Using spontaneous child data as a proxy for frequencies in child-directed speech

Ghada Khattab and Shaima Al Qattan
Newcastle University

Outline

I. Frequency in language/phonological development

II. Gaps in data on (absolute and relative) frequency in Arabic

III. Preliminary data using spontaneous child corpora showing:
   a. Strong effects of frequency on order of C acquisition
   b. Difficulty in disentangling frequency effect from that of complexity

IV. Work underway

I. Role of frequency in language acquisition

- Frequency effects have proven to be pervasive in all aspects of language acquisition (LA)

e.g. Ambridge, 2010; Arnon & Snider, 2010; de Boysson-Bardies & Vihman, 1991; Dell, 1990; Edwards & Beckman, 2008; Edwards et al, 2004; 2015; Gulzow & Gagarina, 2007; Matthews et al 2007; Kirk & Demuth, 2003; Ota, 2013

- high frequency forms:
  - are acquired early
  - exhibit fewer errors
  - cause errors in context with a competing lower freq form
  - interact with other factors (e.g. syllable- and/or word-position; word/utterance length; articulatory and/or morpho-phonological complexity; etc.)

I. Role of frequency in LA (cont.)

- Types of frequency:
  - Type (e.g. nb of different words in which sound occurs) vs Token (overall occurrence regardless of type)
  - Absolute (type and token) vs relative (when compared with competing forms)

- In work on phonology:
  - Type has shown to be a better predictor of age of acquisition than tokens (Edwards and Beckman, 2010; Rose, 2011; Vihman, 2014)
  - Language specific frequency effects are evident as early as the babbling stage (e.g. de Boysson-Bardies & Vihmanm 1991) and throughout early phonological dev. (e.g. Edwards et al, 2004; 2015; Kirk & Demuth, 2003; Ota, 2013; Sosa & Stoel-Gammon, 2012)
  - But: individual differences in rate and order of acquisition

I. Role of frequency in LA (cont.)

- The interaction with other factors is a double-edged sword: it helps us understand the mechanisms that underpin learning but also makes it hard to detect frequency-specific effects:

- Frequency *can* fall out of formal models:
  - In a Chomskyan rule-based theory, more natural rules (with fewer features) are encountered more frequently.
  - In a structuralist model, unmarked sounds are less complex, are targeted more frequently and are acquired earlier than more complex, less frequent sounds.

➢ crosslinguistic tendency for 'simple' sounds to be more frequent, influencing order of acquisition

I. Role of frequency in LA (cont.)

- In functional models the role of input frequency is more pronounced and interacts with articulatory + cognitive constraints in acquisition (e.g. Menn & Stoel-Gammon, 2000; Pierrehumbert, 2003)

➢ Where the above process converges with universal tendencies the outcome is the same, making it hard to disentangle the effect of frequency from complexity/universal constraints

➢ Stronger arguments for frequency effects can be found where language-specific frequency, articulatory complexity, and phonotactic patterning override expected order of acquisition

(e.g. Edwards & Beckman, 2008; Edwards et al, 2004; 2015; Kirk & Demuth, 2003; Li et al, 2009; Morrisette et al, 2003; Stokes and Surendran, 2005; Zamuner, Gerken and Hammond, 2005 Tsurutani, 2007; Van Severen et al, 2012)

- But here, too, individual and other confounding factors can lead to conflicting results (e.g. Levelt and van Oostendorp, 2007; Levelt & Fikkert, 2011; Menn and Vihman, 2011; Sosa & Stoel-Gammon, 2012)

I. Role of frequency in LA (cont.)

- When frequency is not enough:
  - input vs intake, e.g. storage of low freq desirable objects (e.g. *cake*) vs high freq functional ones (e.g. *the*)
    - ➢ Lexical vs phonological frequency
  - age of acquisition: role of frequency may be more obvious in later words
  - Individual preference/experience

➢ Whether or not the role of frequency effects is crucial for explaining the mechanisms that underpin language learning - and therefore for any theory of LA - is still a matter of debate

e.g. review article in JCL by Ambridge et al (2015), with commentaries from both sides of the debate

II. Frequency information for Arabic dialects

- Non-existent (cf. Boudela, 2010's Aralex: MSA, adult, written)
- Research on phonological acquisition in Arabic is scarce

cf. Amayreh, 1994; 2003; Amayreh and Dyson 1998; 2000; Ammar & Morsi, 2006; Dyson and Amayreh, 2000; Faraj, 1988; Hamdan and Amayreh, 2007; Omar, 1973; Saleh et al, 2007

- Yet: Arabic spoken by ~280 million people as a 1st language around the Middle East and North Africa (Procházka, 2006) + as a heritage language for millions more around the world

II. What Arabic might add

- Rich sound inventory (~32-37 Cs) containing many of the 'marked' consonants that vary in their frequency → can contribute to the universal vs language-specific debate on language acquisition

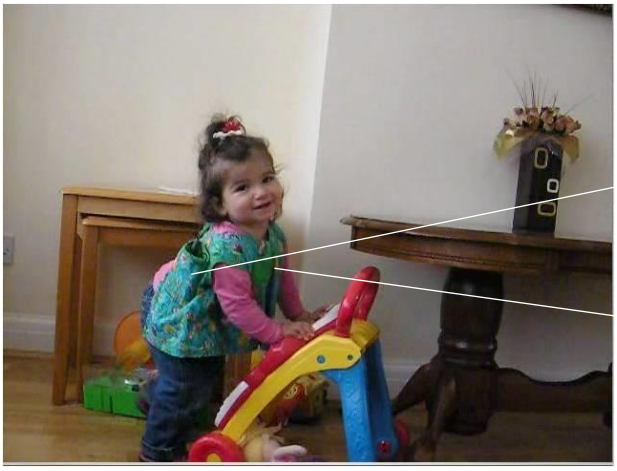
	Bila-bial	Labio-dental	Dental	Alveo-lar	Palato-alv.	Velar	Uvular	Pharyn-geal	Glot-tal
Stop	p* b			t d tˤ dˤ		k (g)	q		ʔ
Nasal	m			n		ŋ*			
Trill/tap				ɾ r					
Fricative		f v*	θ ð (ðˤ)	s z sˤ zˤ	ʃ ʒ (tʃ) (ʤ)	x ɣ	(χ) (ʁ)	ħ ʕ	h
Approxi mant	w	ɹ*			j				
Lateral				l (lˤ)					

III. Current study

Participants		
Age Groups	**Fieldwork locations**	**Sample**
1:4 - 1:7 1:8– 1:11 2:0 – 2:3 2:4 -2:7 2:8 - 2:11 3:0 – 3:3 3:4 – 3:7	**Kuwait** In progress: comparable data from Lebanon; Jordan; Palestine; Qatar	• 140 per site • 10 boys + 10 girls in each group **Total: 700 children across all sites**

Data analysed for this presentation: 70 children, equally distributed across age and gender

III. Audio-visual recordings



Back pocket for mike transmitter

Front pocket for lapel mike

- 30-min spontaneous mother-child interaction
- Hearing screening and speech and language history

III. Transcription and analysis

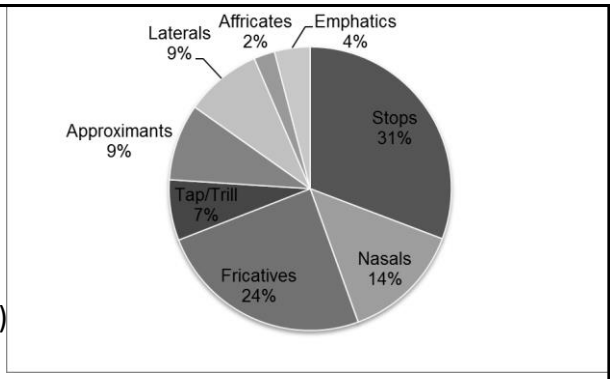
- Orthographic and phonetic transcription in PHON and then conversion to CHAT for future morphosyntactic analyses
- Word ID for early sessions following Vihman & McCune (1994)
- For frequency analyses, all Cs targeted by the children were included at this stage, stratified into word position + type/token
- Typical phonological acquisition and developmental patterns: (Bankson & Bernthal, 1998; Beckman, 2008; Dodd, 1995; Stoel-Gammon & Dunn, 1985; Smit, Hand, Freilinger, Bernthal, & Bird, 1990)
  - ages of *customary production (50%), acquisition (75%)* and *mastery (90%)* for consonants
  - error patterns (age-appropriate, delayed, unusual)

Research questions

- What is the frequency distribution of targeted consonants in a corpus of Arabic child data?
- Is there a relationship between phoneme occurrence frequency and rate of acquisition by Kuwaiti-speaking children aged 1;4 to 3;7?
- How early are language-specific patterns evident in Arabic-speaking children's production both in terms the type of consonants acquired and order of acquisition?

Results

Target consonants (type)
N = 2,806



Target consonants (tokens)
N = 20,044

	Type	Token
Stops	31%	29%
Fricatives	25%	31%
Nasals	14%	16%
Approximants	9%	6%
Laterals	9%	6%
Trill/tap	7%	5%
Emphatics	4%	4%
Affricates	2%	2%
Total	**100%**	**100%**

- General agreement between the two

(*related t test (7) = 0.23; p = 0.824* two tailed)

➢ tendency to select familiar and well practiced/achievable targets that are part of the children's lexical repertoire

Most Frequent → Least frequent

Frequency		Type	Token
>5%	Stops:	/b, t, ʔ/	/b, ʔ/
	Nasals:	/m, n/	/m, n/
	Tap/Trill:	/ɾ/	
	Fricative:		/h, ð/
	Approximant:	/j/	/j/
	Lateral:	/l/	/l/
	Affricate:		
	Emphatic:		
1-5%	Stops:	/d, k, g/	/t, d, k, g/
	Nasals:		
	Tap/Trill:		/ɾ/
	Fricative:	/f, s, z, ʃ, x, ħ, ʕ, h/	/f, s, ʃ, ħ, ʕ/
	Approximant:	/w/	/w/
	Lateral:		
	Affricate:	/ʧ/	/ʧ/
	Emphatic:	/tˤ, sˤ/	/tˤ/
<1%	Stops:	/p, q/	/p, q/
	Nasals:	/ŋ/	/ŋ/
	Tap/Trill:	/r/	/r/
	Fricative:	/v, θ, ð, ɣ /	/v, θ, z, ʒ , x, ɣ/
	Approximant:	/ɹ/	/ɹ/
	Lateral:	/ɬ/	/ɬ/
	Affricate:	/ʤ/	/ʤ/
	Emphatic:	/ðˤ, dˤ, zˤ/	/ ðˤ, dˤ, sˤ, zˤ/

Most frequent:
- "early sounds" (bilabial + coronal stops and nasals; glottal + glide)
- language-specific effect: /l ɾ/
- lexical effects: /h ð/

Least frequent:
- complex sounds (emphatics; fricatives, esp. voiced; trill
- foreign language effect (e.g. /v, p, ɹ, ŋ/)
- lexical effects: /ŋ/
- dialectal variants: /dˤ, dʒ/

Order of acquisition of consonants

	1;4-1;7	1;8-1;11	2;0-2;3	2;4-2;7	2;8-2;11	3;0-3;3	3;4-3;7
Mastery Age (90% of targets)		ʔ		n	k, m	b, k, h, l, w	b, t, d, k, g, ʔ, m, n, f, s, w, j, l
Acquisition Age (75% targets)	b, ʔ, m, n, l	b, t, m, n, j	p, b, k, ʔ, m, n, s, w, j	b, t, d, k, ʔ, m, n, s, w, j, l	b, t, d, ʔ, n, f, s, h, j, l, w	p, t, d, tˤ, g, ʔ, n, ŋ, ðˤ, f, s, z, ħ, j, l, ɬ, ʧ	p, tˤ, ŋ, r, sˤ, z, ʃ, x, ɣ, ʕ, h, ɬ, ʧ, ʤ
Customary Age (50% of targets)	t, d, ʃ, w	d, k, g, f, s, h, w, l	t, d, ʃ, ħ, h, l	g, f, v, sˤ, z, ʃ, x, ħ, ʕ, ʤ, ʧ	g, r, ðˤ, z, x, ħ, ʕ	r, θ, ð, sˤ, x, ɣ, ʕ, ʤ	q, r, ðˤ, zˤ
Not acquired	g, ð, f, s, ħ, h, ʕ, j	tˤ, r, ð, ʃ, ħ, ʕ, ʤ	tˤ, r, ð, f, z, x, ʕ, ʤ	tˤ, q, r, θ, ð	tˤ, r, θ, ð, sˤ, ʃ, ɣ, ɬ, ʧ, ʤ	q	ð, θ
Excluded consonants	p, dˤ, g, q, ɹ, ŋ, θ, ðˤ, v, sˤ, z, zˤ, ʒ, x, ɣ, ɬ, ʤ, ʧ	p, tˤ, dˤ, q, ŋ, θ, ðˤ, sˤ, v, z, zˤ, ʒ, x, ɣ, ɬ, ɹ, ʧ,	dˤ, q, ŋ, ðˤ, sˤ, zˤ, ʒ, ɣ, ɬ, ɹ, ʧ, θ	p, dˤ, ŋ, ðˤ, zˤ, ʒ, ɣ, ɬ, ɹ	p, dˤ, ŋ, v, zˤ, ɹ	dˤ, v, zˤ, ʒ, ɹ	dˤ, v, zˤ, ʒ, ɹ

Percentage of consonant production accuracy

	32	54	64	79	77		89
	79	79	84			87	
	83	85	84	81	83		
	86	86	78	88			
	78	59	63	87	86	89	
	1	5	23	37	46	64	47
	4	7	18	42	56	75	66
	73	67	89	87			
	63	60	69	78	77	89	
	36	76	77	78	82	89	
	86		89	87	89	90	
	7	57	38	68	75	87	
	65	80	75	84	77	89	
	30	52	79	82	76	79	
	26	10	36	51	57	71	85
	30	48	56	65	73	76	89
	71	59	87	89	89		
	50	49	56	55	34	86	82
	3	11	7	33	31	81	84
	29	65	51	67	74	82	
	7	10	38	54	44	78	82
	0	0	46	54	22	57	75
	30	60	26	60	58	67	87
	0	23	35	51	55	77	81
	0	1	43	57	15	61	80
	0	100	0	48	16	44	52
	0	83	0	41	17	89	83
	0	0	0	37	40	75	89
	0	0	0	74	58	78	59
	100	43	89	69	86	88	82
	0	0	33	27	23	68	46
	0	25	0	26	13	50	69
	0	0	0	57	0	67	89
	92	0	50	50	0	44	0
	0	0	0	100	17	86	67
	0	0	0	0	0	100	25
	50	0	100	0	0	0	0

> 90%

> 75%

> 50%

< 50%

exclu-ded

Syllable structure: frequency *vs* error

	SIWI	SIWW	SCWW	SCWF
Total occurrences	19782	18943	5178	4666
Total correct	13840	13819	3788	3620
Percent correct	70%	73%	73%	78%

IV. Summary

- Based on an exploration of targeted words in an Arabic child language corpus:
  - Frequent consonants in early word productions reflect perceptual and motoric abilities → universal tendencies in attention to and accurate production of Cs despite potential of Arabic C inventory
  - A close match between observed type and token frequencies reflects selectivity in early attention and production
  - High frequency consonants are acquired earlier → repetition of neuromotor routines can improve articulatory accuracy and may enhance the phonological memory of learned words (Keren-Portnoy et al, 2010)
  - But: potential confound with perceptual and articulatory salience makes it hard to isolate frequency as the main factor.

Future directions

- Obtain frequencies from CDS (and ADS)
- Look at relative frequency measures
- Include factors known to co-vary with frequency in acquisition such as: age of lexical acquisition; complexity index for consonants and word size; word/syllable position, etc.

Thank you

<u>acknowledgements</u>:

- **Shaima Al-Qattan; Kimary Shahin; Yvan Rose; fieldworkers**
- **The families who have given their time to the project**