

ARTICLE

Input effects in the acquisition of verb inflection: Evidence from Emirati Arabic

Marta SZREDER¹, Laura E. DE RUITER² , and Dimitrios NTELITHEOS¹

¹Department of Speech Language Pathology, United Arab Emirates University and ²Eliot-Pearson Department of Child Study and Human Development, Tufts University, USA

Address for correspondence: Marta Szreder, United Arab Emirates University College of Medicine and Health Sciences, Department of Speech Language Pathology, PO Box 15551 Al Ain, United Arab Emirates. E-mail: marta.szreder@uaeu.ac.ae

(Received 28 January 2020; revised 12 August 2020; accepted 19 January 2021; first published online 20 May 2021)

Abstract

This study investigates the acquisition of the Imperfective verb inflection paradigm in Emirati Arabic (EA), to determine whether the learning process is sensitive to the phonological and typological properties of the input. We collected data from 48 participants aged 2;7 to 5;9 years, using an elicited production paradigm. Input frequencies of inflectional contexts, verb types and tokens were obtained from corpora of child-directed and adult EA. Children's accuracy was inversely related to the input frequency of inflectional contexts, but not related to type and token frequency or phonological neighborhood density. Token frequency interacted with age, such that younger children performed considerably worse on low-frequency tokens, but older children performed equally well on high- and low-frequency tokens. We conclude that learning is input-driven, but that a sufficiently regular paradigm allows children to eventually generalise across all items earlier than in previously studied European languages.

Keywords: verb inflection; morphological acquisition; acquisition of Arabic; input-driven learning; frequency effects

Introduction

Background

Studies of children's acquisition of subject-verb agreement often report high production accuracy early in development (Smoczyńska, 1985), low rates of commission errors (Hoekstra & Hyams, 1998), and cross-linguistic regularities in the order of acquisition, leading some accounts to emphasise maturational factors (Hoekstra & Hyams, 1998; Wexler, 1998) and universal features of grammatical structures (Ackema & Neeleman, 2018; Harley & Ritter, 2002) as the primary determinants of the learning process. However, the expanding range of research methods and cross-linguistic data continue to reveal patterns of asymmetries in children's performance, which varies significantly across inflectional contexts and languages. These asymmetries highlight the role of multiple language-specific factors.

© The Author(s), 2021. Published by Cambridge University Press

One such factor is the distributional properties of the ambient language, the effects of which are often obscured in naturalistic studies due to the fact that children self-select words to attempt, and therefore forms with low frequency and low familiarity are difficult to assess. Once this is controlled for, even when overall accuracy is high, particular morphological endings exhibit much higher error rates, often inversely proportional to the input frequency of the given inflectional context and token (Aguado-Orea & Pine, 2015; Räsänen, Ambridge & Pine, 2016, Engelmann, Granlund, Kolak, Szreder, Ambridge, Pine, Theakston & Lieven, 2019). Indeed, these effects are also robust in the acquisition of noun morphology (Dąbrowska, 2008; Granlund, Kolak, Vihman, Engelmann, Ambridge, Pine & Lieven, 2019; Savičiūtė, Ambridge & Pine, 2018), as well as in many other areas of language development (Ambridge, Kidd, Rowland & Theakston, 2015; Bybee & Hopper, 2001; Lieven, 2010).

Another determinant of cross-linguistic variation in the acquisition of grammatical structures is cue reliability, meaning that highly consistent and distinctive form-function pairings are learnt more quickly (Bates & MacWhinney, 1989; MacWhinney, 1987). For example, the comparatively late acquisition of English 3rd person singular -s has long been attributed to its relative opacity (Brown, 1973; De Villiers & Johnson, 2007). Similarly, cue reliability has been invoked to account for differences in the development of agreement comprehension in English, Spanish, and French (Legendre et al., 2014). Cue reliability has also been shown to explain many cross-linguistic differences in the acquisition of the transitive construction, related to the relative weightings of the cues of word order, and agreement and case marking (Abbot-Smith & Serratrice, 2015; Chan, Lieven & Tomasello, 2009; Krajewski & Lieven, 2014).

Finally, phonological factors can also explain some of the differences in morphological development across languages. Children's production of functional morphemes has been shown to be affected by utterance position and phonological complexity (Song, Sundara & Demuth, 2009; Theodore, Demuth & Shattuck-Hufnagel, 2011), and word position (Kuczaj, 1979; Santelmann, 1998). This is further complicated by prosodic factors (Demuth, 2018; Lleó & Demuth, 1999), meaning that different languages lead children to become sensitive to different phonological cues. For example, word onsets are particularly salient in English, but often not so in languages with different accentual and phonotactic patterns, as evidenced in both children's word recognition (Hallé & de Boysson-Bardies, 1996; Vihman, Nakai, DePaolis & Halle, 2004) and production (Savinainen-Makkonen, 2007; Szreder, 2013).

The distributional, phonological, and typological properties of the ambient language thus have complex effects on learning, such that they affect the saliency and availability of the target structures, as well as shaping children's sensitivity to particular features. In order to understand how these language-specific factors interact in the acquisition of agreement, crosslinguistic evidence is needed. The current study sets out to provide such evidence from Emirati Arabic (EA), a so far understudied language, where agreement is expressed with highly reliable and phonologically salient syllabic prefixes and circumfixes.

Emirati Arabic

Emirati Arabic (EA) is a variety of Arabic spoken in the region of the United Arab Emirates (UAE), which belongs to a continuum of varieties spoken in the broader

region, grouped under the dialectal name of Gulf Arabic. The latter is considered a koine dialect intelligible in and around both shores of the Arabian Gulf including the countries of Kuwait, Bahrain, Qatar, the United Arab Emirates, and parts of Saudi Arabia and Oman. This is quite a large area with considerable variation between local varieties, which has not received much discussion in the Arabic dialectal literature. It is thus perhaps more accurate to think of Gulf Arabic as a dialect continuum with some core similarities rather than as a single dialect. The term Emirati Arabic (as used here) refers to a roughly-defined group of varieties that share core characteristics with specific phonological, lexical, and morphosyntactic idiosyncrasies and a certain degree of intra-dialectal variation which is mostly geographically defined. It incorporates grammatical properties of smaller varieties within the UAE, mainly of tribal nature, including the broader varieties of Dubai and Abu Dhabi (cf. Johnstone, 1967), as well as a number of other varieties with minor, mainly phonological, distinctive properties.

As with other Arabic varieties and dialects, EA verbs are built on a usually trilateral root (three root consonants) and less frequently quadrilateral (four-consonant) ones, by adding one of nine possible templates (sequences of vowels and possibly affixes), deriving nine forms with certain morphosyntactic properties, e.g., exhibiting causative, applicative/associative, inchoative, mediopassive, or reflexive/reciprocal interpretations (Leung, Ntelitheos & Al Kaabi, 2021). In addition, each of the nine verb forms appears in two different stems with distinct aspectual properties: a perfective stem for completed (past) actions and an imperfective stem for present and future ongoing actions.

Table 1 presents the EA verb forms for the perfective stem and the Modern Standard Arabic (MSA) forms for comparison. Following tradition, we represent different forms using Latin numerals and the trilateral root *f-ʕ-l* ‘to do’. It must be noted that the list of meanings of the forms is not exhaustive, as they exhibit some variability. Furthermore, many verbs do not appear in all forms. Form I is considered the basic form of each verb. As regards MSA Form IV, it is not used in EA.

EA verbs are further traditionally classified according to the phonological properties of the root. In keeping with the tradition, we distinguish strong verbs (which do not have a vowel as part of the root), doubled verbs (in which two consonants in the root are identical), and weak verbs (whose root involves a vowel, glide, or the glottal stop) (Feghali, 2008). While this classification may determine the vowel template of the verb, it does not affect affixation in Present Imperfect.

Subject agreement in EA expresses gender, number and person (Qafisheh, 1977; Holes, 1995). The two aspectual stems (perfective and imperfective) exhibit different subject agreement patterns with respect to the affixes they take. All subject agreement for the perfective stem of the verb in EA is expressed with syllabic suffixes. In contrast, the imperfective form is always marked by a syllabic prefix or circumfix.

Table 2 illustrates the verb inflection paradigm for the root *k-s-r* ‘to break’. While the imperfective affixes are not strictly analytical, the general pattern is that the prefix carries the information of person (with the exception of 3SG.F), and the suffix – number and gender. A separate set of circumfixes exists for 2PL.F and 3PL.F, but they are almost never used in the dialects, and instead the traditionally masculine affixes are used with both genders.

There is some variability in the realisation of these forms in casual speech and some imperfective stem suffixes may be omitted. Some speakers drop the final /-n/ of the plural suffix and feminine singular suffix in the imperfective form. The final /-n/ of the plural and feminine singular suffix may also be deleted before the indirect object

Table 1. EA verb forms for the root f-ʕ-l 'to do', with Modern Standard Arabic (MSA) equivalents for comparison.

Form	Meaning/Use	MSA	EA
I	regular	faʕal	faʕal
II	causative/denominative	faʕʕal	faʕʕal
III	reciprocal	faaʕal	faaʕal
IV	causative	ʕafʕal	---
V	reflexive of II	tafaʕʕal	tfaʕʕal
VI	reflexive of III	tfaaʕal	tfaaʕal
VII	passive of I	(ʔi)nfaʕal	nfaʕal
VIII	reflexive of I	(ʔi)ftaʕal	ftaʕal
IX	color/defects	(ʔi)ffʕal	ffʕal
X	causative reflexive	(ʔi)stafʕal	(stafʕal)

Based on Al Kaabi & Ntelitheos (2019).

Table 2. Subject agreement affixes on the Imperfective and Perfective verb stem of k-s-r 'to break' in EA.

Context	Imperfective		Perfective	
	Affix	Verb Form	Affix	Verb Form
1SG	ʔa-	ʔaksir	-t	kisart
1PL	nə-	nəksir	-na	kisarna
2SG.M	tə-	təksir	-t	kisart
2SG.F	tə- -in	təksirrin	-ti	kisarti
2PL.M	tə- -uun	təksiruun	-tu	kisartu
3SG.M	jə-	jəksir	-∅	kisar
3SG.F	tə-	təksir	-at	kisrat
3PL.M	jə- -uun	jəksiruun	-aw	Ksiraw

marker /l-/. The short vowels in affixes also tend to vary due to general instability of vowels in Arabic, and may be realised as more or less open. In addition, weak verbs with a glottal stop or glide root consonant exhibit stem alternation – one stem is used in front of third person suffixes and an alternate stem in front of all other affixes. However, these are less common and will not be investigated in the current study.

Previous studies in the acquisition of verb inflection in Arabic

While there is considerable work on the acquisition of noun plural morphology in Arabic (see Albirini, 2017 for an overview), the development of verbal morphology has only received limited attention. Furthermore, a great majority of the studies are

naturalistic, rather than experimental, investigations. However, we have some knowledge from these studies regarding the age of acquisition and the types of errors observed, which we discuss below.

Age of acquisition

The most comprehensive overview of the acquisition of an Arabic dialect is perhaps the study of 37 children (ages 0;6–15;0 years) acquiring Egyptian Arabic by Omar (1973), based mainly on language samples from elicitation and imitation tasks. In her discussion of morphological development, Omar includes data only from the six children in the earlier stages of acquisition (1;11–3;6), following the babbling stage. While Omar's younger subjects (1;11, 2;3) produce few inflected verb forms apart from imitations, very few such uninflected forms are found in the speech of the older toddlers (2;8, 3;6). Omar (1973) concludes that most inflections are acquired by the age of 2;0–2;6 years. Interestingly, Omar observes that the older children still tend to have a better mastery of verbal inflections in spontaneous production than in imitation trials, where verbal agreement is frequently dropped. For example, Child 1 (2;8) accurately produces 1st person agreement, but omits verbal agreement morphology frequently in imitation trials. The same pattern was observed in Child 2 (3;0) and Child 4 (3;6).

Similar findings are presented in another investigation of spontaneous Egyptian Arabic production by Fahim (2017), investigating the development of verb morphology of three children with language impairment (LI) and a group of six typically developing (TD) children with ages ranging between 2;03 and 4;06. The correct provision of subject agreement for person, number and gender appears to develop in the TD group between the ages of two and four years, with the youngest child producing 28% of their forms with agreement errors, and the oldest TD child at 4;06 producing fully grammatical verbal forms.

Only two studies investigate acquisition of agreement inflections using experimental data. The first one, by Basaffar and Safi (2012), discusses the acquisition of verb inflections in the speech of 32 monolingual children aged two to four years, acquiring Hijazi Arabic (a variety of Arabic spoken in parts of Saudi Arabia). The experimental data consists of comprehension and production of real and nonce inflected verbs. In addition, spontaneous speech and data elicited using a video-clip description and a story retell task is also used. The results show steady improvement in performance with age, with a clear difference across conditions, such that error rates are lower in spontaneous speech than in elicited speech, and lower in real words than in nonce words. While spontaneous production shows near-perfect accuracy early on, elicited production reveals errors in 10%–25% of forms, depending on the person-number-gender context and the task, by the age of four. The second experimental study, by Moawad (2006), is an investigation of gender and number agreement in older, school-aged children, in the nominal and verbal systems of Saudi Arabic (in the Abha variety, spoken in the southwest region of Saudi Arabia, in the city of Abha). Picture-based comprehension and production experiments of 98 children between the ages of 6 and 12, as well as a small longitudinal study of three children reveal that even at this older age, accuracy is significantly lower than in adults.

Error types

The most commonly reported error types are errors of omission: that is, substitutions of an uninflected form for an inflected form, and gender confusion. Omar (1973) reports

that most verb inflection errors were errors of omission, and that commission errors usually involved a masculine affix replacing a feminine affix, especially in imperative contexts. Fahim (2017) also reports mostly errors in gender, such that target feminine forms were erroneously produced for both naturally and grammatically gendered subjects. In addition, children produced both feminine and masculine forms in alternation for the same person or object in successive utterances.

Similarly, Moawad (2006), in her study of the agreement between nouns, and adjectives and verbs observed masculine forms substituting for feminine ones. In addition, Moawad (2006) reports that singular forms are often found to substitute for plural and dual (for nouns) forms (as well as plural substituting for dual). This is further confirmed by Abdalla (2002) and Abdou and Abdou (1986; as reported in Basaffar & Safi, 2012), who present a discussion (in Arabic) of morphological development of two children acquiring Palestinian Arabic, from the first words to five years of age. They find that the singular affixes appear before the plural, a difference they attribute to the complexity of the form. Furthermore, they observe a common developmental trajectory for both children, whereby 3rd person is acquired before 1st, which in turn is acquired before 2nd, although this pattern only holds for perfective forms.

By far the most detailed discussion of subject agreement substitution errors in an Arabic dialect can be found in Aljenaie's (2001, 2010) work on the development of Kuwaiti verbal forms, investigating spontaneous language samples in the speech of three children (ages 1;8–3;1). One of the main observations in her imperfective stem data is that masculine forms substitute quite often for feminine forms (19 such errors in the data) while the reverse substitution pattern is very infrequent (only three errors) in the imperfective form of the stem. A second finding in the Kuwaiti imperfective stem data was a pattern of pronoun reversal (Chiat, 1986; Clark, 1978), despite the fact that Arabic is a pro-drop language, in that the children frequently produced the 2nd person singular (feminine or masculine form, depending on the child's gender) instead of the 1st person singular. The two female children used the 2SG.F form in the place of 1SG (three substitutions), and the male child substituted the 2SG.M for the 1SG in 24 instances.

Summary

Overall, the findings from the acquisition of Arabic are compatible with those in other languages. Specifically, naturalistic studies report early high accuracy in spontaneous production, with apparent adult-like proficiency achieved by the age of four. Nonetheless, the increased error rates in imitated speech (Omar, 1973) and persistent non-adult-like performance in experimental conditions (Basaffar & Safi, 2012; Moawad, 2006) suggest that the high accuracy in spontaneous production may in fact obscure a much lower underlying proficiency, which only surfaces when children cannot self-select the forms they attempt.

Furthermore, the error patterns found in children's production of agreement in Arabic mirror the findings from other languages. Firstly, errors of omission are much higher than errors of commission, suggesting that most of children's difficulties result simply from their lack of familiarity with inflectional affixes, rather than their incorrect application. Secondly, masculine forms are often used to replace feminine forms – a pattern also observed in studies of another Semitic language, Hebrew (Berman, 1985; Levy, 1983). Thirdly, singular forms appear before plural forms, and 3rd person before 2nd person forms. As regards the age of acquisition,

despite the high regularity of the paradigm, the findings appear to be different depending on the study. However, none of the studies employ an experimental paradigm allowing for direct comparison with children acquiring other languages.

Research aims and hypotheses

Engelmann and colleagues (2019) reported the results of an experimental investigation of the acquisition of agreement in two highly inflected European languages, Finnish and Polish. Using an elicitation paradigm, they found effects of type and token frequency in both accuracy rates and substitution patterns (i.e., higher-frequency forms substitute for lower-frequency ones), as well as effects of phonological neighborhood density (PND), such that verb types from higher-density neighborhoods were produced with higher accuracy. We decided to use the same experimental paradigm to study the acquisition of agreement in Emirati Arabic, to enable a direct comparison between the developmental paths of children acquiring typologically different languages.

We had three main research aims. The first aim was to establish the distribution of error rates across the different inflectional contexts. We predicted that error rates would be low overall, as universally found in cross-linguistic research, but also that the errors would not be evenly distributed across inflectional contexts. Specifically, we expected to see higher accuracy in singular than in plural, higher accuracy in masculine than in feminine, and higher accuracy in 3rd person than 1st and 2nd person forms, as in previous studies of Arabic and other languages. Since the experimental paradigm is designed for relatively older children (three to six years), we did not aim to address the question of when agreement first develops. However, we predicted that at least some lasting effects of the asymmetries between different contexts would be observed even at the later ages.

The second aim was to determine whether differences in accuracy, if found, are related to the input frequency of inflectional contexts. We provide a first description of EA child-directed speech, and we predicted that we would find the same positive relationship between input frequency and accuracy as previously shown for Polish and Finnish, using the same methodology.

The third and final aim was to determine whether differences in accuracy are related to the input frequency of individual word forms, i.e., type frequency, token frequency, or PND. We wanted to determine whether there are differences in the acquisition of Arabic vs. Polish and Finnish that could be attributed to the typological differences between the languages, affecting cue reliability and accessibility. We hypothesised that given the extremely regular inflectional paradigm, which is encoded in highly phonologically salient syllabic prefixes and circumfixes, children would be able to generalise their knowledge of the paradigm to all verbs early on, and verb identity would not play a role, especially in older children.

Underlying all the above aims is the intention to provide easily comparable data from a typologically different language, using a methodology developed in previous research of highly inflected European languages, to allow for reliable cross-linguistic comparisons.

Method

Participants

We recruited 48 participants, 19 males and 29 females, aged 2;7 to 5;9 years (mean = 4;7). All participants were native speakers of Emirati Arabic with no known language

impairment. The children were tested by native Emirati Arabic-speaking experimenters at home or at school.

Due to the specificity of the local context, it was impossible to ensure that any of the children were monolingual, nor was it realistic to measure the children's exact exposure to different languages. In the UAE, non-native residents constitute 88% of the population, and English is the lingua franca of public life, meaning that all children have regular exposure to the language. Furthermore, many households employ domestic workers, often from South and Southeast Asia, and East Africa, which further affects the linguistic landscape in which the children grow up. Since this multilingual context is common to all Emirati children's experience, it should not be regarded as an interfering factor, but rather as a constant element, inherent to the process of acquisition of the dialect. In this sense, the participants constitute a fairly homogeneous group, in that they all have contact with multiple languages, but not in a systematic way (i.e., they do not attend an English-speaking school or otherwise receive explicit foreign language instruction).

Input-based predictor variables

We set out to investigate the effects of input frequencies of verb types and tokens, and of the different inflectional contexts. To establish the properties of the input, we used the Emirati Arabic Language Acquisition Corpus (EMALAC, Ntelitheos & Idrissi, 2017). The corpus consists of 24,695 utterances (78,326 words), and is based on 41 half-hour fortnightly recordings of naturalistic interactions collected over two years. The participants are an Emirati woman and six children, aged 1;8–3;10 at the start of the data collection to 3;4–5;9 at the end. The children and the adult are from the same family (first- and second-degree relations). Approximately one-third of the utterances come from the adult participant (8,512 utterances comprising of 29,478 words), while the rest from the children (cf. Ntelitheos & Idrissi, 2017, for detailed information about the composition of the corpus.)

We include the sibling input in our measures of the input for three reasons. First of all, previous studies have suggested that sibling CDS plays an important part in language acquisition (Barton & Tomasello, 1994; Mandle & Tomasello, 1987; Snow, 1995), and it is also often included in CDS corpora, for example the CDS corpus of Polish used by Engelmann et al. (Haman et al., 2011). Second, we believe that the corpus provides the closest approximation of the actual language input for children in the UAE, who typically grow up in large, multi-generational families. This is especially true for the relatively older children, such as the participants in our study, who are unlikely to experience any significant amount of one-on-one interaction with adults on a regular basis. Third, and perhaps most importantly, unlike in studies of early infant development, our aim is not to link initial language learning with specific properties of CDS (which would necessitate a careful selection of age-matched adult CDS). Rather, our purpose is to estimate how what children hear often may affect the ease of retrieval or generalisation of structures that they had already learnt. Therefore, we believe that a naturalistic corpus based on family interactions provides the ideal approximation of input for our purposes, regardless of which family members contribute to those interactions. Since we are not comparing error rates between our participants and the EMALAC children, but rather comparing our participants' error rates to how often certain forms are used around them, we are confident that this does not skew our results. However, in our statistical

analysis, we also include an adult corpus of Gulf Arabic GUMAR (Khalifa *et al.*, 2018) for comparison. GUMAR is the only currently available annotated corpus of 200,381 words of adult EA, and it is based on an internet database of conversational novels – a genre specific to the region, written by internet users in casual language. Overall, we believe that in the absence of a dedicated adult CDS corpus, EMALAC and GUMAR corpora provide a reliable approximation of the properties of the input.

Based on the corpora, we calculated the absolute frequencies of verb types and tokens, as well as of person-number-gender contexts, as a proportion of all words in each corpus. In addition, we also used the EMALAC corpus to establish the PND of all tokens. Engelmann *et al.* (2019) used a measure of PND in Polish and Finnish that was based on dictionary phonological classification of verbs. However, Arabic has only three very general phonological verb classes (strong, weak, doubled), which do not determine the inflection in the imperfective verb form. Furthermore, to ensure comparability, we restricted ourselves to the same animations as used by Engelmann *et al.* (2019), which made it impossible to fully control for the phonological similarities across verbs. PND was therefore instead calculated for each token based on the EMALAC verb lexicon, using the PND calculator CLEARPOND (Marian, Bartolotti, Chabal & Shook, 2012). The PND of each token was thus the number of all EMALAC verb forms which could be derived by means of addition, subtraction or substitution of a single phoneme. While such measure is imperfect in that it partially correlates with type frequency (since different forms of the same verb are often also its phonological neighbors), it provides an estimate of the availability of phonologically similar words, which the child could potentially use as a basis for analogy across forms.

Thus, the four input-based predictor variables were: verb type frequency, token frequency, person-number-gender frequency (all calculated separately based on EMALAC and GUMAR corpus) and PND (calculated based on EMALAC only, since the GUMAR data are not phonologically transcribed).

Design

The study used the elicited production paradigm developed by Engelmann *et al.* (2019), using a subset of the original testing materials available online at <https://osf.io/uepz9/>.

All of the stimuli were cross-checked with the EMALAC corpus. Out of the 32 verbs that were used for Polish in Engelmann *et al.* (2019), 14 were selected, based on their adaptability to the local context and EMALAC frequency, such that half were common (0.03–0.13%), and half fairly uncommon (0–0.003%) in the corpus. Almost all of the verbs were in the default form (Form I), with the exception of *ǰakwi* ‘iron’, which is a Form VII verb.

Table 3 presents the verbs used in the study. All verbs were presented in seven person-number-gender combinations: 1SG, 2SG.F, 3SG.F, 3SG.M, 1PL, 2PL, and 3PL. 2nd person singular masculine (which is syncretic with 3SG.F) was excluded for practical reasons, as all experimenters were female, meaning that a pragmatically natural context could not have been created for the use of the inflection. 2nd and 3rd person plural feminine were excluded because of their very low frequency in modern spoken EA (0.0009% and 0.0076% in the EMALAC corpus, respectively).

We created eight pseudo-randomized lists, each containing 7 out of the 14 verbs. Each participant completed one list: that is, 7 verbs in all 7 person-number-gender combinations, amounting to 49 test items in total per trial.

Table 3. List of verbs used in the study. All verbs given in 3SG.

Lemma	Gloss	EMALAC frequency %	Lemma	Gloss	EMALAC frequency %
jəlʃab	play	0.1321	jəbni	build	0.0026
jəktib	write	0.0692	jəlħag	chase	0.0026
jərsem	draw	0.0564	jəkil	eat	0.0026
jəsmɪʃ	listen	0.0462	jəmʃətʰ	comb	0.0012
jətʰi:r	fly	0.0372	jərges	dance	0.0005
jəʃil	carry	0.0333	jəkwi	iron	-
jəngiz	jump	0.0282	jədig	knock	-

Procedure

The experimental procedure was divided into three stages: training, practice and test. All three stages were completed in a single session, which lasted approximately 30–40 minutes. The first stage was intended to ensure that the participants were familiar with the target verbs. In this stage, the child was shown animations of all seven verbs, and the experimenter provided the gerund form of the verb (e.g., ‘Hatha jari’, *This is running*), to then elicit it back from the child. In the second stage, the test procedure was practiced, and the experimenter elicited all target person-number-gender combinations using animations of a verb not used in the test trial. This was to ensure that the child understood the task. In the test phase, the target verbs were presented in all person-number-gender combinations.

Figure 1 presents a still from one of the videos used in the study. The animations were presented on laptops, using Java-based Processing software (www.processing.org). For 1st and 2nd person contexts, the faces of the characters were digitally replaced with those of the child and the experimenter, to create a pragmatically valid context for the use of these inflections. Each context was thus presented in an animation with different actor/s:

- 1SG: child
- 2SG.F: experimenter (female)
- 3SG.M: male actor 1
- 3SG.F: female actor 1
- 1PL: child and experimenter
- 2PL: experimenter (female) and male actor 2
- 3PL: male actor 3 and female actor 2

After each animation, the child heard the pre-recorded pronoun (e.g., /hi/ *she*), and was asked to repeat it and complete it with the correct form of the verb (e.g., /hi tabni/ *she builds*). While the overt pronoun would not always be natural in spontaneous speech, it was nevertheless included to minimise the chances of the child misinterpreting the target person-number-gender context. This was compatible with the original procedure, as Finnish and Polish are also pro-drop languages.



Figure 1. A sample still from the animation for the verb ‘carry’.

Coding

The responses were recorded on-line on a scoring sheet by the experimenter, as well as audio-recorded using a mobile phone. We had to exclude 123 trials due to a child refusing to continue the experiment or technical problems. There were 2223 usable responses in total.

The responses were coded as correct, incorrect or unscorable. Unscorable responses were those in which the error could not easily be attributed to an issue with the person-number inflection. These included the child using an incorrect verb, incorrect aspectual form, or providing the English equivalent of the word. The rationale for counting these responses separately in the analysis is that it allows for the most conservative estimate of error rates, since only transparent inflectional errors (i.e., correct verb with incorrect agreement marking) are included in the count. The remaining responses were those which involved a correctly inflected form of the verb (correct) and those where the child used the correct verb in the correct aspectual form but with incorrect person-number-gender marking (incorrect).

Results and discussion

We first present descriptive analysis of response types and errors. We then report the results of the statistical analysis using a mixed-effects model, which investigated to what extent children’s proficiency can be explained by input properties.

Response types

The proportions of different types of responses – correct, incorrect and unscorable – were very close to Engelmann *et al.*’s (2019) study on Finnish and Polish. Unscorable responses constituted 34.5% (N = 767) of all data and most commonly involved the child using an incorrect verb (see Table 4). In contrast, errors involving inflectional affixes accounted for only 4% (N = 89) of the responses.

Table 4 presents the distribution of the different response types, together with the percentages from Finnish and Polish for comparison.

Previous research found that children were more likely to make errors of omission than commission. Although Arabic does not have infinitives per se, children are found to produce imperfective bare stems (Aljenaie, 2010) or imperative forms (Abdalla, 2002; Qasem & Sircar, 2017). As Table 4 shows, there were very few such errors in our data. Among the unscorable errors which involved the use of wrong aspectual form ($N = 10$), only four were substitutions of the imperative form, while the remaining six were of the perfective form. Therefore, the children in our study did not make many classic omission errors, and instead produced more errors of commission (all scorable errors).

As regards the unscorable responses involving the use of the wrong verb or an English equivalent, the difficulty in interpreting these errors is that it cannot be known to what extent they constitute an avoidance strategy, and to what extent they reflect issues with word retrieval, attention, or simply faulty experimental materials. However, it is worth noting that the number of unscorable responses for each verb did show a weak positive correlation with the accuracy rates based on scorable responses ($r = 0.53$, $p < .05$). That is, verbs which were often not attempted were also more likely to be produced incorrectly when they were attempted, suggesting that at least some proportion of unscorable responses might have been due to avoidance of challenging targets.

Table 5 presents the percentage of unscorable and accurate responses per verb. The unscorable responses are counted as a percentage of all responses, whereas the accuracy is calculated as a percentage of correct responses out of only the scorable responses. It is worth noting that there were two verbs (*build* and *chase*) which had rates of unscorable attempts that were very high (65% and 85%, respectively), but the scorable attempts showed 100% accuracy. In these cases, it is likely that the unscorable responses reflected genuine lack of familiarity with the verb (which was not remedied by training) or misinterpretation of the video.

There were small differences in the distribution of errors by age of participants. While age was not related to accuracy of scorable responses ($r = 0.01$, $p > .01$), it showed a weak but statistically significant positive correlation with absolute accuracy: that is, including unscorable responses ($r = 0.32$, $p < .05$).

Table 6 presents accuracy rates by age, with the participants grouped into age categories for ease of illustration. Younger children were more likely to not attempt the target, but when they did, they were no less accurate than their older peers. This further supports the conclusion that some proportion of unscorable errors were in fact an avoidance strategy, which masked lack of proficiency in the target inflections. We return to the differences in performance across the age groups in a later section.

Inflectional errors

To avoid speculation about the nature of children's errors in unscorable responses, all analyses in this section include only scorable responses: that is, responses in which the target verb was attempted ($N = 1456$). Since the lack of age effect on accuracy in scorable responses and the even distribution of errors among different age groups suggest that the children were at a similar level of proficiency in the use of the inflections, these analyses are performed on the entire cohort.

In order to understand the patterns of errors, we divided them by the target context and the substituted form, and then looked separately at children's proficiency in person, number and gender, and their markers (cf. Appendix 1 for the full list of errors.)

Table 4. Distribution of scorable and unscorable responses for Emirati Arabic (EA), Finnish (FI), and Polish (PL).

	EA (%)	FI (%)	PL (%)
Correct	61.5	57.9	71.9
Incorrect	4.0	3.6	3.0
Unscorable			
Wrong Verb	26.4		
No Response	4.8		
English	2.9		
Wrong Tense	0.4		
Total	34.5	38.5	25.1

Percentages for Finnish and Polish are taken from Engelmann *et al.* (2019)

Table 5. Types of responses by verb.

Lemma	No. Responses	% Unscorable	% Correct	Lemma	No. Responses	% Unscorable	% Correct
build	153	65	100	fly	154	28	98
carry	154	31	95	iron	179	27	92
chase	176	85	100	jump	153	37	92
comb	156	35	91	knock	154	23	96
dance	153	16	93	listen	154	40	97
draw	159	45	92	play	162	2	91
eat	154	6	95	write	162	39	90

Table 6. Accuracy rates by age group.

Age group	Age range	No. Participants	% Accurate – scorable	% Accurate – all	No. Errors
3-year-olds	2;9–4;0	15	89	48	27
4-year-olds	4;1–5;0	16	98	68	32
5-year-olds	5;1–5;11	17	89	69	30

Errors were not distributed evenly across contexts: while the overall percentage of scorable errors was 6.1%, it ranged from 2.4% to 11.8% depending on the context, which appeared to be inversely related to context frequency. Furthermore, 73% of errors involved a substitution of a more frequent for a less frequent form.

Table 7 presents the percentage of errors in each person-number-gender combination, together with the context's frequency in the EMALAC and GUMAR corpora, ordered by % of errors. Singular contexts have higher frequency than plural contexts in both corpora, and this corresponds to generally lower error rates in singular than in plural contexts. 3SG.M and 3SG.F have high frequency in both corpora, while 1PL and 2PL are very infrequent, and this, too, is reflected in the inverse accuracy of these four contexts. In contrast, there are differences between the two corpora when it comes to the frequencies of 2SG.F and 3PL, such that these contexts are much more frequent in EMALAC than in GUMAR, while the opposite is true for 1SG. The sharp asymmetry between 1SG and 2SG in CDS and adult speech is cross-linguistically common (Ambridge et al., 2015), and confirms that the EMALAC corpus exhibits the properties typically associated with CDS. In our data, the children's error rates appear to reflect the pattern of CDS more closely than that of adult speech. Specifically, 2SG.F and 3PL have high accuracy, as predicted by their high frequency in EMALAC, despite the low frequencies in the adult corpus. Conversely, 1SG is the least frequent form in EMALAC, and the least accurate in the data, despite it being extremely common in GUMAR. These patterns were further confirmed by the statistical model, which we discuss in a later section.

Person

Errors usually did not preserve target person (e.g., 3PL for 3SG), with only 30% of errors being person-congruent. Instead, a great majority of errors involved a substitution of 3rd person markings (plural or singular) for other forms. In fact, these constituted 85% of all errors, with 43% using 3PL as a substitute, and 21% each using 3SG.M and 3SG.F forms. One reason for this may be input frequency, as these forms are by far the most frequent in CDS (cf. Table 7).

Sixty percent of all person errors were due to a prefix error: that is, a substitution with a form which differed ONLY by prefix. There were no cases of prefix omission. This is perhaps unsurprising, given that person is marked on the prefix in a mostly predictable manner, and all forms have prefixes.

Table 8 presents the percentage of all prefix errors. It is apparent that the errors tend to go both ways: that is, prefixes substitute for one another. However, there are differences in the direction of the errors, which appear to be attributable to the difference in frequency between the two forms. For example, 3PL substituting for 2PL is the single most common error, while the reverse error is only attested once in the data, which can be explained by the large difference in the frequency of the forms (cf. Table 7). A similar explanation can be applied to the substitutions between 1SG and 1PL. In general, the more frequent prefix most often substituted for the less frequent one.

Gender

The only context tested in our study which distinguished between genders was 3SG. Based on previous studies, we expected the masculine form to show more accuracy than the feminine form, especially considering 3SG.F is also the only 3rd person form which takes the prefix /tə-/ instead of /jə-/. Indeed, errors were over twice as common in 3SG feminine than masculine targets (7.6% vs. 3.1%, respectively).

Table 9 presents all substitution of both the forms. Almost half (44%, N = 7) of the errors in 3SG.F targets were due to substitution with the masculine equivalent (i.e., /jə-/). However, the reverse pattern (3SG.F substituting for 3SG.M) was also found in 57%

Table 7. Errors by inflectional context.

person/number	2PL	1PL	3SG.F	1SG	3SG.M	2SG.F	3PL
% errors	11.8	10.7	7.6	5.2	3.1	2.4	2.4
GUMAR frequency %	0.12	0.32	1.86	3.05	2.59	0.71	0.3
EMALAC frequency %	0.21	0.17	0.75	0.08	1.18	1.05	0.64

Table 8. All errors due only to incorrect prefix.

Error	Count	% all errors	Substitution	Target
3PL instead of 2PL	17	19.1	jə- -u:n	tə- -u:n
3SG.M instead of 3SG.F	7	7.8	jə-	tə-
3SG.F instead of 1PL	6	6.7	tə-	nə-
3SG.F instead of 3SG.M	4	4.5	tə-	jə-
3SG.M instead of 1SG	4	4.5	jə-	?a-
3SG.F instead of 1SG	4	4.5	tə-	?a-
1SG instead of 1PL	4	4.5	?a-	nə-
1SG instead of 3SG.F	3	3.3	?a-	tə-
3SG.M instead of 1PL	2	2.2	jə-	nə-
1PL instead of 1SG	1	1.1	nə-	?a-
2PL instead of 3PL	1	1.1	tə- -u:n	jə- -u:n
TOTAL	53	60		

($N = 4$) of 3SG.M errors. Since both forms are very frequent in the input and both were equally likely to be used as a substitution for other targets, it is likely that the slightly higher rate of masculine to feminine substitutions is primarily due to the latter's exceptional marking in the paradigm. The overall proportion of feminine to masculine forms outside of 3.SG was not measured, as the gender among stimuli was not balanced (there were more feminine than masculine contexts). Finally, there was no effect of participant sex on the gender errors.

Number

Errors were almost twice as likely to affect plural targets (8.2%) as singular ones (4.6%), in line with the forms' frequency in the input (cf. Table 6). However, children were equally likely to replace the plural target with another plural form (e.g., 3PL for 1PL) as they were to replace it with a singular form (51% vs. 49% of all plural errors, respectively), but were slightly more likely to substitute singular forms for other singular forms (63% of all singular errors).

Number (and sometimes gender) is encoded in the suffix. The suffix is different from the prefix in that it is not present in all forms. Rather, it is null for most singular forms (except 2SG.F /-i:n/) and present in most plural forms (except 1PL). In our study, there were only three contexts which called for the use of a suffix:

Table 9. Substitutions in 3rd person singular by gender.

3SG.M		3SG.F	
Substitution	Number	Substitution	Number
3SG.F	4	3SG.M	7
3PL	3	3PL	5
		1SG	3
		2SG.F	1

2SG.F, 2PL and 3PL. However, despite the relative opacity of the suffix, errors could not be easily attributable to problems with only the suffix, whether in instances of commission or omission.

Table 10 presents all errors that involved the child producing the wrong suffix, or adding a suffix where the target form had none. For each error, the 'Target' column presents the correct adult affix, and the 'Substitution' column presents the child's response. In contrast with the pattern of prefix errors, there were only two errors of commission that affected only the provision of an incorrect suffix, and three further instances of an error in both prefix and suffix.

Table 11 presents the number of suffix omissions. As can be seen from the table, six of these errors were due to suffix omission without changing the prefix, four of which involved the substitution of 2SG.F for 2PL. Given that 2PL was generally the least accurate and least frequent context, it is unlikely that many errors could be attributed solely to problems with suffixation.

Overall, the children's errors were perhaps qualitatively best characterised as reverting to 3rd person affixes – both singular and plural, masculine and feminine – as default forms. This can be attributed to both the very high relative frequency of these forms in the input, and to the pragmatic context. It could be argued that when describing the actions of a character in the video, using 3rd person is always to some extent justified. However, seeing as children did not generally maintain number in these cases, the pragmatic context cannot explain all of the errors. Rather, the input frequency provides a more reliable explanation, and indeed 73% of all errors involved a substitution of a higher-frequency for a lower-frequency form.

Statistical analysis of input-based and age predictors

In order to analyze the extent to which the characteristics of the input may affect children's proficiency in producing verbal morphology, we used Bayesian statistics. Bayesian models have several advantages over traditional approaches. Most importantly, Bayesian inference specifies the probability of an effect being true given the data. This is different from the traditional null hypothesis significance testing (NHST) approach, where the *p*-value specifies the probability of an effect being at least as large as observed, assuming that the null hypothesis is true. A challenge with the NHST approach is that non-significant results cannot be interpreted (see e.g., Dienes, 2011, 2014); the results cannot tell whether there is indeed no effect, or whether there is simply not enough data.

In our analysis, we fitted generalised linear mixed-effects models (glm) in R (R Core Team, 2019), using the rstan package (Stan Development Team, 2020) with the

Table 10. Errors of suffix commission.

Error	Count	% all errors	Substitution	Target
3PL instead of 2SG.F	3	3.4	jə- -u:n	tə- -i:n
2SG.F instead of 2PL	1	1.1	tə- -i:n	tə- -u:n
2SG.F instead of 3SG.F	1	1.1	tə- -i:n	tə-
2SG.F instead of 1PL	1	1.1	tə- -i:n	nə-

Table 11. Errors of suffix omission.

Error	Count	% all errors	Substitution	Target
3SG.F instead of 2PL	4	4.4	tə-	tə- -u:n
3SG.M instead of 2PL	3	3.4	jə-	tə- -u:n
3SG.M instead of 3PL	2	2.2	jə-	jə- -u:n
1SG instead of 3PL	1	1.1	?a-	jə- -u:n
3SG.F instead of 3PL	1	1.1	tə-	jə- -u:n
3SG.M instead of 2SG.F	1	1.1	jə-	tə- -i:n

rstanarm extension (Gabry & Goodrich, 2016; see also Nicenboim & Vasisth, 2016). Because the scorable responses are binary (correct/incorrect), we implemented the models with a binomial link function. We used the models to calculate credible intervals. Unlike confidence intervals, which are often misunderstood (see Morey *et al.*, 2016), credible intervals provide the range within which the true effect lies with a certain probability, given the data.

We fitted models with accuracy (0, 1) as the dependent variable, and Age (in months), VerbFreq (verb type frequency), TokenFreq (token frequency), PNfreq (person-number-gender frequency) and PND (phonological neighborhood density), and the interaction of Age TokenFreq, Age and PND, and PND and TokenFreq as fixed factors. All variables were scaled and centered. In addition, we included random intercepts for Participant and Verb, and by-participant slopes for Verb.

We used weakly informative priors for fixed and random effects, which means that we did not make any strong assumptions about the data. Specifically, the priors for the intercept and slope were a Student *t*-distribution with two degrees of freedom and a mean of zero. For the random effects correlation matrix, a so-called LKJ prior was used (see Sorensen, Hohenstein & Vasisth, 2016, for a tutorial). We report the mean estimate β , the lower and upper limits of the 95% credible interval, and the probability *P* of the effect being smaller than zero (for negative estimates) or greater than zero (for positive estimates). Unlike with *p*-values, there is no binary decision threshold in the Bayesian approach to accept an effect as true. We use the following principles for interpretation:

- If the 95% credible interval does not span zero, we interpret this as strong evidence for an effect.
- If *P* is smaller than .95, we conclude there is no evidence for an effect.

We conducted the analyses separately on EMALAC and GUMAR frequencies, and separately for all responses and scorable responses only.

In the model based on EMALAC frequencies and including only scorable responses (Table 12), we found strong evidence for an effect only for person-number-gender frequency (PNFreq). There was no evidence for effects of Age, VerbFreq, TokenFreq or PND, but there was an interaction between TokenFreq and Age in the predicted direction: TokenFreq played less of a role for older than for younger children.

To illustrate this effect, we again divided the children into age groups (see Table 6) and split the tokens into high- and low-frequency items using a mean split.

Figure 2 illustrates the accuracy rates by age group and centred token frequency, split into high- and low-frequency categories (mean-split), if only scorable responses are included. Counterintuitively, the oldest children are the least accurate, especially on higher-frequency tokens. However, the predicted effect of improvement with age becomes evident when all responses (i.e., both scorable and unscorable ones) are considered ($N = 2223$). In this model (Table 13), there is the same effect of PNFreq and the same interaction between Age and TokenFreq: younger children are much more accurate with high-frequency tokens compared to low-frequency tokens, whereas older children do not show this difference. In addition, the impact of age becomes more prominent, in that the model suggests a high probability of an effect, although the 95% credible interval does marginally span zero.

Figure 3 presents the percentage of accurate, scorable responses out of all responses (both scorable and unscorable). It illustrates children's performance by age group and token frequency (as a binary variable, dividing all centred token frequencies into high or low using a mean-split) when unscorable responses are included. It is apparent that younger children were overall less accurate, and specifically less accurate on low-frequency tokens, but that when faced with such tokens they were more likely to not attempt the target at all and hence produce an unscorable response. The absence of an age effect and the poorer performance of older children on high-frequency tokens in the scorable-only model suggests that including only scorable responses may have, in fact, considerably favored younger children. In other words, the younger children produced fewer scorable responses, but those they produced were mostly accurate.

Importantly, there was strong evidence for an effect of PNFreq in both models, further confirming that producing unscorable responses was related to proficiency in agreement markings.

Figure 4 presents performance by PN context by age, including unscorable responses. A comparison of the model based on all responses and the model based on only scorable responses suggests that excluding unscorable errors may not provide the intended benefits of revealing any additional information about the children's proficiency, as the effects of PNFreq are robust and evident regardless of which responses are counted. Instead, excluding those errors may obscure some age-related differences in performance, by significantly overestimating younger children's proficiency. Nonetheless, even in the all-response model, the main effect of Age was only a tendency, and the Age:TokenFreq interaction was more important.

Figure 5 presents the 95% credible intervals in the all-response EMALAC corpus. The main effect of PNFreq and the Age:TokenFreq interaction have intervals that do not span zero and $P > .95$. Although the interval for Age does span zero, it does so only marginally, suggesting that a tendency for a main effect of Age is likely when all responses are included, which confirms the trends found in the descriptive analysis.

Table 12. Summary of the results of the EMALAC-based Bayesian linear mixed effects model using scorable responses only.

	Mean	Lower bound	Upper bound	P(b > 0)	P(b < 0)
Intercept	5.5458	4.2778	7.2782	1	0
Age	0.0258	-1	1.0334	0.53	0.47
VerbFreq	-0.1712	-0.7474	0.3676	0.26	0.74
TokenFreq	0.4659	-0.1964	1.2	0.92	0.08
PNFreq	0.8111	0.4787	1.1648	1	0
PND	0.3703	-0.2427	1.0649	0.86	0.14
Age:TokenFreq	-0.7514	-1.4177	-0.2169	0	1
Age:PND	-0.0415	-0.5896	0.4459	0.45	0.55
TokenFreq:PND	0.0593	-1.2989	1.4434	0.55	0.45

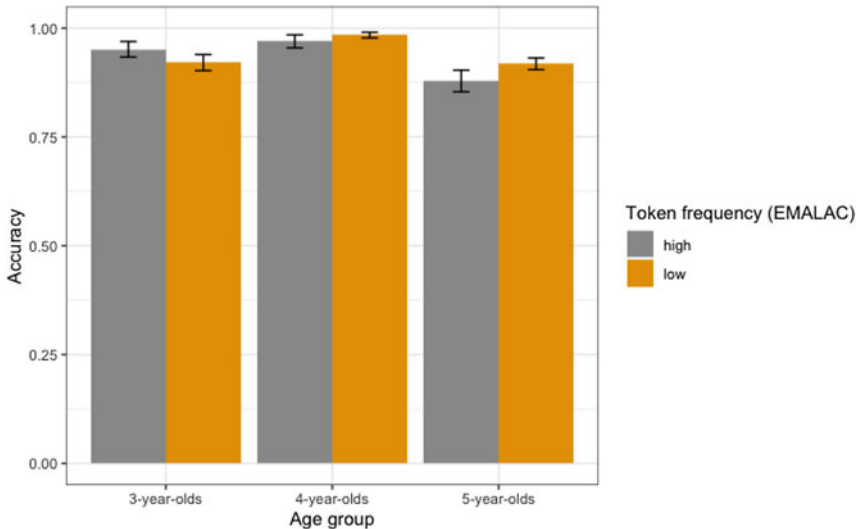


Figure 2. Bar chart illustrating the interaction between Age and TokenFreq for different ages, using only scorable responses. Note that for illustration purposes, the centred token frequency variable has been dichotomised into high and low frequency tokens (mean split), and age is shown in three groups rather than as a continuous, centred variable.

For comparison, we repeated the analyses based on the adult corpus, again separately using only the scorable vs. all responses (Tables 14 and 15). In both the scorable-only model and the all-response model, all of the credible intervals spanned zero – in other words, none of the predictor variables had an unequivocal (positive or negative) effect on accuracy. However, in the all-responses model, there was again a high probability of an effect of both PNFreq and Age, with higher PNFreq and higher ages improving accuracy. The GUMAR all-responses model is thus more in line with the EMALAC models than the GUMAR scorable-only model.

Table 13. Summary of the results of the EMALAC-based Bayesian linear mixed effects model using all responses.

	Mean	Lower bound	Upper bound	P(b > 0)	P(b < 0)
Intercept	1.3212	-0.1163	2.7404	0.96	0.04
Age	0.7237	-0.0556	1.5433	0.97	0.03
VerbFreq	0.5985	-0.6004	1.7787	0.86	0.14
TokenFreq	0.0184	-0.2496	0.3042	0.55	0.45
PNFreq	0.1948	0.0452	0.3482	0.99	0.01
PND	0.3057	-0.0583	0.6823	0.94	0.06
Age:TokenFreq	-0.4442	-0.6766	-0.2335	0	1
Age:PND	-0.0531	-0.2326	0.1241	0.27	0.73
TokenFreq:PND	0.3687	-0.4561	1.1968	0.81	0.19

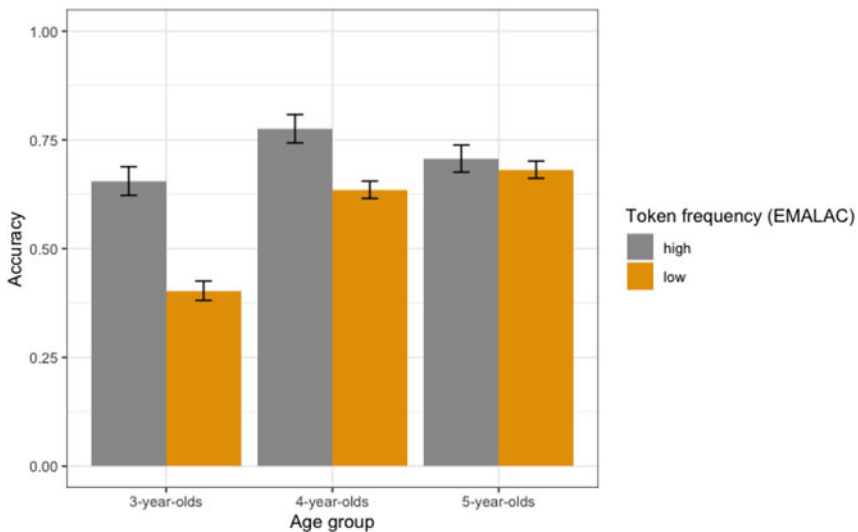
**Figure 3.** Bar chart illustrating the interaction between age and token frequency for different ages, using scorable and unscorable responses.

Figure 6 presents the 95% credible intervals in the GUMAR all-response model. The Age and PNFreq effects both have $P > 95\%$, and although the intervals marginally span zero, they mirror the results of the all-response EMALAC model (cf. Fig. 5).

The difference in the results between the EMALAC and the GUMAR models is most probably related to the different nature of the two corpora. This includes a slightly different distribution of PN frequencies (cf. Table 6), which are in line with commonly reported differences between child- and adult-directed speech, as well as different type and token frequencies. It is likely that the GUMAR corpus, as a collection of adult written – albeit casual – language may be less representative of the

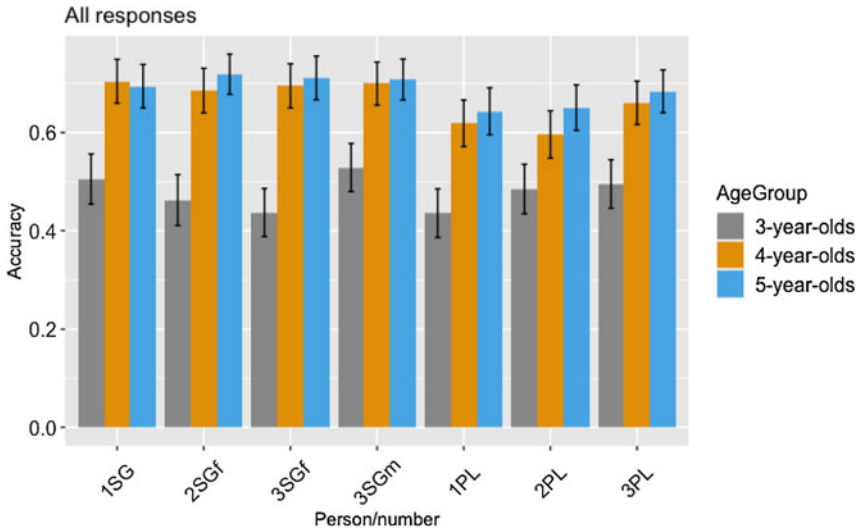


Figure 4. Accuracy by inflectional context and age group, based on scorable and unscorable responses.

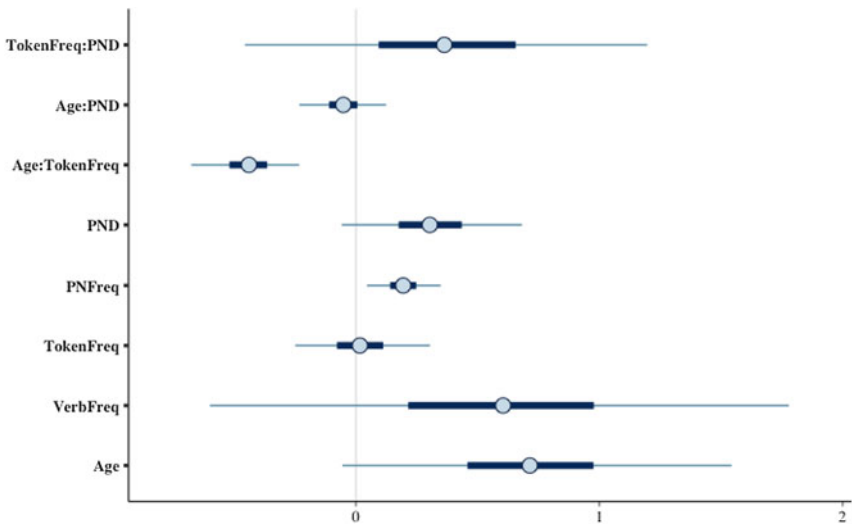


Figure 5. 95% credible intervals for the all-response EMALAC corpus.

language children are typically exposed to than the EMALAC corpus. However, the two are not entirely divergent, as apparent from the fact that the all-response GUMAR model does closely mirror the effects found in the EMALAC models.

Overall, across the four models, we found robust evidence for effects of inflectional context frequency and age:token frequency interaction, and a marginal effect of age. In other words, while we did not find an absolute improvement with age, we found an

Table 14. Summary of the results of the GUMAR-based Bayesian linear mixed effects model using scorable responses only.

	Mean	Lower bound	Upper bound	P(b > 0)	P(b < 0)
Intercept	5.0981	3.9311	6.6239	1	0
Age	0.049	-0.9367	1.0336	0.54	0.46
VerbFreq	-0.1003	-0.6626	0.4566	0.36	0.64
TokenFreq	0.125	-0.4297	0.6894	0.67	0.33
PNFreq	0.2881	-0.0359	0.6187	0.95	0.05
PND	0.0239	-0.4787	0.5951	0.51	0.49
Age:TokenFreq	-0.0536	-0.4951	0.3297	0.41	0.59
Age:PND	-0.1272	-0.5531	0.216	0.26	0.74
TokenFreq:PND	-0.0218	-0.7879	0.7989	0.47	0.53

Table 15. Summary of the results of the GUMAR-based Bayesian linear mixed effects model using all responses.

	Mean	Lower bound	Upper bound	P(b > 0)	P(b < 0)
Intercept	1.3658	-0.0339	2.7705	0.97	0.03
Age	0.731	-0.0492	1.4616	0.97	0.03
VerbFreq	0.3793	-0.77	1.4648	0.76	0.24
TokenFreq	0.093	-0.1987	0.3928	0.73	0.27
PNFreq	0.1614	-0.0159	0.3407	0.96	0.04
PND	0.0068	-0.2668	0.2969	0.52	0.48
Age:TokenFreq	0.1498	-0.0381	0.3517	0.94	0.06
Age:PND	-0.1229	-0.2965	0.0412	0.07	0.93
TokenFreq:PND	-0.1127	-0.4727	0.2495	0.28	0.72

improvement on low-frequency tokens with age. These results suggest that at the age that was tested in our study, all children were already familiar with the paradigm, and thus when they attempted the targets, they were highly accurate. However, their proficiency in the use of the affixes was still sensitive to the inflectional context's frequency, and token frequency – although only at the younger ages.

General discussion

We hypothesised that error rates would significantly vary across contexts, that the variability would be attributable to the contexts' input frequency, and that it would not show effects of verb identity. The first hypothesis was confirmed: although the general error rate was low at 6%, it was almost double that for the least accurate person-number-gender context (2nd person plural, 11.8%), and less than half for the

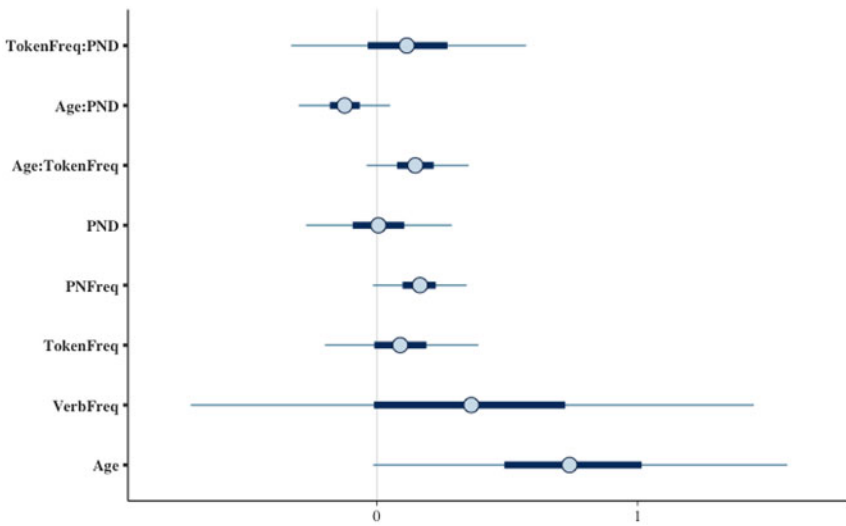


Figure 6. 95% credible intervals in the GUMAR all-response model.

most accurate one (2nd person singular, 2.4%). This is in line with previous experimental findings (Aguado-Orea & Pine, 2015; Basaffar & Safi, 2012; Engelmann *et al.*, 2019; Räsänen *et al.*, 2016), which show that high overall accuracy often conceals unevenly distributed patterns of errors. The distribution of errors across the contexts was also compatible with previous studies on Arabic and other languages, which showed that 3rd person, singular, and masculine inflections show higher rates of accuracy than plural, feminine, 1st and 2nd person inflections (Abdou & Abdou, 1986 as reported in Basaffar & Safi, 2012; Aljenaia, 2010). The inflectional contexts previously shown to be acquired earlier showed higher accuracy in our data and had an effect on the pattern of substitutions. This was true especially of 3rd person forms, which substituted for other forms in 85% of the errors.

Our second hypothesis was that the uneven distribution of errors could be explained by the inflectional context frequency and this was also confirmed: the statistical analysis of the data revealed robust evidence for a main effect of context frequency on response accuracy, such that the contexts with the highest frequency were the most accurate. The effect held whether all or only scorable responses were included, and was also visible in the all-response model using the adult corpus frequencies, confirming that it was not an artifact of the input data selected. This result is in line with the findings of Engelmann *et al.* (2019), and suggests that children's proficiency in verb inflections is related to their exposure to those inflections. In other words, children are better able to provide the affixes they hear most often.

Alternative explanations sometimes offered in the literature for the differences in the order of acquisition and accuracy across different contexts postulate that there are certain linguistic properties inherent in the inflections that determine the order in which they are acquired. Most of these accounts are rooted in the theory of linguistic markedness (Greenberg, 1966; Jakobson, 1963), according to which some inflectional contexts are more marked than others, and hence acquired later than the less marked ones. For example, the markedness hierarchy predicts that 3rd person

contexts will be acquired before 1st person contexts due to the former being less marked than the latter (Ackema & Neeleman, 2018; Hanson, 2000). One of the arguments often used to provide support to these privative feature systems against competing input-frequency based accounts is the late acquisition of 2nd singular pronouns and agreement markers despite their relatively high frequency in child-directed speech (see, for example, Laakso & Smith (2007), where the 2nd singular pronoun is shown to be the most frequent subject in child-directed speech). Despite this high frequency in the input, 2nd person singular pronouns and agreement markers seem to be acquired late in many languages (Ackema & Neeleman, 2018). What these accounts fail to predict are cases where high-frequency 2nd person pronominal agreement in the input corresponds to relatively accurate use of such agreement by children at early stages. In our study, frequency seems to make slightly better predictions with respect to 2nd person agreement development. The 2SG.F context exhibited the smallest number of errors (2.4%, equal to that of 3PL), which inversely corresponds to the second highest frequency in the input, as measured in the EMALAC corpus. This result cannot readily be explained by feature-geometric/privative feature bundle accounts of agreement development, which take 2nd singular and feminine agreements to be ‘marked’.

As Ambridge et al. (2015) note, in the case of low accuracy of 2nd person pronouns and agreement in child data from other languages, longitudinal studies cannot distinguish between failure to learn specific forms despite their high frequency and usage choices in which children “have learned these forms, but find little use for them (e.g., young children are not interested in talking about what their listener is doing)” (p. 245). An interpretation based on pragmatic, rather than grammatical, factors, may be better able to account for cross-linguistic differences in the acquisition of 2SG. Future studies may determine whether perhaps the cultural context in Emirati Arabic could have an effect on the increased use of 2SG by children, as compared to other languages studied.

Finally, it should be noted that while 2SG.F was produced with the highest accuracy, it was nonetheless very rarely used as a substitution for other forms. We propose that this further strengthens the argument for pragmatic factors playing a role in children’s performance. That is, while children know the 2SG.F affixes well, they also understand that it is not appropriate when describing an animation (which is naturally a 3rd person context). This explanation would account for the fact that even though 3SG.F and 3PL were not as accurately produced as 2SG.F, they were at the same time more likely to be used as replacements for other forms. In the context of describing the actions of characters on the screen, if the child is not able to provide the expected person-number-gender affixes, 3rd person affixes are the next, most pragmatically justifiable solution. In contrast, it is possible that the higher number of 2nd person substitutions in naturalistic studies may be associated with that being a more pragmatically justified response in a one-to-one conversation between the child and the experimenter.

Our third hypothesis was that children would be able to generalise the acquired affixes to all verbs early on, since the high availability, reliability, and phonological saliency of EA Imperfective inflections would not require them to rely on item-by-item analogies. Contrary to previous findings in Polish and Finnish, this hypothesis was partially confirmed, as there was no main effect of type frequency, token frequency, or phonological neighborhood density on accuracy in our data. Children performed equally well with high- and low-frequency verb types, with

dense as well as sparse phonological neighborhoods. Nonetheless, contrary to our predictions, there was also a strong interaction between age and token frequency – in scorable and unscorable responses, using CDS and adult corpus frequencies. Furthermore, the effect of age was marginal and was only observable when unscorable responses were considered. While unscorable responses cannot be easily interpreted, there is strong evidence in our data that at least some of them are avoidance errors – their prevalence in the data from younger participants and a reverse correlation between accurate and unscorable responses within each verb suggest that these errors are at least partially related to lower proficiency in the inflections.

These results suggest that all the children already knew the paradigm fairly well, as evident from their high accuracy in scorable responses, but the younger children were more likely to avoid attempting the target and less able to apply their knowledge to low-frequency tokens. This effect was hypothesised by Engelmann *et al.* (2019) but not found in their Polish and Finnish data. It implies that even at a fairly high level of proficiency, in a language which provides a highly regular and accessible paradigm, children do not fully abstract the paradigm away from individual stored exemplars early on.

On the one hand, such long-lasting effects of token frequency are incompatible with maturational approaches, which propose that grammatical structures become fully productive as soon as they are learnt (Radford, 1990; Wexler, 1998). They appear to confirm the predictions made by usage-based approaches which rely on exemplar learning (e.g., Bybee, 2010; Tomasello, 2003). According to these approaches, even when children eventually generalise the grammatical patterns away from strictly item-based constructions, grammatical accuracy is expected to exhibit strong and long-lasting effects of frequency due to the process of analogy being conducted on-line, over all stored exemplars. On the other hand, the interaction of token frequency and age with a simultaneous lack of a main effect of token frequency does suggest that token frequency plays a smaller role in the acquisition of EA than in Polish and Finnish (where a main effect of token frequency is observed but no interaction with age). Indeed, the token frequency does not have an effect for the oldest children in our study. Overall, since the token frequency:age interaction had mostly been hypothesised but not found in previous studies, and since this is the first study of the acquisition of agreement in EA, further research will be needed to investigate the reliability and the nature of this interaction. Nonetheless, it is possible that it indicates that the phonological saliency, reliability and regularity of the paradigm leads to earlier generalisations than in other languages, but later than predicted by maturational approaches.

As regards the effects of PND, while our results cannot illuminate the earliest stages of development, they do suggest that at least by age three children acquiring EA do not rely on phonological similarity in analogizing across forms. We propose that this is due to the properties of the system – unlike in Polish, where phonological similarity determines the inflectional class and therefore the affix that needs to be used, the uniformity of affixes across the EA paradigm does not require reliance on the phonological form of the verb in determining the inflection. Furthermore, the high phonological saliency of the affixes, including the reliable use of syllabic prefixes and less reliable but longer suffixes (i.e., including a long vowel) could explain why EA inflections are learnt earlier than those from even more regular languages, such as Finnish, and why they are learnt independently from the phonological form of the

verb. Our results demonstrate that children rarely omit either prefixes or suffixes: there were only four instances of children substituting an imperative form, only two cases of errors explainable entirely as suffix omission, and none of prefix omission. Further research is also needed to clarify whether an effect of phonological neighborhood density may be found if a different measure of similarity is used. It is possible that, due to Arabic's templatic morphology, analogies are based on different units than the whole word forms, such as consonantal roots.

Our findings further illuminate how data collection and analysis methods can affect the conclusions we draw about child language proficiency, and therefore highlight the need for complementing naturalistic observations with experimental data. Experimental studies consistently reveal lower proficiency level than the results obtained from raw production studies. In other words, when children do not self-select the targets they want to say, they make more errors, which further strengthens the argument for conducting experimental (rather than observational) studies in the acquisition of inflection. Furthermore, our data suggest that focusing exclusively on 'pure' inflectional errors may significantly overestimate the proficiency levels of younger children, while not necessarily providing additional insights into the learning process. The patterns of unscorable responses in our study strongly suggest that these 'unscorable' errors of omission or avoidance often do, in fact, reflect issues with proficiency in the target structures.

Conclusion

Since languages employ a range of grammatical forms and structures which present distinctive challenges, comparing cross-linguistic developmental paths can help us determine to what extent the learning process is shaped by the input. Our results from the acquisition of Emirati Arabic confirm previously reported effects of input frequency on the acquisition of verb inflection, while also suggesting that children are able to generalise learnt forms across types and tokens when the system being acquired is regular and easily accessible. While frequency is unquestionably important, high phonological saliency and regularity of the system may eventually lead to across-the-board generalisations which are no longer susceptible to frequency effects, although this was only true for our oldest participants. The results are compatible with an approach to language development which emphasises the properties of the input as factors determining the trajectory of the learning process, but does not preclude an ability of the child to eventually fully generalise grammatical processes, although such generalisations may take a considerable time to develop.

References

- Abbot-Smith, K., & Serratrice, L.** (2015). Word order, referential expression, and case cues to the acquisition of transitive sentences in Italian. *Journal of Child Language*, 42(1), 1–31.
- Abdalla, F.** (2002). Specific language impairment in Arabic-speaking children: deficits in morphosyntax. Unpublished PhD thesis, McGill University, Canada.
- Abdou, D., & Abdou, S.** (1986). *Dirasa fi mufradat tiflayn*. (A study on two children). Kuwait City: Dhaat al-Salaasil Publication.
- Ackema, P., & Neeleman, A.** (2018). *Features of person: From the inventory of persons to their morphological realization*. MIT Press.

- Aguado-Orea, J., & Pine, J. M.** (2015). Comparing different models of the development of verb inflection in early child Spanish. *PLoS ONE*, *10*(3).
- Albirini, A.** (2017). The acquisition of Arabic as a first language. In E. Benmamoun & R. Bassiouney (eds.), *The Routledge Handbook of Arabic Linguistics* (pp. 227–248). Routledge.
- Al Kaabi, M., & Nteliheos, D.** (2019). Rethinking templates: A syntactic analysis of verbal morphology in Emirati Arabic. *Glossa: a Journal of General Linguistics*, *4*(1), 132.
- Aljenaie, K.** (2010). Verbal inflection in the acquisition of Kuwaiti Arabic. *Journal of Child Language*, *37*, 841–863.
- Aljenaie, K.** (2001). The emergence of tense and agreement in Kuwaiti Arabic children. Unpublished doctoral dissertation, University of Reading.
- Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L.** (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239–273.
- Barton, M. E., & Tomasello, M.** (1994). The rest of the family: The role of fathers and siblings in early language development. In C. Galloway & B. J. Richards (eds.), *Input and interaction in language acquisition* (p. 109–134). Cambridge University Press.
- Basaffar, F. M., & Safi, S. M.** (2012). The Acquisition of Verb Inflections in Hijazi Arabic. *Arab World English Journal*, *3*(2), 266–304.
- Bates, E., & MacWhinney, B.** (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (eds.), *The crosslinguistic study of sentence processing*, 3–76. Cambridge University Press.
- Berman, R.** (1985). The acquisition of Hebrew. In D. Slobin (ed.), *The crosslinguistic study of language acquisition*, 255–63. Hillsdale: Lawrence Erlbaum Associates.
- Brown, R. A.** (1973). *First language*. Allen & Unwin Ltd.
- Bybee, J.** (2010). *Language, usage and cognition*. Cambridge University Press.
- Bybee, J., & Hopper, P.** (2001). *Frequency and the Emergence of Language Structure*. John Benjamins, Amsterdam.
- Chan, A., Lieven, E., & Tomasello, M.** (2009). Children's understanding of the agent-patient relations in the transitive construction: Cross-linguistic comparisons between Cantonese, German, and English. *Cognitive Linguistics*, *20*(2).
- Chiat, S.** (1986). Personal pronouns. In P. Fletcher & M. Garman (eds.), *Language acquisition: Studies in first language development* (pp. 339–55). Cambridge: Cambridge University Press.
- Clark, E. V.** (1978). From gesture to word: On the natural history of deixis in language acquisition. In J. S. Bruner & A. Garton (eds.), *Human growth and development: Wolfson College lectures 1976* (pp. 85–120). Oxford: Oxford University Press.
- Dąbrowska, E.** (2008). The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language*, *58*(4).
- De Villiers, J.G., & Johnson, V.** (2007). The information in third-person /s/: acquisition across dialects of American English. *Journal of Child Language*, *34*, 133–158.
- Demuth, K.** (2018). Prosodic constraints on children's use of grammatical morphemes. *First Language*, *39*, 80–95.
- Dienes, Z.** (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science* *6*(3), 274–90. doi: 10.1177/1745691611406920.
- Dienes, Z.** (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology* *5*(July), 1–17. doi: 10.3389/fpsyg.2014.00781.
- Engelmann, F., Granlund, S., Kolak, J., Szreder, M., Ambridge, B., Pine, J., Theakston, A., & Lieven, E.** (2019). How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology*, *110*, 30–69.
- Fahim, D.** (2017). Verb Morphology in Egyptian Arabic Developmental Language Impairment. *Arab Journal of Applied Linguistics*, *2*(1), 49–73.
- Feghali, H. J.** (2008). Gulf Arabic: The dialects of Kuwait, Bahrain, Qatar, UAE, and Oman. Dunwoody Press.
- Gabry, J., & Goodrich, B.** (2016). Rstanarm: Bayesian Applied Regression Modeling via Stan. *R Package Version 2*(1).
- Granlund, S., Kolak, J., Vihman, V., Engelmann, F., Ambridge, B., Pine, J., & Lieven, E.** (2019). Language-general and language-specific phenomena in the acquisition of inflectional noun

- morphology: A cross-linguistic elicited-production study of Polish, Finnish and Estonian. *Journal of Memory and Language*, 107, 169–194.
- Greenberg, J.** (1966). Language universals, with special reference to feature hierarchies. The Hague, Netherlands: Mouton & Co.
- Hallé, P. A., & de Boysson-Bardies, B.** (1996). The format of representation of recognized words in the infants' early receptive lexicon. *Infant Behavior and Development*, 19, 463–481.
- Haman, E., Etenkowski, B., Luniewska, M., Szwabe, J., Dąbrowska, E., Szreder, M., & Łaziński, M.** (2011). The Polish CDS corpus.
- Hanson, R.** (2000). Pronoun acquisition and the morphological feature geometry. *Calgary Working Papers in Linguistics* 22, 1–14. Calgary: Dept. of Linguistics, University of Calgary.
- Harley, H., & Ritter, E.** (2002). A feature-geometric analysis of person and number. *Language*, 78, 482–526.
- Hoekstra, T., & Hyams, N.** (1998). Aspects of root infinitives. *Lingua*, 106(1–4), 81–112.
- Holes, C.** (1995). Modern Arabic: Structures, functions and varieties. Longman.
- Jakobson, R.** (1963). Implications of language universals for linguistics. In J. Greenberg (ed.), *Universals of language* (pp. 263–278). Cambridge, MA: MIT Press.
- Johnstone, T.M.** (1967). *Eastern Arabian dialect studies*. Oxford University Press.
- Khalifa, S., Habash, N., Eryani, F., Obeid, O., Abdulrahim, D., & Al Kaabi, M.** (2018). A Morphologically Annotated Corpus of Emirati Arabic. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan.
- Krajewski, G., & Lieven, E.** (2014). Competing cues in early syntactic development. In B. MacWhinney, A. Malchukov, & E. A. Moravcsik (eds.), *Competing motivations in grammar and usage*. Oxford: Oxford University Press.
- Kuczaj, S.A.** (1979). Evidence for a language learning strategy: On the relative ease of acquisition of prefixes and suffixes. *Child Development*, 51(1), 1–13.
- Laakso, A., & Smith, L.** (2007). Pronouns and verbs in adult speech to children: A corpus analysis. *Journal of Child Language*, 34, 725–763.
- Levy, Y.** (1983). It's frogs all the way down. *Cognition*, 15, 75–93.
- Legendre, G., Culbertson, J., Zaroukian, E., Hsin, L., Barrière, I., & Nazzi, T.** (2014). Is children's comprehension of subject-verb agreement universally late? *Comparative evidence from French, English, and Spanish*. *Lingua*, 144, 21–39.
- Leung, T., Ntelitheos, D., & Al Kaabi, M.** (2021). *Emirati Arabic: A comprehensive grammar*. Routledge.
- Lieven, E.** (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, 120, 2546–2556.
- Leó, C., & Demuth, K.** (1999). Prosodic constraints on the emergence of grammatical morphemes: Crosslinguistic evidence from Germanic and Romance languages. In Greenhill, A., Littlefield, H., Tano, C. (eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development* (pp. 407–418). Somerville, MA: Cascadilla Press.
- MacWhinney, B.** (1987). The Competition Model. In B. MacWhinney (ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Lawrence Erlbaum.
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A.** (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7(8).
- Mannle, S., & Tomasello, M.** (1987). Fathers, siblings, and the bridge hypothesis. In K. E. Nelson & A. van Kleeck (eds), *Children's Language* (Hillsdale, N.J.: Erlbaum), Vol. 6, 23–42.
- Moawad, R.A.** (2006). *The Acquisition of the Arabic Gender and Number Systems*. Unpublished PhD Dissertation, University of Wales, Bangor.
- Morey, R.D., Hoekstra, R., Rouder, J.N., Lee, M.D., & Wagenmakers, E.J.** (2016). The Fallacy of Placing Confidence in Confidence Intervals. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-015-0947-8.
- Nicenboim, B., & Vasisht, S.** (2016). Statistical Methods for Linguistic Research: Foundational Ideas - Part II. *Language and Linguistics Compass*, 10(11), 591–613. doi: 10.1111/lnc3.12207.
- Ntelitheos, D., & Idrissi, A.** (2017). Language growth in child Emirati Arabic. In H. Ouali (ed.) *Perspectives on Arabic Linguistics XXIX: Papers from the annual symposium on Arabic Linguistics, Milwaukee, Wisconsin, 2015*. John Benjamins.
- Omar, M.** (1973). The acquisition of Egyptian Arabic as a native language. The Hague: Mouton.
- Qafisheh, H.** (1977). A short reference grammar of Gulf Arabic. Tucson: The University of Arizona Press.

- Qasem, F., & Sircar, S.** (2017). Imperative as root infinitive analogue in Yemeni Ibbi Arabic: Two case studies. *Arab Journal of Applied Linguistics*, 2(1), 1–19.
- R Core Team.** (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Radford, A.** (1990). Syntactic theory and the acquisition of English syntax: The nature of early child grammars of English. Oxford: Blackwell.
- Räsänen, S. H. M., Ambridge, B., & Pine, J. M.** (2016). An elicited-production study of inflectional verb morphology in child Finnish. *Cognitive Science*, 40, 1704–1738.
- Santelmann, L.** (1998). The acquisition of definite determiners in child Swedish: Metrical and discourse influences on functional morphology. In A. Greenhill, M. Hughes, H. Littlefield & H. Walsh (eds.), *BUCLD 22 Proceedings*, 651–662. Somerville, MA: Cascadilla Press.
- Savičiūtė, E., Ambridge, B., & Pine, J. M.** (2018). The roles of word-form frequency and phonological neighbourhood density in the acquisition of Lithuanian noun morphology. *Journal of Child Language*, 45(3), 641–672.
- Savinainen-Makkonen, T.** (2007). Geminate Template: A model for first Finnish words. *First Language*, 27(4), pp. 347–359.
- Smoczyńska, M.** (1985). The acquisition of Polish. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition*, Vol. 1. The data; Vol. 2. Theoretical issues (p. 595–686). Lawrence Erlbaum Associates, Inc.
- Snow, C.** (1995). Issues in the study of input: Finetuning, universality, individual and developmental differences, and necessary causes. In P. Fletcher & B. MacWhinney (eds.) *The handbook of child language*. Blackwell.
- Song, J., Sundara, M., & Demuth, K.** (2009). Phonological constraints on children's production of English third person singular *-s*. *Journal of Speech, Language, and Hearing Research*, 52:623–642.
- Sorensen, T., Hohenstein, S., & Vasishth, S.** (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *Tutorials in Quantitative Methods for Psychology*, 12(3), 175–200. doi: 10.20982/tqmp.12.3. p. 175.
- Stan Development Team** (2020). RStan: The R Interface to Stan.
- Szreder, M.** (2013). The acquisition of consonant clusters in Polish – A case study. In M. Vihman, & T. Keren-Portnoy. (eds.), *Child Phonology: Whole word approaches, cross-linguistic evidence*. Cambridge: Cambridge University Press.
- Theodore, R.M., Demuth, K., & Shattuck-Hufnagel, S.** (2011). Acoustic evidence for positional and complexity effects in children's production of plural *-s*. *Journal of Speech, Language, and Hearing Research*, 54, 539–548.
- Tomasello, M.** (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Vihman, M. M., Nakai, S., DePaolis, R. A., & Halle, P.** (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50, 336–353.
- Wexler, K.** (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, 106(1–4), 23–79.

Appendix

The raw numbers of all errors by target and substituted form.

Target	Substituted form						
	1sg	2sgf	3sg	3sgf	1pl	2pl	3pl
1sg			4	4	1		2
2sgf			1				3
3sg				4			3
3sgf	3	1	7				5
1pl	4	1	2	6			8
2pl		1	3	4			17
3pl	1		2	1		1	

Cite this article: Szreder M, De Ruiter LE, Ntelitheos D (2022). Input effects in the acquisition of verb inflection: Evidence from Emirati Arabic. *Journal of Child Language* 49, 684–713. <https://doi.org/10.1017/S0305000921000155>