Language Growth in Child Emirati Arabic

Dimitrios Ntelitheos and Ali Idrissi

Abstract

We calculate developmental indices of language growth in Emirati Arabic based on a two-year longitudinal corpus of six Emirati children[1]. The target indices include Mean Length of Utterance in morphemes and words (MLUm and MLUw), utterance per turn counts (UoT), type-token-ratio (TTR) and D (an index of vocabulary diversity). Spearman correlation tests show significant correlations between MLUm, MLUw and age. A slightly weaker correlation is found between age and D but not with TTR. Finally, a strong correlation is found between age and UoT. The results provide an important new data point in the body of knowledge about language growth, and show that language acquisition displays similar developmental patterns for linear morphological processes across languages.

Introduction

While English and most Indo-European languages have a long tradition of examining aspects of child language production by computing different developmental indices from spontaneous language samples, other widely studied languages, including Arabic, are lacking in this valuable area of research. This may be partially because of the lack of longitudinal corpora of Arabic child language or the appropriate set-up for experimental studies.

Addressing this gap, this paper provides the first systematic longitudinal study of the validity of a number of developmental indices in Arabic language acquisition (but see Shaalan and Khater 2006; Khater and Shaalan 2007; Abdulla 2002, for earlier attempts to check MLUm and MLUw validity in Arabic based on data from spontaneous language samples). We focus on the relation of MLUm with other developmental indices such as age (Miller and Chapman 1981) or lexicon size (Bates, Bretherton, and Snyder 1988). Our tests are based on transcripts from a longitudinal study of six Emirati children over a period of two years. The calculation of MLU scores is based on a number of rules we developed for Arabic and adopted, from Dromi and Berman's (1982) rules for Hebrew and Shaalan and Khater's (2006) rules for Gulf Arabic.

The paper is organized as follows: In Section 2, we discuss the properties of a number of indices that have been developed in order to measure language growth in typical and atypical child language. Section 3 provides a description of the longitudinal corpus on which our study is based and the coding and analysis methods used in retrieving the relevant statistical data. Section 4, lists the results obtained from the study and discusses the significance of correlations between developmental indices and the age of the target children. Section 5 provides a brief discussion of the implications these results have for the study of Arabic child language, as well as the broader contribution to the areas of language assessment and the diagnosis of language delay in children. Finally, we provide some concluding remarks in Section 6.

Developmental Indices

Since Brown's (1973) seminal work in using corpus studies in the field of language acquisition, a number of indices have been developed in order to measure different aspects of grammatical development. Thus, phonological development may be measured as number of segments (e.g. pMLU, (Ingram 2002)); morphosyntactic development is predominately measured as an increase in the length of the child's utterances in terms of morphemes or words (Mean Length of Utterance in morphemes or words (MLUm, MLUw)); lexical development (i.e. productive use of new words) is measured as an increase in lexical diversity, for example an increase in the type/token ratio of a corpus file (TTR) or statistically extracted corpus samples (D); syntactic development is measured by evaluating the growth of "complexity" in the syntactic structures that the child uses (e.g. the Index of Productive Syntax (IPSyn), (Scarborough 1990)); and finally pragmatic and discourse development is measured by different indices that take into account the use of referential expressions and the relation between number of utterances and turns that the child takes in a conversation.

Mean Length of Utterance

Mean Length of Utterance (MLU) is a developmental index that computes morphosyntactic development. MLU counts are calculated by dividing the number of morphemes (MLUm) or words (MLUw) in a spontaneous speech sample or narrative provided by a child by the number of complete child utterances in the sample. Thus, if a sample containing 100 child utterances has in total 300 morphemes, the MLUm for this sample is 300/100 = 3.0 morphemes per utterance. This basic formula has been shown (e.g. Brown 1973) to correlate with morphosyntactic development in children, and a range of MLUm counts can therefore be reliably assigned to a specific stage of morphosyntactic development. For English, Brown (1973) has proposed the following stages:

Stage	MLUm
I	1.75
II	2.25
III	2.75
IV	3.5
V	4.0

1. Stages of morphosyntactic development in English (Brown, 1973)

Since their first use, MLUm counts have been shown to correlate highly with age and developmental stage (Miller & Chapman 1981; Parker & Bronson 2005). Thus, the MLU value of a child language sample should accurately reflect a specific stage in the child's linguistic development as well as help pinpoint the child's actual age (Miller 1981). This, even though supported to a degree by subsequent studies (e.g. Klee 1992; De Thorne et al 2005), has however been shown not to hold across the board. More precisely, at later stages of development MLUm counts stop increasing at the expected rate, and the correlation with age is lost (Bol 1996; Klee 1992). Blake et al (1993) suggest a 4.5 cut-off point for MLUm values with respect to the calculation of clausal complexity in child language. Additional problems have been mentioned with respect, for example, to the ad-hoc, non-standardized decisions involved in utterance segmentation (c.f. Crystal 1974). Finally, the validity of MLUm for measuring morphosyntactic development in other languages has also been challenged, because of the additional ad hoc decisions and other difficulties in calculating MLUm in languages with complex morphological systems (see for example Thordardottir & Weismer (1998), for Icelandic).

Before moving on to explore these issues, based on our EA data, let us discuss first a number of additional developmental measures that we aim to investigate.

Lexical Growth and Turn Taking

Lexical development is usually measured on the number of new words entering a child's vocabulary as they acquire a language. Statistical information is usually computed from spontaneous language samples of children in conversation or narrating a story. One of the first measures used in this context is the type-token ratio (TTR), or the ratio of new words (types) over the total number of words (tokens) in a speech sample. Templin (1957), introduced the index to child corpora and found a consistent ratio of around one different word for every two words uttered, independently of variables such as age range and gender.

However, later work has shown that TTR depends on the size of the input transcript. That is, language samples which contain larger numbers of tokens give lower values for TTR and vice versa (see discussion in Richards 1987; Richards and Malvern 1997). As the children start producing longer utterances and language samples, a greater part of their acquired lexicon emerges and, as a result, the number of available new word types that could potentially be introduced decreases.

In order to avoid sample-size confounds, Richards and Malvern (1997) introduced a new measure of lexical diversity, D. The CLAN software (Computerized Language Analysis) widely used in language development studies incorporates the program *vocd* (MacWhinney 2000; McKee et al, 2000) which calculates D as follows. First, a number of random tokens are selected from the text without being replaced and the TTR is estimated. This process is repeated 100 times for the same number of tokens, and the whole process is repeated three times with increasing numbers of selected tokens. Then, the average TTR for each of these samples is plotted to form the empirical curve. Finally, D is calculated as an estimate that maximizes the fit to the empirical TTR curve. Lower D values indicate poorer lexical diversity while higher D values indicate a rich vocabulary.

We will calculate both TTR and D values for all Emirati Arabic child language samples and try to establish whether there is any correlation of these indices with age and morphosyntactic development.

Finally, we will look briefly at the ratio of utterances over turns (UoT), i.e. the number of utterances the children produce in each turn. This index measures the number of complete ideas expressed by the child during each turn taken and it is predicted that as children grow older they should exhibit longer conversational turns.

Methodology

The EMALAC Database

The EMALAC corpus is based on 41 half-hour recordings of six Emirati children, three girls and three boys, taken roughly every two weeks, for a period of two years. Most of the children were available for the majority of the recordings, while two children participated in around half of the available 43 sessions.

All children come from a middle class socio-economic background with high school and university educated parents employed in government positions or staying at home. The language in their home environment is Emirati Arabic, while they are also minimally exposed to a pidgin variety of Arabic and English, spoken by the domestic helpers in the house.

The investigator visited the children's house and recorded interactions between the children themselves as well as between the children and the investigator. The investigator and the children were related and familiar to each other (first or second degree relations), thus excluding any possible low volubility effect because of lack of familiarity[2].

After completion of the recording phase, the project assistants transcribed the recorded session in broad IPA transcription. Transcription was entered directly by an Emirati Arabic native speaker into a customized database platform. The transcription and coding format followed a simplified version of Codes for the Human Analysis of Transcripts (CHAT, MacWhinney 1991). A second native speaker checked each transcribed file for accuracy.

We used standard rules for utterance segmentation (e.g. Brown, 1973). We followed Miller & Chapman (2004) in establishing utterances as communication units (C-units), i.e. utterances that include only one main clause with attached possible dependent clauses. Utterances that allow only one independent clause seem to be more representative of child speech as they disallow the conjoining of several clauses without pause (Loban, 1976). Answers to yes/no questions, coordinating structures, utterances with very long pauses, usually accompanied by hesitation, incomplete utterances, and so on, were segmented as separate utterances.

The following table provides some basic statistic information of the database, including child names and age-ranges with number of words and number of utterances per child, as well as the total of adult, child-directed utterances:

Child Name	**Gender**	**Age (months.days)**	**Utterances**	**Words**
Fatima	F	46.14-69.21	4183	11326
Abdulaziz	M	42.03-65.09	4737	17017
Mohammed	M	45.05-68.06	4853	15418
Alreem	F	31.19-54.05	1215	2636
Hind	F	21.11-42.17	367	569
Hamad	M	20.00-40-18	824	1873
Child-Directed Adult Speech			8512	29478
Total			24695	78326

2. EMALAC Basic Statistics

Coding and Analysis

The transcribed files were subsequently coded using the CLAN software (Computerized Language Analysis) designed and written by Leonid Spektor at Carnegie Mellon University, which is tailored towards working with the CHAT format.

The annotation process involved the creation of a minMOR grammar, a semi-automatic parser which is a simplified version of the MOR grammar used in the CHILDES project (MacWhinney 2000). Each word in the text was coded following a format that includes lexical and morphological information. For each token of a word the following information was entered in its lexical entry in the minMOR grammar:

1. The surface IPA form of the word.
2. The Grammatical Category.
3. Lemmatization of all forms. For each IPA token we coded the citation form of the stem followed by morpheme separators and the grammatical morphemes it contains.
4. The English translation of the word.

Some examples of lexical entries are provided in (1):

(1) a. damiila{[scat adj&suff]} "dʒamiil-F" =beautiful=
    b. albait {[scat det&n]} "DET#bajt" =the_house=

The rules we used for morphological segmentation are adapted from the ones developed in Dromi and Berman (1982) for Hebrew and Shaalan and Khater (2006) for Gulf Arabic. Morphemes were considered as acquired by the children when they appeared productively in the data in different forms. Thus, the feminine suffix *–a* in adjectives was considered acquired when both feminine and masculine forms appeared in the child data. Prior to that stage, feminine adjectival forms were coded as single morphological units.

An important choice, shared with Dromi and Berman (1982) for Hebrew and Shaalan and Khater (2006) for Gulf Arabic, was to assume that for language development purposes, the root-pattern complexes in EA are considered monomorphemic units. This choice is based mainly on the fact that we do not have any strong experimental evidence that children of these ages are aware of the root-pattern entities as being different building blocks for Arabic morphology (see Berman 1982 for similar observations in Hebrew). In addition, children up to these ages do not productively use roots with a variety of patterns (e.g. to derive causative verbs, or nominalizations). Some acquisition studies (see Aljenaie, Abdalla & Farghal (2011), and references therein) have shown that, while non-linear plural patterns (broken plurals) in Arabic appear relatively early, they are produced in target-like frequency and accuracy later than the "default" feminine affixal plural. Because of the lack of solid experimental evidence on children awareness of root-pattern distinctions, as well as the additional work needed to reliably enter root-pattern distinctions in the annotation scheme adopted for the statistical analysis software used in our study, the study of morphosyntactic development in this paper will concentrate on increase in linear complexity. We leave a more detailed investigation of the development of non-linear patterns for future research.

Once the database was annotated, CLAN allows for a number of tests which automatically calculate developmental indices. In the following section we will discuss the results we obtained by applying the CLAN set of programs on the language samples in the database.

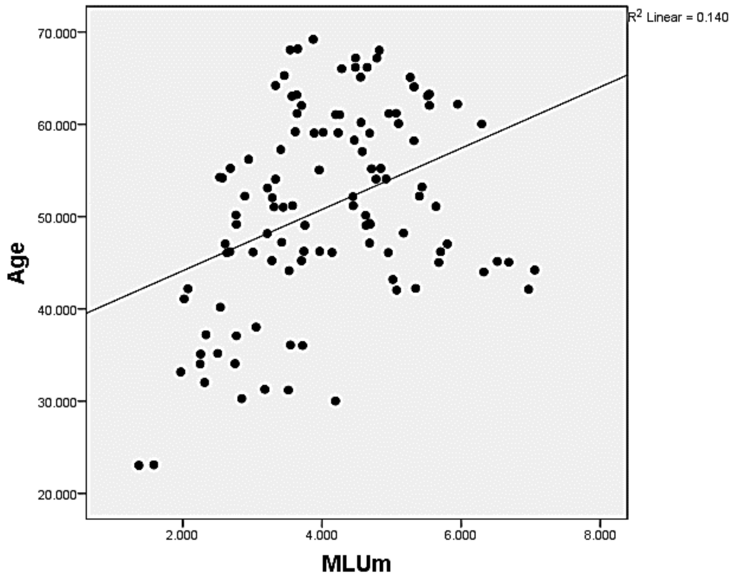
Data and Analysis

MLU Results

MLUm and MLUw were calculated on the EMALAC child files by using the CLAN commands *mlu* for MLUm and *mlt* for MLUw. The program excludes automatically unintelligible utterances or utterances with unintelligible segments, designated in the text as xxx and xx. Full repetitions of the previous word or set of words or of the previous utterance are also excluded. Finally, the program ignores rote passages such as nursery rhymes and songs.

The remaining utterances were scanned for morphological segmentation information, and the total number of morphemes was divided by the total number of relevant utterances, to produce MLUm counts for all children. Once all MLUms were calculated for all data points, we ran a series of Spearman two-tailed correlation tests to assess the relationship between child age and the target developmental indices.

Firstly, the correlation between age and MLUm was found to be highly significant (s(112)= .343, p<0.01, see Scattergram 3).



3. Age – MLUm Correlation

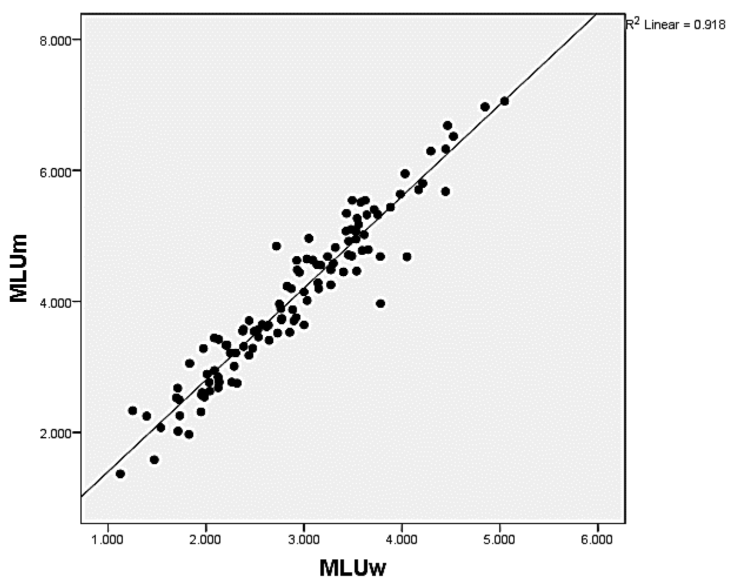
This is an important finding as it matches findings in English and other languages. It is by no means a surprising result, given that similar results have been discovered in a number of other languages (see for example Arlman-Rupp, de Haan & van de Sandt-Koenderman, 1976, for Dutch; Hickey, 1991, for Irish; Linares-Orama & Sanders, 1977, for Spanish; Thordardottir & Weismer, 1998, for Icelandic; Dromi & Berman, 1982, for Hebrew, a Semitic language; and Shaalan and Khater, 2006, for Gulf Arabic). However, this is the first time that such results have been confirmed for an Arabic dialect, based on a longitudinal study.

An important pattern that emerges from the data is that the relation between age and MLUm growth is higher in earlier stages but slows down to non-significant levels at around 4 years of age. Thus, if we collapse together the data from the three younger children (age range 21-54 months), we get a strong correlation between the two variables (s(31)= .446, <0.05). However, when we follow the same process for the three older children (age-range 42-69 months) we find no correlation at all (s(86)= -.129). This is also consistent with findings in other languages showing that MLUm becomes less closely associated with grammatical growth as the child language proficiency develops (Brown, 1973; Miller and Chapman, 1981; Scarborough et al. 1986).

A second issue relevant to the discussion here has to do with the relationship between the two different MLU measures. It has been noted in the literature that MLUm correlates highly with MLUw (.91 to .99) in different languages (Arlman-Rupp et al., 1976; Hickey, 1991; Thordardottir & Weismer, 1998). Shaalan and Khater (2006) investigated the correlation between MLUw and MLUpm[3] (mean length of utterance in productive morpheme, an adaptation of MLUm that, like Dromi and Berman's (1982) approach on Hebrew, takes into account the complex morphological structure of Arabic). Based on language samples obtained from eight Qatari Arabic speaking children aged 2.6-4.5, they found a very strong correlation (r = 0.98) between the two measures. In later work they expanded the sample to 40 children and the new results confirmed their initial findings (Khater & Shaalan 2007).

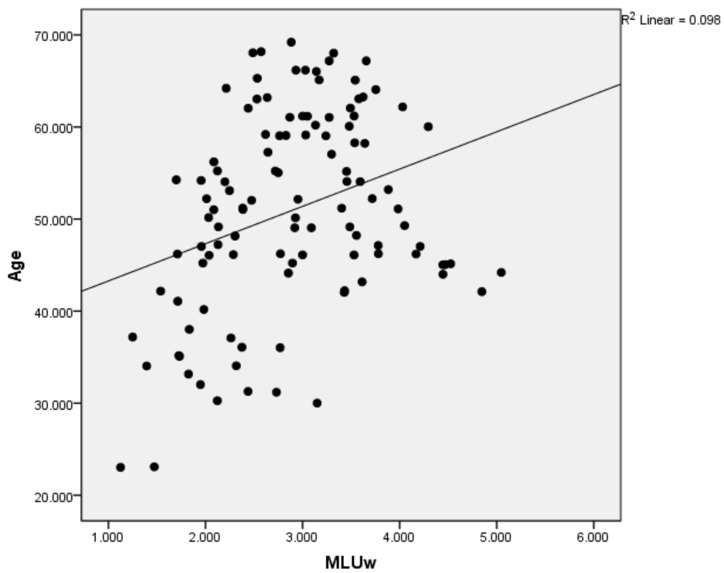
We have seen that the calculation of MLUm presents certain difficulties as, for example, one needs to attest that children productively use masculine-feminine noun and/or adjective alternations in order to start calculating the feminine affix *–a* as a separate morpheme (see discussion in the section on Methodology). As a result, such strong correlations have led investigators to suggest dropping MLUm as the default index of language development, since the calculation of MLUw, based on counting words, does not involve ad hoc decisions on morpheme segmentation and *"is easier, faster, more reliable, and theoretically more sound …. the high correlation between MLUm and MLUw suggests that it is unnecessary to use MLUm as a means of calculating MLU, especially given the uncertain nature of morpheme development"* (Hickey, 1991: 565)[4].

Going back to our data, we found a similar result. MLUm and MLUw show an extremely strong correlation (s(112)= .958, <0.01), as the following scattergram illustrates:



4. MLUm – MLUw Correlation

We checked the correlation between MLUw and age, and we also found a significant correlation: (s(112)= .288, <0.01):



5. Age – MLUw Correlation

It seems, therefore, to be advantageous, especially in clinical contexts, to restrict the diagnosis of language development to MLUw tests (see for example Shaalan and Khater, 2006, for Gulf Arabic). However, a comparison of graphs (3) and (5), indicates that while the correlation between MLUw and age is quite strong, it is weaker than the one between MLUm and age. We assume therefore that more work needs to be done in this area, preferably with a wider pool of subjects for more accurate statistical information, before a decision is made regarding the necessity or redundancy of using MLUm along with MLUw.

Lexical Diversity

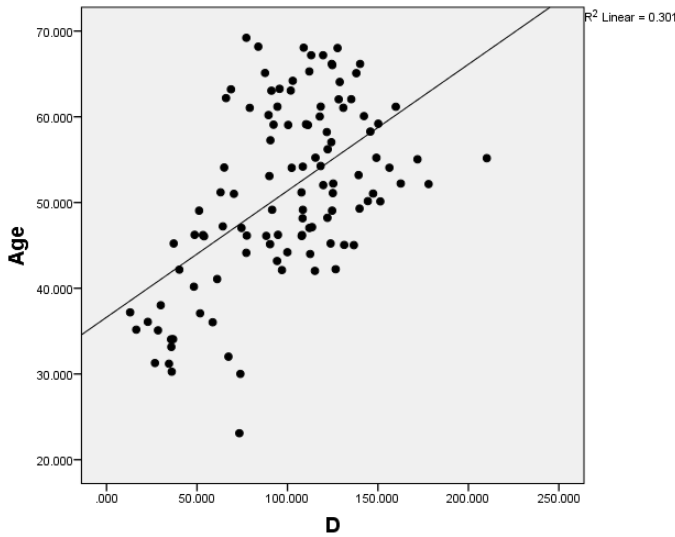
Let us now turn to our results on lexical development. One of the main problems that we faced when coding the data is our decision to use IPA transcription for the recordings. As a result, similar tokens of words were transcribed slightly differently. For example, the children pronounced the adjective *beedˤa* "white.FEM" in all the following forms:

(2)	a. beedˤa	b. beeðˤa
	c. ʔabeedˤa	d. ʔabeeðˤa
	e. ʔabyad	f. beeðˤha
	g. ʔebeyyadah	h. abeedˤa

All these forms would be parsed as different 'types' by CLAN. This problem is however only because of the transcription conventions that we used. An additional, more significant problem has to do with the fact that words that are derivationally related, are also parsed as different types. Thus, *ʔabyaðˤ* "white.MSC" and *labyaðˤ* "the white one (msc.)" are both based on the stem *abyadˤ*, but CLAN parses them as different types. In order to avoid these problems, after completing the transcription of the EMALAC corpus, each transcript was lemmatized in minMOR (see Treffers-Daller, 2013 and earlier work for a similar approach). Thus all word forms in (2) were coded in the lexicon as based on the lemma *beedˤa*. A second lemma *abjadˤ* was assumed for the masculine forms of the adjective. Pending further experimental research, at this point we could not establish whether children were aware of a deeper derivational connection between the masculine and feminine forms of the adjective, and thus we chose to treat the two forms as different lemmas as they are built on separate templates.

Once lemmatization was completed, we ran lexical diversity measuring commands on the MOR tier in CLAN, i.e. the tier that lists all words with lemma, grammatical category, and morphological information. We first computed TTR counts of all children in all files. The results show no correlation between age and TTR ((s(112)= .062). This is not an unexpected result, given our discussion in section 2. The size of different files in the corpus is extremely diverse and thus for each child, texts with few utterances show a higher TTR count than texts with many utterances, at the same age level. For example, in her 5^{th} recording, at age 4;1.15, Fatima produces 173 utterances with a TTR count of 0.491 types over tokens. In the immediately subsequent file, at age 4;2.04 (i.e. 19 days later), she only produces 23 utterances and her TTR count jumps to 0.825. This is a massive difference, but the second count is augmented only because of the small size of the input file.

As was discussed in section 2, a way out of this problem is to use the D index (Richards and Malvern, 1997), which is an indicator of the aggregate probabilities of word occurrences in a text and is independent of the size of the transcript. We ran the vocd command of CLAN on the MOR line, targeting the lemmas of words and ignoring morphological information. Our results show that the correlation between age and D counts is significant (s(112)= .549, <0.01):



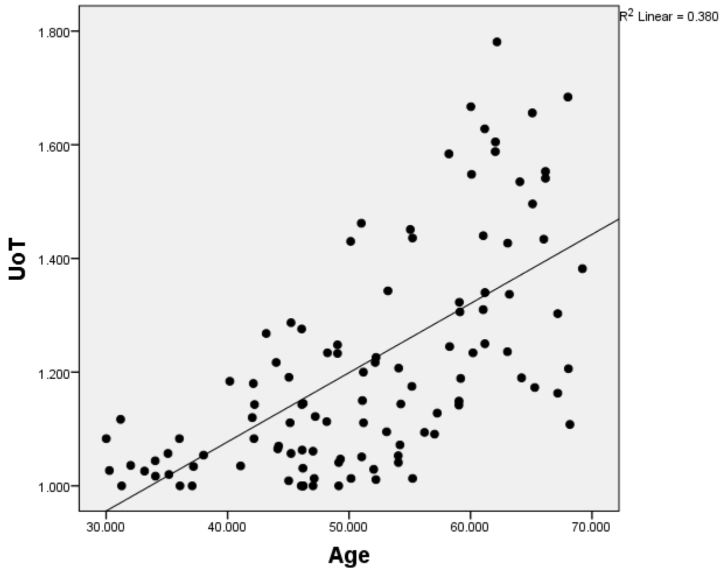
6. Age – D Correlation

This confirms that D is a better indicator of lexical development than TTR in Arabic as well. In addition, a significant correlation is found between MLUm and D (s(112)= .525, <0.01), which seems to indicate that morphosyntactic and lexical development follow similar and parallel growth patterns in the development of Emirati Arabic.

Utterance over Turn Counts

Finally, we looked into the correlation between age and Utterance over Turn (UoT) counts in the data. A turn here is defined as continuous speech by the child, which is not interrupted by another child or adult speaker, over one or more utterances. As children become older and their language grows, they learn how to contribute to the conversation, collaborating with the other speakers and maintaining the conversation topic. McTear (1985) shows that children of up to around 3;8-4;0 mainly contribute to the conversation with initiation-response pairs and thus each turn usually consists of a single utterance. At later stages, and especially with the acquisition of modal auxiliaries and negation, greater topic continuity is achieved.

The UoT index measures the number of utterances in each turn that the child takes, in other words, the number of complete ideas expressed by the child during each turn taken. Based on McTear's (1985) observations, it is predicted that as the children grow older they should exhibit longer conversational turns (i.e. they should hold the floor for a longer period of time during conversation). This is also confirmed by our data. We found a high correlation between age and UoT counts (s(112)= .617, <0.01, Scattergram 7):



7. Age – UoT Correlation

Discussion

The results discussed in the previous sections are a first attempt to explore the developmental stages of Emirati Arabic based on a longitudinal corpus of the language. While not surprising, given their similarity to results obtained in other languages, these results can form the foundation on which further research on figuring out the developmental stages of Emirati Arabic (and Arabic in general) can be based. This is an important research project because of the fundamental lack of work in this area of Arabic linguistics.

The fact that the results are not surprising, may be related to the fact that the current study has concentrated on the acquisition of linear patterns of morphology in EA. However, Arabic has a unique word structure (nonconcatenative morphology) commonly analyzed in terms of a unique inventory of morphological units (the discontinuous roots and word patterns) and processes (root-and-template mapping). In addition, it is characterized by intriguing syntactic constructions (nominalizations, construct state), the acquisition of which has not yet been adequately explored. Arabic morphology is peculiar in that word constituents are not appended one after the other, like in English; rather, they are intertwined in a way that makes them discontinuous and highly abstract. Most morpheme-based analyses of Arabic morphology agree that consonantal roots and word patterns (also called prosodic templates) constitute the building-blocks of Arabic words, but there has been no serious study of what role, if any, these units play in language acquisition. If abstract consonantal roots and prosodic templates turn out to be active in child language, then this would corroborate recent neuropsychological data (Prunet, Béland and Idrissi, 2000; Idrissi, Prunet and Béland 2007), morphological priming data (Boudelaa &

Marslen-Wilson 2001), and recent neurophysiological data showing brain correlates of morphological decomposition to the root (Al-Kaabi 2015, Williams & Marantz 2015).

In analyzing the results obtained in our study, we have not taken into account the development of non-linear morphological processes in EA because we have not yet determined how to implement non-linear patterns in the coding method used by CLAN, the software used for statistical analysis of our data. Further work is needed in order to establish ways to check how non-linear patterns develop in children. We leave this for future research.

Given the different levels of morphological complexity involved in different languages, and different conventions in word segmentation practices, indexes of morphosyntactic development such as MLUm and MLUw cannot readily be used for crosslinguistic comparison of developmental paths. Trying to adapt word-segmentation rules to allow for greater convergence of MLU scores between languages can lead sometimes to ad hoc rules and does not necessarily guarantee that MLU measures are equivalent across languages (see Eisenberg, McGovern Fersko, and Lundgren, 2001). In fact, as Eisenberg et al (2001: 331) show, even in bilingual children, because of L1 effects on L2, MLU data from monolingual speakers would not be appropriate for children with the same L1, learning a second language.

However, MLU counts can be used to show similar developmental paths. Ignoring differences in actual numbers of MLU scores that make direct comparison impossible, one can still track similar patterns in MLU growth, showing that such similarities correspond to similar developmental paths. We believe that our results indicate that Emirati children follow similar developmental trajectories as children speaking other, better-studied languages. The strong MLUm and MLUw correlations with age as well as with other developmental indices such as D and UoT, show that Emirati children follow very similar developmental patterns with, for example, children acquiring English.

If we split the ages of our target children into three stages, with the first stage corresponding to 23-36 months, the second to 37-48 months, and the third 48 months and older, a pattern emerges. The MLUm of children in the first stage ranges between 1.368-4.194 morphemes per utterance. The range is 2.826 morphemes per utterance, showing a variance of 0.696. In later stages, MLUm counts range between 2.019-7.057 in Stage 2 and 2.530 and 6.295 in Stage 3 showing greater variance, as was discussed in Section 4. The mean value of MLUm for stage 1 is 2.71 morphemes per utterance, while in stage 2 it increases to 4.24 and remains the same for Stage 3. Thus, from around the second to the third year of age, children exhibit an increase of 1.53 morphemes per utterance in mean MLUm value. This is compatible with results in studies on other languages. For example, Klee et al (1989) found that the MLUm of their sample of normally developing English children increases by an average of 0.085 morphemes a month (1.02 morphemes per year).

Obviously, the statistical results reported here are not sufficient to establish robust developmental stages like those established for English in Brown (1973) (see Table 1, in Section 2). In order to achieve a high level of accuracy in pinpointing exactly these correlations between MLUm ranges and certain developmental stages of Emirati children, we need to broaden considerably the pool of target children, including more children representative of each age and from different demographics (gender, linguistic background, geographical location, and so on)[5].

Such work can further contribute to the development of diagnostic tests for atypical language development in Arabic. Language Sample Analysis (LSA) has been used in speech pathology for a long time as an alternative to standardized tests (see Eisenberg et al 2001 for a review). As Eisenberg et al (2001: 323) report, a number of surveys in the USA have shown that LSA use in diagnosing atypical language development constitutes an increasing practice. A survey in 2000 (Loeb et al 2000) shows that 93% of speech-language pathologists reported using LSA, with MLUm reported as the most frequently listed LSA index, with 91% usage.

In Arabic, MLUm counts have also been used to establish different developmental trajectories between normal language development and populations with a form of atypical language (see Abdalla 2002; Shaalan 2010). However, these counts have focused on small populations, tailored towards specific studies of atypical language development, and based on spontaneous language samples. As far as we know, there has been no systematic study to establish robust developmental stages for typically developing Arab children. To this respect, a thorough description of typical phonological, morphological and morphosyntactic development of Arabic will undoubtedly be of great use to clinicians. We hope that the study reported here will provide the foundation for providing such a description.

Finally, establishing stages of morphosyntactic development is a process that cannot solely rely on MLU counts. In his original work, Brown (1970) uses MLU counts to set the stages listed in Table 1 of Section 2 here, but he provides additional evidence from the development of certain morphosyntactic properties to support this division. Thus, in Stage I the children seem to develop semantic roles and syntactic relations; in Stage 2 a number of grammatical morphemes appear and meaning gets modulated; in Stage III the children develop modalities and simple sentences[6]; and so on. However, as Brown (1970:59) observes, *"… the whole development of any one of the major constructional processes is not contained within a given stage interval. Semantic roles go on developing after Stage I; the modulations of meaning extend from Stage II to beyond even Stage V."* We hope our study will set the foundation for establishing similar stages for EA. Obviously, these stages will need to be reinforced by establishing the periods when similar morphosyntactic processes develop in EA. However, this is a long-term research project beyond the scope of this paper.

Taking into account current ongoing research projects (e.g. Ntelitheos, Idrissi and Tamimi, 2009; Ntelitheos, 2013), we can get an idea on how certain grammatical properties fit with possible MLU–defined stages in EA. Thus, in Brown (1970:271) articles (definite determiner "the" and indefinite determiner "a(n)") are shown to develop in Stage III for Sarah (age 3:1) and Stages IV (age 3;8) for Adam and V (age 2;3) for Eve. In Ntelitheos et al (2009) it is similarly shown that the definite determiner in EA does not appear almost at all in ages 1;6-2;5, and appears in only around 28% of the total of noun phrases in ages 2;6-3;0. Children reach adult-like targets after the age of 3;1 and onwards.

A second case can be drawn from the development of possessive structures in EA (Ntelitheos 2013). Children seem to use the free state, analytic possessive structure, headed by the possessive particle *ma:l*, in the very early stages of grammatical development (1;23-3;00). In later stages, (3;1-5;00) children reach target frequencies, where the construct state overwhelmingly outnumbers the free/analytic state, as is the case with adult-to-adult exchanges. These stages are somewhat different than the corresponding stages of possession acquisition in English (stages II-IV, ages 1;11-3;2) but this may be due to the complexity encoded in the construct state in EA (see Ntelitheos 2013 for discussion).

It is obvious that further work needs to be done in order to gain a better understanding of the hierarchical order of morphosyntactic acquisition in EA. This is an important step towards establishing robust developmental stages in EA language acquisition. We hope that the work reported here sets the foundation for further development in this important subfield of Arabic linguistics.

Conclusion

We tested the validity of a number of developmental measures in Emirati Arabic, based on conversations of 6 Emirati children over a period of 24 months. We found positive correlations between MLUm/MLUw/D/UoT indices and age for all 6 children, while no correlation between age and TTR was found, a possible effect of the size of input data files. Besides providing a new data point in the body of knowledge about language growth and development in general, the present results show that language acquisition displays the same developmental patterns cross linguistically when linear patterns of morphological development are considered.

References

Abdalla, F. A. (2002). Specific language impairment in Arabic-speaking children: deficits in morphosyntax. Ph.D. Dissertation, School of Communication Sciences and Disorders, McGill University.

Al-Kaabi, M. (2015). The Anatomy of Arabic Words: The Role of the Root in

Representations and Processing. Ph.D. Dissertation, Department of Linguistics, New York University.

Aljenaie, K., Abdalla, F. & Farghal, M. (2011). Developmental changes in using nominal number inflections in Kuwaiti Arabic, *First Language 31*, 222-239.

Arlman-Rupp, A., de Haan, D. & van de Sandt-Koenderman, M. (1976). Brown's Early Stages: Some evidence from Dutch, *Journal of Child Language*, *3*, 267-274.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language*, *20*, 139-152.

Bol, G. W. (1996). Optional subjects in Dutch child language. In Ch. Koster &F. Wijnen (Eds.), *Proceedings of the Croningen Assembly on Language Acquisition* (pp. 125-133). Groningen, The Netherlands: Centre for Language and Cognition, Groningen.

Boudelaa, S., and Marslen-Wilson, W. D. (2000). Non-concatenative morphemes in language processing: Evidence from Modern Standard Arabic. In J. McQueen and A. Cutler (eds.) *Proceedings of the Workshop on Spoken Word Access Processes,* (p.p. 23–26), Nijmegen: Max Planck Institute for Psycholinguistics.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Crystal, D. (1974). [Review of the book] A First Language: The Early Stages. *Journal of Child Language*, *1*, 289-334.

DeThorne, L.S., Johnson, B.W., & Loeb, J.W. (2005). A closer look at MLU: What does it really measure*? Clinical Linguistics & Phonetics, 19, 8*, 635- 648.

Dromi, E. & Berman, R. A. (1982). A morphemic measure of early language development: Data from Modern Hebrew. *Journal of Child Language*, *9*, 403-424.

Eisenberg, S. L., McGovern Fersko, T., and Lundgren, C. (2001). The use of MLU for identifying language impairment in preschool children: A review. *American Journal of Speech-Language Pathology, 10*, 323-342.

Gwilliams, L. & Marantz, A. (2015). Non-linear processing of a linear speech stream: The influence of morphological structure on the recognition of spoken Arabic words. *Brain and Language*, *147*, 1–13.

Hickey, T. (1991). MLU and the acquisition of Irish. *Journal of Child Language*, *18*, 553-569.

Idrissi, A., Prunet, J-F. & Béland, R. (2007). On the Mental Representation of Arabic Roots. *Linguistic Inquiry, 39, 2*, 221–259.

Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language*, *29*, 713-733.

Khater, M., & Shaalan, S. (2007). Reporting norms for mean length of utterance (MLU) in words and morphemes for Qatari Speaking Children. Paper presented at Linguistics in the Gulf Conference, University of Qatar, Doha, Qatar.

Klee, T., Schaffer, M., May, S., Membrino, I. and Mougey, K. (1989). A Comparison of the Age-MLU relation in Normal and Specifically Language-impaired preschool children. *Journal of Speech and Hearing Disorders*, *54*, 226-233.

Klee, T. (1992). Measuring children's conversational language. In S. F. Warren & J. Reichle (Eds.), *Causes and effects in communication and language intervention* (pp. 315–330). Baltimore, MD: Brookes.

Linares-Orama, N., & Sanders, L. J. (1977). Evaluation of syntax in three-year-old Spanish speaking Puerto Rican children. *Journal of Speech, Language, and Hearing Research, 20*, 350–357.

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.

Loeb, D. F., Kinsler, K., & Bookbinder, L. (2000). Current language sampling practices in preschools. Poster presented at the Annual Convention of the American Speech-Language Hearing Association, Washington, D.C.

MacWhinney, B. (1991). *The Childes Project: Tools for Analyzing Talk*. Hills- dale, NJ: Erlbaum.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk, Vol. 2: The database*. 3rd ed. Mahwah, NJ: Lawrence Erlbaum

Malvern, D.D. and Richards, B.J. (1997). A new measure of lexical diversity. In Ryan, A. and Wray, A., (Eds.), *Evolving models of language* (p.p. 58-71). Clevedon: Multilingual Matters.

McCarthy, P. M. and Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing 24*, *4*, 459–488.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring Vocabulary Diversity Using Dedicated Software, *Literary and Linguistic Computing, 15, 3,* 323-337.

McTear, M. (1985). *Children's Conversation*. Oxford: Blackwell.

Miller, J. F. (1981). *Assessing language production in children: Experimental procedures*. Baltimore, MD: University Park Press.

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, *24*, 154-161.

Miller, J. F., & Chapman, R. S. (2004). *Systematic analysis of language transcripts (SALT, v8.0)* [Computer software and manual]. Madison, WI: Language Analysis Laboratory, Waisman Center, University of Wisconsin-Madison.

Ntelitheos, D., Idrissi, A. & Tamimi, O. (2009). Acquisition of definiteness in Emirati Arabic. Proceeding of the 10th Annual UAE University Research Conference, 1235-1243.

Ntelitheos, D. (2013). A Corpus Study of the Distribution of Possessives in Child and Adult Emirati Arabic. Paper presented in Experimental Arabic Linguistics (EXAL 2013), UAEU, Al Ain, 6-7 November, 2013.

Parker, M. D. and Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, *25*, *3*, 365-376.

Prunet, J-F., Béland, R. and Idrissi, A. (2000). The mental representation of Semitic words, *Linguistic Inquiry, 31*, 609–648.

Richards, B. J. (1987). Type/Token Ratios: what do they Really Tell us? *Journal of Child Language, 14*, 201-209.

Richards, B. J., and Malvern, D. D. (1997). Quantifying Lexical Diversity in the Study of Language Development. *The University of Reading New Bulmershe Papers*.

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11*, 1-22.

Scarborough, H. S., Wyckoff, J., & Davidson, R. (1986). A reconsideration of the relation between age and mean utterance length. *Journal of Speech and Hearing Research, 29*, 394–399.

Shaalan, S. (2010). Investigating Grammatical Complexity in Gulf Arabic Speaking Children with Specific Language Impairment (SLI). Ph.D. Dissertation, University College London.

Shaalan, S. and Khater, M. (2006). A Comparison of Two Measures of Assessing Spontaneous language Samples in Arabic Speaking Children, Poster presented at Child Language Seminar, University of Newcastle Upon Tyne.

Templin M. C. (1957). *Certain language skills in children*. Minneapolis: University of Minnesota Press.

Thordardottir, E. T., and Weismer, S. E.. (1998). Mean length of utterance and other language sample measures in early Icelandic. *First Language, 18*,1-32.

Treffers-Daller, J. (2013) Measuring lexical diversity among L2 learners of French: an exploration of the validity of D, MTLD and HD-D as measures of language ability. In: Jarvis, S. and Daller, M. (eds.) Vocabulary knowledge: human ratings and automated measures. Benjamins, Amsterdam, pp. 79-104.

[1] This project was financially supported by UAEU individual grants #1492-01-01-10, 02-06-1–11/08, and 01-06-1–11/09. We would like to express our gratitude to the children who participated in this study and their parents for granting their consent and to our RAs Othaiba Tamimi, Sara Al Qahtani, Fatima Al Eisai and Aamna Shemaili for transcribing, translating, and coding the data as well as their contribution to discussions within the EMALAC group and the write-up of various papers. All errors are of course our own.

[2] Thanks to an anonymous reviewer for commenting on this issue.

[3] Our MLUm counts are calculated in the same way as MPU (morpheme-per-utterance) counts for Hebrew (in Dromi and Berman (1982) MLUpm for Arabic in Shaalan and Khater (2006). Thus, for this stage of calculations we exclude root-pattern morphological distinctions and we take the root-pattern complex to be a single morpheme for developmental measures.

[4] Similar strong correlations have been found between MLUm and other measures of utterance length such as syllable counts but measuring syllables for child speech runs into problems with children's tendency to duplicate syllables and to use diminutives (Hickey, 1991).

[5] The EMALAC team has initiated a project of language sample collection from child centers across the UAE, funded by UAEU individual grant 31H074/2015.

[6] Thanks to an anonymous reviewer for bringing this to our attention.