Journal of Child Language (2025), 1–16
doi:10.1017/S030500092510007X

CAMBRIDGE UNIVERSITY PRESS

BRIEF RESEARCH REPORT

A corpus analysis of child and child-directed speech in Palestinian Arabic: A first approach to syntactic development

Tala Nazzal[1,2] and Anna Gavarró[2]

[1]Department of Applied and Allied Medical Sciences, Faculty of Medicine and Health Sciences, An-Najah National University, Nablus, Palestine and [2]Departament de Filologia Catalana, Universitat Autònoma de Barcelona, Barcelona, Spain
Corresponding author: Tala Nazzal; Email: t.nazzal@najah.edu

(Received 23 March 2024; revised 05 May 2025; accepted 15 May 2025)

Abstract
We present a new corpus of child and child-directed speech (CDS) in Palestinian Arabic. It includes transcriptions following the CHILDES guidelines and features recordings of 16 monolingual Palestinian Arabic-speaking children with an age range of 19–58 months and their adult interlocutors. We analyse the children's morphosyntactic development and identify a variety of target word orders (45 in child speech, 50 in CDS), with prevalent SV(O) structures; we also found high rates of null subjects in both populations, marginal errors in children's verbal agreement morphology, and early emergence of serial verb constructions, observed from 23 months of age.

Keywords: child spontaneous production; corpus; early acquisition; Palestinian Arabic; word order; null subjects; verbal morphology; serial verb construction

إحنا بنعرض مجموعة جديدة من البيانات اللغوية اللي بتخص حكي الاطفال والحكي الموجه للأطفال باللهجة الفلسطينية. وتسجيلات ل 16 طفل لغتهم الوحيدة اللهجة الفلسطينية، CHILDS, المجموعة بتتضمن نسخ مكتوبة حسب توجيهات وعمرهم بين 19 و58 شهر، وكمان تسجيلات للأشخاص البالغين للي بحكوا معهم. بعد م حللنا التطور النحوي حددنا ترتيب للكلمات)45 في حكي الأطفال و05 في حكي البالغين(، وشفنا ان التركيب اللغوي الشائع أكثر شي هو فاعل - فعل - مفعول به). ولاحظنا نسب عالية من الضمير المستتر في حكي الفئتين. وكمان أخطاء بسيطة في تصريف الأفعال عند الأطفال وظهور بدري للتراكيب اللي بتضمن اكتر من فعل ورا بعض من عمر 23 شهر.

الكلماتالمفتاحية: كلامالأطفال; البياناتاللغوية; اكتسابااللغةالمبكر ; اللهجةالفلسطينية; ترتيب الكلمات; الضمير المستتر; صرف الأفعال; التراكيب; الفعليةالمتسلسلة

1. Introduction

There have been notable efforts in developing different methods to study language acquisition in children. These include resources for naturalistic/spontaneous diaries, corpora, elicited production, and imitation for language production, and the preferential

looking paradigm, act-out tasks, and truth value judgement tasks for language comprehension. Cross-sectional experiments in which several children are tested at specific ages are useful to see if they have acquired a particular grammatical structure at a given age. This type of data is extremely useful for providing norms of typical development. However, it does not address one of the field's goals, which is to understand how a given child's knowledge of language evolves over time. Longitudinal case studies and corpora can provide the kind of detailed, fine-grained data on the transition from one stage of generalisation to the next and can also reveal the individual variances in the process of learning (Demuth, 2008). This paper focuses on this rich data source, the corpus, in which the natural interactions between interlocutors can be used to test different hypotheses and theoretical claims. Brown (1973) pioneered the use of child–adult interactions to assess various theories and hypotheses related to language acquisition. Since then, there has been a substantial increase in the construction and utilisation of corpora in numerous languages. Linguistic databases are particularly instrumental in investigating the acquisition of lexical, syntactic, and discourse knowledge.

Corpora derived from formal, written adult data are inadequate for drawing any conclusions about the input children draw on to acquire language, although they may provide a measure of adult grammar. Conversely, corpora of child-directed speech (CDS) play a crucial role in examining the characteristics of the input that children receive. Therefore, there is a pressing need for corpora containing child and CDS data for various languages and language varieties (Dash & Arulmozi, 2018; Demuth, 2008).

Despite their advantages, corpora are not without challenges. A language corpus cannot fully encompass the limitless variations in language usage by its speakers across diverse situations and contexts (Dash & Arulmozi, 2018). Many existing child corpora are limited by the inclusion of data from a small number of children, and they may be biased by the contexts in which they are collected (Soderstrom, 2007). Small sample sizes and lack of diversity in some CDS corpora can hinder the generalisability of findings (Lieven, 2010). CDS corpora recorded in laboratory settings can lack the naturalness of everyday interaction, which may lead to skewed representations of language use (Snow, 1995). Furthermore, corpora may misrepresent the grammatical knowledge of children or fail to capture grammatical well-formedness accurately (Demuth, 2008).

In spite of these limitations, corpora remain a necessary source of information on child language development. While no corpus can fully represent the vast variability of language in real-world contexts, CDS corpora provide invaluable insights into the linguistic input that shapes children's language acquisition. A notable database in the field of child language acquisition is CHILDES (MacWhinney, 1996; MacWhinney, 2000; MacWhinney & Snow, 1990), a comprehensive resource comprising naturalistic data from over 28 languages, totalling around 44 million spoken words. Developed over the years, CHILDES gathers spontaneous speech from diverse languages, from English to Chinese, from the Romance languages to the Germanic and Japanese. The database, built with the contribution of researchers conducting their own studies, predominantly features adult–child interactions (Corrigan, 2012).

Researchers interested in the study of the acquisition of Arabic had, up to this date, some resources, although all of them dwelled on adult varieties. For example, there is one morphologically annotated corpus available for Palestinian Arabic, the Curras Notebook (Jarrar et al., 2017). It consists of 56,000 tokens from written sources. However, to the best of our knowledge, no child corpora for any Arabic variety is available except for Egyptian Arabic, the Salama corpus, which can be found on the CHILDES platform.

The lack of a Palestinian Arabic child corpus motivated us to build one for child and CDS. The aims of this paper are twofold: we aim to present the new corpus of child and CDS for Palestinian Arabic and also to conduct a first analysis of the spontaneous productions of children and compare it to that of adults. In particular, the hypothesis we consider is whether the development of morphosyntactic features is as early as Very Early Parameter Setting (Wexler, 1998) establishes (see also Hoekstra & Hyams, 1998). In particular, we investigate subject–verb agreement, the presence of null subjects, and word order alternations. This research is, therefore, part of the collective effort to outline what very young children have established about their target grammar by the time they produce their first syntactic productions.

2. Method

The corpus presented in this paper consists of the transcripts of the early acquisition of Palestinian Arabic based on the recordings of child–adult interactions collected at different sites (Hebron, Taybeh, Tulkarm, Jenin, Nablus, Qalqilya, Ramallah, Jaljuliya) so that the corpus constitutes a representative sample of Palestinian Arabic. The recordings include the spontaneous productions of 16 healthy monolingual Palestinian Arabic-speaking children aged between 19 and 58 months at the time of recording (mean age in months = 36; 50% females) and those of their adult interlocutors (mean age in years = 28; 84% males). None of the children recorded were premature, had a hearing impairment or had any other health issues.

To obtain the data, families were recruited within the personal and professional networks of the first author. Parents, having signed informed consent forms, agreed to record 30-minute face-to-face spontaneous interactions with their child every 2 weeks. Adults were encouraged to engage in a variety of play activities, ask open-ended questions, and discuss life events with their children to promote active conversation. These interactions were recorded between February and August 2021 using Apple iPhones. The smartphone was placed in the room where the recording was taking place, with children moving around it within a range of 3 m. Adult participants provided additional details for each recording, including which people were present in each recording and where the recording took place (e.g., the child's home, the grandparents' house). Audio files were sent to the authors by an adult family member via WeTransfer. Only recordings with clear and high-quality sound were considered, resulting in the exclusion of one recording.

Under the supervision of the authors and two speech-language pathologists based in Palestine, 59 recordings were obtained from the 16 families who agreed to participate. Each child was recorded 2–5 times, resulting in a total duration of 1,387 minutes of recordings. The mean recording duration was 23.52 minutes, ranging from 7 to 35 minutes. While the target recording duration was 30 minutes, some sessions ended up being as short as 7 minutes due to the children's lack of cooperation. In those instances, the children lost interest and disengaged from the activity with their parents. However, since these shorter recordings still captured valuable natural interactions in Palestinian Arabic, we decided to include them in our analysis.

This resulted in 9,285 utterances of child speech and 10,496 utterances of CDS. Table 1 presents the characteristics of the children (identified by number), the number of their productions, and their mean length of utterance in words. MLUw was calculated

Table 1. Characteristics of children's recordings and production

Child	Gender	Language input	Socioeconomic status	Age (yy;mm;dd)	MLUw	No. of obtained recordings	No. of child utterances
1	Female	Arabic	Middle[1]	2;05.25–2;08.00	2.18–2.37	5	889
2	Male	Arabic	Middle	1;09.05–2;02.02	1.35–1.9	5	589
3	Male	Arabic	Middle	4;01.10–4;02.07	1.67–1.77	2	586
4	Male	Arabic	Middle	1;11.06–1;11.08	1.05–1.56	4	309
5	Male	Arabic	Middle	3;05.21–3;10.20	1.75–2.41	5	799
6	Male	Arabic	Middle	3;07.29–3;11.02	2.18–2.63	5	1,118
7	Female	Arabic	Middle	4;07.28–4;10.07	2.54–3.36	4	917
8	Female	Arabic	Middle	2;09.00–3;01.06	1.11–1.29	5	896
9	Male	Arabic	Middle	3;11.14–4;02.26	1.27–1.47	5	477
10	Female	Arabic	Middle	2;10.08–3;00.28	1.06–1.09	3	608
11	Male	Arabic	Middle	3;11.14–4;01.00	1.25–1.69	4	366
12	Female	Arabic	Middle	2;08.12–2;09.05	2.09–2.15	2	458
13	Female	Arabic	Middle	2;10.15–2;11.14	1.47–1.48	3	288
14	Female	Arabic	Middle	2;00.10–2;00.25	1.62–1.63	2	361
15	Female	Arabic	Middle	1;07.21–1;08.03	1.06–1.07	2	186
16	Male	Arabic	Middle	04;01.00–04;02.03	1.88–1.92	3	438
Total						59	9,285

manually over the first 100 spontaneous utterances by each child. The characteristics of the adults interacting with the children are shown in Table 2.

Notice that the MLU of two children, 3 and 16, were noticeably lower than those of their age peers (MLU of 3 was between 1.67 and 1.77, and between 1.88 and 1.92 for 16). Under closer scrutiny, the productions of 16 became similar to those of his peers when a larger sample of his productions were taken into account. The productions of 3, under the same method, remained low for his age, and therefore one might consider the possibility that he is affected by language delay or even disorder.[1]

2.1. Coding

Collected recordings were transcribed in Arabic as well as a romanised system. The transcription and coding for the data of both children and adults followed the same method. Unintelligible utterances were transcribed as xxx, while incomplete words were

[1]All participants were from middle-class backgrounds, as determined on the basis of parental occupation, educational level, family size, and household income. The average estimated monthly household income for middle-class families was 4,500 New Israeli Shekels (NIS), derived from the 2020 Socio-Economic Conditions Survey by the Palestinian Central Bureau of Statistics. Under these criteria, individuals from middle-class environments comprise approximately 80% of the Palestinian population.

Table 2. Characteristics of the adults' recordings

Child	Interlocutor	Age (years)	No. of adult utterances
1	Mother	31	292
	Aunt	23	888
2	Mother	28	713
	Father and mother	29 and 28	154
3	Brother	20	610
4	Mother	26	103
	Aunt	21	15
	Father	35	313
5	Aunt	20	1,181
6	Mother and sister	45 and 20	375
	Sister	20	779
7	Aunt	33	73
	Mother	21	151
8	Mother	23	995
9	Mother	34	722
10	Aunt	23	210
	Mother	23	618
11	Aunt	21	254
	Mother	34	257
12	Brother	20	490
13	Mother	27	353
14	Mother	26	468
15	Father	32	258
16	Mother	35	858
Total			10,496

transcribed with the omitted part in parenthesis as *(ban)do:ra* for *bando:ra* "tomato." Furthermore, special markers were used to indicate special forms of speech, such as @d for dialect form, @f for family form, @s\$n for second-language form, and @si for singing (MacWhinney, 2000).

The corpus created, named the [Nazzal] corpus, was manually transcribed following the CHILDES manual (MacWhinney, 2000) and checked using the CLAN software. The corpus can be found under the subdirectory of Arabic in the Other directory. Only the parents of 11 participants of those reported above gave permission for their transcripts to be part of the CHILDES platform, whereas the other five participants did not because personal family issues were discussed in the recordings. The children and parents whose interactions are currently available online in CHILDES are detailed in the Appendix A1 and A2.

The totally of the recordings (that is those detailed on Tables 1 and 2) were used for the purposes of the morphosyntactic analysis presented in the remainder of this paper. A total of 3,370 utterances from children's utterances and 5,681 adult utterances were included in the analysis reported. Single-word utterances including yes/no answers, and utterances such as recite the months in a year or a number series were excluded, as well as passive sentences. When utterances consisted of more than one clause, each was analysed separately for the purposes of the analysis of word order and subject production.

Furthermore, for the purposes of the analysis of the productions, and to establish if there was any change in the children's productions over the course of development, the children's productions were divided into three age groups: 19–26 months (mean age in months = 22.7; 50% females), 29–37 months (mean age in months = 33.5; 100% females), and 41–58 months (mean age in months = 48.6; 14% females). These groupings allowed for roughly equal intervals between age ranges and age groups of similar length. In addition, this division provided the most balanced distribution of participants within each group. This grouping meant that some age groups were not represented in the sample: there is a gap between 26 and 29 months and another between 37 and 41 months. Lacking information on these two periods does not seem to compromise the validity of the results, as argued in the next section.

3. Results

In this section, we present a quantitative study of the productions of children and their adult interlocutors in the corpus built (including all the children and adults reported, not only those whose transcripts appear in CHILDES). Our goal was to consider the macro-parameters that characterise Palestinian Arabic; we focused on the presence of null subjects, subject–verb agreement and word order. We included in our analysis declarative, imperative, and interrogative clauses, that contain a verb, modal, or auxiliary verb, as well as verbless copular sentences in the present tense, where "be" is phonetically null and the predicate can be a noun phrase, an adjective phrase, or a prepositional phrase (Aoun et al., 2009; Benmamoun, 2000). No difference in the word order structure was found when the subject or the object appeared as a noun or as a pronoun, and for that reason they were not coded differently. Therefore, "S" denotes a subject, and "O" denotes an object (direct or indirect, including reflexives), and may refer to full Determiner phrases, proper names, and pronouns. Since the analysis focuses on word order, wh- elements are coded as "S" or "O" according to their function. Finally, the term auxiliary verb "aux" has been used to refer to both auxiliary verbs, such as *kana* "was" and *sˤa:ra* "become" and modals verbs such as *biddi* "I want," *la:zim* "have to," *yimken* "may, could," *baqdar* "can," considered modal verbs in Palestinian Arabic (Alharbi, 2002; Aoun et al., 2009).

3.1. Null subject

We measured the incidence of null and overt subjects in the productions of children and adults; they are exemplified in (1). Since imperatives consistently present null subjects, only declarative and interrogative clauses, along with verbless copular sentences, were included in the analysis.

(1a) ʃtare:na: bu:za. (child 12, 2;08.12)
Bought–1PL ice-cream
"We bought ice-cream."

Table 3. Distribution of overt and null subjects by children and adults

	Feature of subjects				
	Overt subjects		Null subjects		
Participants	Count	%	Count	%	Total
Adults	1,311	34.8	2,460	65.2	3,771
All children	964	28.7	2397	71.3	3,361
Children 19–26 months	101	31.6	219	68.4	320
Children 29–37 months	221	29.2	537	70.8	758
Children 41–58 months	595	28	1530	72	2125

(1b) ʔana ʔakalet kaʕke. (child 14, 1;09.28)
I ate–1SG cake
"I ate a cake."

The results appear in Table 3.

We calculated the proportions of null subjects over null and overt subjects in the speech samples of the four age groups (three groups of children, 19–26, 29–37, and 41–58 months and one group of adults). A Wilcoxon Signed Ranks Test was used to assess whether there were significant differences between the performances of the three child groups compared to adults. The results showed no significant difference between adults and the youngest age group ($Z = .365$, $p = .715$), nor between adults and the middle age group ($Z = -.135$, $p = .893$), or adults and the oldest age group ($Z = -.338$, $p = .735$). To compare the performance differences among the three age groups (young, middle, and old), a one-way analysis of variance (ANOVA) was conducted. The dependent variable was the performance score, and the independent variable was age group. Since the overall ANOVA was not significant, a post hoc test (Tukey's HSD) was performed to further examine pairwise comparisons among the groups. The one-way ANOVA revealed that there was no statistically significant difference in performance across the three age groups, $F(2,13) = .475$, $p = .632$. The post hoc Tukey's HSD test confirmed that none of the pairwise comparisons reached statistical significance ($p > .05$ for all comparisons). These findings indicate that age does not have a significant effect on performance in this dataset.

3.2. Subject–verb agreement

We then considered subject–verb agreement, which appears in the form of discontinuous morphology in Palestinian Arabic, as in the Semitic languages in general; agreement was considered for all inflected forms, main verbs and auxiliaries. The results appear in Table 4. Children in the youngest age group (19–26 months) produced a noticeably higher rate of agreement errors (9.3%) compared to the two older groups (1.7% for 29–37 months and 0.8% for 41–58 months). Most of these errors involved incorrect marking of numbers (29 out of 30), with only one person error and no gender errors, for reasons that remain for future research.

Table 4. Subject–verb agreement errors in child production

		Agreement errors		Error type		
Age group	No. of verbs	Count	%	Number	Gender	Person
19–26 months	322	30	9.3	1	29	0
29–37 months	830	14	1.7	8	4	2
41–58 months	1962	16	0.8	6	8	2
Total	3114	60		15	41	4

Notice that copular sentences with *ka:n* 'be' present the overt form of the verb only in some tenses, but not in the present tense (Aoun et al., 2009; Benmamoun, 2000). According to our recounts, children produced copular sentences with *kana* "was" or *raħ* "will" in the past and future tenses but never in the present tense. A total of 376 copular sentences were produced by children, 260 (69%) in the present tense, all with null "be," 116 in the past and future tenses, all with overt "be." No errors were found. Examples of child production are given in (2a) and (2b). For adults, 524 copular sentences were found, of which 399 (76%) presented null "be" in the present tense.

(2a) Ma:ma: beʃ-ʃu ɣul. (child 13, 2;11.14)
Mama at the-work.
"My mom is at work."

(2b) Ma:ma ka:nat fel-ħisˤa. (child 5, 3;06.29)
mama Was in-the-class
"My mom was at the class."

3.3. Word order

The frequency of different word orders was manually analysed; declarative, imperative, and interrogative clauses were coded (with a total of 5,634 sentences for adults and 3,298 for children). No variation in word order structure was observed whether the subject or object was expressed as a noun or a pronoun, as mentioned above. For sentences with an overt subject and an overt object, SVO (3a) was the predominant order in adult (71.1% of sentences) and child production (91.55% of sentences), while VSO accounted for 2.39% of sentences in adults and 4.21% of sentences in children. Non-canonical but grammatical word orders OVS, OSV, $VO_{Clitic}S$, VOS, and $OSVO_{Clitic}$ (3b) represented 26.51% of adult production and 4.24% of child production.

(3a) Ba:ba: rasam da:ʔira (child 16, 04;01.00)
daddy drew–3 M.SG circle
"Daddy drew a circle."

(3b) l-may ʔana ʃribt-ha:. (child 12, 2;08.12)
the-water I drank–1SG-it
"The water I drank it."

Table 5. Frequency and percentage of sentences with S, V, O, children, and adults

	All children	19–26 months	29–37 mo	41–58 months	Adults
Word order	Count (%)	Count (%)	Count (%)	Count (%)	Count (%)
S V O	125 (47.89)	12 (52.17)	27 (44.26)	86 (48.58)	203 (40.43)
S V O_{Clitic}	94 (36.01)	9 (39.13)	23 (37.7)	62 (35.08)	134 (26.69)
O V S	2 (0.76)	0 (0)	1 (1.63)	1 (0.56)	72 (14.34)
V O_{Clitic} S	0 (0)	0 (0)	0 (0)	0 (0)	22 (4.38)
O S V	2 (0.76)	0 (0)	1 (1.63)	1 (0.56)	18 (3.58)
V O S	6 (2.29)	1 (4.34)	0 (0)	5 (2.82)	16 (3.18)
S Aux V O	15 (5.74)	1 (4.34)	5 (8.19)	9 (5.08)	12 (2.39)
V S O	11 (4.21)	0 (0)	2 (3.27)	9 (5.08)	12 (2.39)
S V V O	5 (1.91)	0 (0)	2 (3.27)	3 (1.69)	8 (1.59)
O S V O_{Clitic}	1 (0.38)	0 (0)	0 (0)	1 (0.56)	5 (0.99)
Total	261	23	61	177	502

Abbreviations: V = Verb, S = Subject, O = Direct or Indirect Object, PP = Prepositional Phrase, Aux.v = Auxiliary Verb.

Word order distribution for adults and children is presented in Table 5 (sentences with overt S and O) and Table 6 (sentences with null arguments included). (The order of presentation of different word orders is based on their incidence in adult production.)

In total, adults exhibited 50 different word orders, while children produced 45.

We found consistent word order structures in children and adults, with no ill-formed sequences in either case.[2] Following a reviewer's suggestion, we calculated the frequency of occurrence of SV versus VS structures for the three age groups separately to explore the developmental trajectory of children. The results confirmed that SV was the predominant structure across all three age groups of children, with the mean frequencies as follows: youngest group (19–26 months) 0.81, middle group (29–37 months) 0.87, and older group (41–58 months) 0.78. Similarly, the adult group also showed a mean of 0.78 for SV.

As may be observed, sentences with more than one verb (VV, VVO, etc.), known as serial verb constructions (Altakhaineh & Zibin, 2017; Hussein, 1990), were found in the corpus (see (5)). Serial verb constructions were found in 4.53% of adults' sentences (featuring various word orders: VV, VVO, VV PP, SVV, SVVO, Aux VVO, VVS, and VVV) and 2.06% of child sentences (from 8 children, with an age range of 23 to 56 months); the first occurrence was found at 23 months.

(4) Ra:ħat tʒi:b ʔed-daftar. (child 4, 1;11.08)
Went.2FEM.SG bring–2FEM.SG the-notebook
"She went to bring the notebook."

[2]Given the large number of words orders encountered, we do not attempt a statistical comparison of the different age groups. A more detailed examination of word order production in an expeirmental setting remains for future research.

Table 6. Frequency and percentage of different word orders, children and adults

	All children	19–26 months	29–37 months	41–58 months	Adults
Word order	Count (%)	Count (%)	Count (%)	Count (%)	Count (%)
V	815 (24.7)	135 (38.35)	238 (27.38)	442 (22.16)	1026 (18.21)
V O	489 (14.83)	45 (12.78)	100 (11.65)	344 (16.47)	743 (13.18)
V PP	244 (7.39)	5 (1.42)	45 (5.24)	194 (9.66)	389 (6.9)
V O_{Clitic}	243 (7.37)	29 (8.23)	77 (8.97)	137 (6.56)	387 (6.87)
S Pred.	241 (7.3)	32 (9.09)	97 (11.3)	112 (5.57)	369 (6.55)
S V	193 (5.85)	29 (8.23)	45 (5.24)	119 (5.69)	236 (4.19)
S V O	125 (3.79)	12 (3.4)	27 (3.31)	86 (4.11)	203 (3.6)
S V PP	103 (3.12)	5 (1.42)	12 (1.39)	86 (4.11)	112 (1.99)
S V O_{Clitic}	94 (2.85)	9 (2.55)	23 (2.68)	62 (2.96)	134 (2.38)
Aux V O	93 (2.82)	2 (0.56)	29 (3.37)	62 (2.96)	87 (1.54)
Aux V	88 (2.67)	7 (1.98)	11 (1.28)	70 (3.35)	116 (2.06)
V S	87 (2.64)	15 (4.26)	8 (0.93)	64 (3.06)	84 (1.49)
O V	23 (0.7)	0 (0)	2 (0.23)	21 (1)	373 (6.62)
V O_{Clitic} O	51 (1.55)	2 (0.56)	19 (2.21)	30 (1.49)	268 (4.75)
V O_{Clitic} PP	42 (1.27)	1 (0.28)	10 (1.16)	31 (1.49)	47 (0.83)
Aux	40 (1.21)	5 (1.42)	28 (3.26)	7 (0.33)	11 (0.2)
Aux V PP	39 (1.18)	1 (0.28)	11 (1.28)	17 (0.81)	63 (1.23)
V PP O	35 (1.06)	1 (0.28)	16 (1.86)	18 (0.86)	45 (0.8)
Aux O	32 (0.97)	11 (3.12)	6 (0.69)	15 (0.71)	49 (0.87)
V V	28 (0.85)	0 (0)	12 (1.39)	16 (0.76)	123 (2.18)
S Aux V	25 (0.76)	2 (0.56)	8 (0. 93)	15 (0.71)	43 (0.76)
O V O_{Clitic}	23 (0.7)	1 (0.28)	2 (0.23)	20 (0.95)	31 (0.55)
Pred.	19 (0.58)	0 (0)	3 (0.34)	16 (0.76)	30 (0.53)
V V O	15 (0.46)	1 (0.28)	4 (0.46)	10 (0.47)	63 (1.12)
S Aux V O	15 (0.46)	1 (0.28)	5 (0.58)	9 (0.43)	12 (0.21)
V PP S	12 (0.36)	0 (0)	2 (0.23)	10 (0.47)	11 (0.2)
V S PP	12 (0.36)	0 (0)	1 (0.14)	11(0.52)	5 (0.09)
V S O	11 (0.33)	0 (0)	2 (0.23)	9 (0.43)	12 (0.21)
V V PP	9 (0.27)	0 (0)	2 (0.23)	7 (0.33)	21 (0.37)
V O PP	7 (0.21)	0 (0)	1 (0.14)	6 (0.28)	20 (0.35)
O Aux V PP	6 (0.18)	0 (0)	2 (0.23)	4 (0.19)	28 (0.5)
S V V	6 (0.18)	0 (0)	3 (0.34)	3 (0.14)	18 (0.32)
V O S	6 (0.18)	1 (0.28)	0 (0)	5 (0.24)	16 (0.28)
S V V O	5 (0.15)	0 (0)	2 (.23)	3 (0.14)	8 (0.14)

Table 6. *(Continued)*

	All children	19–26 months	29–37 months	41–58 months	Adults
Word order	Count (%)	Count (%)	Count (%)	Count (%)	Count (%)
Aux V V O	4 (0.12)	0 (0)	0 (0)	4 (0.19)	8 (0.14)
Aux V S	4 (0.12)	0 (0)	1 (0.11)	3 (0.14)	6 (0.11)
O V PP	4 (0.12)	0 (0)	1 (0.11)	3 (0.14)	77 (1.37)
O Aux	2 (0.06)	0 (0)	0 (0)	2 (0.09)	58 (1.03)
O S V	2 (0.06)	0 (0)	1 (0.11)	1 (0.04)	18 (0.32)
O V S	2 (0.06)	0 (0)	1 (0.11)	1 (0.04)	72 (1.28)
V V S	1 (0.03)	0 (0)	0 (0)	1 (0.04)	6 (0.11)
O S V O_{Clitic}	1 (0.03)	0 (0)	0 (0)	1 (0.04)	5 (0.09)
PP V	1 (0.03)	0 (0)	1 (0.11)	0 (0)	31 (0.55)
PP V O	1 (0.03)	0 (0)	0 (0)	1 (0.04)	2 (0.04)
PP V O_{Clitic}	1 (0.03)	0 (0)	0 (0)	1 (0.04)	3 (0.05)
Aux V	0 (0)	0 (0)	0 (0)	0 (0)	39 (0.69)
O Aux V	0 (0)	0 (0)	0 (0)	0 (0)	94 (1.67)
O V V	0 (0)	0 (0)	0 (0)	0 (0)	6 (0.11)
V O_{Clitic} S	0 (0)	0 (0)	0 (0)	0 (0)	22 (0.39)
Aux V O_{Clitic}	0 (0)	0 (0)	0 (0)	0 (0)	2 (0.04)
V V V	0 (0)	0 (0)	0 (0)	0 (0)	2 (0.04)
Total	3,298	352	858	2,088	5,634

Abbreviations: V = Verb, S = Subject, O = Direct or Indirect Object, PP = Prepositional Phrase, Aux.v = Auxiliary Verb, *Pred* = *Predicate.*

4. Discussion

In this study, we examined several grammatical features in the speech of Palestinian Arabic-speaking children across different age groups. Overall, child production in the domains investigated shows a strong tendency towards adult-like patterns, though some variability remains, particularly in the youngest age group. The results show a clear developmental trajectory in subject–verb agreement and word order. The youngest group (19–26 months) had a higher error rate in agreement (9.3%), particularly with number, compared to the older groups (1.7% and 0.8%). This is the one domain where the younger group is barely above 90% adult-like performance. SVO word order was consistently used across age groups, with the youngest group (0.81%) already aligning with the adult rate (0.78). Serial verb constructions, though more frequent in adults (4.53%), were also present in children (2.06%), and attested from 23 months of age. Palestinian Arabic is a null subject language and, as such, it allows for phonetically null object when the discourse context allows the speakers to retrieve the relevant information (Albirini et al., 2011; Kenstowicz, 1989; Rizzi, 1982). We have shown that children produce null subjects at the same rate as adults do. This is consistent with previous studies in other null subject languages. For Romance, the results of null subject production are 62%, 70%, and 67% for Catalan (Bel, 2003; Cabŕe-Sans & Gavarŕo, 2006), Italian (Lorusso et al., 2005), and Spanish

(Bel, 2003), respectively; child null subject rates are not significantly different from them. A Yemeni Ibbi Arabic study by Qasem (2020) reported an 86–87% of null subjects in children's production, although no results were given for adult production. These studies converge in the idea that the null subject parameter is set very early (Wexler, 1998).

Palestinian Arabic presents person, number, and gender agreement, as exemplified in (3a) above. The error rate in production of subject–verb agreement in children was 1.92% for the age range of 18 to 56 months. The near-perfect subject–verb agreement observed in further attests to the early mastery of agreement morphology in Palestinian Arabic. Given the non-concatenative nature of Arabic morphology, this challenges any claims that morphological complexity hinders early language development (Dromi et al., 1999). The non-concatenational character of Arabic morphology and its complexities (T et al., 2021) are no obstacle for early attainment. The absence of "be" in copular sentences in the present tense among Palestinian Arabic-speaking children aligns with Schütze's (2004) claim that children have early command of the realisation of Tense, since these children do not insert a copula where it is not present in the adult language, and systematically insert it when is required.

Regarding word order, the current study found SVO as the predominant order in Palestinian Arabic speech production, whether adult or child. This observation is in line with Benmamoun (1997), Shlonsky (1997), Mohammad (2000), and Saiegh–Haddad (2003), who assert that SVO is the default word order in Palestinian Arabic, while VSO is the basic word order in Standard Arabic. Similar word order preferences, favouring SVO, were identified in various spoken Arabic varieties, such as Jordanian (El-Yasin, 1985), Egyptian (Albirini et al., 2011), and Moroccan (Announi, 2021).

In contrast to our findings, Friedmann and Costa (2011) found a preference for VS order as opposed to SV in their study of child Palestinian Arabic, using a repetition task with 20 children of ages 1;9 to 3;0. Similarly, Khamis-Dakwar (2011) found a preference for VSO as opposed to SVO in another repetition task run with Palestinian Arabic children in the same age range. The source of this contrast may be in the methods used in those two studies; in particular, the fact that children chose to change the word order in the repetition tasks reported may indicate that the discourse setting invited a given word order over another; the discrepancy remains for future research. On the other hand, the current study's findings are in line with Abboud et al.'s (2022) research on Lebanese Arabic-speaking children, indicating simultaneous emergence of SV and VS orders. Overall, children's spontaneous production indicates knowledge of numerous word order alternations, with no deviant word orders attested. These word order alternations imply the resource to various syntactic operations (wh-movement in questions, object dislocations).

Serial verb constructions have received no attention in the literature on the acquisition of Arabic. These are sentences where multiple verb forms appear consecutively in a single clause, denoting a complex action or event (Altakhaineh & Zibin, 2017; Hussein, 1990). The absence of research on the acquisition of serial verb constructions in Arabic leaves an open avenue for future investigation.

The results of the analysis of some of the core properties of Palestinian Arabic in the children's early productions align with the predictions of Very Early Parameter Setting (Wexler, 1998) or Early Morphosyntactic Convergence (Hoekstra & Hyams, 1998). The grammatical phenomena examined range from the production of null subjects and subject–verb agreement to word order distribution, absence/presence of copular "be" and production of serial verbs. While in this last case, the findings may be nearly anecdotal, while for null subjects and agreement, naturalistic data provide abundant

evidence for grammatical acquisition. Moreover, when we consider the children grouping them in three age subgroups, we find that early production does not differ from that of the older children (with the exception of subject–verb agreement errors, which are slightly higher for younger children and may be attributed to the acquisition of morphological exponents). Overall, in our interpretation, our results point to continuity in early development. Other domains in which child grammar is generally agreed to be delayed, as for example passive voice, have not been considered, and have been left for later work.

The observations so far were possible thanks to the collection of child and adult interactions in a naturalistic setting. The resulting corpus has been made available to the community through the CHILDES platform serving as a valuable resource for researchers, educators, and practitioners alike. While corpora have their limitations, in the case of child language they provide abundant information on grammatical phenomena. These findings, and other drawn from the corpus, can be used in comparative work with other languages, can serve as reference in language impairment studies, and can inform experimental design.

Acknowledgements. The authors wish to thank the Palestinian children and their parents for their participation in this research. The authors also thank An-Najah National University (www.najah.edu) and Universitat Autònoma de Barcelona for the technical support provided to publish the present manuscript.

Funding statement. This study received financial support through the project Development and acquisition of preverbal syntax and semantics (DAPSS), PID2022-138413NB-100, Ministerio de Ciencia e Innovación.

Competing interests. The authors declare none.

References

Abboud, L., **Choueiri, L.**, **Seifeddine, N.**, & **Tuller, L.** (2022). The emergence of subjects in Lebanese two-year-olds. *Journal of Child Language*, **49**(1), 1–14. https://doi.org/10.1017/S0305000922000587.

Albirini, A., **Benmamoun, E.**, & **Saadah, E.** (2011). Grammatical features of Egyptian and Palestinian Arabic heritage speakers' oral production. *Studies in Second Language Acquisition*, **33**(2), 273–303. https://doi.org/10.1017/S0272263110000768.

Alharbi, A. (2002). Verbal modals. *Majalat Jami'at Um Alqura*, **14**(1), 1–28.

Altakhaineh, A., & **Zibin, A.** (2017). Verb+verb compound and serial verb construction in Jordanian Arabic (JA) and English. *Lingua*, **201**, 45–56. https://doi.org/10.1016/j.lingua.2017.08.010.

Announi, I. (2021). The problem of word order and verbal movement in Moroccan Arabic. *International Journal of Linguistics, Literature and Translation*, **4**(4), 34–54. https://doi.org/10.32996/ijllt.2021.4.4.6.

Aoun, J., **Benmamoun, E.**, & **Choueiri, L.** (2009). *The syntax of Arabic*. Cambridge University Press.

Bel, A. (2003). The syntax of subjects in the acquisition of Spanish and Catalan. *Probus*, **15**(1), 1–26. https://doi.org/10.1515/prbs.2003.003.

Benmamoun, E. (1997). Licensing of negative polarity items in Moroccan Arabic. *Natural Language and Linguistic Theory*, **15**(2), 263–287. https://doi.org/10.1023/A:1005727101758.

Benmamoun, E. (2000). *The feature structure of functional categories: A comparative study of Arabic dialects*. Oxford University Press.

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

Cabŕe-Sans, Y., & **Gavarŕo, A.** (2006). Subject distribution and verb classes in child Catalan. In A. Belikova, L. Meroni, & M. Umeda (Eds.), *GALANA2- proceedings of the conference on generative approaches to language acquisition - North America (GALANA 2)* (pp. 123–133). Cascadilla Press.

Corrigan, R. (2012). Using the CHILDES database. In E. Hoff (Ed.), *Research methods in child language: A practical guide* (pp. 271–284). Wiley Online Library.

Dash, N. S., & **Arulmozi, S.** (2018). Limitations of language corpora. In: *History, features, and typology of language corpora*. Springer, Singapore. https://doi.org/10.1007/978-981-10-7458-5_15

Demuth, K., (2008). Exploiting corpora for language acquisition research. In *Corpora in language acquisition research: Finding structure in data* (pp. 199–205).

Dromi, E., **Leonard, L. B.**, **Adam, G.**, & **Zadunaisky-Ehrlich, S.** (1999). Verb agreement morphology in Hebrew-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **42**(6), 1414–1431. https://doi.org/10.1044/jslhr.4206.1414.

El-Yasin, M. K. (1985). Basic word order in classical Arabic and Jordanian Arabic. *Lingua*, **65**(1–2), 107–122. https://doi.org/10.1016/0024-3841(85)90022-1.

Friedmann, N., & **Costa, J.** (2011). Acquisition of SV and VS order in Hebrew. *European Portuguese, Palestinian Arabic, and Spanish. Language Acquisition*, **18**(1), 1–38. https://doi.org/10.1080/10489223.2011.530507.

Hoekstra, T., & **Hyams, N.** (1998). Aspects of root infinitives. *Lingua*, **106**, 81–112.

Hussein, L. (1990). Serial verbs in colloquial Arabic. In B. D. Joseph & A. M. Zwicky (Eds.), *When verbs collide: Papers from the 1990 Ohio state mini-conference on serial verbs* (pp. 340–354). The Ohio State University.

Jarrar, M., **Habash, N.**, **Alrimawi, F.**, **Akra, D.**, & **Zalmout, N.** (2017). Cur ras: An annotated corpus for the Palestinian Arabic dialect. *Language Resources and Evaluation*, **51**(3), 745–775. https://doi.org/10.1007/s10579-016-9370-7.

Kenstowicz, M. (1989). The null subject parameter in modern Arabic dialects. In O. Jaeggli & K. Safir (Eds.), *The null subject parameter* (pp. 263–275). Springer.

Khamis-Dakwar, R. (2011). Early acquisition of SVO and VSO word orders in Palestinian colloquial Arabic. In E. Broselow & H. Ouali (Eds.), *Perspectives on Arabic linguistics XXII-XXIII* (Vol. **317**, pp. 281–292). https://doi.org/10.1075/cilt.317.13kha.

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, **120**(11), 2546–2556. https://doi.org/10.1016/j.lingua.2010.06.005.

Lorusso, P., **Caprin, C.**, & **Guasti, M. T.** (2005). Overt subject distribu tion in early Italian children. In *A supplement to the proceedings of the 29th annual Boston University conference on language development*, Cascadilla Press.

MacWhinney, B. (1996). The CHILDES system. *American Journal of Speech-Language Pathology*, **5**(1), 5–14. https://doi.org/10.1044/1058-0360.0501.05.

MacWhinney, B., (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

MacWhinney, B., & **Snow, C.** (1990). The child language data exchange system: An update. *Journal of Child Language*, **17**(2), 457–472.

Mohammad, M. A. (2000). *Word order, agreement and pronominalization in standard and Palestinian Arabic*. John Benjamins.

Qasem, F. A. A. (2020). The acquisition phenomenon of null and overt subjects in the early speech of Arabic-speaking children. *Macrolinguistics*, **8**(1), 68–87. https://doi.org/10.26478/ja2020.8.12.5.

Rizzi, L. (1982). *Issues in Italian syntax*. Foris.

Saiegh–Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied PsychoLinguistics*, **24**(3), 431–451. https://doi.org/10.1017/S0142716403000225.

Schütze, C. T. (2004). The non-omission of nonfinite be. In A. D. P. Svenonius & M. Richardsen (Eds.), *Proceedings of the 19th Scandinavian conference of linguistics, Nordlyd, vol. 31.3: Acquisition* (pp. 606–622). https://doi.org/10.7557/12.46.

Shlonsky, U. (1997). *Clause structure and word order in Hebrew and Arabic: An essay in comparative Semitic syntax*. Oxford University Press.

Snow, C. E. (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* (pp. 180–193). Blackwell Publishing.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, **27**(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002.

Taha, J., **Stojanovik, V.**, & **Pagnamenta, E.** (2021). Expressive verb morphology deficits in Arabic-speaking children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, **64**(2), 561–578. https://doi.org/10.1044/2020_JSLHR-19-00292.

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. https://doi.org/10.1016/S0024-3841(98)00029-1.

Appendix

Table A1. Characteristics of the children's recordings in CHILDES

Child	No. of files (identifier)	Age (yy;mm;dd)	MLUw	No. of child utterances
Child 1	20525, 20606, 20625, 20710, 20800	2;05;25–2;08;00	2.18–2.37	889
Child 2	10905, 10928, 11025, 11116, 20202	1;09;05–2;02;02	1.35–1.9	589
Child 3	40110, 40207	4;01;10–4;02;07	1.67–1.77	586
Child 4	10614, 10814, 10829, 11108	1;11;06–1;11;08	1.05–1.56	309
Child 5	30521, 30615, 30629, 30725, 31020	3;05;21–3;10;20	1.75–2.41	799
Child 6	30729, 30815, 30900, 30919, 31102	3;07;29–3;11;02	2.18–2.63	1,118
Child 7	40728, 40818, 40913, 41007	4;07;28–4;10;07	2.54–3.36	917
Child 8	20900, 20920, 21006, 21024, 30106	2;09;00–3;01;06	1.11–1.29	896
Child 9	31114, 31130, 40017, 40100, 40226	3;11;14–4;02;26	1.27–1.47	477
Child 10	21008, 21100, 30028	2;10;08–3;00;28	1.06–1.09	608
Child 11	31114, 31124, 40017, 40100	3;11;14–4;01;00	1.25–1.69	366
Total		1.86		7,554

Table A2. Characteristics of the adults' recordings in CHILDES

Child	No. of files (identifier)	Interlocutor	Age (years)	No. of adult utterances
Child 1	20525, 20606	Mother	31	292
	20625, 20710, 20800	Aunt	23	888
Child 2	10905, 11025, 11116, 20202	Mother	28	713
	10928	Father and Mother	29 and 28	154
Child 3	40110, 40207	Brother	20	610
Child 4	10614	Mother	26	103
	10814	Aunt	21	15
	10829, 11108	Father	35	313
Child 5	30521, 30615, 30629, 30725, 31020	Aunt	20	1,181
Child 6	30729	Mother and Sister	45 and 20	375
	30815, 30900, 30919, 31102	Sister	20	779
Child 7	40728	Aunt	33	73
	40818, 40913, 41007	Mother	21	151
Child 8	20900, 20920, 21006, 21024, 30106	Mother	23	995
Child 9	31114, 31130, 40017, 40100, 40226	Mother	34	722
Child 10	21008	Aunt	23	210
	21100, 30028	Mother	23	618
Child 11	31114, 40100	Aunt	21	254
	31124, 40017	Mother	34	257
Total				8,703

Cite this article: Nazzal, T., & Gavarró, A. (2025). A corpus analysis of child and child-directed speech in Palestinian Arabic: A first approach to syntactic development. *Journal of Child Language* 1–16, https://doi.org/10.1017/S030500092510007X