CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Distribution of Nominal Word-Patterns and Roots in Palestinian Arabic: A Developmental Perspective in Early Childhood

Yasmin SHALHOUB-AWWAD[*],[**] (iD) and Maram KHAMIS-JUBRAN[*],[**]

*Department of Learning Disabilities, University of Haifa and **The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa
Address for correspondence: Yasmin Shalhoub-Awwad, Department of Learning Disabilities, Faculty of Education, University of Haifa, Haifa, Israel. Email: yasmin@edu.haifa.ac.il

Abstract

This study investigated the acquisition of word-patterns and roots in the nominal system of the spoken language of Palestinian Arabic (PA) and its distance from Standard Arabic (StA). It described, analyzed, and quantified the nominal system (roots and word-patterns) as reflected in the language corpus of Palestinian-Arab kindergarteners 3 to 6 years old. The results showed that NON-LINEAR DERIVED NOUNS (deverbal nouns) are the most frequently used category (49.5%). PRIMITIVE NOUNS comprise 43.1% of the nouns, whereas LINEAR DERIVED NOUNS barely exist before children start school (0.3%). Additionally, the results showed that half of the nouns were built from common word-patterns and roots between PA and StA, whereas 30% of the nouns were constructed from different word-patterns with common roots. Although PA and StA have much in common morphologically, there exists a significant degree of divergence.

Introduction

The basic vocabulary acquired by preschool-age children is primarily derived from spoken language. This core lexicon develops during the school years to become the mental lexicon of literate adults (Ravid, 2004). The words in the mental lexicon are stored in a sophisticated, multidimensional network, rather than as separate units. They are also interconnected via various associative relationships according to the different semantic, syntactical, morphological, and phonological representations, which differ in strength (Ravid, 2004). Morphology is considered one of the major organizing devices of the mental lexicon of many languages, which relates structural and semantic constituents within words (Marslen-Wilson, 2007). This morphological organization is particularly important in Semitic languages: Arabic and Hebrew as

CrossMark

synthetic languages with rich morphology where many grammatical and lexical notions are encoded in word-internal structures (Boudelaa, 2014; Ravid, 2012). This morphological organization is grounded by strong evidence mainly from priming studies of word structure and meaning, which has been conducted in these languages (e.g., Boudelaa & Marslen-Wilson, 2005, 2011; Deutsch, Frost & Forster, 1998; Velan, Frost, Deutsch & Plaut, 2005). The current study presents new data and analyses of the core morphological structures of the Arabic nouns (roots and word-patterns), as reflected in the language corpus of native Arabic-speaking 3- to 6-year-old children in the regional dialect known as "Palestinian Arabic" in the north of Israel. It describes, analyzes, and quantifies these morphological structures as a window on linguistic knowledge and usage across development.

Root and Pattern Lexical Organization

Derivation deals with fundamental word structure and new word-formation. In Arabic, it mainly emerges in two ways: non-linear derivation and linear derivation. Much of the Arabic lexicon is organized by the non-linear combination of two sub-lexical morphological primes: root and word-pattern (e.g., Boudelaa & Marslen-Wilson, 2005, 2011; Shalhoub-Awwad & Leikin, 2016). The root usually consists of three consonants; the word-pattern is made up of vowels but can contain consonants (Wright, 1995). The tri-consonantal root carries the semantic core meaning shared to varying degrees by most of the derivatives of a given root (Boudelaa & Marslen-Wilson, 2015), whereas the word-pattern provides information about the phonological structure of the surface form and its morpho-syntactic properties (Deutsch & Malinovitch, 2016). The insertion of the root consonants within the word-pattern produces a specific lexical item with a unique meaning and a well-defined grammatical category (Saiegh-Haddad & Henkin-Roitfarb, 2014). For example, the surface form مَلْعَب /malʕab/'playground' consists of the root ل.ع.ب /l.ʕ.b/ with the general semantic load 'playing' and the word-pattern maCCaC[1] denoting the place noun. This process of combining roots with word-patterns in Semitic languages is considered effective and productive, operating on a limited number of common roots to encode notions belonging to the same semantic field (unlike languages with concatenative morphology, such as English, that use many different lexemes for semantically related words). For example, all of the Arabic words /miftaːħ/, /faːtiħ/, /fatħa/, /maftuːħ/, /tafattaħ/, /ʔiftataħ/, and /munfatiħ/ have in common the root /f.t.ħ./ with the semantic field of opening, which makes them morphologically, phonologically, and semantically (to a degree) related. In contrast, their respective English counterparts – key, bright, slot, opened, bloomed, launched, broad-minded – do not share morphological or phonological relationships, although they may be semantically/associatively related. Furthermore, the word-pattern links all word-pattern-related lexical items as having the same prosodic structure (e.g., stress, vowels) and a similar categorical meaning and part of speech (Saiegh-Haddad, 2018). For example, all of the words – /maktab/ 'office', /malʕab/ 'playground', /maṭbax/ 'kitchen', /masnaʕ/ 'factory', /malʒaʔ/ 'shelter' – have in common the word-pattern maCCaC, with the same prosodic structure and place noun meaning. The second type of derivation is the LINEAR DERIVATION, which is limited to one morpheme and manifested by "pasting" it linearly, For example, adding the derivational suffix يّ /iːyy/ to شُمس /ʃams/ 'sun' creates the adjective شَمسِيّ /ʃamsiːyy/'sunny' (Shalhoub-Awwad & Leikin, 2016).

---

[1]The letter "C" represents the position of a root consonant.

There are two types of roots in Arabic: STRONG OR SOUND ROOTS, whose three consonants remain phonologically stable and systematically surface in every derived or inflected form feature in every complex form. These contrast with the WEAK ROOTS whose radicals contain the glides /y/ or /w/, or both. They are considered 'weak' since they are prone to morpho-phonological changes (Saiegh-Haddad & Henkin-Roitfarb, 2014). For instance, when the weak root w.f.q 'consent' is combined with the word-pattern ʔiCtaCaCa, the glide /w/ undergoes regressive assimilation from the word-pattern consonant /t/ immediately following it resulting in the allomorphic form /ʔittafaqa/'agree'. Most of the roots in the Arabic language are strong, and only 10% are weak (Boudelaa & Marslen-Wilson, 2004).

There are two main lexical categories in Arabic: verbs and nouns. Since our research focused on nouns, we will expand our description of the nominal system. Nouns may be morphologically sub-classified into two classes: deverbal nouns and primitive nouns. DEVERBAL NOUNS are those thought to be directly related to verbs. When the surface form is a DEVERBAL NOUN, it is usually straightforward to factor out the respective contributions of the root and the word-pattern (Boudelaa & Marslen-Wilson, 2015, p. 2–3). For instance, the meaning of the surface form /minʃafat-un/'a towel' derives from the combination of the meaning of the root /n.ʃ.f/ 'drying/' 'wiping' with the morpho-syntactic value of the word-pattern miCCaCa indicating a singular, feminine instrument noun (Boudelaa & Marslen-Wilson, 2015). Unlike deverbal nouns, PRIMITIVE NOUNS are widely held to be basic non-derived constructs (Wright, 1995). Even though, they are traditionally analyzed in the same way as deverbal nouns and verbs – by roots and word-patterns – the contribution of the word-pattern to the meaning, however, is much less constrained than in deverbal nouns (e.g., the noun /ʃams/ 'sun' is made up of the root /ʃ.m.s/ but the word-pattern CaCC which conveys, in this case, a feminine reading, can also have a masculine reading, as in /kalb/ 'dog' (Boudelaa & Marslen-Wilson, 2015).

The current study was based on the MORPHEME-BASED APPROACH that conceives of every Semitic surface form as a combination of a root and a word-pattern and the lexicon as a repository of roots and word-patterns with a set of rules to associate them (e.g., Bohas & Guillaume, 1984; Cohen, 1951; Hilaal, 1990; McCarthy, 1982). This approach is in sharp contrast with the stem-based approach, which views the Semitic lexicon as being built around processes that take the stem as a basic unit (e.g., Bat-El, 2017; Benmamoun, 2003; Heath, 2003; Ratcliffe, 2004). This view subsumes Semitic morphology within the same set of universal rules applied across the world languages. However, our decision to adopt the morpheme-based approach stemmed from substantial empirical support mainly from priming studies showing strong priming by roots and word-patterns (e.g., Boudelaa & Marslen-Wilson, 2005, 2011; Deutsch & Malinovitch, 2016; Deutsch, Velan, & Michaly, 2018; Frost, Forster, & Deutsch, 1997; Shalhoub-Awwad & Leikin, 2016; Shalhoub-Awwad, 2019). This pattern of results is clearly inconsistent with the strong stem-based view, which rejects the role of the root and word-pattern as relevant morphological units in processing (e.g., Bat-El, 2017). It also shows that Semitic mental lexicon is fed mainly by the non-linear combination of the root and word-pattern (Boudelaa, 2014), and points to their central role in lexical learning and poses the question as to how these structures are learned. Understanding the developmental trajectory and acquisition process of these structures (roots and word-patterns) is essential for insight into lexical and morphological development. It needs to be based on reliable knowledge as to the number and types of roots and word-patterns constituting the Arabic

lexicon – information that is currently unavailable. The current study takes one step in this direction by focusing on morphological structures of the Arabic nouns (roots and word-patterns) as reflected in the spoken language corpus of native Arabic-speaking 3- to 6-year-old children.

Acquisition of the Nominal System in Arabic

There is a lack of empirical studies tracing the acquisition of Arabic as a first language. Most of the research has looked at the acquisition process of inflectional morphology, mainly the acquisition of noun pluralization. One of these studies specifically investigated the acquisition of noun pluralization in Palestinian Arabic (Ravid & Farah, 1999; Saiegh-Haddad, Hadieh & Ravid, 2012). Alongside the work on inflectional morphology, two studies have looked at the acquisition of derivational morphology. However, they examined the distribution of the verbal patterns and their semantic-syntactic functions in Palestinian Arabic. The first study by Tarabani (2006) focused on the acquisition of the verbal patterns in the spoken language of Palestinian 2–6-year-old preschool children which indicated that the most frequently used verbal pattern was the non-augmented verbal pattern CaCaC, followed by CaCCaC and then ʔitCaCCaC, while the rest of the patterns were rare. Similar findings were obtained from a recent study of Laks, Hamad, and Saiegh-Haddad (2019) that focused on spoken narrative text production by adult native speakers. Their results showed that the most frequent verbal pattern is the CaCaC followed by tCaCCaC and CaCCaC as the second most frequent patterns. The rest of the patterns are far less frequent, and some are hardly used at all. Two main conclusions may stem from these two studies: the distribution of the Palestinian Arabic verbal patterns is maintained over time and does not change from infancy to adulthood. Second, the three verbal pattern forms differ in size from the ten productive verbal patterns of the StA reported by Holes (2004). No other studies examined the Arabic derivational system. The current study, therefore, will be the first to investigate the acquisition of word-patterns and roots in the nominal system of the spoken language of Palestinian Arabic in northern Israel.

Hebrew is a Semitic language that is structurally similar to Arabic, and due to the lack of studies regarding the nominal system acquisition in Arabic, we looked into the acquisition of Hebrew derived nouns (HDNs) in the hope of gaining insight into this question. It was found that HDNs appear only occasionally in the language of those age 3. At age 8–9, the acquisition begins apace but is not completely established until age 15 (Mayrose, 1998, Ravid & Avidor, 1998). One explanation as to why the acquisition of HDNs is delayed until grade school is that they belong to the domain of derivational morphology, which is mostly non-linear and acquired later than the inflectional categories, which are mainly linear except for tense shifting in verbs. The inflection in Arabic, similar to Hebrew, also involves linear and non-linear procedures. However, the non-linear inflection procedure can be quantitatively more productive than the linear process, because, in addition to the tense shifting in verbs similar to Hebrew, it also appears in the most frequent pluralizing procedure in Arabic: namely, broken plural[2] (Boudelaa & Gaskell, 2002).

---

[2]*Broken Plural* (BP) is a simultaneous root-and-pattern affixation, that is, interdigitating consonant-vowel patterns on the root radicals of the singular noun. It involves changing the form of the singular noun through various morpho-phonological processes, such as long vowel insertion, consonant gemination, and the affixation of consonants besides those of the root (Holes, 2004).

Hence, this earlier exposure to the non-linear processes for broken plurals and temporal inflection alongside its existence in the derivational domain may lead to earlier emergence of derived nouns in Arabic relative to Hebrew.

Given the paucity of Arabic studies, we sought to check the expected developmental trajectory of the nominal system acquisition in Arabic, as stated in the Arab curriculum in Israel for literacy preparation in preschool (3- to 6-year-olds). According to the curriculum (Foundation of Reading and Writing in Arabic as a Mother Tongue: A Preschool Curriculum, Ministry of Education, 2008), the most pivotal aspect of development occurs in acquiring nouns and adjectives. Between 4 and 6 years of age, kindergarteners are expected to use common nominal patterns in different words, such as the patterns of: professions/occupations: CaCCa:C (e.g. /tˤˤabbax/ 'cook'), instruments: miCCa:C (e.g. /mifta:ħ/ 'key'), miCCaCa (e.g. /miknasat-un/ 'sweeper'), Ca:Cu:C (e.g. /ħa:su:b/ 'computer') and places: maCCaC (e.g. /maktab/ 'office'), maCCaCa (e.g. /madrasat-un/ 'school'). At 3 to 6 years of age, children can use passive participles /ʔism-l'mafˤu:l/ such as maCCu:C (e.g. /maktu:b/ 'written/letter'). This contributes to children's vocabulary expansion, as well as their appreciation of words in the same morphological family. Moreover, the gradual process of acquisition is seen in the creative errors of children, such as the use of /munaʒʒir/ instead of /naʒʒa:r/ 'carpenter'), which demonstrates an understanding of the meaning of the nominal pattern (professions), as well as the relationship between the root and the pattern.

The Acquisition of Palestinian and Standard Arabic

The most conspicuous hallmark of the Arabic language is diglossia (Maamouri, 1998) in which two varieties, the Spoken (or Colloquial) Arabic (SpA) and Modern Standard Arabic or Standard Arabic (StA), exist and are used side-by-side by the same speakers (Ferguson, 1959; Maamouri, 1998; Myhill, 2014). SpA is the variety that is used in daily conversation and everyday communication. It is not uniform across the Arabic-speaking world[3]; each region has its unique variation, and each geographical area has its characteristic vernacular. Also, Arabic vernaculars may vary substantially in proportion to their geographic distance from one another (Ryding, 2005). Palestinian Arabic (PA) is the original naturally-acquired mother tongue of approximately 1.8 million Palestinian Arab citizens of Israel. It is spoken to the Palestinian child from birth, with the child going through the usual stages of untaught early language acquisition (Laks & Berman, 2014). Thus, it is undoubtedly the primary spoken language, i.e., the first language.

Moreover, the vast majority of Israeli Arabs attend Arabic-medium schools and study all school subjects in Arabic (Saiegh-Haddad & Everatt, 2017). StA, in contrast, a modern descendant of Classical Arabic and Literary Arabic, is highly uniform across the entire Arabic world and is the only language variety that has a conventional written form. StA is typically not acquired as a first language. However, children are exposed to it as soon as they start watching satellite TV (in particular children's TV channels and dubbed series), begin attending nursery school, or when they pray. According to the new curriculum for literacy preparation in preschool (3 to 6 years old), most words at these ages are acquired from spoken language, but

---

[3]Arabic world from North Africa to the Arabian Gulf: Morocco, Algeria, Tunisia, Libya, Egypt, Sudan, Dibouti, Somalia, Mauretania, Jordan, Lebanon, Palestine, Syria, Iraq, Kuwait, Bahrein, Qatar, UAE, Oman, Saudi Arabia, Yemen, and Comoros.

also a significant contribution comes from the written StA – absorbed from storybooks and TV programs. Hence, native speakers of Arabic in Israel from an early age are actively and constantly engaged with StA before they start formal schooling (Badry, 2001; Saiegh-Haddad & Everatt, 2017).

Nonetheless, formal exposure starts at the onset of elementary school, when the child became 6 or 7 years old. Therefore, it is considered the primary language of literacy: namely, the language children are taught to read and write at school (Saiegh-Haddad & Everatt, 2017). Thus, native-speaking children are required to master the use of StA as a linguistic system, which differs greatly from their everyday experience as native speakers of Palestinian Arabic (Laks & Berman, 2014). Hence, linguistic proficiency in Arabic involves concurrent competence in using StA alongside proficiency in Palestinian Arabic (Saiegh-Haddad & Henkin-Roitfarb, 2014).

Linguistic Distance between Standard Arabic and Palestinian Dialect

The linguistic disparities between the written standard form and the spoken form have been referred to by Saiegh-Haddad (2003) as "linguistic distance" which is manifested in Arabic in all domains – semantics, phonology, morphology, and syntax (Abu-Rabia, Share & Mansour, 2003; Holes, 2004; Saiegh-Haddad, 2018). Notably, the degree of linguistic distance varies in each Arabic region, according to the distance from or proximity of each dialect from StA. This affects the degree to which the same StA word may be identifiable by children from different SpA backgrounds (Saiegh-Haddad & Spolsky, 2014). However, Boudelaa and Marslen-Wilson (2013) claimed that the dialects are different in similar ways from StA. A series of studies by Saiegh-Haddad (2003, 2004, 2005, 2007a; Saiegh-Haddad, Levin, Hende & Ziv, 2011) has examined the direct effect of phonological distance between Palestinian Arabic and StA on the development of phonological awareness in Arabic and early literacy acquisition.

The findings demonstrate that novel StA phonemes and their counterparts in the spoken dialects do not necessarily have identical phonemic inventories and are significantly more challenging to access than phonemes from a child's spoken dialect. For example, the StA voiceless uvular plosive /q/ is realized as a voiceless glottal plosive /ʔ/ in the northern Palestinian urban dialect. This phonological distance complicates the development of phonological awareness and likely poses challenges for learning to read in Arabic. Another phonological difference between StA and Palestinian Arabic relates to the architecture of the syllable. Saiegh-Haddad and Spolsky (2014) analyzed the spoken corpus of 5-year-old children who speak the central Palestinian dialect and the lexical basis of first and second-grade textbooks (which represents the StA corpus). Results established that the predominant SpA syllable structures were CVC (51.8%), followed by CCVC (26.8%). However, in StA, the most common syllable structures were CVCC (46%) and CVC (42%). Additionally, they examined the lexical distance between SpA and StA by counting the distribution of three types of words: 1) "unique words" – words that have distinct forms in SpA and StA, 2) "cognates" – words which keep partially overlapping phonological forms in SpA and StA, and 3) "overlapping words," which maintain their surface phonological form in SpA and StA. The results showed that the most common type of lexical items in the child's lexicon are cognates (40.6%), followed by unique words (38.2%), and finally identical words (21.2%). In essence, SpA and StA lexicons overlap only partially.

In recent years, the Arabic division in the Israeli Ministry of Education has mandated capitalizing on the spoken vernacular of children by leveraging literacy acquisition in StA. The new curriculum for literacy preparation in preschool (ages 3 to 6) relies on the assumption that the starting point in the development of basic literacy skills in Arabic is spoken language. Saiegh-Haddad and Everatt (2017) present the reasons for this assumption: a) StA and SpA share many structural linguistic features; b) the linguistic knowledge and representations that the child brings to the literacy learning task are almost wholly in SpA; c) some aspects of the linguistic reservoir that the child has in SpA can be used to leverage StA literacy acquisition; and d) some basic literacy skills in SpA may be conducive to the acquisition of literacy in StA (Foundation of Reading and Writing in Arabic as a Mother Tongue: A Preschool Curriculum, Ministry of Education, 2008, p. 9).

To date, the morphological knowledge in SpA that children bring to literacy acquisition, its distance from its corresponding forms in StA, and the variety that children are required to handle in their literacy acquisition have not been systematically studied. Although the basic interplay between root and word-pattern morpheme seems intact, several differences can be pointed out between StA and SpA. Boudelaa and Marslen-Wilson (2013) brought up some of the changes in the regional dialect of Southern Tunisia from StA related to the word-pattern morpheme, both at the level of form and function. They show that, at the form level, StA word-patterns, in many cases, undergo changes in Southern Tunisian Arabic, which is illustrated by reduced or deleted vowels and/or consonants. These changes also exist in PA relative to StA. For example, the StA word-pattern CaCi:C corresponds to the word-pattern CCi:C in PA (where the vowel /a/ has been omitted) as in /kabi:r/- /kbi:r/ 'big'. In addition to these reduced vowels, Holes (2004) refers to additional types of changes in word-patterns in several dialects relative to StA word-patterns; these too correspond to PA in terms of "regularizing" the doubled verb participle – such as in the form pattern I[4] double verb (e.g., /ħt̪ˤt̪ˤ/ 'put'). In PA, it has the derived form Ca:CiC (/ħa:t̪ˤit̪ˤ/'putting') rather than Ca:CC (/ħa:t̪ˤt̪ˤ/) in StA), which results in regularizing a morphological anomaly – such as in pattern I hollow verb passive participle of roots in which C2= /y/, such as /mabyu:ʕ/ compared to /mabi:ʕ/ in StA. In addition, some relational adjectives formed by attaching the termination -i:yy to a noun can be alternated with -a:ni such as /awwala:ni/'first' in PA rather than /awwali:yy/ in StA. Functionally, Boudelaa and Marslen-Wilson (2013) claimed that some of the word-patterns in Southern Tunisian Arabic differ significantly from their StA counterparts, which shares parallels with PA. The category CaCCa:n has been particularly productive in PA and functions as an active participle in verbs denoting temporary states that do not have the Ca:CiC word-pattern. However, a number of verbs have both types of participle, which usually have different types of referent, with CaCCa:n being predicated of animate beings and Ca:CiC of inanimate entities, abstract or concrete. Thus /b.r.d/ ('be cold') /barda:n/ is used to mean 'feeling cold' (by human beings), whereas /ba:rid/ means 'cold' as applied to the weather, drinks, or temperament (Holes, 2004, p.158). This study is a step toward exploring this distance in derivational morphology, focusing on nouns as one of the main lexical categories.

---

[4]Pattern I- The root pattern C1C2C3. it is an unaugmented verbal root, is the basis of the Arabic derivational morphology. The root becomes a Pattern I verb through the interdigitation of short vowels between it consonants. For example: /katab/, /daxal/.

In light of the above literature review, we investigated a) the process of acquisition of the nominal patterns and their distribution among 3- to 6-year-old native Arabic-speaking children; b) the developmental trajectory of the morphological structures (roots and word-patterns) in the nominal system; and c) the linguistic distance between nominal patterns in PA and their form in StA.

Method
Participants
Fifty-four native Arabic-speaking children from three age groups: ages 3 to 4, 4 to 5 (pre-kindergarten), and 5 to 6 (kindergarten) were recruited from three public northern Israel preschools (the Galilee region). All participants were children on track developmentally from average socioeconomic status (SES) who speak the regional dialect of the north (Galil) Palestinian Vernacular.

Procedure
The participants in the study were recorded for approximately two hours, during spontaneous conversations in different settings in the kindergarten's dining area and playground. The communication between the children in the dining area focused mainly on food, comparing their meals, teasing each other (ages 3–4), using different nicknames mostly for obese children (age 4–6). Other contexts were about the teacher, table manners (age 3–4), extra-curriculum activities (age 4–5), singing rhyme songs, and producing more peer conversations (ages 5–6). However, during the playtime, the children in all the age groups occupied themselves by creating a play, a process that was full of arguments/debates regarding the role of each player. The 3–4 pre-kindergarteners prepared a cake and had a birthday party, which was completed successfully and smoothly, while the older age children chose to run a play about night beasts and daemons.

Following informal meetings that allowed the children to become accustomed to the presence of the researcher, recording commenced. Subsequently, the recordings were transcribed, and the nouns in the corpora were identified and coded manually for the following nouns' classifications:[5]

1) NON-LINEAR DERIVED NOUNS: Nouns that draw on a non-concatenative word building principle, whereby morphemic units (root and word-pattern) are not linearly strung together one after the other, but intertwined with one other;
2) LINEAR DERIVED NOUNS: Nouns that draw on concatenative word building whereby morphemic units are linearly strung together one after the other;
3) NON-LINEAR AND LINEAR DERIVED NOUNS: Nouns that include non-linear derivation process followed by linear derivation;
4) PRIMITIVE NOUNS: The category of nouns that are not verbal derivatives (Holes, 2004); and
5) FOREIGN NOUNS: Nouns borrowed/loaned into Arabic.

Nouns tokens were counted as all occurrences of inflected and noninflected noun forms. The analysis disregards both inflectional and clitics structures of the nouns, to

---

[5]The nominal patterns that appear in spontaneous conversations among kindergarteners aged 3 to 6 years old were based on Holes (2004).

focus on noun lemmas and their derivational components – roots and nominal word-patterns. All nouns were coded manually according to their pattern, root and semantic class (e.g., verbal noun, active participle, noun of profession). Roots constitute different structural skeletons so that all nouns sharing the same consonantal root were considered as one root. Root types, strong or weak, as defined above (see page 5, 2nd paragraph). Noun types constitute noun lemmas: namely, unique combinations of roots and patterns yielding nouns. For example, the combination of the root *k-b-r* with the CCi:C pattern represents one noun type (/kbi:r/'big'), while the combination of the same root with the ?aCCaC pattern constitutes another type (/ʔakbar/'biggest').

Otherwise, the study examined the morphological distance between PA and StA at the level of form, in which the nominal word-patterns and roots were classified into five categories:

1) COMMON WORD-PATTERNS AND ROOTS (with/without phonological changes in roots' consonants), e.g., the word /ʔakbar/'bigger' appears without any phonological changes in roots' consonant. However, /ʔatʔal/ in PA and /ʔaθqal/'heavier' in StA have identical word-pattern ?aCCaC. However, the root /θ.q.l/ has undergone phonological changes in its first (C1) and second (C2) consonants, in which the /θ/ and /q/ in StA have been replaced, respectively by /t/ and /ʔ/;

2) COMMON WORD-PATTERNS WITH DIFFERENT ROOTS, e.g. /ʔaħla:/ in PA and /ʔaʒmal/ 'most beautiful' in StA have identical word-pattern ?aCCaC but different roots /ħ.l.y/ in PA versus /ʒ.m.l/ in StA;

3) PARTIALLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS (with/without phonological changes in roots' consonants), e.g., /kb:ir/ in PA corresponds to /kabi:r/ 'big' in StA, CCi:C in PA versus CaCi:C in StA without any phonological change in the root's consonants. However, /kti:r/ in PA corresponds to /kaθi:r/ in StA *'many'*, CCi:C in PA versus CaCi:C in StA and with phonological change in the root's second consonant, /θ/ in StA versus /t/ in PA;

4) FULLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS, *e.g.* /ʒuʕa:n/ in PA correspond to /ʒa:ʔiʕ/ in StA 'hungry', CuCa:C in PA versus Ca:CiC in StA; and

5) Different word-patterns with different roots, e.g., /xitya:r/ in PA versus /ʕaʒu:z/ in StA 'old man' CiCCa:C (with quadriliteral root) in PA versus CaCu:C in StA).

The final step was to examine the lexical distance between PA and StA by tallying the distribution of three categories of nouns: IDENTICAL NOUNS (words that have identical forms in the two varieties), COGNATE NOUNS (words which share partially overlapping phonological forms in PA and StA) and UNIQUE NOUNS (nouns that have totally distinct forms in PA and StA).

Results

The first aim of the study was to investigate the acquisition of the nominal system among 3- to 6-year-old native Arabic-speaking children. The database of two hours, spontaneous conversations of children (aged 3–4, 4–5 and 5–6) in the kindergarten's dining area and playground, yielded a total of 4289 word tokens and 614 noun tokens accounting for 14.32% of all word tokens. Broken down, 5- to 6-year-old kindergarteners produced 2240 total word tokens and 320 noun tokens (14.29% of

Table 1. Distribution of nouns by frequencies (N) and percentages (%) among kindergarteners ages 3–6

| | Group Age | | | | | | | |
| Type of nouns | 3–4 | | 4–5 | | 5–6 | | Total | |
	N	%	N	%	N	%	N	%
Non-linear derivation	47	47	94	48.4	163	50.9	304	49.5
Primitive nouns	46	46	89	45.8	130	40.6	265	43.1
Foreign nouns	5	5	8	4.1	17	5.3	30	4.8
Non-linear and linear derivation	2	2	3	1.5	8	2.5	13	2.1
Linear derivation	0	0	0	0	2	0.6	2	0.3
Total	100	100	194	100	320	100	614	100

all word tokens), whereas the 4- to 5-year-old kindergarteners produced 1750 total word tokens and 194 noun tokens (11.09% of all word tokens) and those 3 to 4 years of age created the 499 word tokens and 100 noun tokens (20.04% of all word tokens), which constitutes the highest portion of noun tokens between the other two elder age groups.

The distribution of nouns in the speech of kindergarteners ages 3 to 6 was categorized into five groups (see Table 1): non-linear derived nouns, linear derived nouns, non-linear and linear derived nouns, primitive nouns, and foreign nouns. The results revealed that the most common type of noun used was the NON-LINEAR DERIVED NOUNS (e.g.,/maksu:r/ 'broken' with the root /k.s.r/ and the word-pattern maCCu:C), making up 49.5% of the nouns in the speech of the kindergarteners. The second most common category was PRIMITIVE NOUNS, comprising 43.1% of nouns (e.g., /ʔab/ 'father'). It is important to emphasize that the analysis of this category adhered to the method we used to analyze the deverbal nouns– by roots and word-patterns. However, in several nouns ranging between 2%-3%, we were unable to determine their roots or word-patterns (e.g., /zaʕtar/ 'thyme', /tum/ 'mouth').

The FOREIGN NOUNS category (e.g., /belefon/ 'moblie') covered 4.8% of the total, while NON-LINEAR AND LINEAR DERIVED NOUNS (e.g., /ħasa:si:yyi/ 'sensitivity' with the root /ħ.s.s./ and the word-pattern CaCCaC intertwined with one other and with the morpheme /:yyi/ attached to them linearly at the end) made up 2%. LINEAR DERIVED NOUNS (e.g., /dahabi/ 'golden' with the primitive noun /dahab/ and the morpheme /i:/ attached to it linearly), making up 0.3% of the nouns, hardly exist in the speech of kindergarteners ages 5 to 6, whereas kindergarteners aged 3 to 4 and 4 to 5 do not use nouns in this category at all.

A chi-square test of independence was performed to examine the relationship between age groups and the use of the different categories of nouns. The analysis produced a non-significant relationship ($p = 0.92 > 0.05$), confirming that the distribution of noun types does not change with age, only the overall frequency.

The second aim of the study was to investigate the developmental trajectory of roots and word-patterns in the Arabic nominal system. The objective was to describe, analyze, and quantify the morphological structures (roots and patterns) in the Arabic nominal system:

Table 2. Distribution of strong and weak roots by frequencies (N) and percentages (%) among kindergarteners ages 3–6

| | Age Group | | | | | | | |
| | 3–4 | | 4–5 | | 5–6 | | Total | |
Root type	N	%	N	%	N	%	N	%
Strong root	32	69.6	64	71.1	114	76	210	73.4
Weak root	14	30.4	26	28.9	36	24	76	26.6
Total	46	100	90	100	200	100	336	100

The distribution of the roots in the nominal system among kindergarteners aged 3 to 6 was classified by their type, i.e., strong or weak. Table 2 presents the frequencies and percentages of each type within the three age groups, as well as the total frequency and percentages. As shown, strong roots (e.g., /ʕ.l.m/ 'flag' (were the most prevalent, making up 73.4% of the types of the total roots, whereas weak roots (e.g., /n.w.m/ 'sleeping') comprised just over the remaining one-fourth (26.6%) of the distribution.

A chi-square test of independence was performed to examine the relationship between age and root type. The analysis produced a non-significant relationship ($p = 0.57 > 0.05$) indicating, once again, that the 3:1 distribution of strong to weak roots was stable across age (see Figure 1).

Table 3 shows the nominal patterns that appear in the spontaneous conversations of preschool children:

The analysis of the distribution of nominal word-patterns by their semantic class and types among kindergarteners aged 3 to 6 yielded several main findings: as shown in Figure 2, the most frequently used is that of the VERBAL NOUNS /ʔalmaṣˁdar/ which refer either to the actual occurrence of an action or to the possibility of its occurrence, making up 27.2% of the distribution with the word-patterns CaCC, CaCaC, CaCa:C, and CCu:Ci. The second most frequent nominal pattern is the ACTIVE PARTICIPLES /ʔismu-lfaʕil/ which describes the state in which the subject of the verb from which it is derived finds itself as a result of the action or event that the verb describes, with the word-patterns Ca:CiC, muCaCCiC, and mitCa:CiC, making up 17.8%. This is followed by the nominal word-patterns of NOUNS/ADJECTIVES DESCRIBING STATES/QUALITIES: CCi:C, CiCiC, CaCi:C, and CaCCa:n comprise 17.1% of the distribution. In addition to these most common categories, there are INSTANCE NOUNS /ʔisimu-l- marrati/, which are formed by adding the feminine suffix –a to the verbal nouns with the word-patterns CaCCi, CaCCa or the collective nouns with the word-patterns CaCaCa and CaCaCi. They comprise 8.1% of the distribution, followed by ADJECTIVES THAT EXPRESS A COMPARATIVE RELATIONSHIP with the nominal word-pattern ʔaCCaC (7.7%), and PASSIVE PARTICIPLES, which refer to the state in which the object or completion of the verb from which it is derived finds itself after completion of the action/event, with the nominal word-patterns maCCu:C and muCaCaC, which comprise 6.7% of the distribution. These were followed by NOUNS OF PLACE AND TIME, making up 5% with the patterns maCCaC and maCCaCi, and ABSTRACT NOUNS which are formed by attaching the termination -iyyi to a nominal base, making up 4.4% of the distribution. The least commonly used is the NOUNS
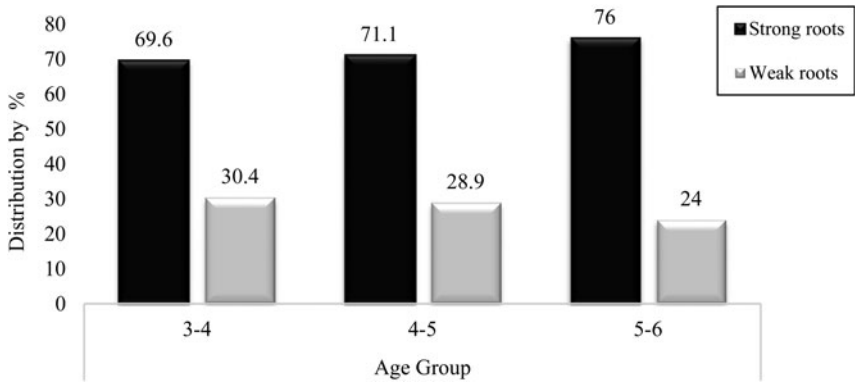
Figure 1. Distribution of the roots in the nominal system among kindergarteners ages 3–6

Table 3. The nominal word-pattern by semantic class and types that appear in spontaneous conversations among kindergarteners ages 3–6

| Semantic class	Nominal word-pattern in PA	Example
Verbal noun	CaCC CaCaC CaCa:C CCu:Ci	/ʕaks/ 'against' /habal/ 'foolishness' /ʔasˤasˤ/ 'punishment' /sxu:ni/ 'warmth'
Passive participles	maCCu:C muCaCaC	/maksu:r/ 'broken' /muxayyam/ 'camp'
Active participles	Ca:CiC muCaCCiC mitCa:CiC	/sˤa:ħib/ 'friend' /mʕallim/ 'teacher' /mitdˤa:yiʔ/ 'upset'
Adjectives that express comparative relationship	ʔaCCaC	/ʔakbar/ 'bigger'
Nouns /adjectives describing states/qualities	CCi:C CiCiC CaCi:C CaCCa:n	/kti:r/ 'many' /ħiliw/ 'sweet/nice' /sari:ʕ/ 'fast' /malya:n/ 'full', /xarba:n/ 'rotten'
Nouns of profession and Inanimate nouns of instruments	CaCCa:C CaCCaCa	/kazza:b/ 'liar' /sayya:ra/ 'car'
Nouns of place and time	maCCaC maCCaCi	/masraħ/ 'theatre' /madrasi/ 'school'
Instance nouns	CaCCi CaCCa CaCaCa CaCaCi	/dˤarbi/ 'a blow' /natˤtˤa/ 'a jump' /ʃajara/ ' a tree' /samaki/ 'a fish'
Nouns of instruments	maCCa:C maCCaCa miCCaCi	/mħħa:i/ 'eraser' /maʕlaʔa/ 'spoon' /miħrami/ 'handkerchief'
abstract nouns	nominal base+ iyyi	/ħasa:siyi/ 'sensitivity'
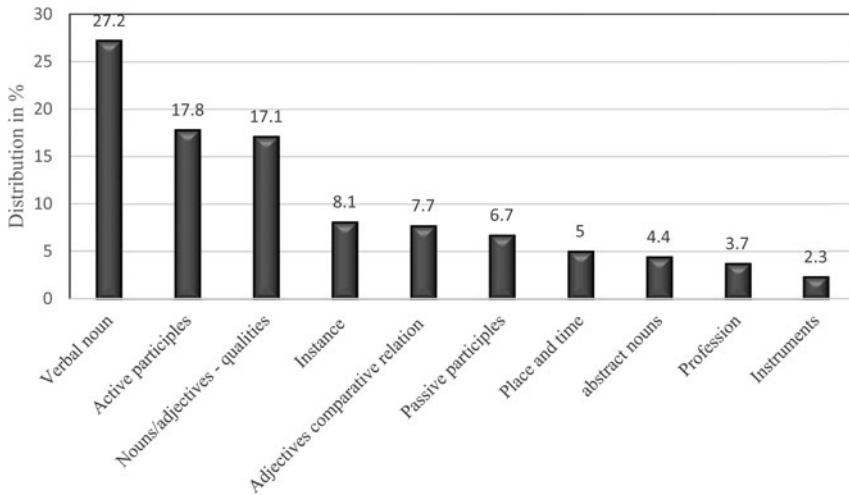
Figure 2. The total distribution of nominal word-patterns by percentages among kindergarteners ages 3–6

PROFESSION/OCCUPATION with the word-patterns CaCCa:C and CaCCa:Ca, making up to 3.7% and NOUNS OF INSTRUMENTS with the patterns maCCa:C, maCCaCa, and miCCaCi, making up 2.3% of the distribution among the three age groups.

The analysis of the distribution between the three age groups indicates a common finding between the two older groups, those aged 4 to 5 and 5 to 6, as compared to the youngest group 3 to 4 years old. Figure 3 shows that the VERBAL NOUNS and NOUNS/ADJECTIVES DESCRIBING STATES/QUALITIES are the most frequent nominal word-patterns with the same prevalence 19.6% in the speech of kindergarteners ages 3 to 4 years old, followed by the ACTIVE PARTICIPLES and ADJECTIVES THAT EXPRESS A COMPARATIVE RELATIONSHIP (with the same frequency at 15.2%), and succeeded by PASSIVE PARTICIPLE patterns at 13%. Among the other two age groups, on the other hand, ADJECTIVES THAT EXPRESS A COMPARATIVE RELATIONSHIP represent only 7.5%-7.7% of the distribution, and the PASSIVE PARTICIPLE patterns represent 4.3%–6.3%. The VERBAL NOUNS are the most frequent pattern in the speech of the two groups of older kindergarteners, ages 4 to 5 (where it is 31%) and 5 to 6 (where it is 26.9%). The ACTIVE PARTICIPLES are the second most frequent nominal pattern among these two groups, 20.7% among ages 4 to 5 and 16.9% among ages 5 to 6. Moreover, NOUNS/ADJECTIVES DESCRIBING STATES/QUALITIES account for 16.3% of the distribution on nominal word-patterns among kindergarteners 4 to 5 years old, and 16. 9% among kindergarteners 5 to 6 years old.

The results show a progression in using the nominal word-patterns of nouns of instance and nouns of instruments. INSTANCE NOUNS were used with increasing frequency as children were older, such that 6.5% demonstrated a distribution of nominal patterns among the age groups 3 to 4 and 4 to 5 years old, but 9.4% of those 5 to 6 years old demonstrated the same. However, the NOUNS OF INSTRUMENTS were only somewhat present in the speech of kindergarteners aged 4 to 5, making up 2.2%, and in the speech of children 5 to 6, at 3.1%, but not appearing at all in the speech of kindergarteners aged 3 to 4.
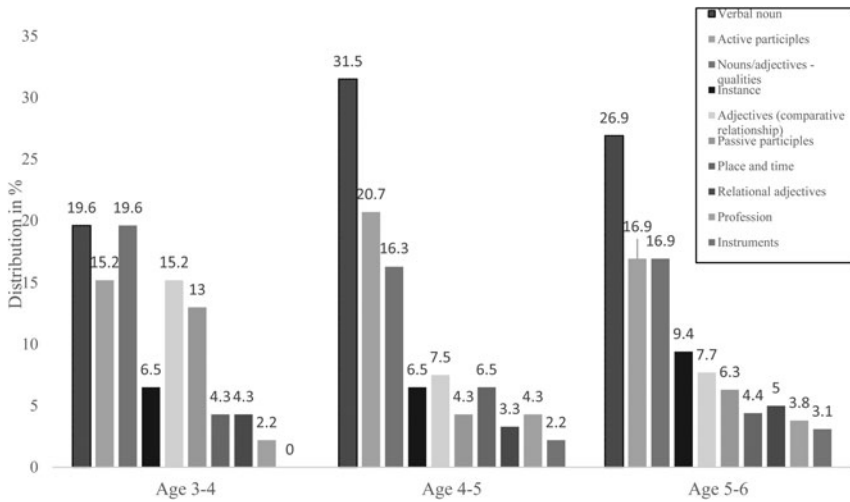
Figure 3. The distribution of the nominal word-patterns within the three age groups

A chi-square test of independence was performed to examine the relationship between age groups and the different nominal patterns. Yet again, this test produced a non-significant relationship ($p = 0.71 > 0.05$), reaffirmed the stability of form distribution across age.

The final objective of the study was to examine the linguistic distance between PA and StA. The study investigated the morphological distance between PA and StA across five categories of nominal word-patterns and roots. Figure 4 shows that COMMON WORD-PATTERNS AND ROOTS category (with/without phonological changes in roots' consonants) is the most frequently used overall, comprising 47.6%, (e.g.,/ ʔatʔal/ in PA and /ʔaθqal/'heavier' in StA with the word-pattern ʔCCaC). The second most frequently used category is that of PARTIALLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS (with/without phonological changes in roots' consonants), making up 29.1% (e.g., /kti:r/ with the word-pattern CCiC in PA vs. /kaθi:r/ with the word-pattern CaCCiC) These categories were followed by DIFFERENT WORD-PATTERNS WITH DIFFERENT ROOTS making up 13% of the distribution (e.g.,/xitya:r/ in PA vs. /ʕaʒu:z/ in StA' old man' (and 7.5% were FULLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS (e.g., /ʒuʕa:n/ in PA vs. /ʒa:ʔiʕ/ in StA' hungry'). However, COMMON WORD-PATTERNS WITH DIFFERENT ROOTS category (e.g., /ʔaħla:/ in SpA vs. /ʔaʒmal/ in StA' most beautiful' with the pattern ʔaCCaC) was produced relatively rarely, making up only 2.7%.

The analysis of the distribution, as presented in Figure 4, revealed several main findings in the three age groups. COMMON WORD-PATTERNS AND ROOTS category (with/ without phonological changes in roots' consonants) is the most frequent in the speech of the three age groups (3 to 4 years: 50%, 4 to 5 years: 44.5%, and 5 to 6 years: 48.7%), PARTIALLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS category is the second most frequent among the three age groups (3 to 4 years: 29.5%, 4–5 years: 31.5% and 5–6 years: 27.6%), followed by DIFFERENT WORD-PATTERNS WITH DIFFERENT ROOTS (3–4 years: 11.4%, 4–5 years: 10.9% and 5–6 years: 14.7%). However, FULLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS category also accounts for 10.9% of the
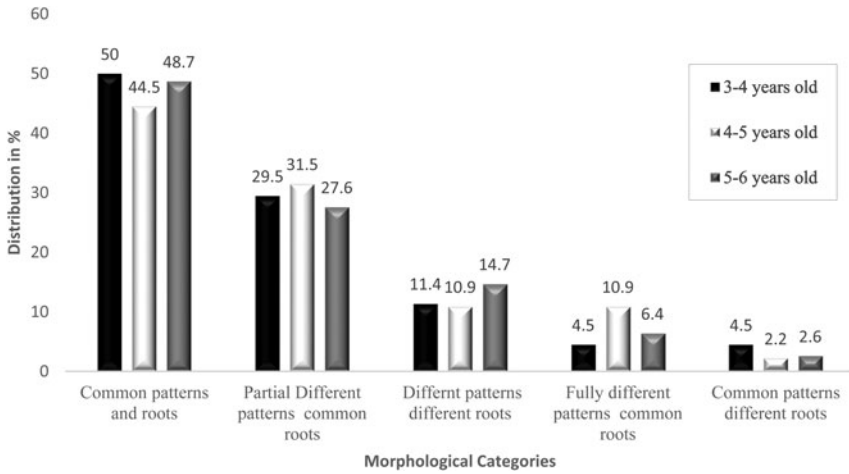
Figure 4. The distribution of morphological distance categories between PA and StA among Arab kindergarteners ages 3–6
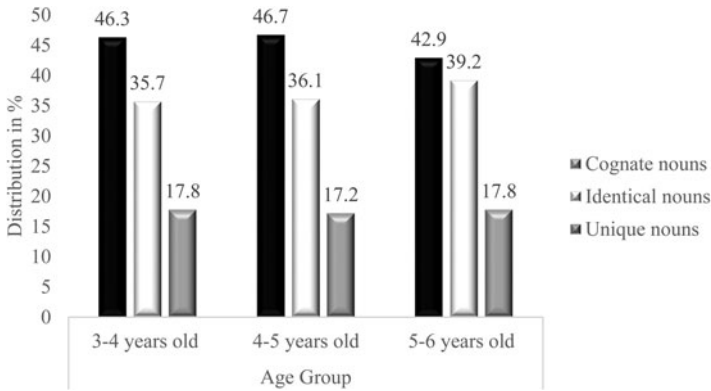

Figure 5. The distribution of lexical categories between PA and StA among kindergartners ages 3–6

distribution among kindergarteners aged 4 to 5 years old. Besides, among the three age groups, COMMON WORD-PATTERNS WITH DIFFERENT ROOTS category (3 to 4 years: 4.5%, 4 to 5 years: 2.2%, and 5 to 6 years: 2.6%) is relatively rare, as is FULLY DIFFERENT PATTERNS WITH COMMON ROOTS category (4.5%) among the youngest group.

A chi-square test of independence was performed to examine the relationship between age and morphological distance. The test was non-significant ($p = 0.84 > 0.05$).

Finally, the current study examined the lexical distance between PA and StA in the spontaneous speech of this sample.

According to Saiegh-Haddad and Spolsky (2014), lexical distance is classified according to three categories: identical nouns (nouns that have a similar lexical form in both Standard and Spoken Arabic), cognate nouns (nouns which are used in both Standard and spoken Arabic but have different surface phonological forms on the

two varieties), and unique nouns (nouns that only exist in spoken Arabic and are not used in Standard Arabic and do not have, therefore, a conventional written form). The analysis of our current corpus, according to Saiegh-Haddad and Spolsky's three categories, as presented in Figure 5, revealed that the most frequent category in all three groups was cognate nouns, making up 44.6% (e.g., -/mʕallim/ = /muʕallim/ teacher) among the three different groups. The second most frequent category was identical nouns, comprising 37.6%, whereas only 17.6% of the spoken nouns were unique.

A chi-square test of independence examining the relationship between age and lexical distance was non-significant (p = 0.93 > 0.05).

Discussion

This study investigated the developmental trajectory of the morphological structures of Arabic nouns (roots and word-patterns) as reflected in the language corpus of native Arabic-speaking 3- to 6-year-old children in the regional dialect known as Palestinian Arabic in the north of Israel. The first aim of the study was to investigate the acquisition of the nominal system among 3- to 6-year-old Arabic-speaking children. The results showed that the older kindergarten group (aged 5 to 6) produced over three times more nouns (n = 320) than the youngest group (aged 3 to 4) (n = 100). This finding is in accordance with the Arab curriculum in Israel for literacy preparation in preschool (3- to 6-year-olds), which indicates that between ages 4 and 6 kindergarteners are expected to demonstrate awareness of common nominal patterns in different words and that the most important aspect of development occurs in nouns and adjectives. However, the highest proportion of the nouns out of the total word tokens was obtained among those of 3–4 years of age (20.04%) relative to the 4 to 5 (11.09%) and 5–6-year-old kindergarteners (14.29%). This finding is in line with Tarabani's (2006) findings that children produce more verbs with increasing age.

The distribution of the nouns in the speech of kindergarteners aged 3 to 6 was across five different categories: NON-LINEAR DERIVED NOUNS, LINEAR DERIVED NOUNS, NON-LINEAR AND LINEAR DERIVED NOUNS, PRIMITIVE NOUNS, and FOREIGN NOUNS. The results showed that while the most frequently used category is that of NON-LINEAR DERIVED NOUNS (e.g., /maksu:r/ 'broken' with the root /k.s.r/ and the pattern maCCu:C), making up 49.5% of nouns, PRIMITIVE NOUNS which are also traditionally analyzed in the same way as the non-linear derived nouns (Boudelaa & Marslen-Wilson, 2015), are almost as common, comprising 43.1% of the nouns (e.g., /ʔab/ 'father'). On the other hand, LINEAR DERIVED NOUNS barely exist before children start school, making up 0.3% in the speech of kindergarteners aged 5 to 6 years. In contrast, kindergarteners aged 3 to 4 and 4 to 5 do not use nouns in this category at all. The remaining categories were FOREIGN NOUNS, which comprised 4.8% of the total, while 2% of the nouns produced were NON-LINEAR AND LINEAR DERIVED NOUNS (e.g., /ħasa:siyyi/ 'sensitivity'). These findings show that almost 90% of nouns are built according to non-linear root-plus-pattern word formation, whereas linear derivation constitutes only 0.3% of the nouns in the speech of kindergarteners aged 5–6 years and does not exist before this age. The preference for the non-linear word formation in preschool peer-talk over linear concatenation is in line with findings from Hebrew, which reveal non-linear formation as the basic, earliest-acquired structure of content words relative to linear concatenation, which emerges as a later device (Ben-Zvi & Levie,

2016). Although the linear concatenation is more transparent and appears structurally simpler than root-plus-pattern derivation and is expected to be easier to learn (Clark & Berman, 1984) in some sense, it is more complex. According to Ravid (2006), these are sophisticated nouns acquired mostly through written language and media. They rely on a pool of basic words expanded by suffixation with abstract meanings. Manipulating the linear lexicon in production requires a large inventory of items. It also requires precise knowledge about stem types and the way they changed under suffixation alongside the categorical functions of these linear suffixes. Hence, linguistic knowledge of the Palestinian Arabic-speaking children is affected by the distributional properties of the language, which correspond with the non-linear root-plus-pattern default formation. That is to say, Palestinian Arabic-speaking children start their morpho-lexical journey with the Semitic tool of non-linear root-and-pattern morphology. This tool constitutes a major typological factor of organizing the early Arabic nominal lexicon, which underlies word learning in the language.

One of the most exciting and interesting findings is related to PRIMITIVE NOUNS, which are almost as common as the non-linear derived nouns, comprising 43.1%. Our analysis of these nouns adhered to the method we used to analyze the deverbal nouns – by roots and word-patterns, even though the contribution of the word-pattern to the meaning is much less constrained (Boudelaa & Marslen-Wilson, 2015). Nonetheless, for 2%-3% of these nouns, we were unable to determine their roots and/or word-patterns (e.g., /zaʕtar/ 'thyme', /tum/ 'mouth'). These two issues, of which the word-pattern of most of its words has minimal contribution to the meaning together with part of its nouns with no internal structure, raise the possibility that this large category falls somewhere "in-between" the two competing views of the Semitic morphology, and can be described as a category moving along a continuum axis of productivity. This continuum begins with isolated nouns with low-productivity that do not belong to a morphological family, up to derived nouns (with roots and patterns) with high productivity. A similar line of argument has been obtained from four experiments by Velan and Frost (2011) using rapid serial visual presentation (RSVP), transposed-letter (TL) priming, and form-priming manipulations. They contrasted the processing of Hebrew words, which have an internal root-and-pattern structure and consist of typical non-linear derived words, with Hebrew words that are morphologically simple with no internal structure and resemble base-words in European languages. Words with root and pattern structure do not reveal form-priming or TL priming effects, suggesting that these words are lexically organized by neighborhoods of root morpheme. In contrast, morphologically simple words revealed form and TL priming effects, suggesting that they are lexically organized by orthographic neighborhoods. Their findings indicate that Hebrew readers display remarkable flexibility and hint at two parallel lexical systems. Future research will need to analyze these primitive nouns based on both views of Semitic morphology and explore whether these nouns are stem-based units or are structured by intertwining a word-pattern with a consonantal root.

Most of the roots in the Arabic language are strong roots, and only 10% are weak roots (Boudelaa & Marslen-Wilson, 2004). The analysis of the root system in the speech of kindergarteners aged 3 to 6 years old in our study confirmed the dominance of strong roots, making up 73.4% of the distribution, but also indicated that weak roots are not rare, comprising just over one-fourth of the distribution. This finding shows that native Arabic-speaking children can handle variability in morphological input from early ages (3 years of age).

Analysis of the distribution of the nominal patterns revealed that the VERBAL NOUNS with the word-patterns CaCC, CaCaC, CaCa:C and CCu:Ci were the most frequent making up 27.2% followed by ACTIVE PARTICIPLES with the word-patterns Ca:CiC, muCaCCiC, and mitCa:CiC, comprising 17.8%. As well as by NOUNS/ADJECTIVES DESCRIBING STATES/QUALITIES, making up 17.1% with the word-patterns CCi:C, CiCiC, CaCi: C, CaCCa:n. In addition to these most common categories, INSTANCE NOUNS (with the word-patterns CaCCi, CaCCa, CaCaCa, and CaCaCi) comprise 8.1% of the distribution, as well as ADJECTIVES THAT EXPRESS A COMPARATIVE RELATIONSHIP with the nominal word-pattern ʔaCCaC making up 7.7%, and PASSIVE PARTICIPLES, with the nominal word-patterns maCCu:C and muCaCaC, comprising 6.7% of the distribution. These were followed by NOUNS OF PLACE AND TIME, making up 5% with the patterns maCCaC and maCCaCi, and ABSTRACT NOUNS, which are formed by attaching the termination -iyyi to a nominal base making up 4.4% of the distribution. The least commonly used are the NOUNS OF PROFESSION/OCCUPATION with the word-patterns CaCCa:C and CaCCa:Ca, making up 3.7% *and* NOUNS OF INSTRUMENTS with the patterns maCCa:C, maCCaCa, and miCCaCi, making up 2.3% of the distribution among the three age groups. Examination of these distributions across the three age groups indicates that, remarkably, the category frequencies (as percentages) did not change across age, only in their overall frequency. It might be that there are no developmental changes in the structure of the system, but rather an increase in the number of nouns. This finding, however, seems to be a result of working with a small corpus. Another possible explanation for the lack of development is that the early stages of the acquisition of the nominal system of PA in a category-specific morphological system can be seen to derive from the early lexical acquisition and to yield generalizations that can aid in the learning of new words (Ravid, 1995a). This finding is in line with the distribution of the Palestinian Arabic verbal patterns that are maintained over time and do not change from infancy to adulthood (Tarabani, 2006; Laks *et al.*, 2019).

The ultimate purpose of this study was to investigate the linguistic distance between PA and StA. The study, as noted above, is a first step in exploring the morphological distance between SpA and StA. It investigated five categories of patterns: 1) COMMON WORD-PATTERNS AND ROOTS 2) COMMON WORD-PATTERNS WITH DIFFERENT ROOTS, 3) PARTIALLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS, 4) FULLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS, and 5) DIFFERENT WORD-PATTERNS WITH DIFFERENT ROOTS. The results indicate that nouns with COMMON WORD-PATTERNS AND ROOTS (e.g., PA: /ʔatʔal/ vs. StA: /ʔaθqal)/ 'heavier) account for approximately half of the nouns in the speech of kindergarteners aged 3 to 6 years (47.6%). The second most frequent category is the PARTIALLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS (29.1%), which reflects only minor distance between the two forms of word-patterns (e.g. PA: /kti:r/ *vs.* StA: /kaθi:r/ 'many'). Additionally, 13.1% of the nouns are nouns with DIFFERENT WORD-PATTERNS WITH DIFFERENT ROOTS (e.g.,/xitya:r/ in PA vs. /ʕaʒu:z/ in StA 'old man' (and 7.5% are nouns with FULLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS (*e.g.* PA: /ʒuʕa:n/ vs. StA: /ʒa:ʔiʕ/ 'hungry'). This important finding may lead to two opposing conclusions: if the "PARTIALLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS" category belongs more to the "common" category than "different" categories, this means that close to 77% of the nominal word-patterns in preschooler's speech have an equivalent in StA. If this assumption is true, it may attest that SpA (here PA) and StA are more similar than dissimilar. Hence, this conclusion will be consistent with Holes (2004), who claims that the word-patterns in spoken Arabic

represent a subset of those of the StA system. Similar claims had been made that the distance between spoken and standard Arabic may be larger in the phonological than the morphological domains (Saiegh-Haddad & Taha, 2017; Schiff & Saiegh-Haddad, 2018).

The other possible conclusion leans on the opposite assumption, i.e., that the "PARTIALLY DIFFERENT WORD-PATTERNS WITH COMMON ROOTS" category belongs more to "different" categories than the "common" category. This assumption leads to the conclusion that approximately 50% of the nominal word-patterns in preschoolers' speech are dissimilar from StA system, thus affirming that PA and StA are more lexically dissimilar. Hence, contrary to Holes (2004), a substantial number of the nominal word-patterns are not a subset of the StA system. This conclusion may further illuminate the extent of the spoken/written divide in Arabic. It will be consistent with the study of Boudelaa and Marslen-Wilson (2013), which shows that there are differences between MSA and STA (Southern Tunisian Arabic) in the morphological domain. Yet, it remains to be established whether or not this "PARTIALLY DIFFERENT word-patterns (between PA versus StA)" category is actually perceived by young children learning to read as variants of each other. The answer is related to children's phonological sensitivity or memory and their ability to restrain these differences between the different nominal word-patterns.

Finally, our study examined the lexical distance between PA and StA. The lexical distance was distinguished by Saiegh-Haddad and Spolsky (2014) according to three categories of nouns: IDENTICAL, COGNATE, and UNIQUE NOUNS. Among all three age groups, the most frequently used category was that of COGNATE NOUNS, making up 44.6% of the distribution, followed by IDENTICAL NOUNS making up 37.6%. These findings attest that by a vast majority PA lexemes are either identical or near-identical to the StA lexicon. On the other hand, 17.6% of the PA forms are UNIQUE. Once again, therefore, this indicates substantial overlap alongside a non-trivial proportion of unique non-overlapping items that do not constitute a subset of the StA lexicon. These current findings are partly consistent with Saiegh-Haddad and Spolsky's (2014) findings for 5-year-old children who speak the central dialect of Palestinian Arabs living in Israel. They found that the most common type of lexical item in the child's lexicon is the cognate (40.6%), followed by the unique word (38.2%), and finally the identical word (21.2%). These differences may point to regional variation but may also be a result of different methods of data analysis. It is noteworthy that, although the most frequent category of nouns in the current study was that of COGNATES, thereby confirming a certain degree of lexical distance between nouns in PA and StA, the second most frequent category was IDENTICAL NOUNS, implying that the lexicons of nouns in SpA and StA substantially overlap.

This study may contribute to apprising educational practice and policy in Arabic. This is because it deepens understanding about the linguistic reservoir that children bring to school and hence the ability to make informed decisions about language and literacy instruction. This information should help in designing any scientifically sound, evidence-based language/reading education program, and developing curricula for the preschool Arabic-speaking population. More direct and specific steps may be trying to raise the awareness of preschoolers about the distance between PA nominal word-patterns and their correspondence in StA. This awareness could reduce the impact of these diglossic structures in the acquisition of basic reading processes in first grade.

Additional suggestions may be updating the preschool curriculum by adding nominal word-pattern structures (/ʔalmasˤdaru/ – VERBAL NOUNS and /ʔassifatu-l-muʃabbahatu/ – ADJECTIVES DESCRIBING STATES/QUALITIES), which our study shows as having the highest frequency in word-patterns but are currently not mentioned at all in the preschool curriculum.

Finally, it should be noted that this study included a respectable sample of 54 kindergarteners, but that its total corpus was not large. A larger sample and more recordings from a wider variety of settings would help establish the reliability of its findings. Besides, the sample included children who speak a single dialect only and may not be generalized to other dialects.

References

Abu-Rabia, S., Share, D., & Mansour, M. S. (2003). Word recognition and basic cognitive processes among reading-disabled and normal readers in Arabic. *Reading and Writing: An Interdisciplinary Journal*, *16*(5), 423–442.

Badry, F. (2001). *Acquiring the Arabic Lexicon: Evidence of Productive Strategies and Pedagogical Implications*. Palo Alto, CA: Academic Press, LLC.

Bat-El, O. (2017). Word-based items-and processes (WoBIP): Evidence from Hebrew morphology. In C. Bowern, L. Horn, and R. Zanuttini (eds), *On Looking into Words (and beyond)*, 115–135. Berlin: Language Sciences Press.

Benmamoun, E. (2003). The role of the imperfective template in Arabic morphology. *Language Acquisition and Language Disorders*, *28*, 99–114.

Ben-Zvi, G., & Levie, R. (2016). Development of Hebrew derivational morphology from preschool to adolescence. *Acquisition and development of Hebrew: From infancy to Adolescence*, 135–173.

Bohas, G., & Guillaume, J.-P. (1984). *Etudes des theˊories des grammairiens Arabes*. I Morphologie et Phonologie. Damascus, Syria.

Boudelaa, S. (2014). Is the Arabic mental lexicon morpheme-based or Stem-based? Implications for Spoken and Written word recognition. In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic Literacy* (pp. 31–54). New York: Springer.

Boudelaa, S., & Gaskell, M. G. (2002). A re-examination of the default system for Arabic plurals. *Language and cognitive processes*, *17*(3), 321–343.

Boudelaa, S., & Marslen-Wilson, W. D. (2004). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language*, *90*(1–3), 106–116.

Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes*, *20*, 207–260.

Boudelaa, S., & Marslen-Wilson, W. D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes*, *26*(4–6), 624–652.

Boudelaa, S., & Marslen-Wilson, W. D. (2013). Morphological structure in the Arabic mental lexicon: Parallels between standard and dialectal Arabic. *Language and Cognitive Processes*, *28*, 1453–1473.

Boudelaa, S., & Marslen-Wilson, W. D. (2015). Structure, form, and meaning in the mental lexicon: Evidence from Arabic. *Language, cognition and neuroscience*, *30*(8), 955–992.

Clark, E. V., & Berman, R. A. (1984). Structure and use in the acquisition of word formation. *Language*, *60*, 90–542.

Cohen, M. (1951). Langues chamito-sémitiques et linguistique historique. *Scientia*. 86, 304–310.

Deutsch, A., Frost, R., & Forster, K. (1998). Verbs and nouns are organized and access differently in the mental lexicon: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1238–1255.

Deutsch, A., & Malinovitch, T. (2016). The role of the morpho-phonological word pattern unit in single-word production in Hebrew. *Journal of Memory and Language*, *87*, 1–15.

Deutsch, A., Velan, H., & Michaly, T. (2018). Decomposition in a non-concatenated morphological structure involves more than just the roots: Evidence from fast priming. *Quarterly Journal of Experimental Psychology*, Vol. *71*(1) 85–92.

Ferguson, C. A. (1959). Diglossia. *Word*, *14*, 47–56.

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 829–856.

Heath, J. (2003). Arabic derivational ablaut, processing strategies, and consonantal "roots". In J. Shimron (ed), *Language processing and acquisition in languages of Semitic, root-based, morphology*, (pp. 100–115). Amsterdam: John Benjamins Publishing.

Hilaal, Y. (1990). Deriving from roots and word patterns. *Linguistica Communicatio*, 1, 77–80, (in Arabic).

Holes, C. (2004). *Modern Arabic: Structures, functions and varieties*. Washington D.C. : Georgetown University Press.

Laks, L., & Berman, R. A. (2014). A new look at diglossia: Modality-driven distinctions between spoken and written narratives in Jordanian Arabic. In *Handbook of Arabic Literacy* (pp. 241–254). Springer Netherlands.

Laks, L., Hamad, I., & Saiegh-Haddad, E. (2019). Verbal patterns in Palestinian Arabic. *The Mental Lexicon*, 14(2), 209–236.

Maamouri, M. (1998). *Language Education and Human Development: Arabic Diglossia and Its Impact on the Quality of Education in the Arab Region*. World Bank, Mediterranean Development Forum.

Marslen-Wilson, W. D. (2007). Processes in language comprehension. *Oxford handbook of psycholinguistics*, 495–524.

Mayrose, O. (1998). *Acquisition of action nominals in Hebrew*. MA thesis, Tel Aviv University (in Hebrew).

McCarthy, J. J. (1982). Prosodic templates, morphemic templates, and morphemic tiers. In H. Van der Hulst & N. Smith (Eds). *The structure of phonological representations* (pp. 191–223). Dordrecht: Foris Publications.

Ministry of Education (2008). *Foundation of Reading and Writing in Arabic as a Mother Tongue: A Preschool Curriculum*. Centre for Curriculum Planning and Development. Secretariat of Pedagogy. Ministry of Education. Israel.

Myhill, J. (2014). The effect of diglossia on literacy in Arabic and other languages. In E. Saiegh-Haddad and M. Joshi (eds), *Handbook of Arabic Literacy* (pp. 197–223). Dordrecht: Springer.

Ratcliffe, R. R. (2004). Sonority-based parsing at the margins of Arabic morphology: In response to Prunet, Beland, and Idrissi (2000) and Davis and Zawaydeh (1999, 2001). *al-'Arabiyya*, 37, 73–95.

Ravid, D. (1995a). The acquisition of morphological junctions in Modern Hebrew. In H. Pishwa & K. Maroldt (Eds.), *The development of morphological systematicity: A crosslinguistic perspective*. Tubingen: Gunter Narr, 55–77.

Ravid, D. (2004). Derivational morphology revisited: Later lexical development in Hebrew. *Language development across childhood and adolescence*, 53–81.

Ravid, D. (2006). Word-level morphology: A psycholinguistic perspective on linear formation in Hebrew nominals. *Morphology*, 16(1), 127–148.

Ravid, D. D. (2012). *Spelling morphology: The psycholinguistics of Hebrew spelling* (Vol. 3). Springer Science & Business Media.

Ravid, D., & Avidor, A. (1998). Acquisition of derived nominals in Hebrew: Developmental and linguistic principles. *Journal of Child Language*, 25(02), 229–266.

Ravid, D., & Farah, R. (1999). Learning about noun plurals in early Palestinian Arabic. *First Language*, 19, 187–206.

Ryding, K. C. (2005). *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: the case of Arabic diglossia. *Applied Psycholinguistics*, 24, 431–451.

Saiegh-Haddad, E. (2004). The impact of phonemic and lexical distance on the phonological analysis of word and pseudo words in a diglossic context. *Applied Psycholinguistics*, 25, 495–512.

Saiegh-Haddad, E. (2005). Correlates of reading fluency in Arabic: Diglossic an orthographic factors. *Reading and Writing: An Interdisciplinary Journal*, 18, 559–582.

Saiegh-Haddad, E. (2007a). Linguistic constraints on children's ability to isolate phonemes in Arabic. *Applied Psycholinguistics*, 28, 605–652.

Saiegh-Haddad, E. (2018). MAWRID: A model of Arabic word reading in development. *Journal of Learning Disabilities*, 51(5), 454–462.

Saiegh-Haddad, E., & Everatt, J. (2017). Literacy Education in Arabic. In Kucirkova, N., Snow, C., Grover, V., and McBride-Chang, C. (Eds.). *The Routledge International Handbook of Early Literacy Education*. USA: Taylor & Francis Routledge.

Saiegh-Haddad, E., Hadieh, A., & Ravid, D. (2012). Acquiring noun plurals in Palestinian Arabic: Morphology, familiarity, and pattern frequency. *Language Learning*, *62*(4), 1079–1109.

Saiegh-Haddad, E., Levin, I., Hende, N., & Ziv, M. (2011). The linguistic affiliation constraint and phoneme recognition in diglossic Arabic. *Journal of Child Language*, *38*(02), 297–315.

Saiegh-Haddad, E., & Henkin-Roitfarb, R. (2014). The structure of Arabic language and orthography. In *Handbook of Arabic literacy* (pp. 3–28). Springer Netherlands.

Saiegh-Haddad, E., & Spolsky, B. (2014). Acquiring literacy in a diglossic context: Problems and prospects. In *Handbook of Arabic literacy* (pp. 225–240). Springer Netherlands.

Saiegh-Haddad, E., & Taha, H. (2017). The role of morphological and phonological awareness in the early development of word spelling and reading in typically developing and disabled Arabic readers. *Dyslexia*, *23*(4), 345–371.

Schiff, R., & Saiegh-Haddad, E. (2018). Development and relationships between phonological awareness, morphological awareness and word reading in spoken and standard Arabic. *Frontiers in psychology*, *9*, 356.

Shalhoub-Awwad, Y., & Leikin, M. (2016). The Lexical Status of the Root in Processing Morphologically Complex Words in Arabic. *Scientific Studies of Reading*, 1–15.

Shalhoub-Awwad, Y. (2019). The role of nominal pattern in Arabic reading acquisition: Insights from cross-modal priming. *Scientific Studies of Reading*, 1–14.

Tarabani, A. (2006). *The distribution of Palestinian Arabic verb patterns in preschoolers' conversation*. Submitted in partial fulfillment of the requirements for the Master's degree in the Sacker Faculty of Medicine of Health Professionals, Department of Communication Disorders, Tel-Aviv University (In Hebrew).

Velan, H., Frost, R., Deutsch, A., & Plaut, D. C. (2005). The processing of root morphemes in Hebrew: Contrasting localist and distributed accounts. *Language and Cognitive Processes*, *20*(1–2), 169–206.

Velan, H., & Frost, R. (2011). Words with and without internal structure: What determines the nature of orthographic and morphological processing? *Cognition*, *118*(2), 141–156.

Wright, W. (1995). *A grammar of the Arabic language*. Cambridge: Cambridge University Press.