of Type I errors that multiple comparisons make even in mixed models. The addition of a third factor, language, to a model would increase the number of comparisons from 15 to 35. Note in any case that the parameters of FL and frequency are still standardized between languages, so a quantitative comparison between language models can still be made.

The final step before fitting models was to include a parameter controlling for the articulatory demands of the phones. When discussing child speech, an obvious concern about motor limitations arises. No model of emergence is complete without an articulatory complexity metric but it is surprisingly difficult to quantify. A final parameter, **articulatory_complexity**, was modeled as a fixed effect in each of the models to control for its acknowledged role on speech emergence (McAllister Byun, 2012; Smith & Zalaznik, 2004). The scale used for the parameter articulatory_complexity is adopted with modification from Stokes & Surendran (2005) (Table 4).[v]

Table 4: Articulatory complexity metric

LEVEL	DESCRIPTION	PHONES
1	-Rapid/ballistic movement -Slow/progressive movement	p, w, h
2	-Some lingual control for frication -Velar place of articulation -Laryngeal mastery	t, k, b, d, g, j, f, p^h, t^h, k^h
3	-Tongue tip and dorsum manipulation	r, ɹ, ɻ, ɾ, l
4	-Complete lingual control for fricatives -Transition from ballistic to frication	s, z, ʃ, ð, θ, v, ɉ, ʒ, x, ɣ, tʃ, dʒ, ts, ts^h, ʂ, tʂ, $tʂ^h$, ɕ, tɕ, $tɕ^h$

Best model fits resulted in parameter coefficients for **frequency** and **FL** that varied by language (Table 5).[vi] Because the frequency and FL covariates were standardized via z-score normalization, a quantitative comparison can be made although each language was fit to a separate model. Negative linear coefficients signify that as a phone's FL/frequency increases, the AoE decreases: a more frequent phone, or a phone with higher FL, will emerge earlier. Positive linear coefficients indicate that FL still predicts AoE, but FL becomes more relevant as children progress through developmental stages.

Table 5: Frequency and FL predicting AoE by language

	ESTIMATE	S.E.	Z-VALUE	P-VALUE	97.5% CI
ENGLISH					
Frequency	-1.01	0.12	-8.51	***	-1.25, -0.78
Func. Load	0.25	0.09	2.79	**	0.08, 0.43
JAPANESE					
Frequency	-0.22	0.08	-2.66	**	-0.38, -0.06
Func. Load	-0.003	0.06	-0.05	n.s.	-0.12, 0.11
MANDARIN					
Frequency	-1.05	0.06	-18.79	***	-1.17, -0.94
Func. Load	-0.01	0.05	-0.23	n.s.	-0.11, 0.09
SPANISH					
Frequency	-2.89	0.16	-18.09	***	-3.21, -2.58
Func. Load	-0.89	0.09	-9.33	***	-1.07, -0.70
TURKISH					
Frequency	-1.31	0.07	-18.39	***	-1.45, -1.17
Func. Load	0.37	0.05	7.03	***	0.27, 0.48

$p < .01$ **, $p < .001$ ***

The significance of FL and frequency for these models indicates that, alone, articulatory complexity does not explain when children first produce consonants. For three languages, Spanish, English, and Turkish, both frequency and articulatory complexity are insufficient to explain AoE. Spanish-learning children learn higher FL and higher frequency phones before lower FL and lower frequency phones, even after controlling for articulatory complexity (Figures 2, 3). FL was also a significant predictor, along with frequency, for English and Turkish but in a reverse direction (English: β=0.25, ***, CI=0.08, 0.43; Turkish: β=0.37, ***, CI=0.27, 0.48). Phones with greater FL will tend to be acquired at later developmental stages. This could mean that children learning these languages only learn to employ contrast within the lexicon as they age which would coincide with the child's growing receptive and productive vocabularies. Finally, despite the strong relationship between FL and AoE in the single sample correlation from Japanese (r = -.29), after controlling for frequency and articulation, FL does not predict AoE for Japanese. Nor does FL predict AoE in children learning Mandarin. In both cases, less frequent phones are acquired at later developmental stages (Mandarin: β=-1.05, ***, CI=-1.17, -0.94; Japanese: β=-0.22, ***, CI=-0.38, -0.06).

Table 6: Articulatory complexity predicting AoE by language

Measure of articulatory complexity		ESTIMATE	S.E.	Z-VALUE	P-VALUE	97.5% CI
ENGLISH						
	[2]	4.12	0.20	20.42	***	3.73, 4.52
	[3]	9.12	0.36	25.41	***	8.43, 9.84
	[4]	13.58	0.46	29.26	***	12.71, 14.54
JAPANESE						
	[2]	0.15	0.16	0.92	n.s.	-0.17, 0.47
	[3]	2.92	0.28	10.54	***	2.38, 3.47
	[4]	4.18	0.20	21.09	***	3.80, 4.57
MANDARIN						
	[2]	-0.18	0.14	-1.24	n.s.	-0.46, 0.11
	[3]	2.90	0.19	15.41	***	2.53, 3.27
	[4]	0.11	0.14	0.79	n.s.	-0.16, 0.39
SPANISH						
	[2]	-1.34	0.24	-5.51	***	-1.82, -0.86
	[3]	0.30	0.29	1.02	n.s.	-0.27, 0.87
	[4]	-0.98	0.29	-3.44	***	-1.54, -0.42
TURKISH						
	[2]	-0.23	0.17	-1.35	.18	-0.56, 0.10
	[3]	4.16	0.26	15.98	***	3.65, 4.67
	[4]	0.97	0.16	6.26	***	0.67, 1.28

$p < .001$ ***

In sum, frequency predicted AoE, beyond articulatory complexity, for all five languages tested. Beyond these factors, FL also predicted emergence in Turkish, English, and Spanish, albeit in different directions. Yet even in these three languages where FL did play a predictive role upon emergence, frequency was the stronger predictor as demonstrated with the higher normalized coefficients and confidence intervals for each language's model.

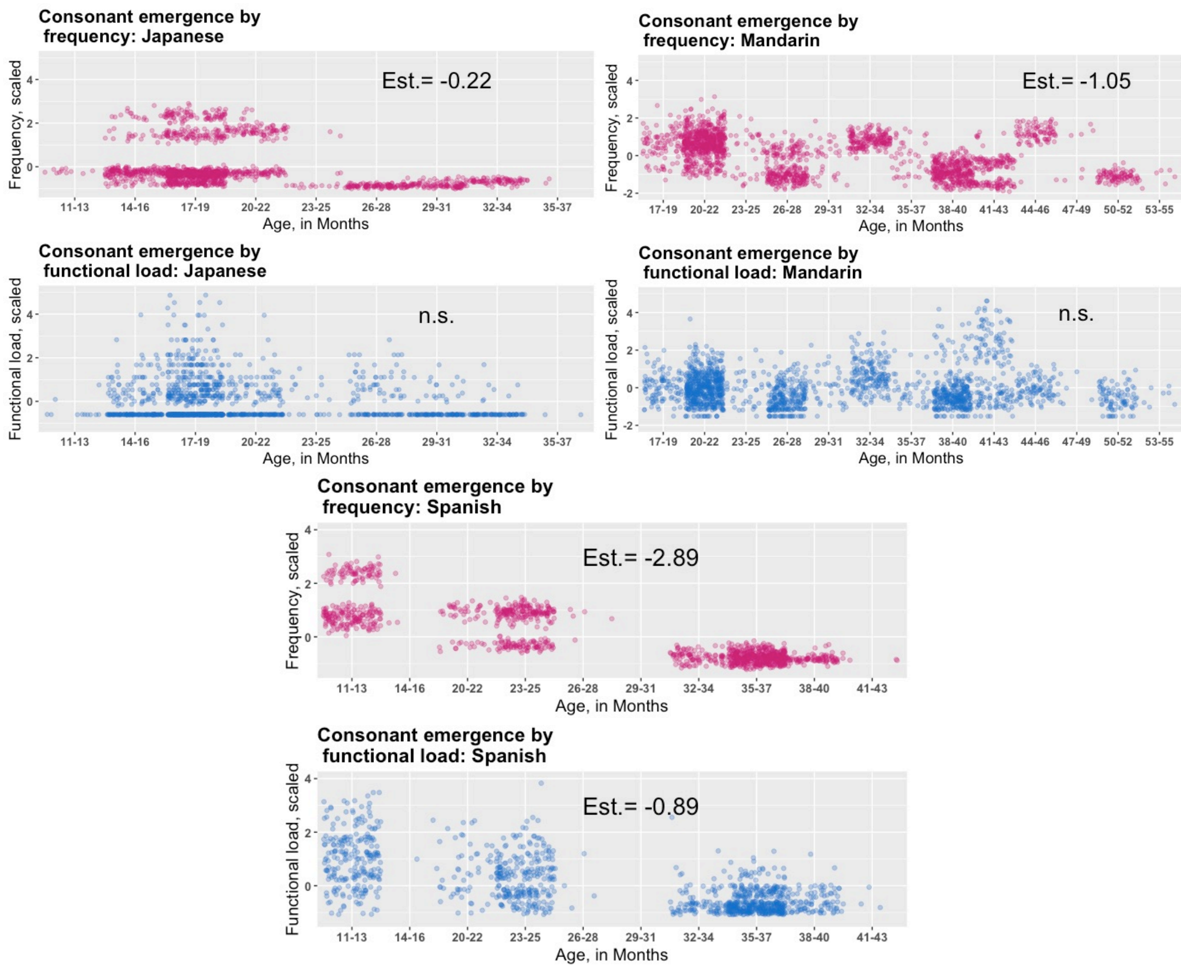

Figure 2. Consonant emergence by functional load and frequency in Japanese, Mandarin, and Spanish. Each point represents the FL or frequency of a phone measured over a 750-word lexicon. ‘Age, in Months’ is modelled 100 times for each phone with μ=mean phone AoE and σ =1.

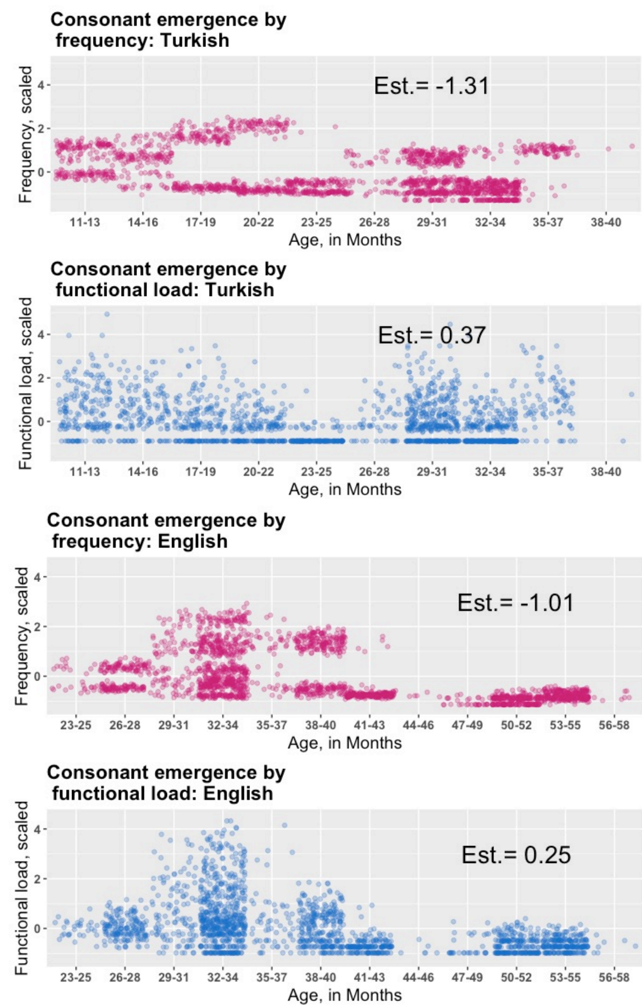

Figure 3. Consonant emergence by functional load and frequency in Turkish and English. Each point represents the FL or frequency of a phone measured over a 750-word lexicon. 'Age, in Months' is modelled 100 times for each phone with μ=mean phone AoE and σ=1.

Discussion

Phoneme frequency influences children's phonological development. The more a child hears a sound, the faster s/he can focus attention to articulating it and producing the corresponding acoustic signal. Yet this intuitive influence of the language environment on development has limitations. Edwards et al. (2015) cite the example of English /ð/. Due to words like 'the' and 'that', /ð/ has exceptionally high token frequency, but low type frequency in English. This explains, in part, why /ð/ emerges relatively late for children learning English. So while frequency is correlated with development, alone, it cannot complete the picture of speech development.

Likewise, many speech sounds emerge late due to non-cognitive factors like children's motor limitations and immature physiology (Green et al., 2000; McGowan et al., 2004; Nittrouer, 1993). But cross-linguistically, the same segment can emerge in child speech at different developmental stages (Edwards & Beckman, 2008). So articulatory demands also cannot fully predict when consonants emerge in child speech.

The model here incorporates both of these factors and tests an additional parameter: functional load. Even when controlling for type frequency in the ambient language and the articulatory complexity of the sounds, children learning English, Turkish, and Spanish manipulate the words in their language environment to inform their timeline of early phone production. This supports previous findings about the importance of the lexicon for speech development in models of infant learners of English (Feldman et al., 2013) as well as conclusions that distinct frequency-based measures make distinct predictions for child phonology (Jarosz, Calamaro, & Zentz, 2017). However, using coefficients that are normalized between the language models, we can conclude that, at least at the segmental level, children acquiring Japanese and Mandarin do not use this ambient information.

There is of course a natural circularity to any argument about ambient language effects and language acquisition. It follows a chicken-or-the-egg logic: do children acquire segments because they are more frequent in the input or more frequently contrast words? Or are those sounds more frequent because they are more "naturally" acquired or easier to articulate? Both explanations are valid and models here suggest that cross-linguistically, universals such as a child's immature motor skills, and language-specific parameters such as frequency, govern speech development. Note, however, that the order of speech development is not the only relevant factor in phonological learning. The age at which a sound is reliably produced is also significant. However such an analysis is not feasible for a typologically-diverse study until we have a larger body of literature measuring children's phonological production with a single, standardized tool. At that time, we can make a comparison between the predictors of consonant emergence order and consonant production reliability. Furthermore, this work built off of previous attempts to quantify the contribution of FL and frequency so ambient language was quantified as CDS. But FL or frequency could differ when computed over the child's *own* lexicon. An important next step is to compare between CDS and, for example, MCDI data over the relevant ages. Additionally, previous work opted to model articulatory complexity, not perceptual confusability; yet, both perception and articulation delimit the order of speech emergence (Smit, et al., 1991).

The languages studied here are diverse; only English and Spanish have any genetic relationship. Structural differences across languages – phoneme inventory, phonotactics, tonal contrasts – may affect the role of frequency/FL. Perhaps the most obvious example is Mandarin, a tonal language. FL did not predict consonant emergence in Mandarin, replicating Stokes & Surendran (2005)'s finding from another tonal language, Cantonese. The authors attributed their conclusion on FL to the high load of tone in Cantonese. The same explanation could be offered for the Mandarin model outlined here because Mandarin's four tones carry a high FL (though not uniformly so, see Surendran & Levow, 2004). In Mandarin, a phonemic form like /ma/ distinguishes between 'mom,' 'horse,' 'hemp,' and 'scold' via tone. This has led many to posit that the unit of representation in some languages such as Mandarin and Cantonese is not an individuated phoneme. Rather, it is a syllable disassociated from tone, in both children and adults (Chen et al., 2002; McBride-Chang et al., 2008; O'Seaghdha et al., 2010; cf. Verdonschot et al., 2015). Further evidence in support of this conclusion is that Mandarin children reliably produce tonal contrasts at an earlier age than vowels or consonants (Hua & Dodd, 2000).

Like Mandarin, FL did not predict emergence in Japanese and, like Mandarin, the representational unit in Japanese may not be the segment. In Japanese the smallest level of phonological representation is the mora, or a weighted syllable, in both adults (Verdonschot et al., 2011) and children (Ota, 2015). If children transact contrasts at the syllabic or moraic level, it follows that FL measured phonemically would not predict speech emergence. This is speculative and only a different instantiation of FL, measured over morae and syllables in Japanese and Mandarin, respectively, could definitively conclude this. (Note that the choice to calculate FL on the basis of phonemes here is in line with previous work on another Chinese language and it standardized the measure across the languages studied.)

In Spanish, English, and Turkish, though FL predicted emergence, frequency was the stronger predictor. This runs counter to Stokes & Surendran (2005)'s conclusion as they found that FL was a better predictor of emergence than frequency. These differences may be due to the measurement of FL over CDS corpora in this study and not adult speech corpora as in that study. Of further interest is the role that FL appears to play at different developmental stages across the languages. In Spanish, phones with higher FL are acquired earlier. But in English and Turkish, though FL did predict emergence, after controlling for articulation and frequency, phones with greater FL were *later* to develop. This suggests that children learning Turkish and English may only employ contrast in the ambient language to inform speech emergence at later developmental stages.

Still, there are several reasons why we interpret these differences in the directionality of FL between Spanish and English/Turkish with caution. First, recall that the G2P converter did not recognize 10.71% of the words in the Turkish corpus. This is anticipated in languages with such a morphologically-complex structure, where the parser will encounter many novel words, but it is nevertheless a limitation. Second, /s/ and /z/, the plural morphemes in English, were so ubiquitous in the English corpus that they risked skewing models and were removed before data analysis. But this is a clear limitation on our English data. A more complete analysis, one computing FL over morphologically decomposed corpora in each of these languages, could include English /s/ and /z/ to illustrate the developmental trajectory of all English phonemes. Only then could we understand which parts of the lexicon children employ as they calculate FL.

There are likewise reasons to hypothesize why we see differences between Turkish/English and Spanish. One explanation for the importance of consonant FL later in development, at least in Turkish, could be Turkish's highly agglutinating structure. In a language like Turkish, the information typically expressed in function words – tense, pronouns, prepositional relations – is expressed in suffixes attached to root words. In more analytic languages, however, these grammatical functions are often separate lexemes. Very little is known about the acquisition or children's representation of morphemes in non-analytic languages (Durrant, 2013). However, instead of computing contrasts over the entire lexicon, children may instead compute primarily across semantically meaningful words such as nouns or action verbs, and focus less upon function and grammatical concepts expressed in morphemes. And of course as a child's vocabulary grows, they continue to acquire nouns and verbs while function morphemes/concepts are relatively finite. In their examination of FL as a cross-linguistic predictor of sound change, Wedel et al. (2013) opted to remove function words entirely from the analysis. Consequently, FL calculation may differ over these separate areas of the Turkish lexicon. Computing FL over datasets with and without grammatical and function words could be a way to test this hypothesis empirically.

In conclusion, replicating some previous results, ambient frequency predicted consonant emergence for all languages studied. In addition, the ability to contrast words, calculated by phone FL, also predicted the order that consonants emerged in Spanish, English, and Turkish. These relationships remain even after controlling for the articulatory demands. However, models suggest that phones with FL emerged later in development for English and Turkish while the opposite was true for Spanish (higher FL phones emerged early). These results reaffirm the contribution of physiology on early speech production, but demonstrate how articulatory universals interact strongly with the language of exposure.

Acknowledgements: For feedback and assistance, thank you to the fall 2017 members of Cognitive Science 190: Computational Models of Cognition (UC Berkeley), especially Michael Levy. Additional thanks to Chris Hench, Ronald Sprouse, Terry Regier, and Jan Edwards.

References

Aguirre, C. (2000). *La adquisición de las categorías gramaticales en español*. Ediciones de la Universidad Autónoma de Madrid.

Amayreh, M. M. (2003). Completion of the consonant inventory of Arabic. *Journal of Speech, Language, and Hearing Research*, *46*(3), 517-529.

Author. (2018). Functional load and frequency as predictors of consonant emergence across five languages. Open Science Framework project: osf.io/284kh/

Bernstein-Ratner, N. (1987). The phonology of parent-child speech. In K. Nelson and A. van Kleeck, editors, *Children's Language*, volume 6. Erlbaum, Hillsdale, NJ.

Brent, M.R., & Cartwright, T.A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61: 93-125.

Cataño, L., Barlow, J. A., & Moyna, M. I. (2009). A retrospective study of phonetic inventory complexity in acquisition of Spanish: Implications for phonological universals. *Clinical Linguistics & Phonetics*, *23*(6), 446-472.

Christensen, R. H. B. (2015). Ordinal - Regression Models for Ordinal Data. R package.

Deng, X., & Yip, V. (2017). A multimedia corpus of child Mandarin: the Tong corpus. *Journal of Chinese Linguistics, 46* (1), 69-92.

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of the Sciences*, *104*(41), 16027-16031.

Dinnsen, D. A., Chin, S. B., Elbert, M., & Powell, T. W. (1990). Some constraints on functionally disordered phonologies: Phonetic inventories and phonotactics. *Journal of Speech, Language, and Hearing Research*, *33*(1), 28-37.

Durrant, P. (2013). Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory*, *9*(1), 1-38.

Edwards, J., Beckman, M. E., & Munson, B. (2015). Frequency effects in phonological acquisition. *Journal of Child Language*, *42*(2), 306.

Edwards, J., & Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development*, *4*(2), 122-156.

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, *120*(4), 751.

Hockett, C. F. (1955). *A manual of phonology* (No. 11). Waverly Press.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language*, *27*(1), 3-42.

Inkelas, S. 2014. *The Interplay of Morphology and Phonology*. Oxford: Oxford University Press.

Itô, J. and Mester, A. (1999). The structure of the phonological lexicon, in N. Tsujimura (ed.), *The handbook of Japanese linguistics*. Blackwell.

Kunnari, S., Nakai, S., & Vihman, M. M. (2001). Cross-linguistic evidence for the acquisition of geminates. *Psychology of Language and Communication*, *5*(2), 13-24.

Jarosz, G., Calamaro, S., & Zentz, J. (2017). Input frequency and the acquisition of syllable structure in Polish. *Language Acquisition*, *24*(4), 361–399.

Lahey, M., Liebergott, J., Chesnick, M., Menyuk, P., & Adams, J. (1992). Variability in children's use of grammatical morphemes. *Applied Psycholinguistics*, *13*(3), 373-398.

Lin, S., & Demuth, K. (2015). Children's acquisition of English onset and coda /l/: Articulatory evidence. *Journal of Speech, Language, and Hearing Research*, *58*(1), 13-27.

Lipski, J. M. (1994). *Latin American Spanish*. London: Longman.

Mathesius, V. (1929). La structure phonologique du lexique du tchèque moderne. *Travaux du cercle linguistique du Prague*, 1:67-84.

McAllister Byun, T. (2012). Positional velar fronting: An updated articulatory account. *Journal of Child Language*, *39*(5), 1043-1076.

McAuliffe, M. Socolof, M., Mihuc, A., Wagner, & M., Sonderegger, M. (2017). Montreal Forced Aligner [Computer program]. Version 1.0.0.

McBride-Chang, C., Tong, X., Shu, H., Wong, A. M. Y., Leung, K. W., & Tardif, T. (2008). Syllable, phoneme, and tone: Psycholinguistic units in early Chinese and English word recognition. *Scientific Studies of Reading*, *12*(2), 171-194.

McGowan, R. S., Nittrouer, S., & Manning, C. J. (2004). Development of [ɹ] in young, Midwestern, American children. *Journal of the Acoustical Society of America*, *115*(2), 871-884.

Miyata, S. (2012). Japanese CHILDES: The 2012 CHILDES manual for Japanese.

Nakanishi, Y. (1982). K_on hattatsu. In R. Uchisugawa and Y. Yasuko (Eds.) *K_za gengo sh_gai chiry_ky_iku 4: K_onsh_gai*. Tokyo: Fukumura Shuppari.

Nittrouer, S. (1993). The Emergence of Mature Gestural Patterns Is Not Uniform Evidence From an Acoustic Study. *Journal of Speech, Language, and Hearing Research*, *36*(5), 959-972.

Oh, Y. M., Pellegrino, F., Coupé, C., & Marsic, E. (2013). Cross-language comparison of functional load for vowels, consonants, and tones. In *Proceedings of Interspeech 2013* (pp. 3032-3036).

O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115*(2), 282-302.

Ota, M. (2015). L1 phonology: phonological development. In H. Kubozono (Ed.), *Handbook of Japanese Language and Linguistics, vol. 2*. Berlin: Mouton de Gruyter.

Prather, E. M., Hedrick, D. L., & Kern, C. A. (1975). Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders*, *40*(2), 179-191.

Pye, C., Ingram, D., & List, H. (1987). A Comparison of Initial Consonant Acquisition in English and Quiché, in Keith Nelson and Anne van Kleeck (Eds.), *Children's Language*, *Vol. 6*. Erlbaum.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423. 623–656.

Slobin, D. (1982). Universal and particular in the acquisition of language. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art*. New York: Cambridge University Press.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1991). The Iowa Articulation Norms Project and Its Nebraska Replication. *Journal of Speech Language and Hearing Research*, *34*(2), 446.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, *45*(1), 22-33.

So, L. K., & Dodd, B. J. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, *22*(3), 473-495.

Stokes, S. F., & Surendran, D. (2005). Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children. *Journal of Speech, Language, and Hearing Research*, *48*(3), 577-591.

Surendran, D., & Niyogi, P., (2003). *Measuring the functional load of phonological contrasts (Tech. Rep. No. TR-2003-12).* Chicago: Department of Computer Science, University of Chicago.

Surendran, D., & Niyogi, P. (2006). Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In O. N. Thomsen (Ed.), *Competing models of linguistic change: Evolution and beyond (pp. 43–58).* Amsterdam: John Benjamins.

Surendran, D., & Levow, G. (2004). The Functional Load of Tone in Mandarin is as High as that of Vowels. In *Proceedings of the International Conference on Speech Prosody 2004* (pp. 99–102). Nara, Japan.

Topbas, S. (1997). Phonological acquisition of Turkish children: implications for phonological disorders. *European Journal of Disorders of Communication*, *32*(4), 377-396.

Trubetzkoy, N. (1939). Grundzüge der phonologie. *Travaux du cercle linguistique du Prague: 7*.

Türkay, A. F. (2005). *Children's Early Lexicon in terms of Noun-Verb Dominance*. Unpublished PhD Dissertation. Cukurova Univ., Adana, Turkey.

Wedel, A., Jackson, S., & Kaplan, A. (2013). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech*. *56*(3), 395-417.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, *128*(2), 179-186.

Van Severen, L., Gillis, J. J., Molemans, I., Van Den Berg, R., De Maeyer, S., & Gillis, S. (2013). The relation between order of acquisition, segmental frequency and function. *Journal of Child Language*, *40*(4), 703-740.

Verdonschot, R. G., Kiyama, S., Tamaoka, K., Kinoshita, S., La Heij, W., & Schiller, N. O. (2011). The functional unit of Japanese word naming: evidence from masked priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1458.

Verdonschot, R. G., Lai, J., Chen, F., Tamaoka, K., & Schiller, N. O. (2015). Constructing initial phonology in Mandarin Chinese: Syllabic or subsyllabic? A masked priming investigation. *Japanese Psychological Research*, *57*(1), 61-68.

Vihman, M. M. (2014). *Phonological development: The first two years*. 2nd edn. Hoboken: Wiley-Blackwell

[i] These forms are annotated in the corpora according to CHAT transcription standards (e.g. second-language form=@s:*) (MacWhinney, 2000), so removal is replicable. Additional cleaning procedures carried out in R are also available in the OSF project affiliated with this work (Author, 2018).

[ii] The only developmental data available was acoustic (Kunnari, Nakai, & Vihman 2001). For consistency, emergence calculation was limited to diary data.

[iii] Hua & Dodd (2000) report development in Putonghua Mandarin (standard Beijing). The Mandarin CDS corpus documents development in Shenzhen Mandarin, a southern variety.

[iv] In Japanese, samples < 750 resulted in a FL of 0 for almost all segments/sample. This is likely due to heterogeneous phonotactics and distinct lexical strata in Japanese so a sample size of at least 750 is required to gauge FL (Inkelas, 2014; Itô & Mester, 1999).

[vi] The coefficients in Table 5 report on models without Mandarin /ts/ and Spanish /ɾ/. Alone, these phones changed their respective language model fits, as verified in systematic leave-one-out tests. The coefficients for a Mandarin model with /ts/ were (**frequency**=-1.30 ***, **FL**=0.43 ***) meaning that the significance of FL was due to the presence of /ts/. Similarly, the coefficients for the Spanish model with /ɾ/ were **frequency**=-2.14 *** and **FL**=1.17 *** meaning that the positive relationship of emergence and FL was due to /ɾ/.