Canadian Journal of Linguistics/Revue canadienne de linguistique, 63(4): 527–555, 2018
doi: 10.1017/cnj.2018.9

Development of phonetic complexity in Arabic, Berber, English and French

FRÉDÉRIQUE GAYRAUD
Laboratoire Dynamique du Langage UMR 5596 CNRS and University of Lyon
Frederique.Gayraud@univ-lyon2.fr

MELISSA BARKAT-DEFRADAS
Institut des Sciences de l'Évolution de Montpellier UMR5554 CNRS and University of Montpellier
melissa.barkat-defradas@um2.fr

MOHAMED LAHROUCHI
Laboratoire Structures Formelles du Langage UMR 7023 CNRS and University Paris 8
mohamed.lahrouchi@cnrs.fr

and

MAHÉ BEN HAMED
Laboratoire Base, Corpus, Langage, UMR 7320 CNRS, University of Nice Sofia Antipolis, UMR5554 CNRS, and University of Montpellier
mahe.benhamed@gmail.com

Abstract

The goal of this study is to provide crosslinguistic data on the acquisition of phonetic complexity among children acquiring four different languages: Tunisian Arabic, Tashlhiyt Berber, English, and French. Using an adaptation of Jakielski's (2000) Index of Phonetic Complexity (IPC), we carried out an analysis to assess phonetic complexity of children's early vocabulary in the four languages. Four different samples from each language were analyzed: 50 words selected from an adult dictionary of each language, 50 words from child-directed speech, 50 words targeted by the child, and the child's actual pronunciations of those 50 words. Globally, we hypothesized that children's early productions would be shaped by universal articulatory constraints, but also by the language they are exposed to, depending on its phonological complexity. Our findings show that Arabic displays higher degrees of complexity compared to Berber, English and French, and that children acquiring Arabic target and produce more complex words than children learning Berber, English and French.

Keywords: acquisition, crosslinguistic, phonetic complexity, universals, language specificities

Résumé

L'objectif de cette étude est d'analyser, dans une approche translinguistique, l'acquisition de la complexité phonétique dans quatre langues : l'arabe tunisien, le berbère tachelhit, l'anglais, et le français. À partir de l'adaptation de l'Index de Complexité Phonétique (IPC) de Jakielski (2000), nous avons conduit une analyse évaluant la complexité phonétique des premiers mots produits par les enfants de ces quatre langues. Pour chacune des langues, quatre échantillons ont été examinés : 50 mots aléatoirement sélectionnés dans un dictionnaire, 50 mots du langage adressé à l'enfant, 50 mots ciblés par les enfants, et enfin la production effective par les enfants de ces mêmes 50 mots. L'hypothèse générale était que les productions précoces des enfants seraient influencées par des contraintes articulatoires universelles, mais aussi par la complexité phonologique de la langue à laquelle ils sont exposés. Nos résultats montrent que l'arabe est la langue qui manifeste le plus de complexité phonologique comparé au berbère, à l'anglais et au français, et que les enfants en cours d'acquisition de l'arabe ciblent et produisent des mots plus complexes que les enfants acquérant le berbère, l'anglais ou le français.

Mots-clés: acquisition, translinguistique, complexité phonétique, universaux, spécificités langagières

1. Introduction

Children typically produce their first words around 12 months of age. However, the full mastery of an adult-like sound system is not achieved before 8 (Sander 1972) or even 10–12 years of age (Smith and McLean-Muse 1986)[1]. In English for instance, some phonemes such as /s/ or consonant clusters such as /spl/ are not acquired before 7 and 9 years of age (Smit et al. 1991). This extended period required for the mastery of the speech sound system of the adult language is due to the fact that young children are neither endowed initially with an adult-like vocal tract configuration nor with the neuromuscular control for producing the range of sounds of their ambient language (Kent and Murray 1982, Green et al. 2002, Stoel-Gammon and Sosa 2007). These anatomical and neurophysiological constraints result in a restriction on children's early phonetic inventories (Nip et al. 2009, Green et al. 2010). Children's phonetic inventory is initially composed of sounds produced primarily by the jaw (MacNeilage et al. 2000), on which they have a better muscle control compared to the motion of lips and tongue movements (Green et al. 2002). As a consequence of these universal biological constraints, children acquiring different languages show a similar restricted inventory of sounds (Locke 1983, 1995). Indeed, babbling and first-word productions demonstrate universal patterns: children show a preference for labials and coronals, stops, nasals and glides, open syllables, short utterances, few consonant clusters (and if any, they tend to be homorganic), and more reduplication than variegation (Vihman et al. 1985, Oller et al. 1994, MacNeilage

[1]Abbreviations used : CDS : child-directed speech ; H : hypothesis ; IPC : Index of Phonetic Complexity ; L1 : first language.

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

et al. 1999). Furthermore, these preferences have been shown to influence the words that children select to produce. Thus, the inventory of children's early vocabulary is not composed of randomly selected words. Rather, it has been suggested that children select words with phonetic characteristics that are already present in their own phonological systems (Ferguson and Farwell 1975, Vihman et al. 1985, Schwartz et al. 1987). Other studies on lexical selectivity have shown that children attempt more complex words targets according to age (Dobrich and Scarborough 1992). An Index of Phonetic Complexity (henceforth IPC), based on the phonetic regularities observed in the babbling and the first-word period, was proposed to assess children's phonetic development (Jakielski 2000, 2002). The IPC has proven to be a valuable tool for different purposes such as assessing phonological skills in toddlers (Morris 2009), comparing speech acquisition in bilingual vs. monolingual children (Gildersleeve-Neumann and Wright 2010), or exploring the relationships between phonetic complexity and stuttering (Howell et al. 2006, Howell and Au-Yeung 2007). The IPC, which considers productions composed of less preferred segments and segment associations as more complex, permits one to measure the development of phonetic complexity in words both targeted and produced by children. In addition, it was shown that IPC scores at 12 months predicted speech and language skills at 18 months (Furey 2003). Biomechanical constraints of the production system (MacNeilage and Davis 1990) and lexical selectivity are both universal tendencies.

Crosslinguistic studies hence provide support for a strong determination of early phonetic inventories by biological constraints. However, they do not rule out an influence of the ambient language. Languages differ to a large extent in terms of their phonological inventories and phonotactics, making the input more or less difficult to acquire for children. Crosslinguistic analysis of diverse languages enables us to distinguish between potentially universal and language-specific patterns (Stoel-Gammon 2011). In fact, previous analyses have shown that segmental development, namely word shapes and CV co-occurrences, are influenced by input frequency in the ambient language (Saffran et al. 1997), as well as by the functional load of segments in the language, that is, how much use a language makes of its available contrasts (Stokes and Surendran 2005). As languages vary on those parameters, previous findings suggest that some languages may be acquired at a faster rate than others. For instance, So and Dodd (1995) showed that Cantonese children acquire phonology at a faster rate than English-speaking children as they master the contrastive use of tones and vowels by two years and that few phonological errors occur after age four. Similarly, Caselli et al. (1995) observed that Italian children were slower in vocabulary acquisition compared to English-speaking children: Italian children lag behind the English group in total vocabulary size at most ages between 8 and 16 months. Another crosslinguistic study, comparing L1 vocabulary growth at 16–30 months of age, found that Galician children produce fewer words than Basque, French and Mexican-Spanish-learning children (Pérez-Pereira et al. 2007). In brief, phonetic development seems to be strongly influenced by universal biological constraints but also by the characteristics of the ambient language.

However, considering the ambient language, one must keep in mind that the type of language to which children are exposed (referred to as Child-Directed Speech,

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

henceforth CDS) differs in several important ways from the adult language. When addressing children, caregivers adjust their language by simplifying and clarifying the linguistic material (Ferguson 1964) in order to engage children's attention and facilitate language acquisition (Snow 1977, Werker et al. 2007)[2]. CDS is characterized by simplified syntax, shorter utterances, restricted vocabulary, repetitions, phonetic modifications (Kuhl 2000), increased variations in fundamental frequency and longer pauses (Ferguson 1977, Papoušek et al. 1991, Albin and Echols 1996, Andruski and Kuhl 1996). Moreover, this specific register used by parents to promote infants' language learning has been shown to be almost universal (Ferguson 1964, Kitamura et al. 2001, Kuhl et al. 1997, Monnot 1999, Kitamura et al. 2001).

The current study aims at examining the phonetic complexity of words produced by children acquiring four different languages: Arabic (Tunisian vernacular), Berber (Tashlhiyt variety), English (American) and French (France). These languages show different phonetic and phonological characteristics of interest for early language development, such as word length, word complexity (syllable types, consonant clusters) and phonemic inventory diversity. For example, the Arabic lexicon, which is largely derived from basic consonantal roots, includes many polysyllabic words. In terms of syllable types, French shows a strong preference for open syllables (76%), English exhibits more closed syllables (60%) (Delattre 1965) and Arabic displays 49,92% of closed vs. 50,08% of open syllables (Hamdi et al. 2005). These languages also differ in consonant cluster[3] requirements: In Tashlhiyt Berber, clusters are found in any position within the word and in Arabic, they are rarely found in word-initial and word-final position. In French, both positions are permitted but a bias towards word initial position is attested. Phonemic inventories are also quite diverse. Indeed, when computing the consonant/vowel ratio, two groups emerge: Berber and Arabic are highly consonantal languages (Hamdi et al. 2005, Ridouane and Fougeron 2011) whereas vowels are more frequent than consonants in French and, to a lesser extent, English; Berber and Arabic display a significant proportion of fricatives as compared to English and French. Moreover, the phonological inventories of Arabic and Berber put forward a large number of back consonants (i.e., uvulars, pharyngeals and glottals) that are known to be acquired rather late (Omar 2007). In sum, this study aims to provide informative contribution comparing the phonetic development in four languages including Berber and Arabic, languages in which studies of phonetic development are rare.

We elaborated six interrelated hypotheses for our crosslinguistic study:

H1. The different languages should display different degrees of complexity. The dictionary words in Berber and Arabic, which are highly consonantal languages, should have higher complexity scores compared to the more vocalic English and French.

[2]However, some studies suggest that CDS is not necessary facilitating as it introduces more variability in input (Sundberg and Lacerda 1999, Dodane and Al-Tamimi 2007, Benders 2013, Dilley et al. 2014)

[3]In the present study, the term 'cluster' refers to consonant sequences that occur both within and across syllables.

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

The following IPC parameters are of special interest for our cross linguistic analysis:

1. *Place of articulation*: More complexity expected in Arabic, due to the many back (i.e., dorsal) consonants, and in French, due to the frequent /ʁ/ (Gromer and Weiss 1990).
2. *Clusters*: More complexity expected in Berber and Arabic, in which clusters are more frequent.
3. *Complex articulation*[4] should contribute to complexity in Berber and Arabic.
4. *Word length*: More complexity on this parameter is expected in Arabic and Berber, which show many polysyllabic words due to the insertion of vocalic patterns and affixes into the root for lexical derivation.
5. *Final Consonant*: Less complexity is expected in French, due to its preference for open syllables.
6. *Variegation (Place)*: More variegation is expected in Arabic, due to the non-homorganic consonantal rule, which constrains the root-skeleton.
7. *Variegation (Manner)*: in Arabic (likewise in all Afro-Asiatic languages) it is well known that the co-occurrence of identical and homorganic consonants in the root sequence is restricted by the Homorganic Cooccurence Constraint (Greenberg 1950). This restriction leads to expect Arabic will attest more variegation than the other languages under study.
8. *Rhoticity*[5] will contribute to complexity in American English only.

H2. In each language, we expect the words that mothers use when addressing their children (CDS) to be less complex than the words used in the "adult" language (i.e., represented here through dictionary words).

H3. According to the lexical selectivity hypothesis, children should attempt words (targets) that are less complex to produce than many other words in the adult language (Dictionary).

H4. Given the biomechanical constraint hypothesis, we expect children's actual productions to be less complex than the targets they attempt.

H5. However, we expect an effect of the ambient language: children acquiring a phonologically more complex language should target and produce more complex forms. Hence, we expect Arabic- and Berber-speaking children's IPC scores to be higher than those observed for English- and French-speaking children.

H6. In sum, if a complex parameter is frequent in the ambient language, children should use it or attempt it more often than if it is not present in the ambient language.

Table 1 recapitulates the different hypotheses.

[4]When consonant articulation occurs at the same time as another articulation is being made at a different place in the vocal tract, the consonant is said to form a complex articulation (Ladefoged and Maddieson 1996, Barkat-Defradas and Embarki 2009).

[5]Rhoticity in English (the pronunciation of the historical rhotic consonant /r/) is one of the most prominent features distinguishing varieties of English. Here we studied the acquisition of American English, which belongs to the rhotic varieties.

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

H1	Dictionary	Globally, Berber and Arabic should show higher complexity because they are highly consonantal languages
		a. More complexity in Arabic and French for place of articulation
		b. More complexity in Berber and Arabic for clusters
		c. More complexity in Berber and Arabic for complex articulations
		d. More complexity in Arabic and Berber for word length
		e. Less complexity in French for final consonant
		f. More complexity in Arabic for variegation (Place)
		g. More complexity in Arabic for variegation (Manner)
		h. More complexity in English for Rhoticity
H2	CDS	Less complexity in CDS compared to Dictionary
H3	Targets	Less complexity in Targets compared to Dictionary
H4	Actual	Less complexity in Actual compared to Targets
H5	Targets & Actual	Berber and Arabic Targets and Actual more complex than French and English
H6	Targets & Actual	Targets and Actual should reflect the ambient language

Table 1: Predictions for Dictionary, CDS, Target and Actual words in the four different languages

2. METHOD

In this section, we first describe the participants in the study (section 2.1), followed in section 2.2 by the procedures used to gather the data, and then turn to the analysis in section 2.3.

2.1 Participants

Sixteen children from four linguistic communities: Arabic (Tunisian vernacular), Berber (Tashlhiyt variety), American English and French (France) were included in the study (see Table 2). The parents did not report any concerns about the

Language	N	Sex	Age Range (months;days)
Arabic	4	2M/2 F	8;09–24;06
Berber	4	2M/2 F	7;21–18;12
English	4	2M/2 F	11;27–20;05
French	4	3M/1 F	9;09–22;22

Table 2: Participants' demographic information

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

children's language development, hearing status, or general development. The Arabic and Berber data are part of The PREMs Project[6] (Principal Investigator: Sophie Kern). The French data are part of the French Kern corpus (Kern et al. 2009), and the English data comes from the Providence Corpus (Demuth et al. 2006).

2.2 Procedures

The children were recorded[7] in natural settings at home in interaction with their mother. The recording sessions took place twice a month starting slightly before the onset of first word production until a few months after the lexical spurt. As Jakielski (2000) recommends using 50 words to compute the IPC, four lists of 50 words were analyzed for each language: (i) the first 50 words actually produced by each child (hereafter referred to as Actual), (ii) the targets corresponding to these first 50 words, (iii) 50 words randomly extracted from CDS produced by each mother and (iv) 50 words from the adult language randomly selected from dictionaries[8] (See Appendix 1 to 4 for an illustration). For each list, an adaptation of the Index of Phonetic Complexity (Jakielski, 2000) was computed (see Table 3). The IPC is based on the phonetic regularities observed during the babbling and the first-words periods. Vocal outputs that are composed of the less preferred segments (or segment associations) in early development are rated as more complex in the IPC. This allows measuring the development of phonetic complexity in both word targets and words actually produced by the children. The IPC consists of eight parameters: consonants by place and manner, vowels by class, word shape and word length (in syllable type and number), singleton consonants by place variegation, contiguous consonants and cluster by type (i.e., homo- vs. hetero-organic). However, as this measure was initially designed to capture the phonetic complexity of English, it must be adapted for crosslinguistic comparison in order to account for other determining features of phonetic complexity exhibited in the four languages under examination.

2.3 Data analysis

In order to take into consideration the typological peculiarities of our linguistic sample, we first had to integrate into the original IPC model a new parameter relative to consonantal articulation. We called it *consonant by articulation class* in reference to Jakielski's first two parameters (i.e., Consonant by place and manner class and

[6] <http://www.ddl.ish-lyon.cnrs.fr/projets/prems/index.asp?Langue=ENPage=Presentation>

[7] Audio-video data was obtained using a Sony© HDR CX Camera 740VE. The audio-video recording material was placed so as to capture the profile of the child/mother dyad. In addition to the acoustic signal recorded via the internal microphone of the camera, we used an additional external recording equipment (multidirectional microphone Tascam© DR07 MK2) which was placed in front of the children at about 1 meter.

[8] Dictionaries and databases: for Arabic, we used Cheraifi (2005); for French, Database Corpaix (2000) http://sites.univ-provence.fr/~veronis/data/freq-oral.txt); for Berber, El Mountassir (2003) and Chafik (1990/1999); and for English, The Robert and Collins French-English / English-French Dictionary (1993).

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

Parameter		No points for:	One point each for:	Max. points
1	Consonant by articulation class	simple	complex	1 each
2	Consonant by place class	labials, coronals, glottals	dorsals, pharyngeals	1 each
3	Consonant by manner class	stops, nasals, glides	fricatives, affricates, liquids	1 each
4	Vowel by class	monophthongs, diphthongs	rhotics	1 each
5	Word shape	ending in a vowel	ending in a consonant	1 each
6	Syllable count	monosyllabic and disyllabic	≥ trisyllabic	1 each
7	Consonants by place variegation	reduplicated	variegated	1 each
8	Consonants by manner variegation	reduplicated	variegated	1 each
9	Contiguous consonants	words with no clusters	words with clusters	CC: 1pt. >2C: 2pts.
10	Cluster by type	homorganic	heterorganic	1 each

Table 3: Adapted Index of Phonetic Complexity Scoring Scheme (based on Jakielski, 2000)

Consonant by place class). This new parameter was created in order to discriminate between simple vs. complex articulations (see parameter 1 in Table 3). In Arabic and Berber, two types of secondary (or complex) articulation are phonemically attested: pharyngealization (both in Arabic and Berber) and labialization (in Berber only). Basically, during the realization of a pharyngealized consonant (for example [$t^{ʕ}$, $d^{ʕ}$, $ð^{ʕ}$]), the pharynx is constricted and the root of the tongue is retracted. Such consonants, which require a skillful control of the back of the vocal tract (Barkat-Defradas and Embarki 2009, Lahrouchi and Ridouane 2016), are acquired very late in development (Omar 1973). We therefore added 1 point for such complex segments, that are typical of Afro-Asiatic languages (Hetzron 1997). The same rationale was applied for labialization, which consists of adding lip rounding to the principal articulation. We considered radical consonants (i.e., pharyngeals) as particularly complex in terms of the consonant by place class, since they are, on the one hand, even more posterior than dorsals (that are themselves considered as complex in the original IPC model) and, on the other hand, since the mastery of production for these consonants is reported to occur rather late (Amrayeh 1994, Amrayeh and Dyson 1998). Lastly, considering the fact that consonantal clusters are frequent in the Western varieties of colloquial Arabic, under the influence of Berber, and can thus be very long, we decided to add up to 2 points when more than two consonantal

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

segments are contiguous. Table 3 recapitulates the different parameters included in our IPC adaptation.

Our purpose here is to compare the total phonetic complexity of children's production cross-linguistically, as well as the relative contribution of each of these ten parameters to the overall complexity. Translating each word into a phonetic sequence in which each element is described (as a vowel, a consonant, or a secondary articulation, with consonants being identified in terms of place and mode of articulation), we can identify and compute each parameter, and the IPC as their total sum. Table 4 provides an example of IPC scoring for four dictionary words, one from each language, for the numbered parameters listed in Table 3.

3. Results

The dictionary sample is a window on the expected mean word phonetic complexity of the languages under study. Since we are interested in the differences/similarities between languages or samples rather than in the absolute distribution of the IPC values, we compute the 95% confidence interval error bar (5% type I error, computed through bootstrapping) rather than the standard variation interval, to allow for visual comparison of sample means. As for the actual discussion of results, differences in behavior are supported by a Kruskal-Wallis H-test to test the null hypothesis that the population medians of all of the samples (Dictionary, CDS, Actual and Target) are equal. This test is the non-parametric version of a one-way ANOVA, and is required in our case because data distributions are not normal. This test assumes a chi-square distribution but provides the advantage of allowing both fine grain comparisons between two samples, even samples of different sizes (thus allowing for repeated sampling), and a generalization over more samples.

Figure 1 shows that the four languages have indeed different mean word IPCs, with a decline from Arabic > Berber > English > French. However, confidence intervals are quite large, which was to be expected from the small sample sizes, which casts doubt on the validity of this decreasing pattern. Indeed, inter-language H-test comparisons only support the difference between Arabic and the other languages ($p < 0.05$), but not among Berber, English and French.

This for the overall IPC. However, each of the 10 phonetic parameters used to compute the IPC could contribute differently – depending on the language – while

Parameter:		1	2	3	4	5	6	7	8	9	10	Total IPC
French	fork: [fuʁʃɛt]	0	1	1	1	0	0	1	1	1	1	7
English	mixture: [mɪkstʃər]	0	1	3	1	1	0	1	1	2	1	11
Berber	girl: [tafruxt]	0	1	2	0	1	1	1	1	2	2	11
Arabic	he fell: [sqətˤ]	1	1	1	0	1	0	1	1	1	1	8

Table 4: Index of Phonetic Complexity Scoring for one word in each language

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

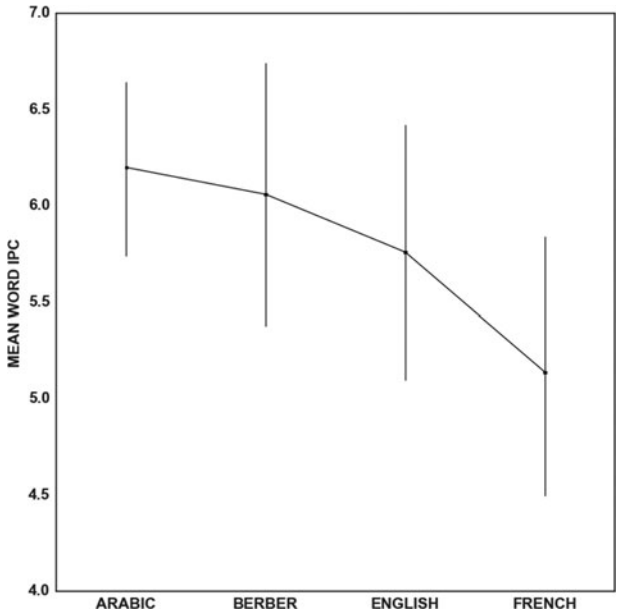

Figure 1: Mean word IPC showing a decreasing decline from Arabic > Berber > English > Vertical bars show the 95% confidence interval for the means for the dictionary as a whole.

being leveled out by the others, with the specific expectations formulated under H1. Such individual contributions for the Dictionary sample are displayed in Figure 2 as *specific contributions,* that is, the difference between the proportion of the overall complexity (total IPC) the parameter actually accounts for, and what would be expected from a uniform contribution of all parameters. Since there are ten of those, the uniform contribution null hypothesis would set each parameter's contribution to a tenth of the overall complexity (in other words, the baseline corresponds to a 10% contribution of the parameter). Specific parameter contributions should thus vary theoretically between −0.1 and 0.9.

To explore the sub-hypotheses outlined for H1, we need to proceed parameter by parameter, with specific expectations for each language. To test whether the mean tendency for the language supports our expectations, we proceed to inter-language T-tests to assess whether the mean behaviors of languages taken two-by-two are statistically significantly different.

As expected in H1.a, Arabic and French display significantly greater contributions ($p < 10^{-3}$) of place of articulation (parameter 2: 12–14% contribution to the overall IPC) than Berber and English (6–7%). As for Manner (parameter 3), it accounts for some 20% of the total IPC for all four languages, with no significant inter-language differences. In terms of variegation, Arabic indeed exhibits the highest contributions for both place (parameter 7: 16.2%) and manner (parameter

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

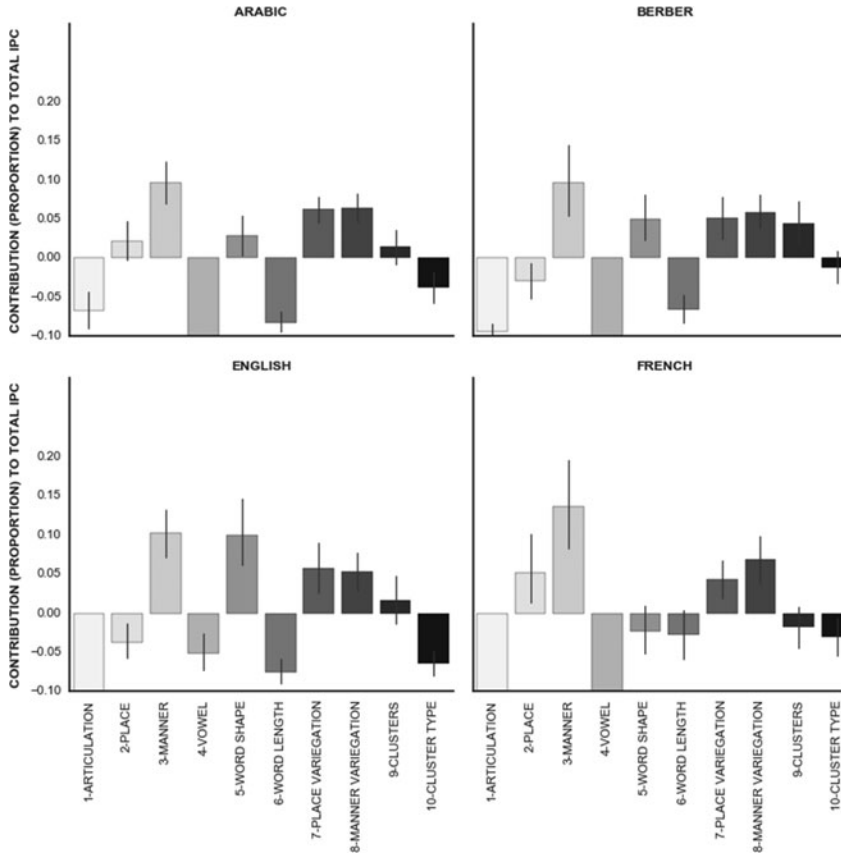

Figure 2: Specific parameter contribution per Language (with respect to the uniform contribution null hypothesis). For each parameter, values could theoretically range between −0.1 and 0.9.

8: 16.3%), but they are not significantly higher than the contributions of these parameters for the other three languages, where values range from 13.8 to 16.1%, invalidating the expectations of hypotheses H1.f and H1.g. The Rhoticity parameter (parameter 4, H1.h), which is exclusive to English, only accounts in this language for 4.7% of the total IPC, but is still significantly different from zero ($p < 10^{-4}$).

Moving on to complex articulations: contrary to our assumption in H1.b concerning consonantal clusters (parameter 9), Berber and Arabic (14.1% and 11.4% respectively) differ significantly only from French (7.8%, $p \leq 0.05$) but not from English (11.5%), with Arabic and English only slightly above the 10% baseline of a uniform contribution of all parameters to the total IPC. However, Arabic and Berber significantly differ from English in terms of cluster composition (parameter 10: $p \leq 0.05$, 6.1%, 8.6% and 3.5% respectively), but not from French (6.7%), as predicted by H1.c.

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

To explore the effect of word structure and shape, we first distinguished monophthongs from diphthongs in our data description, then computed word length as the number of vowels, assumed to act as syllabic nuclei (except for Berber, where, null syllabic nuclei were also identified, using Dell and Elmedlaoui's (2002) syllabification algorithm). For this parameter (parameter 6, H1.d), all four languages exhibited less than the 10% contribution anticipated if all parameters had an equal contribution to the overall IPC, with values ranging from 1.7% to 6.9%, and an unexpected low value for Arabic. However, as expected in H1.e, French exhibits significantly less complexity (7.4%, $p < 0.01$) than all three other languages in word shape (parameter 5: 12–20%) due to a preference for open final syllables.

Complexity measures on the dictionary samples provide the baseline against which children's production and lexical selectivity, but CDS can also be analyzed. Figure 3 shows the mean IPC values for the different samples in the various languages.

Overall, the four samples differ significantly for all four languages ($p < 10^{-6}$), with a general tendency to a decreasing complexity from Dictionary to CDS to Target to Actual child productions. However, major differences exist between languages. Regarding H2, for instance, we find that the lesser complexity expected for CDS compared to the Dictionary sample is supported only for Berber (4.34 vs. 6.06) and French (3.92 vs. 5.14) with $p < 0.01$, but not for Arabic (5.42 vs. 6.2) or English (5.4 vs. 5.76). As for H3, while Arabic, English and French exhibit significant IPC differences between Dictionary or CDS and Target (p at least < 0.02), the Berber data only supports a difference between Dictionary and Target (4.26), but not between CDS and Target, suggesting that adults actually pre-sample their

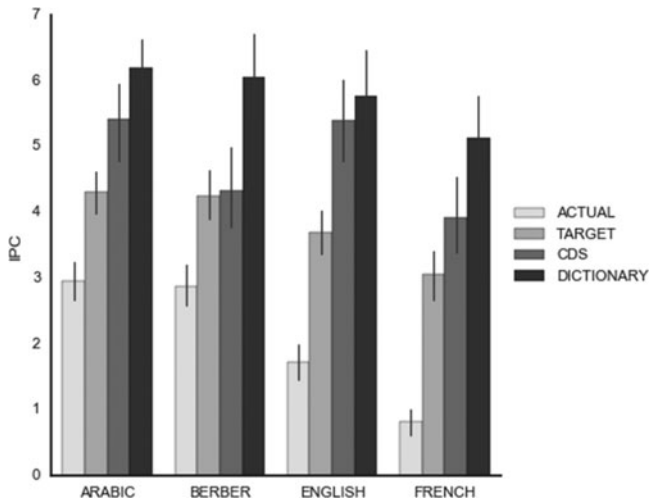

Figure 3: Mean word IPC and 95% confidence intervals for the Dictionary/adult language, CDS, actual child production and target for Arabic, Berber, English and French

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

language to select for appropriate complexity in the speech they address to the child. Moving on to H4, all languages exhibit significantly lower complexity of the Actual production of the child with respect to the intended Target ($p < 10^{-6}$). Arabic and Berber show similar Target complexity and Actual-to-Target complexity differential (2.96–4.31 and 2.87–4.26 respectively). Nevertheless, while English and French display lower Target complexity (3.7 and 3.06 respectively), and lower Actual complexity as predicted by H5 (1.72 and 0.81 respectively), they also display larger Actual-to-Target complexity differentials especially for French, prompting the question of accuracy.

Given that the individual parameters focus primarily on consonants, we computed the Actual-to-Target accuracy in the realization of consonants for the four languages (see Figure 4) as the proportion/percentage of conformity with target. Despite the larger overall Actual-to-Target complexity differential, French children achieve better accuracy than English children, typically, and fall within the range of accuracy of Berber kids ($p < 0.02$). More importantly though, and despite larger Target complexity, Arabic and Berber children show better accuracy scores (66% and 59% respectively vs. 53% for French and 44% for English). These results show that these Actual-to-Target complexity differentials do not necessary translate into Actual-to-Target accuracy, but more crucially they suggest that the typological characteristics of the language do indeed play a role in the acquisition process, calling for

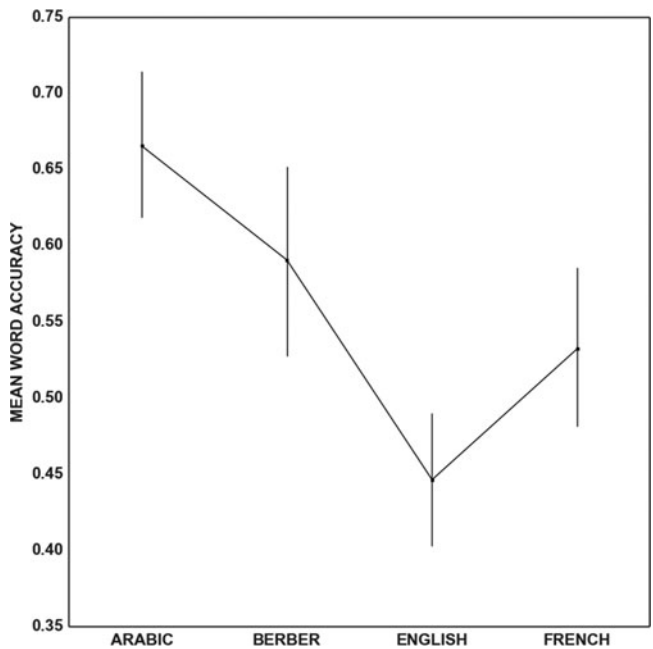

Figure 4: Mean Actual-to-Target accuracy in consonant production for Arabic, Berber, English and French. Vertical bars display the 95% confidence interval

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

a detailed examination of the individual contribution of each parameter to the overall IPC score of children's actual production, as compared to the ambient language (dictionary) parameter contribution landscape.

Relevant to H6, Figure 5 shows that not all parameters contribute equally to the mean IPC of the child's actual production, but also that, depending on the language, they do not depart equally from what is expected from other samples, and especially the dictionary language. For instance, Arabic Actual (16.2%) shows an enhanced tendency for children to produce closed final syllables, irrespective of what is displayed in either the Target (9.1%) or the CDS (9.6%), thus mirroring more of the dictionary tendency (12.9%) than these two latter samples do. On the other hand, clusters, which were shown previously to have a significant contribution to the overall IPC in the dictionary (14.1%), behave differently ($p < 10^{-4}$) in the Targets selected by children (7.7%) and in their Actual productions (6%), while these latter do not depart significantly from the CDS (5%) sample. This suggests that both CDS and Targets selected

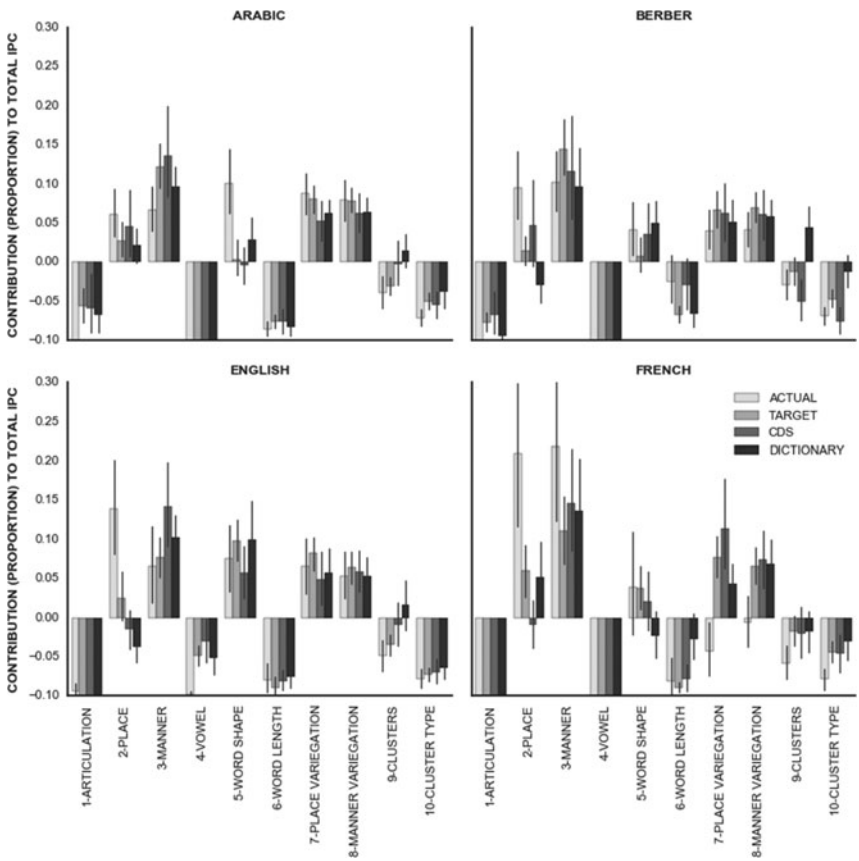

Figure 5: Mean parameter contribution by sample (Dictionary, CDS, Target and Actual) and by language (Arabic, Berber, English, French)

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

by the children tend to misrepresent clusters during this phase of acquisition, thereby reducing the complexity of their production.

As for manner of articulation (parameter 3), Arabic and Berber tend to have comparable complexity in the Dictionary, CDS and Target samples, but this complexity is not achieved in the Actual children productions ($p < 10^{-3}$). In other words, while the selected targets mirror the complexity of the ambient language for this parameter, children's productions do not. Conversely, in English and French, the CDS does mirror the complexity of the Dictionary for this parameter, but neither the children's Targets nor Actual productions do ($p < 0.01$) as they display similar levels of complexity for this parameter. In terms of place of articulation (parameter 2), Targets and Actual reflect the characteristics of both Dictionary and CDS for Arabic, Berber and English but not for French. In French, Target complexity for this parameter mirrors the CDS, but not the Dictionary, and Actual does not reflect either of the Dictionary, the CDS or the Target. Conversely, place variegation (parameter 7) for this language shows that the Target's complexity mirrors the Dictionary rather than the CDS. In addition, while place and manner contribute the most to Actual productions in this language, their variegation contributes the least, contrary to Arabic, Berber and English.

4. Discussion

Using an adaptation of Jakielski's (2000) Index of Phonetic Complexity (IPC), we carried out an analysis to assess phonetic complexity of children's early vocabulary in four languages: Arabic, Berber, English and French. Globally, we hypothesized that children's early productions would be shaped by universal articulatory constraints, but also by the language they are exposed to, depending on its phonological complexity.

Considering language-specific aspects, we observed that as hypothesized, some languages of our samples, namely Arabic and Berber, show higher IPC scores (complexity) than English and especially French (H1). Our results also reveal the different IPC parameters contribute differently to overall phonetic complexity depending on language specificity. Manner and place of articulation contribute significantly to complexity in all four languages, and not only in Arabic and French as predicted by H1.a. In other words, children in these languages do not systematically avoid fricatives or liquids, which are frequent in the ambient language (dos Santos 2007 for French, Nahar et al. 2012 for Arabic), and to which they are consequently frequently exposed to via CDS. Similarly, for place of articulation, although children prefer labials and coronals to dorsals in the early period of production, they are able to produce posterior consonants (like dorsals and/or pharyngeals) when these types of sounds are frequently encountered in the ambient language (i.e., typically in Afro-Asiatic languages where they are rather frequent as shown by Basset (1946) and Bonnot (1977). However, some parameters clearly differ across languages. Clusters are significant contributors to phonetic complexity in Berber, and, to a lesser extent, in Arabic and English (H1.b). In Tunisian Arabic, as in all Western

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

Arabic vernaculars, the prevalence of complex consonantal clusters is mainly due to the loss of unstressed vowels in open syllables under the influence of linguistic substrates (i.e., namely Berber), and this contributes to the increased number of clusters. As for final consonant, French is confirmed to be easier than the other languages, as predicted by H1.e. In contrast, some other parameters that were expected to play a role in complexity scores (H1.c, d, f, g, h), in particular word length and complex articulation (H1.c and d), do not seem to contribute significantly to the global complexity score. Moreover, the expected decreasing complexity: Berber < Arabic < English < French which led us to hypothesize that these differences in phonetic complexity in the adult languages would influence phonetic development in children acquiring these languages. However, our results did not show this expected pattern: only Arabic is significantly more complex than the three other languages. An obvious assumption is that a phonetically complex language such as Arabic would take longer to acquire, and would be more challenging to reach accuracy. Nevertheless, our findings suggest that this is not the case. Although our sample is too small to draw solid conclusions, our data indicate that Arabic infants are not delayed in the timing of first words production, given that first words are produced before the age of twelve months in our sample. Similarly, children acquiring Arabic and Berber show the best accuracy scores compared to the *a priori* articulatory less complex languages, English and French.

Turning to universal aspects of phonetic development (H2, H3 and H4), our findings are in line with previous work underlying the universal tendency for caregivers to modify their speech when addressing children (H2) (Ferguson 1964, 1978; Kuhl et al. 1997; Monnot 1999; Kitamura et al. 2001). Some studies had investigated CDS in Arabic and Berber and cultural differences have been documented in the range of modifications of CDS (Ferguson 1956, Gumperz and Hymes 1964, Bynon 1968, Omar 1973, Haggan 2002, Al-Shatty 2003, Ferguson 2004), and the fact that caregivers use a special register to address young children remains indisputable cross-linguistically. The vast majority of CDS studies are focused on semantic, syntactic or prosodic characteristics. Fewer address the phonetic characteristics of CDS, but Kuhl et al. (1997) for instance showed that mothers in different languages (American English, Russian and Swedish) produce vowels that are acoustically more extreme when addressing their young children, thus providing information about the sound system of the infant's native language in an exaggerated form. CDS thus promotes language learning by separating sounds into contrasting categories. The present study did not focus on potential modifications of segments in CDS, but our findings show that caregivers use words that are less complex than in the adult language. This suggests the operation of lexical selectivity in CDS just as lexical selectivity is hypothesized in children's productions. Thus, one can assume that caregivers reduce the gap between children's restricted articulatory capacities and the necessary capacities to produce phonetically complex words of the adult language.

As predicted by H3, children also seem to select the words they produce or attempt to produce depending on their phonetic complexity (Ferguson and Farwell 1975, Vihman et al. 1985, Schwartz et al. 1987). Indeed, in the four languages

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

under study, we observed that words attempted (Target) or produced (Actual) by children have lower complexity scores than CDS or adult words (H3 and H4). Finally, the fact that in all the four languages, actual productions show reduced complexity compared to attempted words (and obviously to CDS or adult speech) illustrates the fact that during this early period of first word productions, children's early phonetic inventories are strongly limited by anatomical and neurophysiological constraints (MacNeilage et al. 2000, Green et al. 2002, Stoel-Gammon and Sosa 2007, Geen et al. 2010). Looking at the detailed parameters that children either produce or avoid depending on the language, our findings show a mixed influence of language specificity and neurophysiological constraints. Globally, children's actual productions tend to reflect the tendencies displayed in adult productions (H5 and H6). The fact that in Arabic, children's use of closed syllables (cf. actual items) mirrors the adult language (dictionary words, but not targets nor CDS) can be explained by the fact that this process of simplification does not affect the word randomly but rather affects morphemes that are added to the root (for example, in the actual item #48 *tatt* it is the morpheme of the third person singular at the imperfective [ji-] prefixed to the verbal root which is dropped. These morphological elisions also explain why children's actual words are not as long as predicted. Another consequence of this phenomenon is illustrated in # 39 *ʔam:i* 'my uncle' which is actualized *ʔam*: 'uncle' and where the suffixed first person possessive determiner [-i] is dropped, resulting in a CVC sequence. Doing this, the child produces a form that is similar to the one recorded in dictionaries and which corresponds to the radical form of the words. Turning to clusters, which are strong contributors of complexity in Berber (and to a lesser extent in Arabic and English), they seem to be avoided by children acquiring these languages (in Arabic for example, #18 *kaskək* ('your helmet'), #35 *tkasir* ('you break'), #42 *ħizha* ('lift it' (fem.)) and/or #50 *ʔistana* ('wait") are actualized under the following simplified forms: *ka, kasər, tiz:a and stan:æ* respectively. This finding is not surprising, as consonant clusters appear to be especially challenging for children. Indeed, they are not produced before age 2 and their acquisition is one of the longest-lasting aspects of speech acquisition in normally developing children (McLeod et al. 2001). In sum, the different IPC parameters are not equally difficult for children to produce: while some (such as producing fricatives or dorsals) can be overcome by children even at a very early age in spite of biological constraints, others (such as consonant clusters) need more time to master.

https://doi.org/10.1017/cnj.2018.9 Published online by Cambridge University Press

	Actual	Target		CDS		Dictionary	
	IPA	IPA	Translation	IPA	Translation	IPA	Translation
1.	mɛm:æ	mama	mommy	tˤɑ:ħ	to fall	ʕəʃb	medical
2.	itæj:	tˤɑj	peek a boo	ʔi:ʒæ	he comes	ħuku:m	government
3.	pæ	papa	daddy	tˤɑtˤɑ	tata	ħi:la	trickery
4.	dæ	laʔ	no	ʔaħ:æ:	ouch!	farləχla	fieldmouse
5.	ʔi	ʔɛj	yes	kiχ:æ	eew!	msisa	bracelet
6.	tæ:ħ	tˤɑ:ħ	it fell	qu:l	tell!	ħəwəl	to try
7.	bæj	bah	ok	næħ:i:	remove	qa:l	he told
8.	æm	mam:i	mum	ʔahajjæ:	here it is	mraħ:əm	marble
9.	bæħ	baħ	It disappeared	wi:ni	where is it?	zra:ʕi:	agricultural
10.	ʔɛjʒi	ʔiʒa	come here !	ʃu:	what is it?	mdˤəl	straw
11.	bɛj	baj baj	bye-bye	dad:u:ʃ	walk!	mudˤif	stewart
12.	tɛd:	ʃi:dd	hold !	jɛz:i	stop!	butˤain	ventricle
13.	ʔæh:	ʔaħ	ouch !	træh	show me!	dəhʃ	baby donkey
14.	hup	ħutˤ	put down !	ba:ħ	gone (nothing!)	wəsχ	dirtiness
15.	ʒaʒæ	ʤa:ʒa	chicken	alo	hello	səq:a:t	gourmet
16.	ħibɛʃ	mænħibɛʃ	I don't want	ʔaʕmil	do !	kursi	saddle
17.	ktib	ʔiktibli	write (to me)	hæ:t	put !	ʕrəq	sweat
18.	ka	kaskək	your helmet	læ:	no	ħad:	fence
19.	tæ:ʕi	mtæ:ʕi	mine	ʔɛj	go !	rij:a:ga	bib
20.	bebe	bebe	baby	χu:	brother	qəm	estimate
21.	kak:æ	kaka	poo	kaka	poo	məstəɣəl	enjoy
22.	tat:a	ba tˤ:ɑ	duck	jimʃi	he goes/he walks	bəlɣa	oriental slipper
23.	tatæ	tˤɑtˤɑ	beat !	sˤɑħi:t	thank you	ʃkara	bag
24.	ħutæ:	ħuta	fish	ʔaqɛf	get up !	bəqra:ʒ	tea pot
25.	bab	dub	bear	bæʕ	bèè (bleat noise)	ħɛ:l	weather
26.	na:n	ħsˤa:n	horse	ʒi:b	bring	sˤəf:a	clarify

27.	næhnæ	lihnæ	here	hæðıka	this one	məl:ɛ:ħ	seller
28.	mæmæ	mæʔ	water	tħutˁ	you put	da-ijəb	melting
29.	kæk	qird	monkey	ʔistan:æ	wait !	qəfza	jump
30.	tæ:ħ	muftæ:ħ	key	ʔælʕib	play !	da:r	home
31.	tawtaw	taw:a	now	ʃʕær	hair	ma:t	perish
32.	dæb:u:sæ	dab:uzæ	bottle	mtaʕ	his	rʒaʕ	take back
33.	ʃbi:k	ʃbi:k	what happens to you?	sˁɑfiq	applaud !	tˁbi:b	doctor
34.	ʔuxdæn	ʔuxzur	watch !	ʃnuwæ	what is it ?	faħ:əm	explain
35.	kas:ər	tkas:ir	you broke	bismilah	bless you	hdər	argue
36.	bæz	xubz	bread	ʕaqu:la	nice	sqətˁ	fall
37.	fasu:	ʕasˁfu:ra	bird	ʔaʕtˁi	give	ɣri:b	unheard-of
38.	tax:a	kix:a	eew !	matmis:iʃ	don't touch (to her)	bətˁtˁəl	idle
39.	ʕam:	ʕam:i	my uncle	ʔuqʕud	sit down !	mʕal:əm	expert
40.	ku:ra	ku:ra	ball	ɣa:di	there	səndu:q	coffin
41.	di	haði	this one	ʃbi:k	what happens ?	hənd	magnet
42.	tizza	ħizha	lift it up (fem.)	nrakablik	I'll put it together for you	qəmr	moon
43.	laħda	laħðˁa	one minute please	tkal:im	talk !	qəsma	part
44.	kuja	xuja	my brother	nχab:i:h	I hide it	stəχbər	be informed
45.	ku:	qu:m	get up !	ħsˁɑ:n	horse	ʒəʕ:ada	spatula
46.	nan:i	nan:i	sleep !	ʕasˁfu:ra	bird	aθat	furniture
47.	ʔahu	ʔahuwa	here it is !	matibki:ʃ	don't cry	mədwəd	nursery
48.	tatt	jitattti	(he) beats	kifæ:ʃ	how ?	rqi:q	thin
49.	ræjæ	mræjæ	mirror	nhizz	bring	dərbuka	drum
50.	stan:æ	ʔistana	wait !	su:q	market	ʕilaqa	coat rack

Appendix 1: Examples of Arabic data

	Actual	Target		CDS		Dictionary	
	IPA	IPA	Translation	IPA	Translation	IPA	Translation
1.	bːabːa	bajbaj	goodbye	dːaħ	hit (have a pain)	ara	give
2.	babaː	baba	dad	ʕaw	noise	aman	water
3.	amːaː	mama	mom	aratid	give it to me	afus	hand
4.	hbːa	sːˤbːatˤ	shoes	xix:i	poo	ilm	skin
5.	ajalːa	lalːa	auntie	titːi	sit down	krf	tie
6.	haːh	hak	take	xːiʃ	it is bad (thing)	adr	press
7.	manːama	lmunika	toy	ʃaʃːa	a donkey	rgl	close
8.	qːaqːa	ʁika	like this	ʃiʃːi	eat	ifta	he went
9.	tːa	xtːa	this one (fem.)	baʕːa	sheep	usin	they took
10.	bːaħ	bːaħ	disappeared	bːaħ	It's over	awid	bring!
11.	məməm	mumːu	toy (generic)	dːaw	walk / go	inijas	tell him
12.	hatːi	hatːi	here it is (fem.)	dajdːa	fall down	ufiɣ	I found
13.	ajːaː	tajːaː	another one (fem.)	baʕːatin	sheep	lkmɣ	I arrived
14.	ʃɛʃ	ʃaʃːa	horse (baby talk)	ʃtːi	grandfather	tamːnt	honey
15.	bːɛħbːa	aħbːudˤ	tummy	kakːa	chocolate	takurtː	ball
16.	bula	bula	this is a light bulb	etːej	move !	agrzam	lion
17.	mna	magana	this is a watch	buʕːu	monster	tafruxt	girl
18.	qaqːa	kukuʕːu	cock (baby talk)	tinxarin	nose	afrux	boy
19.	alaqa	alqːaʕ	sheep	tabatːˤaħt	foot	imi	mouth
20.	aːlaː	ara	give me	tafust	hand	argaz	man
21.	baʕpːɛʕ	lbanan	banana	tamzːˤuxt	ear	tamɣart	woman
22.	tːɛtetːe	ʁid	here	tagajːut	head	amadl	hill
23.	tʃtiʃk	juʃkad	he came	timimːit	small mouth	amdlu	cloud
24.	aːpbː	ibːi	it is ripped	tiwalːin	eyes	skr	do
25.	acaːca	agajːu	head	mimiʃ	mouth	fl	let
26.	memːe	ʕmti	my aunt	tabitːu	eyes	ls	wear

27.	handbːa	andːu	we are going	tidˤudˤin	fingers	gn	sleep
28.	ʃu	ʃuf	look	ʃkːata	eat (it)	izˤuran	veins
29.	neːnːe	ninːi	sleep (baby talk)	tasːawalt	trousers	azˤalim	onion
30.	aʃːe	adʃːʁ	I will eat	banana	banana	tiwit	you took away
31.	ajːiː	hajːi	here I am	biʒːu	baby	tabrat	letter
32.	haʃtʃkːa	iʃqːa	it is tough	tˤrʃt	break	smun	pick up
33.	ʃual	ʃuwːr	slow down	ʕanː	car	knu	lean
34.	anana	wijnːa	mom's one	babat	shoes	kru	rent
35.	ʃaʃaː	matˤeʃa	tomatoes	fufːu	fire/ hot	igawr	he sit down
36.	bakːaʔ	baraka	that's enough	akːik	give / Show	tkʃmt	you entered
37.	ʁːa	ʁwːa	this one	baja	biscuits	ssnkr	wake
38.	alti	xalti	my aunt	stːii	grandfather	udm	face
39.	mamiandːu	manisrandːu	where	ħmal	donkey	idamːn	blood
40.	d	adˤadˤ	finger	ʕːu	fermented milk	aɣrum	bread
41.	mimːi	amuʃː	cat	daʕ	also	juda	it's enough
42.	bwa	iswa	he drank	xːixːiʃ	dirty	izra	he saw
43.	jːah	jah	yes	kukuʕːu	cok	alim	straw
44.	uːfːuʔ	sfːuʁt	get him out	ħaħːa	boho / cut	afunas	bull
45.	fːufːu	ifːuʁ	he went out	dːaħːati	hit (me)	ajdi	dog
46.	mːam	imːim	it's tasty	bːˤatˤ	shoes	ngr	between
47.	pxːu	buxːuʃ	here's the insect	ħawa	cow	tirmt	food
48.	ʃːaʔ	ʃː	eat	hawhaw	dog	azur	terrace
49.	sːasːa	sħsːa	be quite	tˤanun	yogourt	izgr	he crossed
50.	adːa	tdːˤrˤ	she fell down	tiʒiʒːit	meat	xdmn	they worked

Appendix 2: Examples of Berber data

	Actual	Target		CDS		Dictionary	
	IPA	IPA	Translation	IPA	Translation	IPA	Translation
1.	dʌbɛ	bʌmbəlbiː	bumblebee	ækrəˌbæt	acrobat	əsiːtəs	acetous
2.	dædeɪ	dædiː	daddy	ədɔːr	adore	ɑːrʧduːk	archduke
3.	əzi	dʌrtiː	dirty	ænt	ant	bækəs	bacchus
4.	dʌwɪ	dɑgiː	doggie	əsliːp	asleep	bæstʃən	bastion
5.	dʌdi	dʌkiː	duckie	beɪbisɪdər	babysitter	ˈbætəl	battle
6.	ɑbwiː	beɪbiː	baby	bædəl	battle	Klæːʃ	clash
7.	hʌ	kaʊ	cow	bɪskət	biscuit	kɔɪn	coin
8.	dʌ	dɑt	dot	kɛərfəl	careful	kʌm	come
9.	hʌbu	lɑbstər	lobster	ʧɔːk	chalk	saikliŋ	cyclin
10.	ʌmɛ	mɑmiː	mommy	dɑːrlɪŋ	darling	dælmeɪʃən	dalmatian
11.	mʌm	muːn	moon	dɪfərənt	different	ded	dead
12.	bəbɛ	pʌpiː	puppy	iːzi	easy	dɪsgrʌntəld	disgruntled
13.	ʊˈwʌ	bɑl	ball	ɛnʤɔɪ	enjoy	ɛstʃuəri	estuary
14.	ɛ	aɪ	eye	fɑːrm	farm	friːzɪŋ	freezing
15.	ɑtʃi	ɑntiː	auntie	fɪlθi	filthy	fjuːz	fuse
16.	kizɪ	kɪdiː	kitty	ʤərɑːf	giraffe	heɪz	haze
17.	æː	rɔr	roar	gɜːrl	girl	hɛvi	heavy
18.	diθhɪ	tiːʧər	teacher	greɪt	great	haɪbərneɪt	hibernate
19.	dʌʔdæ	bækpæk	backpack	gɪtɑːr	guitar	ɪnsɛst	incest
20.	bɑ	bɛr	bear	hevi	heavy	ɪntoʊn	intone
21.	peɪ	bai	bye	hɔːrs	horse	lɑgər	lager
22.	dʌkʰ	stʌk	stuck	ɪnsɛkt	insect	lɪv	live
23.	bu	bluː	blue	ɪʧi	itchy	mæʧ	match
24.	bɑk	boʊt	boat	ʤækət	jacket	mɛntəl	mental
25.	bʊk	bʊk	book	ʤʌgəlɪŋ	juggling	mɪkstʃər	mixture

26.	bo	baʊ	bow	kæŋgəruː	kangaroo	peɪpəl	papal
27.	bu	bruːm	broom	kaɪt	kite	roʊmən	roman
28.	gʊkeɪ	kʊkiː	cookie	leɪtər	later	ruːl	rule
29.	duʃ	ʃuːz	shoes	liːv	leave	skɛərsli	scarcely
30.	bu	bɜrd	bird	lɪkɪŋ	licking	tæb	tab
31.	bɑk	blæk	black	lʌv	love	tækəl	tackle
32.	pɪteɪ	pɪglət	piglet	mægnət	magnet	viəməns	vehemence
33.	bu	buːts	boots	miːn	mean	vɜrʤənəl	virginal
34.	bʌbaʊwɪs	bʌbəlz	bubbles	mʌŋki	monkey	vəsɪfərəsli	vociferously
35.	boː	bʌg	bug	neɪm	name	wiːkli	weekly
36.	keɪk	keɪk	cake	nɛkləs	necklace	waɪld	wild
37.	kɑː	kɑr	car	naɪs	nice	kærɪdʒ	carriage
38.	kɑk	klɒk	clock	ɔrənʤ	orange	tɛkst	text
39.	flɛlɑ	floʊwər	flower	peɪntɪŋ	painting	bɑks	box
40.	fuː	fuːd	food	pɪkʧər	picture	aʊtlaɪn	outline
41.	kik	kɪk	kick	ˈreɪni	rainy	pækt	pact
42.	mu	muːs	moose	rɔːr	roar	pʊt	put
43.	pik	pɪŋk	pink	sændi	sandy	ripʌblək	republic
44.	doʃ	stɑrz	stars	skrʌbɪŋ	scrubbing	raɪt	right
45.	sɛː	sʌn	sun	siːt	seat	sædnəs	sadness
46.	fwi	θriː	three	spəgɛdi	spaghetti	snæp	snap
47.	tos	toʊz	toes	tɪkət	ticket	skwiːz	squeeze
48.	wɛt	wɛt	wet	ˈju nɪkɔːrn	unicorn	taɪm	thyme
49.	pʌpo	pɜrpəl	purple	wæks	wax	traɪəl	trial
50.	dʒiː	triː	tree	jɛloʊ	yellow	voʊg	vogue

Appendix 3: Examples of English data

	Actual IPA	Target IPA	Translation	CDS IPA	Translation	Dictionary IPA	Translation
1.	maːma	mamɑ̃	mummy	ʁəgaʁd	look	ɑ̃tɔʁs	sprain
2.	paː	papa	daddy	komɑ̃se	start	bʁutij	trinket
3.	toχto tœː	dodo	sleep	pɑ̃s	think	ɑ̃tuzjasme	enthusiasm
4.	paː	pɛ̃	bread	vjɛ̃	come	eʁezi	heresy
5.	tœdeː	tete	suck	tɛlmɑ̃	so much	ʁətʁɛ	withdrawal
6.	taːte	taʁtin	sandwich	ivɛʁ	winter	tʁuve	find
7.	daːdaː	gato	cake	dusmɑ̃	slowly	kʁəvɛzɔ̃	flat tire
8.	œbiː	abɛj	bee	pø	can	kanaʁ	duck
9.	bɛbeː	bebe	baby	dədɑ̃	inside	duluʁø	painful
10.	teːte	tɛte	head	gʁɑ̃	big	libeʁalism	liberalism
11.	tataː	tata	auntie	fɛ	do	pʁɔvizjɔ̃	supply
12.	bibiː	bebe	baby	vwala	here	ʒødi	Thursday
13.	ɔp	ɔp	hop	kɔm	like	kʁist	Christ
14.	awoːa aalo	alo	hello	atɑ̃	wait	avaʁi	damage
15.	kukuː	kuku	hello	lezaʁ	lizard	ɑ̃gaʁ	shed
16.	apoː	pɔ̃pje	fireman	dakɔʁ	okay	ʁɑzwaʁ	razor
17.	poː	po	pot	ale	go	tɑ̃	time
18.	kokœ	tɔ̃tɔ̃	uncle	aʁiv	arrive	ataʁde	linger
19.	apeː	papje	paper	avɑ̃taʒ	advantage	kuʃe	lying
20.	œmaːʁœː	kameʁa	camera	mwɛ̃	less	maʁʃɑ̃	seller
21.	teteː	tete	sucking	dɑ̃	in	səkwe	shake
22.	kokɛː	elikɔptɛʁ	helicopter	pɔʁtik	crossbar	tʁiko	knitting
23.	ɔvaː	oʁəvwaʁ	bye bye	pusɛ̃	chick	vɑ̃	wind
24.	tatɔː	klakson	horn	amne	brought	deʁegle	upset
25.	baː	bal	ball	ʁapɛl	remember	dɛʁm	dermis
26.	baː	bɛ̃	bath	ʒø	games	ɔtɔmatik	automatic
27.	apaː	lapɛ̃	rabbit	butɔ̃	button	ɑ̃gɔ̃se	constrained
28.	biːboː	bibʁɔ̃	bottle	tele	TV	ɔfʁiʁ	offer

29.	guː	lego	lego	komɑ̃	how	vəly	hairy
30.	mɔː	mɔʁ	bite	lɛs	let	velɛ̃	vellum paper
31.	maː	mal	hurt	lynɛt	glasses	ɛʁb	grass
32.	neːne	mimi	cute	aplɛ	called	defile	parade
33.	apoː	ʃapo	hat	pwasɔ̃	fish	kale	wedge
34.	ameː	fɛʁme	closed	penibl	tiresome	kɔ̃tʁəpɛtʁi	spoonerism
35.	bɛ	tɔ̃b	fall	vaʃ	cow	epanwiʁ	blooming
36.	puː	pul	hen	ɛ̃tɛʁdi	forbidden	dɔsil	docile
37.	fwɑ	fʁwa	cold	ʃosɛt	socks	ʃɑ̃s	chance
38.	vyː	pyl	jumper	gaʁsɔ̃	boy	apsɑ̃	absent
39.	kaː	kask	helmet	pətit	small	aspik	aspic
40.	pyː	py	stink	kat	four	fam	woman
41.	afaː	paʁfœ̃	perfume	pʁɛ	ready	manifɛstasjɔ̃	demonstration
42.	nɔːne	koksinɛl	ladybird	tabl	table	peje	payed
43.	aneː	fɛʁme	close	aspiʁatœʁ	vacuum cleaner	sɛ̃	saint
44.	buː	bul	ball	mɔ̃tʁ	show	tiʁɑjœ	infantryman
45.	af	ʒiʁaf	giraffe	telekomɑ̃d	remote control	ynite	unity
46.	ɔuː	ʃosyʁ	shoe	dəbu	stand	ɔnɔʁe	honored
47.	bubeː	pubɛl	trash	nuʁityʁ	food	etɛ̃	off
48.	dadaː	salad	salad	kanaj	raffish	kɑ̃	camp
49.	ʁoːgwɛ	dekʁoʃe	pick up	pubɛl	trash	seʁɔm	serum
50.	apiː	apɥi	press	mɑ̃ʒe	eat	kɑ̃tɔnad	To no one in particular

Appendix 4: Examples of French data

References

Albin, Drema Dial, and Catherine H. Echols. 1996. Stressed and word-final syllables in infant-directed speech. *Infant Behavior and Development* 19(4): 401–418.

Al-Shatty, Tareq Hamad. 2003. *Differentiating between the interactive styles of two types of caregiver with 11–21 months Kuwaiti children*. Doctoral dissertation, University of Reading.

Amrayeh, Mousa. 1994. *A normative study of the acquisition of consonant sounds in Arabic*. Doctoral dissertation, University of Florida, Gainsville.

Amrayeh, Mousa, and Alice T. Dyson. 1998. The acquisition of Arabic Consonants. *Journal of Speech Language and Hearing Research* 41(3): 642–653.

Andruski, Jean E., and Patricia K. Kuhl. 1996. The acoustic structure of vowels in mothers' speech to infants and adults. In *Spoken Language, 1996. ICSLP 96. Proceedings, Fourth International Conference on Spoken Language Processing* Vol. 3: 1545–1548.

Barkat-Defradas, Melissa, and Mohamed Embarki. 2009. Mécanismes articulatoires et corrélats acoustiques de l'emphase en arabe et en berbère : un état de l'art. In *Mélanges en hommage à Naima Louali*, ed. S. Chaker, Amina Mettouchi, and Gérard Philipson, 21–42. Paris: Peeters.

Benders, Titia. 2013. Mommy is only happy! Dutch mothers enhance affect in infant-directed speech, not didactic intent. *Infant Behavior and Development* 36(4): 847–962.

Basset, André. 1946. Le système phonologique du berbère. *Comptes rendus du Groupe Linguistique d'Études Chamito-Sémitiques* 4: 33–36.

Bonnot, Jean-François. 1977. Recherche expérimentale sur la nature des consonnes emphatiques de l'arabe classique. *Travaux de l'Institut Phonétique de Strasbourg* 9: 47–88.

Bynon, James. 1968. Berber nursery language. *Transactions of the Philological Society* 67(1): 107–161.

Caselli, Maria Cristina, Elizabeth Bates, Paola Casadio, Judi Fenson, Larry Fenson, Lisa Sanderl, and Judy Weir. 1995. A cross-linguistic study of early lexical development. *Cognitive Development* 10(2): 159–199.

Chafik, Mohamed. 1990–1999. *Dictionnaire arabe-amazighe*. Tomes I, II, III. Académie du Royaume du Maroc.

Cheraifi, Claudie. 2005. *L'arabe Maghrébin, Petit dictionnaire français-arabe*. Paris : Slaktine.

Delattre, Pierre. 1965. *Comparing the phonetic features of English, French, German and Spanish*, Heidelberg: Harrap. 118 pp.

Dell, François, and Mohamed Elmedlaoui. 2002. *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Springer.

Demuth, Katherine, Jennifer Culbertso, and Jennifer Alter. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, 49(2): 137–173.

Dilley, Laura, Amanda Millett, Devin McAuley, and Tonya Bergeson. 2014. Acoustic-phonetic variation in word-final alveolar consonants in infant-directed speech over the first two years. *Journal of Child Language* 41(2): 155–175.

Dobrich, Wanda, and Hollis S. Scarborough. 1992. The phonological characteristics of words young children try to say. *Journal of Child Language* 19(3): 597–616.

Dodane, Christelle, and Jalal-eddin Al-Tamimi. 2007. An acoustic comparison of vowel systems in adult directed-speech and child-directed speech: Evidence from French, English and Japanese. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)*, ed. Jürgen Trouvain and William J. Barry, 1573–1576. Saarbrücken: Saarland University.

El Mountassir, Abdallah. 2003. *Dictionnaire des verbes tachelhit-français (parler berbère du sud du Maroc)*. Paris: L'Harmattan.

Ferguson, Charles A. 1956. Arabic baby talk, in *For Roman Jakobson:* Essays on the occasion of his sixtieth birthday, ed. Morris Halle, Horace Lunt, Hugh Mc Lean and Cornelis H. van Schooneveld, 121–128. The Hague: Mouton.

Ferguson, Charles A. 1964. Baby talk in six languages. *American Anthropologist* 66(6), part 2: 103–114.

Ferguson, Charles A. 1977. Baby talk as a simplified register. In *Talking to children*, ed. Catherine E. Snow and Charles A. Ferguson, 209–235. Cambridge: Cambridge University Press.

Ferguson, Charles A. 1978. Talking to children: A search for universals. In *Universals of Human Language: Vol. 1, Method and Theory*, ed. Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik, 203–224. Stanford, CA: Stanford University Press.

Ferguson, Charles A. 2004. Talking to children: A search for Universals. In *First language acquisition: The essential readings*, ed Barbara Lust, and Claire Foley, 176–189. Oxford, Blackwell.

Ferguson, Charles A., and Carol B. Farwell. 1975. Words and sounds in early language acquisition. *Language* 51(2): 419–439.

Furey, Joan Elizabeth. 2003. *The growth of early phonological and lexical development: A longitudinal investigation*. Doctoral dissertation, University of Illinois at Urbana-Champaign.

Green, Jordan R., Christopher A. Moore, and Kevin J. Reilly. 2002. The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research* 45(1): 66–79.

Green, Jordan R., Ignatius S. Nip, Ben Maassen, and Pascal Van Lieshout. 2010. Some organization principles in early speech development. In *Speech Motor Control: New Developments in Basic and Applied Research*, ed. Ben Maassen and Pascal van Leshout, 71–188. Oxford: Oxford University Press.

Greenberg, Joseph. 1950. The patterning of root morphemes in Semitic. *Word* 6: 162–181.

Gildersleeve-Neumann, Christina E., and Kira L. Wright. 2010. English speech acquisition in 3-to 5-year-old children learning Russian and English. *Language, Speech, and Hearing Services in Schools* 41(4): 429–444.

Gromer, Bernadette, and Marlise Weiss. 1990. *Apprendre à lire*. Paris : Armand Colin.

Gumperz, John Joseph, and Dell H. Hymes. 1964. *The ethnography of communication*. Menasha, WI: American Anthropologist Association.

Haggan, Madeline. 2002. Self-Reports and Self-Delusion regarding the use of Motherese: Implications from Kuwaiti adults. *Language Sciences* 24(1): 17–28.

Hamdi, Rym, Salem Ghazali, and Melissa Barkat-Defradas. 2005. Syllable structure in spoken Arabic dialects: a comparative investigation, *Proceedings of Interspeech 2005 – Eurospeech – 9th European Conference on Speech Communication and Technology*, Lisboa, Portugal.

Hetzron, Robert. 1997. *The Semitic Languages*. London: Routledge.

Howell, Peter, James Au-Yeung, Scott Yaruss, and Kevin Eldridge. 2006. Phonetic difficulty and stuttering in English. *Clinical Linguistics and Phonetics* 20(9): 703–716.

Howell, Peter, and James Au-Yeung. 2007. Phonetic complexity and stuttering in Spanish. *Clinical Linguistics and Phonetics* 21(2): 111–127.

Jakielski, Kathy. 2000. Quantifying phonetic complexity in words: An experimental index. Paper presented at the Annual Child Phonology Conference, Cedar Falls, IA.

Jakielski, Kathy. 2002. A new method for measuring articulatory complexity. Paper presented at the Annual Convention of the 2002 American Speech–Language–Hearing Association, Atlanta, GA.

Kent, Raymond D., and Anne D. Murray. 1982. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America* 72(2): 353–365.

Kern, Sophie, Barbara Davis, and Inge Zink. 2009. From babbling to first words in four languages: Common trends across languages and individual differences. In *Becoming Eloquent: Advances in the emergence of language, human cognition, and modern cultures*, ed. Francesco d'Errico and Jean-Marie Hombert, 205–232. Amsterdam: John Benjamins.

Kitamura, Christine, Chayada Thanavishuth, Denis Burnham, and Sudaporn Luksaneeyanawin. 2001. Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development* 24(4): 372–392.

Kuhl, Patricia K. 2000. A new view of language acquisition. *Proceedings of the National Academy of Sciences* 97(22): 11850–11857.

Kuhl, Patricia K., Jean E. Andruski, Inna A. Chistovich., Ludmila A. Chistovich, Elena V. Kozhevnikova, Viktoria L. Ryskina, and Francisco Lacerda. 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science* 277(5326): 684–686.

Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Lahrouchi, Mohamed, and Rachid Ridouane. 2016. On diminutives and plurals in Moroccan Arabic. *Morphology* 26(3): 453–475.

Locke, John L. 1983. *Phonological acquisition and change*. New York: Academic Press.

Locke, John L. 1995. *The child's path to spoken language*. Harvard: Harvard University Press.

MacNeilage, Peter F., and Barbara L. Davis. 1990. Acquisition of speech production: Frames, then content. In *Attention and performance 13: Motor representation and control*, ed. Marc Jeannerod, 453–475. Hillsdale, NJ: Lawrence Erlbaum.

MacNeilage, Peter F., Barbara Davis, Ashlynn Kinney, and Christine L. Matyear. 1999. Origin of serial-output complexity in speech. *Psychological Science*, 10(5): 459–460.

MacNeilage, Peter F., Barbara Davis, Ashlynn Kinney, and Christine L. Matyear. 2000. The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development* 71(1): 153–163.

McLeod, Sharynne, Jan Van Doorn, and Vicky A. Reed. 2001. Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology* 10(2): 99–110.

Monnot, Marilee. 1999. Function of infant-directed speech. *Human Nature* 10(4): 415–443.

Morris, Sherrill R. 2009. Test-retest reliability of independent measures of phonology in the assessment of toddler's speech. *Language, Speech, and Hearing Services in Schools* 40 (1): 46–52.

Nahar, Khalid, Moustafa Elshafei, Wasfi Al-Khatib, and Husni Al-Muhtaseb. 2012. Statistical Analysis of Arabic Phonemes for Continuous Arabic Speech Recognition, *International Journal of Computer and Information Technology* 1(2): 49–61.

Nip, Ignatius S., Jordan R. Green, and David B. Marx. 2009. Early speech motor development: Cognitive and linguistic considerations. *Journal of Communication Disorders* 42(4): 286–298.

Oller, Kimbrough, Rebacca E. Eilers., Michele L. Steffens, Michael P. Lynch, and Richard Urbano. 1994. Speech-like vocalizations in infancy: An evaluation of potential risk factors. *Journal of Child Language* 21(1): 33–58.

Omar, Margaret K. 2007. *The acquisition of Egyptian Arabic as a native language*. Georgetown: Georgetown University Press.

Papoušek, Mechtild, Hanuš Papoušek, and David Symmes. 1991. The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development* 14(4): 415–440.

Pérez-Pereira, Miguel, Margareta Alegren, Mariela Resches, Maria Jose Ezeizabarrena, Carmen Díaz, and Inaki García. 2007. Cross-linguistic comparisons between Basque

and Galician. *Proceedings from the first European network meeting on communicative development inventories*. Gävle: Gävle University Press.

Ridouane, Rachid, and Cécile Fougeron. 2011. Schwa elements in Tashlhiyt word-initial clusters. *Journal of Laboratory Phonology* 2 (2): 275–300.

Saffran, Jenny R., Elissa L. Newport, Richard N. Aslin, Rachel A. Tunick, and Sandra Barrueco. 1997. Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science* 8(2): 101–105.

Sander, Eric K. 1972. When are speech sounds learned? *Journal of Speech and Hearing Disorders* 37(1): 55–63.

dos Santos, Christophe. 2007. Développement phonologique en français langue maternelle : une étude de cas. Thèse de doctorat, Université Lumière Lyon 2.

Schwartz, Richard G., Laurence B. Leonard, Diane M. F. Loeb, and Lori A. Swanson. 1987. Attempted sounds are sometimes not: An expanded view of phonological selection and avoidance. *Journal of Child Language* 14(3): 411–418.

Smit, Ann B., Linda Hand, Joseph J. Freilinger, John E. Bernthal, and Ann Bird. 1991. The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech, Language, and Hearing Research* 34(2): 446–446.

Smith, Bruce L., and Ann McLean-Muse. 1986. Articulatory movement characteristics of labial consonant productions by children and adults. *The Journal of the Acoustical Society of America* 80(5): 1321–1328.

Snow, Catherine E. 1977. Mothers' speech research: From input to interaction. In *Talking to children: Language input and acquisition*, ed. Catherine E. Snow and Charles A. Ferguson, 31–49. New York: Cambridge University Press.

So, Lydia K., and Barbara J. Dodd. 1995. The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*. 22(3): 473–495.

Stoel-Gammon, Carol, and Anna V. Sosa. 2007. Phonological development. *Blackwell Handbook of Language Development*, ed. Erika Hoff and Marilyn Schatz, 238–256. Oxford: Blackwell.

Stoel-Gammon, Carol. 2011. Relationships between lexical and phonological development in young children. *Journal of Child Language* 38(1): 1–34.

Stokes, Stephanie F., and Dinoj Surendran. 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.

Sundberg, Ulla, and Francisco Lacerda. 1999. Voice onset time in speech to infants and adults. *Phonetica* 56(3–4): 186–199.

Vihman, Marilyn May, Marlys A. Macken, Ruth Miller, Hazel Simmons, and Jim Miller. 1985. From babbling to speech: A re-assessment of the continuity issue. *Language* 61(2): 397–445.

Werker, Janet F., Ferran Pons, Christiane Dietrich, Sachiyo Kajikawa, Laurel Fais, and Shigeaki Amano. 2007. Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition* 103(1): 147–162.