Experimental Arabic Linguistics

Studies in Arabic Linguistics

10

Edited by Dimitrios Ntelitheos and Tommi Tsz-Cheung Leung

John Benjamins Publishing Company

Experimental Arabic Linguistics

Studies in Arabic Linguistics

This book series aims to publish original research in all fields of Arabic linguistics, including – but not limited to – theoretical linguistics, historical linguistics, sociolinguistics, pragmatics, typology, and language acquisition. Submissions from all current theoretical frameworks are welcome. Studies may deal with one or more varieties of Arabic, or Arabic in relation to or compared with other languages. Both monographs and thematic collections of research papers will be considered.

The series includes monographs and thematically coherent collective volumes, in English.

For an overview of all books published in this series, please see
benjamins.com/catalog/sal

Volume 10

Experimental Arabic Linguistics
Edited by Dimitrios Ntelitheos and Tommi Tsz-Cheung Leung

Experimental Arabic Linguistics

Edited by

Dimitrios Ntelitheos
United Arab Emirates University

Tommi Tsz-Cheung Leung
United Arab Emirates University

John Benjamins Publishing Company
Amsterdam / Philadelphia

∞™ The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/sal.10

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2021021468 (PRINT) / 2021021469 (E-BOOK)

ISBN 978 90 272 0884 2 (HB)
ISBN 978 90 272 5960 8 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Acknowledgements

The majority of the papers in this volume stem from presentations at the 3rd Meeting of the Experimental Arabic Linguistics Conference (EXAL18), hosted by the Department of Linguistics at the United Arab Emirates University on 28–29 October, 2018. We would like to thank all presenters and participants of the conference. We are also grateful to the College of Humanities and Social Sciences at the United Arab Emirates University for their support in hosting EXAL18. We would additionally like to thank the ever-expanding community of people working on Experimental Arabic Linguistics, many of whom volunteered to externally review the chapters, this volume would not have been possible without their help. Finally, a heartfelt thanks to the tremendous support and help from the *Studies in Arabic Linguistics* Series Editors Elabbas Benmamoun and Enam Al-Wer and the John Benjamins editorial team. We hope that this effort will lay the foundations for many subsequent explorations of the fertile new field of experimental Arabic linguistics.

Introduction

Experimental approaches to Arabic linguistics

Dimitrios Ntelitheos and Tommi Leung
United Arab Emirates University

1. Introduction

The term Arabic Linguistics refers to the study of Classical Arabic (e.g., religious texts), Modern Standard Arabic (the formal variety of mainly written discourse), and various Arabic dialects from the synchronic or the diachronic point of view. While the study of the historical development of Arabic within the rich Classical Arabic linguistic tradition, as well as descriptive studies of different dialects, have received considerable attention from very early on (see Behnstedt & Woidich, 2013), detailed structural analysis of the standard and dialectal varieties, e.g., within the generativist tradition, has shown significant growth since the second half of the twentieth century in parallel to a vast increase in fieldwork and the amount of data accessible to the researchers.

This growth in theoretical analyses of grammatical phenomena in Arabic dialects has sporadically been enriched with experimental work, especially in experimental phonetics (see, e.g., Al-Ani, 1970). However, considerable research on experimental linguistics in Indo-European languages is not readily matched by a similar output in Arabic linguistics. It was not until 2013 that a specialized conference for the field (Experimental Arabic Linguistics (EXAL)) was hosted by United Arab Emirates University. Subsequent EXAL meetings were hosted in different places in the UAE (NYU-Abu Dhabi in 2015 and UAEU in 2018).

2. Experimental Arabic linguistics

The field of experimental linguistics is concerned with the study of linguistic representations and their relation to other cognitive and motor-control mechanisms, based on quantitative data, drawn primarily from specific experimental apparatuses and other user interfaces. We assume a broader definition of experimental

https://doi.org/10.1075/sal.10.int

linguistics (e.g., as stated in Hemforth, 2013), which is defined as the use of the quantitative methods of some sort to investigate linguistic representations. The scope of quantitative methods in experimental linguistics can be all-encompassing, including evidence that stems from a strictly experimental design or corpus studies or a combination of the two. Based on these assumptions, we consider here studies related to the exploration of issues pertinent to the Arabic language in two different ways: approaches that employ novel experimental methodologies and/or design to address long-standing issues in Arabic linguistics and approaches that investigate empirical/methodological issues in typical and atypical development of the Arabic language. The latter research area also includes studies that use experimental data to address linguistic issues related to specific target populations speaking Arabic as their first or second language. This second endeavor may have less of a focus on the "linguistic representation" aspect of study at the outset, yet still fits the broader definition of experimental linguistics as stated above, as long as the studies employ quantitative methods to address issues pertaining to the target language. Thus, experimental studies of typical or atypical language development may contribute to the development of better language rehabilitation strategies and facilitate the mapping of linguistic structural patterns across various populations, leading to a better understanding of the properties of human language and its mental/cognitive representation.

What follows is by no means an exhaustive literature review of all the aspects of experimental Arabic linguistics, as this would probably require a lengthy book treatment of the field. We just list and briefly summarize several areas where Arabic has made contributions to experimental approaches in linguistic research and areas where modern experimental methodologies have facilitated the discussion of important issues in understanding the grammatical properties of Arabic. Sections 2.1–2.5 introduce some recent developments in various areas of the experimental study of Arabic linguistics. A summary of the papers in this volume will be presented in Section 3.

2.1 Experimental phonetics

The most comprehensibly studied area is Experimental Arabic Phonetics, with numerous studies appearing in journals or as book chapters. Hassan and Heselwood (2011), for example, put forth a collection of cutting-edge papers, which tackled various issues related to Arabic phonetics, using multiple experimental methodologies, including acoustic studies, electromagnetic articulography, electropalatography, nasoendoscopy, videofluoroscopy, spectrography, laryngoscopy, and ultrasound.

Issues that are of interest to Arabic phoneticians focus on the articulatory and acoustic properties of sounds that distinguish the Arabic phonemic inventory from other sound systems cross-linguistically. These include pharyngeal fricative sounds, pharyngealized (emphatic) consonants, and geminates. The rapid advancement of audio and imaging technologies has provided a rich set of methodologies in investigating these properties. Consider, for example, the progress in understanding the articulatory and acoustic properties of the voice pharyngeal segment ع, which is transcribed in IPA as [ʕ]. Embarki (2013) provided a snapshot of the historical development of this understanding, starting with Al-Ani's (1970) experimental phonetic work in Arabic. Based on cineradiographic data (x-rays of the vocal tract) combined with acoustic data, Al-Ani shows that the articulation of [ʕ] involves low pharyngeal constriction with retracted tongue body and clear voicing (but no frication) characteristics (as shown with spectrographic studies). Al-Ani considered [ʕ] a pharyngeal stop. Using the same methodology, Ghali (1983, p. 440) preferred a "trill" manner of articulation for the phoneme rather than plosive, implementing a special feature "frequentative." Switching to the endoscopic technique, which visualizes articulatory adjustments during the production of the phoneme in Jordanian Arabic, Zawaydeh (2003) confirmed pharyngeal constriction, showing a reduction in the distance between the epiglottis and the pharyngeal wall. Yet another technology was used by Hassan et al. (2011). Through an ultrafast laryngoscopic study combined with electroglottography (EGG), they confirm the pharyngeal articulation but also point out a voiced aryepiglotto-epiglottal fricative variant in Iraqi Arabic. This example of a study of the articulatory properties of a single segment shows how the gradual introduction of new experimental phonetics technology can help our understanding of the properties of sounds that are exclusive to the Arabic sound system and a limited number of other languages.

The advancement of audio and imaging technologies has led to a blossoming field of study in Experimental Arabic Phonetics. Numerous papers have used these technologies to address long-standing questions on the articulatory and acoustic properties of Arabic sounds. Some of the technologies used include the following: *software-based sound spectrography*, which is one of the cheapest and standard acoustic phonetic research methodologies used in Arabic to study for example the acoustic exponents of emphatic/pharyngealized sounds (see Khattab, Al-Tamimi, & Heselwood (2006) for a review); *nasoendoscopy*, the endoscopic study of the nasal passage and the nasopharyngeal and oropharyngeal regions (Al-Tamimi & Heselwood, 2011; Heselwood & Al-Tamimi, 2011); *electromagnetic articulography (EMA)*, which uses electromagnetic fields around the speech organs of the speaker to detect prepositioned metal sensor-pellets attached to articulators (Embarki, Ouni, Yeou, Guilleminot, & Al Maqtari, 2011; Gafos, Hoole, & Zeroual, 2011; Zeroual, Esling, & Hoole, 2011); *laryngoscopy*, the endoscopic study of the larynx and supraglottal structures (Hassan & Esling, 2011); *electropalatography*, an electrode-based exploration of passive articulators (Heselwood, Howard, & Ranjous, 2011); *aerometry*, the measurement of airflow pressure using a pressure transducer and pneumotachograph, (Bakalla, 1983; Yeou & Maeda, 2011); *videofluoroscopy*, a radiographic method for obtaining the videos of the movement of articulators in the vocal tract (Al-Tamimi & Heselwood, 2011; Heselwood & Al-Tamimi, 2011); and *ultrasound* techniques where an ultrasound scanner obtains the computerized images of articulators as they move (Zeroual et al., 2011).

2.2 Psycholinguistic studies

In the area of Arabic language processing and production, most research has focused on Arabic word structures, especially the representational status of Arabic roots and patterns.

A series of studies in this area led by Boudelaa and Marlsen-Wilson (2004a, 2004b, 2005, 2011) examined whether Arabic roots and patterns can prime, independently of semantic or phonological (i.e., form) effects (see also Mahfoudhi, 2007). Using masked and cross-modal and auditory priming on verbs and nouns, they argued that semantic or phonological cues do not facilitate word processing, whereas morphological roots and patterns do, even when the prime and the target are not semantically related. Further support for the representational status of the consonantal root stemmed from Ussishkin, Dawson, Wedel, and Schluter's (2015) research on Maltese Arabic. They used auditory priming with equally audible primes and targets and masked auditory priming (using volume-attenuation and compression) with speakers of Maltese Arabic. The results confirmed visual masked priming results (e.g., Twist, 2006) in that forms sharing a consonantal root do prime the targets, but forms sharing a word pattern do not. In summary, consonantal roots seem to facilitate visual/auditory word recognition in Maltese Arabic independently of phonological overlap. Subsequent priming experiments by Schluter (2013) for Moroccan Arabic and by Al Kaabi (2015) for Emirati Arabic further verified the representational status of roots in Arabic.

The variation of the priming experiment further supports the status of consonantal roots in Arabic. For instance, Perea, Abu Mallouh, and Carreiras (2010) argued that the "transposed letter effect" is observed on the level of roots. By manipulating lexical primes that contain transposed root vs. non-root letters, they observed that transposed non-root (e.g., pattern) letters facilitated the recognition of the target word (as long as the prime and the target share the same root letters), whereas transposed root letters significantly inhibited the word recognition of the target word. This further supported the proposal that Arabic roots (which consist of a fixed order of consonants) play a crucial role in lexical retrieval.

The cognitive status of root consonants was further supported by Prunet, Béland, and Idrissi's (2000) case study of an Arabic/French bilingual patient with aphasia. The patient in this case produced frequent metathesis errors in word-reading, picture-naming, and repetition tasks, but these errors are more sensitive to root consonants than affixal consonants or vowels. Prunet et al. (2000) took this as evidence for the status of the consonantal root as a distinct abstract entity in the minds of native speakers.

Another area of Arabic psycholinguistics that provides deep insights to a better understanding of language processing is Arabic orthography. Arabic orthography is characterized by its right-to-left writing system, the use of ligatures in which a single letter can assume different shapes depending on its position within the word, and the consonantal writing system. While the "full" system of Arabic orthography consists of diacritical marks for the indication of short vowels and consonant lengthening (geminates), diacritics are dropped in almost all contexts, unless lexical ambiguity cannot be resolved by context.

Research on whether the inclusion of vowel diacritics enhances reading ability has led to mixed results. Roman and Pavard (1987) and Bentin and Ibrahim (1996) showed that the inclusion of diacritics impeded lexical decisions on isolated words but facilitates sentence reading (Roman, Pavard, & Asselah, 1985). In terms of reading accuracy in the development of reading skills, Abu-Rabia and Siegel (1995) and Abu-Rabia (1997, 2001) affirmed that poor and skilled readers benefited in terms of reading accuracy from the presence of diacritics in the text, although this effect might disappear at the later stages of reading development. In fact, reading fluency may be impeded by the presence of diacritics in later stages (Saiegh-Haddad & Schiff, 2016).

While psycholinguistic work on reading processing and its development has been fundamental in resolving issues related to phonological and morphological awareness, research on Arabic reading has been greatly facilitated by the introduction of eye-tracking technology, to which we turn next.

2.3 Eye-tracking studies

The use of eye-tracking and neuroimaging technology has been used extensively in addressing linguistic problems in Indo-European languages but only sporadically in Arabic. Thus, while eye-tracking studies in Arabic have scarcely appeared in the literature after the work by Gray (1956) and Roman and Pavard (1987), it is only recently that reading in Arabic has been at the focus of experimental research.

Earlier studies on eye movement were not conducted by using the noninvasive eye tracker that measures corneal reflection (e.g., Farid & Grainger, 1996). Instead, eye fixations are controlled by the position of the word display on the screen. It was not until the work by Jordan et al. (2014); Paterson, Almabruk, McGowan, White, and Jordan (2015); and Hermena (2016), among others, in which the use of a high-speed eye tracker (e.g., 1000-Hz sampling frequency) was widely used in Arabic research. Overall, eye movement research in Arabic usually focuses on how the right-to-left writing direction impacts the perceptual span (e.g., Jordan et al., 2014) and the relation between Arabic word structure and initial fixation (Farid & Grainger, 1996; Hermena et al., 2017, etc.). Other word-level variables, such as the number of letters, initial bigrams, and diacritical marks, have also been used as the experimental conditions in Arabic eye movement studies (Hermena, 2016; Hermena, Drieghe, Hellmuth, & Liversedge, 2015; Paterson et al., 2015). It is also possible to study Arabic word structure (as part of the mental representation) without using its writing system (i.e., orthography) as the visual interface. For instance, Alamri's (2017) study on eye movement investigated Arabic spoken word recognition using the visual world paradigm (Allopenna et al., 1998; Dahan et al., 2001; Tanenhaus et al., 1995).

Another area where the eye-tracking methodology has been used to investigate Arabic is visual patterns in scanning online materials on web pages. For instance, Remya et al. (2011) investigated visual reading patterns of Arabic script exhibited by web users in their interaction with Arabic interfaces in various types of web materials and found differences in the reading patterns exhibited by the readers compared with the studies in English interfaces because of variations in the direction of Arabic scripts and cultural factors. Other studies that have shown similar results are Al-Salloom et al. (2010), Al-Rashid et al. (2010), and Al-Wabil et al. (2010).

2.4 Neurophysiological studies

Work on the processing (and production) of Arabic has expanded to neurophysiological studies on similar phenomena in Arabic using electroencephalography (EEG) and magnetoencephalography (MEG) techniques, with the former recording electrical activity, and the latter the magnetic fields in the brain during the performance of linguistic tasks. In EEG studies, the methodology looks at Event-Related Potentials (ERPs), i.e., the measurements of electrical activities in the brain time-locked to specific linguistic stimuli.

Neurophysiological data have been collected to better understand the root-pattern representation of Arabic morphology. For instance, Boudelaa et al. (2009) conducted an ERP study to investigate a mismatch negativity (MMN) response to word pairs differing in either a root consonant/pattern vowel in addition to two control conditions formed by pseudowords. The result showed different time-courses in the electrical activity during MMN effects on roots (at 16 ms) than on patterns (much later at 250 ms). In addition, the activity during root negativity effects seems to be located in the fronto-central regions in the right and left hemispheres for the roots. By contrast, for the patterns, the activity location resides above the Perisylvian brain region. The divergent time signatures and topographies of root and pattern activities are taken as evidence for the separate statuses of these two morphological objects.

In two experiments, Al Kaabi (2015) used MEG recordings of a priming study to investigate morphological decomposition in Arabic. In the first experiment, she found priming effects in morphologically and semantically related words but not for the control pseudowords. In addition, MEG data were inconclusive as to the neural activity related to the morphological and semantic conditions. In the second experiment, Al Kaabi used MEG recordings of a visual lexical decision task, measuring the neural activity of the probability of a word given its stem and the probability of a word given its root. The results indicated an M170 response in the left fusiform gyrus pointing toward early word morphological decomposition and a consonantal root frequency M350 response in the left temporal lobe providing further neurophysiological evidence for a root-based abstract representation.

The decompositional account of Arabic was further supported by functional magnetic resonance imaging (fMRI) findings suggesting that not only lexical (i.e., verbs and nouns) but also function words, which are formed on a root-pattern templatic format, elicit increased activation of the same neural area typically involved in morphological parsing processes (Boudelaa, Bozic, & Marslen-Wilson, 2010; Bozic, Tyler, Ives, Randall, & Marslen-Wilson, 2010).

There has also been an increasing interest in the neuropsychological study of Arabic sentence processing. For instance, Westerlund, Kastner, Al Kaabi, & Pylkkänen (2015) investigated the significance of the left anterior temporal lobe in combinatorial syntactic processes, including those of modifier-predicate and predicate-argument composition in English and Arabic. They showed that this specific brain region exhibits increased responses to words in combinatory contexts independently of composition type and word order, indicating an essential role in basic semantic composition. In a different MEG study, Matar, Pylkkänen, & Marantz (2019) manipulated syntactic category predictability within the fully well-formed expressions of Arabic. The results confirmed that predictive mechanisms while reading involve predicting upcoming syntactic categories, with activities in two cortical regions: the left visual and right frontal cortices.

Finally, the neural correlates of Arabic–English bilingual language use were investigated by Blanco-Elorrieta and Pylkkänen (2016). They used MEG to investigate the degree of the neural overlap between language control and domain-general cognitive control in action and perception, i.e., whether the neural activity of bilinguals during language-switching is influenced by whether switching between languages is active or passive. Their results indicated a clear dissociation of language control mechanisms in production versus comprehension. By using MEG technology, the authors obtained the first characterization of the spatiotemporal profile of these effects, establishing that switching processes begin around 400 ms after stimulus presentation.

2.5 Typical and atypical language development

Whether studies on typical and atypical language developments constitute a legitimate subfield of experimental linguistics remains a controversial issue. For instance, research on speech therapy and pathology may not touch upon core linguistics issues at all, focusing instead on the pathological issues of diagnosis and treatment (the same analogy applies to the relation between the actual practice of physiotherapy and clinical research on orthopedics). However, given the all-encompassing nature of experimental linguistics, we include here studies that investigate the language behaviors of certain populations with typical and atypical language developments as long as they follow the rigorous methodology that typical experimental linguistic work does.

In the field of first language (L1) acquisition, Arabic has been the focus of extensive research in corpus-based and experimental studies, e.g., Albirini (2018), Badry (2006), and Walter (2006). Various linguistic modules and categories in Arabic have been studied in the context of L1 acquisition, e.g., phonemic inventories (Ammar & Morsi, 2006; Ayyad & Bernhardt, 2009; Dyson & Amaryeh, 2007; Khattab, 2007; Shahin, 1995), morphological inflections and derivations (Abu-Nofal, 2005; Aljenaie, 2010; Badry, 2005; Basaffar & Safi, 2012; Moawad, 2006; Mourchid, 1998; Ravid & Farah, 1999; Ravid & Hayek, 2003; Rola, 2005), and grammatical categories (Abu El-Haija, 1981; Aftat, 1982; Al-Buainain, 2003; Aljenaie, 2008; Smadi, 1980). There are other studies focusing on the acquisition of syntactic structures and operations, e.g., interrogatives (Aftat, 1982; Al-Buainain, 2003; Smadi, 1980), binding relations (Bolotin, 2002; Mustafawi & Mahfoudhi, 2002), word-order permutation (Aljenaie & Farghal, 2009), relative clauses (Aller, 1978), and negative island constraints (Abdulkarim et al., 1997).

Conversely, the domain of atypical language development and language disorders across the life span has received limited attention (see Safi (2006) and Alhawary (2006) for reviews). One area that has attracted sporadic research is that of hearing-impaired children. For instance, Ammar and Rifaat (1998) examined the speech of the hearing-impaired speakers of Egyptian Arabic and found a statistically significant difference in vowel pitch and duration in their speech when compared with hearing subjects. In a much later study, Friedman and Haddad-Hanna (2014) focused on the comprehension of sentences with wh-movement by 24 Palestinian-Arabic-speaking children with hearing impairment, of which 21 had binaural and 3 had monaural hearing loss. Children with binaural hearing loss presented extreme difficulties in processing object relatives, object questions, and topicalization in SV and VS orders in contrast with monaurally hearing-impaired children, who performed at the same level as hearing children in the experiment.

In a single study on cleft-palate, Shahin (2002) considered data from three Palestinian-Arabic-speaking children. The study results reflected similar results in other languages, in that Arab cleft-palate children produce nasalized vowels and "retracted" or "glottalized" obstruents with parallel devoicing.

An area that has received increasing attention in Arabic speech-language impairment research is Developmental Language Disorders (DLD, previously Specific Language Impairment (SLI)). The phonological aspects of DLD in Egyptian children were explored by Ammar (1992) who found that the typically developing group performed better than the DLD group in the quality of segment articulation and the number of simplifying phonological processes adopted in articulation, a result also confirmed by a later study by Salameh, Nettelbladt, & Norlin (2003) on Bilingual Arabic-Swedish-speaking children. Turning their focus on the grammatical aspects of DLD speech, Salameh, Håkansson, & Nettelbladt (2003) asserted that the former exhibit several deficiencies when it comes to inflectional morphology at the word and agreement at the intra- and inter-phrase levels on the basis of a longitudinal study on DLD and typically developing bilingual children. Abdalla (2002) investigated DLD in children acquiring Hijazi Arabic, with results indicating a language level for DLD children equivalent to that of two years younger typically developing children. Fahim (2005) examined the language of three children with DLD speaking Egyptian Arabic compared with twelve typically developing children and found that the former go through a series of developmental stages that are different from those characterizing the linguistic development of the latter. DLD children dropped functional elements at a greater rate and used fewer verbal forms, substituted simpler imperfective verb forms for inflected perfective ones, and in general made more verb agreement errors compared with typically developing children. Finally, Shaalan (2010) conducted a larger-scale study on 26 DLD Qatari, Gulf-Arabic-speaking children, administering several batteries of tests and checking mastery of morphosyntactic complexity, word order, complex syntactic structures, and phonological awareness. The study results showed deficiencies in the performance of DLD children in all these areas, with significant problems in verb agreement morphology realizing gender and number features, a result further collaborated by Abdalla and Crago (2008) with a study specifically targeting verb morphology.

3. Papers in this volume

For obvious reasons, the papers included in this volume can only touch upon a highly limited number of areas within the broader field of Experimental Arabic Linguistics. Most of the papers have stemmed from the Experimental Arabic Linguistics Conference 2018 (EXAL18), which took place at the United Arab Emirates University in October 2018. The papers belong to three different broad thematic groups.

Hermes, Barlaz, Shosted, Liang, and Sutton investigate how emphatic spread, the spread of the pharyngealization articulatory gesture to adjacent segments, and the phonetic correlates associated with the pharyngealization spread triggered by emphatic consonants as observed in Cairene Arabic. More particularly, they study the pharyngealization spread through different vowel lengths (short and long), examining spread in left-to-right and right-to-left directions. They also combine experimental results from real-time Magnetic Resonance Imaging (rtMRI) with the acoustic measurements of the first and second formant frequency measures.

The data from the articulatory and acoustic investigations indicate varying strengths of pharyngealization spread across vowels of different lengths. This is supported by the variation in the extent of formant frequency modification due to the presence of a pharyngealized consonant in the acoustic data and by the smaller 2D pharyngeal area measurements and the more constricted pharyngeal contours in the articulatory data. In addition, the articulatory data confirm that speech segments preceding pharyngealized consonants are generally more constricted than those following them, suggesting that anticipatory (leftward) spread is stronger than perseveratory (rightward) spread in Cairene Arabic. Finally, they assert that among two minimal pairs containing the same pharyngealized consonant but differing in vowel length, 2D pharyngeal areas are generally more constricted in the speech segments of the word containing the longer vowel, suggesting that the effect of pharyngealization is stronger in longer vowels.

Szreder, Derrick, and Ben-Ammar focus on two phonological alternations in Emirati Arabic, k-affrication [k]~[tʃ] and dʒ-deaffrication [dʒ]~[j]. Using a corpus analysis (based on the EMAC corpus; Ntelitheos & Idrissi, 2017) and elicitation, they claim that the variation is determined by a plethora of factors, e.g., lexical factors that are speaker-specific, the neighboring vowels, and coronal consonants. Other historical and sociophonetic factors may also contribute to the intricate variational system. These intermingling factors confirm the traditional hypotheses that variation in these sounds is influenced by a multiplex of linguistic and social factors.

Two papers focus on the study of eye movement in two domains of linguistics, namely, reading and phonological awareness. AlJassmi, Hermana, and Paterson provide a state-of-the-art overview of eye movement research on reading in Arabic.

In addition to summarizing the most significant spatial and temporal measures when conducting an eye movement experiment (e.g., reading time, fixation duration, refixation probabilities, and word-skipping rate), the authors focus on how the differences between Arabic (or Semitic languages) and English (or Indo-European languages that use the Roman orthography) regarding the word structure and the writing system correspond to their concomitant distinctions in eye movement measures. The Arabic (and Semitic) writing system is orthographically distinctive for its directionality (right-to-left direction), ligatures (the shape of a letter is position-dependent) and letter-to-sound correspondence (Arabic letters are consonantal, whereas vowels are not explicitly expressed). Arabic words are also argued to be more "dense" than Latinate words in that a single Arabic word (written by 4–8 letters) can express the meaning of a complete sentence. These areas of distinction strongly suggest that any scientific conclusions drawn from eye movement studies in Latinate languages cannot be applied to Arabic without a rigorous experimental control. Finally, the aforementioned typological distinctions suggest that the current models of the reading (e.g., E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003) and SWIFT models (Engbert, & Kliegl, 2012; Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005) should be parameterized, which accommodates variation between languages.

Marquis, Al Kaabi, Leung, and Boush illustrate how the phonological awareness of Emirati Arabic could be described by the various measures of eye movement. They also argue that a significant difference existed in terms of fixation patterns between experimental conditions that manipulate onset consonants and rhymes. In addition, whether the experimental subjects receive prior feedback by the experimenter during the practice trials has an impact on the overall eye fixation patterns of the subject. This study elucidates the use of eye movement as an indirect measure of the subject's phonological awareness of Emirati Arabic and, in a deep sense, the interface between vision and phonology.

Marquis introduces the Language Acquisition Test for Arabic (LATFA), an assessment tool based on oral language skills to predict future literacy difficulties in Emirati Arabic (EA)-speaking children. The concept of LATFA, which consists of language perception (phonological awareness and speech perception) and production (morphological awareness and phonological production) tasks, is originally adopted from Québec French (Rvachew et al., 2017). The LATFA is Arabic-specific in the sense that the particular linguistic properties of Arabic are used as the tasks, e.g., the tense and agreement system for grammar and morphology and the rhyme system for phonology. The same group is given a literacy test a year later, which focuses on their command on the writing of Arabic alphabet, dictation, and plurals. A total of 43 Emirati Arabic-speaking first-graders participated in the research.

Marquis claims that the oral language skills tested in LATFA are a good predictor, in the sense that children who have been experiencing difficulties in Arabic literacy could be predicted (with a 70% success rate) by their LATFA performance a year ago.

Al-Hassan and Marinis investigate the morphological competence of Saudi Arabic-speaking children with autism spectrum disorders (ASD). They utilize the Saudi Arabic Sentence Repetition Task (Marinis & Armon-Lotem, 2015) on three groups of Saudi Arabic-speaking children, namely, typically developing (TD) children, children with autism with normal language (ALN), and children with autism with language impairment (ALI). The scoring scheme of the verbal abilities of the children is calculated on the basis of whether children repeat the sentences verbatim (the verbatim scheme) or use the targeted structure (the structural scheme). The children are tested in terms of various grammatical structures, such as wh-questions, topicalizations, and relative clauses. Overall, the TD children perform better than the ALN and ALI children in both scoring schemes. The ALN children have comparable performance with TD in the structural scheme, yet score less than TD in the verbatim schema. The ALI children have the worst overall performance in verbatim and structural schemes.

Shaalan, Egan, Gould, and Olsen's two-year longitudinal study involving 29 children with ASD in the UAE assess their Arabic and English competence regarding their receptive and expressive vocabulary after an intensive behavioral intervention program. They find that most Emirati ASD children who have undergone the intensive behavioral intervention program show positive vocabulary gains in both languages. Moreover, there are no detrimental effects of bilingual intervention in the other language, a result that leads to their recommendation to provide bilingual intervention to bilingually exposed children with autism.

Khater studies how nonword repetition tests, an assessment of short-term memory, and speech perception/production could be used as a diagnostic of speech and language impairment. By focusing on Gulf Arabic children, a series of word and nonword repetition tests that vary in terms of the number of syllables are applied to the language impairment or clinical (CL) group and a TD group. She finds that the CL group performs consistently worse than the TD group in the word and nonword repetition tests across all testing conditions. Moreover, she notices a correlation between the overall performances of the groups in word/nonword repetition and Arabic receptive and expressive vocabulary. The result partially concurs with the phonological short-term memory account (Gathercole, 2006; Gathercole & Baddeley, 1990) and phonological processing skill account (Chiat, 2001; Snowling, Chiat, & Hulme, 1991).

References

Abdalla, F. (2002). Specific language impairment in Arabic-speaking children: Deficits and morphosyntax. PhD dissertation. McGill University, Canada.

Abdalla, F., & Crago, M. (2008). Verb morphology defects in Arabic speaking children with specific language impairment. *Applied Psycholinguistics*, 29, 315–340. https://doi.org/10.1017/S0142716408080156

Abdulkarim, L; Roeper, T., & de Villiers, J. (1997). Negative Islands in Language Acquisition. Paper presented at New Perspectives on Language Acquisition: Minimalism and Pragmatics, University of Massachusetts, Amherst, June 2–4, 1997.

Abu El-Haija, L. A. (1981). The acquisition of the negation system in Arabic as spoken in Jordan, PhD dissertation. The Pennsylvania State University, State College, PA.

Abu-Nofal, H. (2005). Learning to inflect Palestinian Arabic adjectives in plural contexts. Paper presented at the *Workshop on the acquisition of Morphology in Palestinian Arabic: Universal, particular and contextual factors*, Tel Aviv University, Tel Aviv, Israel, Friday, July 29, 2005.

Abu-Rabia, S. (1997). Reading in Arabic orthography: The effects of vowels and context on reading accuracy of poor and skilled Arabic native readers in reading paragraphs, sentences, and isolated words. *Journal of Psycholinguistic Research*, 26, 465–482. https://doi.org/10.1023/A:1025034220924

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing: An Interdisciplinary Journal*, 14, 39–59. https://doi.org/10.1023/A:1008147606320

Abu-Rabia, S., & Siegel, L. S. (1995). Different orthographies, different context effects: The effects of Arabic sentence context in skilled and poor readers. *Reading Psychology: An International Quarterly*, 16, 1–19. https://doi.org/10.1080/0270271950160101

Aftat, M. (1982). The acquisition of negation and wh-questions in a Moroccan Arabic speaking four-year-old child. PhD dissertation. University of Texas at Austin, TX.

Alamri, A. (2017). Phonological, semantic and root activation in spoken word recognition in Arabic: Evidence from eye movements. PhD dissertation. University of Ottawa, Canada.

Al-Ani, S. H. (1970). *Arabic phonology: An acoustical and physiological investigation*. The Hague: Mouton. https://doi.org/10.1515/9783110878769

Albirini, A. (2018). The acquisition of Arabic as a first language. In E. Benmamoun, & R. Bassiouney (Eds.), *The Routledge handbook of Arabic linguistics* (pp. 227–248). Oxford: Routledge.

Al-Buainain, H. (2003). Developmental stages of the acquisition of negation and interrogation by children native speakers of Qatari dialect. *Journal of the Faculty of Humanities and Social Sciences*, 25, 9–45. http://hdl.handle.net/10576/8344

Alhawary, M. T. (2006). Language impairment. In K. Versteegh (Ed.) *Encyclopedia of Arabic language and linguistics* (pp. 674–681). Leiden: Brill.

Aljenaie, K. (2008). Negation in early Kuwaiti Arabic child language. *Arab Journal for the Humanities*, 26,102. Kuwait: Kuwait University.

Aljenaie, K. (2010). Verbal inflection in the acquisition of Kuwaiti Arabic. *Journal of Child Language*, 4, 841–863. https://doi.org/10.1017/S0305000909990031

Aljenaie, K. & Farghal, M. (2009). Comprehension of three word orders in Kuwaiti Arabic child language. *Language Sciences*, 31, 494–506. https://doi.org/10.1016/j.langsci.2007.10.003

Al Kaabi, M. (2015). The anatomy of Arabic words: The role of the root in representations and processing. PhD dissertation. New York University, New York.

Aller, Y. S. K. (1978). The acquisition of relative clause constructions in Arabic. PhD dissertation. Indiana University, Bloomington, IN.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Al-Rashid, W., B. Al-Dosary, & R. Al-Salloom. (2010). Visual attention on search engine results: Information retrieval of Arabic web content. Paper presented in *IADIS 2010 Informatics 2010 Conference*, June 30, 2010.

Al-Salloom, R., W. Al-Rashid, & B. Aldosary. (2010). An exploratory eye tracking study of visual attention on web adverts and banners. Paper presented in *IADIS 2010 Informatics 2010 Conference*, June 30, 2010.

Al-Tamimi, F. and Heselwood, B. (2011). Nasoendoscopic, videofluoroscopic and acoustic study of plain and emphatic coronals in Jordanian Arabic. In B. Heselwood, & Z. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (pp. 165–191). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.319

Al-Wabil, A. H. Al Gibreen, R. George, & B. Aldosary. (2010). Exploring the validity of learning styles as effective personalization parameters in elearning environments: An eyetracking Study. Paper presented in *International Conference on Computer Technology and Development* (ICCTD 2010), Nov 2, 2010.

Ammar, W. (1992). Articulation disorders in Arabic. PhD dissertation. University of Alexandria, Egypt.

Ammar, W., & K. Rifaat. (1998). An acoustic study of vowels of deaf and normal-hearing thirteen to eighteen-year-old Egyptian females. *Jerash for Research and Studies* 2(2), 29–38.

Ammar, W., & Morsi, R. (2006). Phonological development and disorders: Colloquial Egyptian Arabic. In H. Zhu (Ed.) *Phonological development and disorders in children* (pp. 204–232). Bristol, UK: Multilingual Matters Limited. https://doi.org/10.21832/9781853598906-011

Ayyad, H., & Bernhardt, B. M. (2009). Phonological development of Kuwaiti Arabic: Preliminary data. *Clinical Linguistics & Phonetics*, 23(11), 794–807. https://doi.org/10.3109/02699200903236493

Badry, F. (2005). Acquisition of Arabic word formation: A multi-path approach. In M. T. Alhawary, & E. Benmamoun (Eds.), *Perspectives on Arabic Linguistics XVII-XVIII* (pp. 243–272). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.267.13bad

Badry, F. (2006). First language acquisition. In K. Versteegh (Ed.) *Encyclopedia of Arabic language and linguistics* (pp. 96–104). Leiden: Brill.

Bakalla, M. H. (1983). The treatment of nasal elements by early Arab and Muslim phoneticians. In H. M. Cornelis, E. F. Versteegh, K. Koerner, & H-J. Niederehe (Eds.), *The history of linguistics in the Near East* (pp. 49–69). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/sihols.28.04bak

Basaffar, F. M., & Safi, S. M. (2012). The acquisition of verbal inflections in Hijazi Arabic. *Arab World English Journal*, 3–2, 266–304.

Behnstedt, P. & Woidich, M. (2013). Dialectology. In J. Owens (ed.) *The Oxford handbook of Arabic Linguistics* (pp. 300–325). New York: Oxford University Press.

Bentin, S., & Ibrahim, R. (1996). New evidence for phonological processing during visual word recognition: The case of Arabic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 309–323. https://doi.org/10.1037/0278-7393.22.2.309

Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Bilingual language control in perception versus action: MEG reveals comprehension control mechanisms in anterior cingulate cortex and domain-general control of production in dorsolateral prefrontal cortex. *Journal of Neuroscience*, 36(2), 290–301. https://doi.org/10.1523/JNEUROSCI.2597-15.2016

Bolotin, N. (2002). Acquisition of binding in L1 Arabic. In D. B. Parkinson, & E. Benmamoun (Eds.), *Perspectives on Arabic linguistics XIII–XIV* (pp. 211–218). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.230.12bol

Boudelaa, S., & Marslen-Wilson, W. D. (2004a). Abstract morphemes and lexical representation: The CV-Skeleton in Arabic. *Cognition*, 92, 271–303. https://doi.org/10.1016/j.cognition.2003.08.003

Boudelaa, S., & Marslen-Wilson, W. D. (2004b). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language*, 90, 106–116. https://doi.org/10.1016/S0093-934X(03)00424-3

Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes*, 20, 207–260. https://doi.org/10.1080/01690960444000106

Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y., & Marslen-Wilson, W. D. (2009). Arabic morphology in the neural language system: A mismatch negativity study. *Journal of Cognitive Neuroscience*, 22(5), 998–1010. https://doi.org/10.1162/jocn.2009.21273

Boudelaa, S., M. Bozic & W. Marslen-Wilson. (2010). Neuro-cognitive complexity of Semitic morphology: An fMRI investigation of derivational morphology in Arabic. Paper presented in the *Annual Meeting of the Society for Neuroscience*, San Diego, California, USA, Nov 13–17, 2010.

Boudelaa, S., & Marslen-Wilson, W. D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes*, 26, 624–652. https://doi.org/10.1080/01690965.2010.521022

Bozic, M., Tyler, L. K., Ives, D. T., Randall, B., & Marslen-Wilson, W. D. (2010). Bi-hemispheric foundations for human speech comprehension. *Proceedings of the National Academy of Sciences*, U.S.A., 107, 17439–17444.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes*, 16(2–3), 113–142. https://doi.org/10.1080/01690960042000012

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42(4), 317–367. https://doi.org/10.1006/cogp.2001.0750

Dyson, A. & Amaryeh, M. M. (2007). Jordanian Arabic speech acquisition. In S. McLeod. (Ed). *The international guide to speech acquisition* (pp. 288–299). Clifton Park, NY: Thomson Delmar Learning.

Embarki, M. (2013). Phonetics. In J. Owens (Ed.), *The Oxford handbook of Arabic linguistics* (pp. 37–49). Oxford: Oxford University Press.

Embarki, M., Ouni, S., Yeou, M., Guilleminot, C., & Al Maqtari, S. (2011). Acoustic and electromagnetic articulographic study of pharyngealisation: Coarticulatory effects as an index of stylistic and regional variation in Arabic. In B. Heselwood, & Z. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (pp. 193–215). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.319

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42, 621–636. https://doi.org/10.1016/S0042-6989(01)00301-7

Engbert, R., & Kliegl, R. (2012). Parallel graded attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 787–800). Oxford, New York: Oxford University Press.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, 112, 777–813. https://doi.org/10.1037/0033-295X.112.4.777

Fahim, D. (2005). Developmental language impairment in Egyptian Arabic. PhD dissertation. University of London.

Farid, M., & Grainger, J. (1996). How initial fixation position influences visual word recognition: A comparison of French and Arabic. *Brain and Language*, 53(3), 351–368. https://doi.org/10.1006/brln.1996.0053

Friedmann, N., & Haddad-Hanna, M. (2014). The comprehension of sentences derived by syntactic movement in Palestinian Arabic-speaking children with hearing impairment. *Applied Psycholinguistics*, 35, 473–513. https://doi.org/10.1017/S0142716412000483

Gafos, A., Hoole, P., & Zeroual, C. (2011). Preliminary study of Moroccan Arabic word-initial consonant clusters and syllabification using electromagnetic articulography. In B. Heselwood, & Z. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (pp. 27–46). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.319.02gaf

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27, 513–543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, 81, 439–454. https://doi.org/10.1111/j.2044-8295.1990.tb02371.x

Ghali, M. M. (1983). Pharyngeal articulation. *Bulletin of the School of Oriental and African Studies*, 46, 432–444. https://doi.org/10.1017/S0041977X00043159

Gray, W. S. (1956). *The teaching of reading and writing*. Paris: Unesco.

Hassan, Z. M., Esling, J., Moisik, S., & Crevier-Buchman, L. (2011). Aryepiglottic trilled variants of /ʕ, ħ/ in Iraqi Arabic. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 831–834). Hong Kong. http://www.icphs2011.hk

Hassan, Z., & Esling, J. (2011). Investigating the emphatic feature in Iraqi Arabic: Acoustic and articulatory evidence of coarticulation. In B. Heselwood, & Z. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (pp. 218–234). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.319

Hassan, Z. M., & Heselwood, B. (2011). *Instrumental studies in Arabic phonetics*. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.319

Hemforth, B. (2013). Experimental linguistics. In M. Aronoff (Ed.), *Oxford online bibliographies* (pp. 1–16). New York: Oxford University Press.

Hermena, E. W., Drieghe, D., Hellmuth, S., & Liversedge, S. P. (2015). Processing of Arabic diacritical marks: Phonological–syntactic disambiguation of homographic verbs and visual crowding effects. *Journal of Experimental Psychology: Human Perception and Performance*, 41(2), 494–507. https://doi.org/10.1037/xhp0000032

Hermena, E. W. (2016). Aspect of word and sentence processing during reading Arabic: Evidence from eye movements. PhD dissertation. University of Southampton.

Hermena, E. W., Liversedge, S., & Drieghe, D. (2017). The influence of a word's number of letters, spatial extent, and initial bigram characteristics on eye movement control during reading: Evidence from Arabic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 451–471. https://doi.org/10.1037/xlm0000319

Heselwood, B., Howard, S., & Ranjous, R. (2011). Assimilation of /l/ to /r/ in Syrian Arabic: An electropalatographic and acoustic study. In B. Heselwood, & Z. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (pp. 63–98). Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.319.04hes

Heselwood, B., & Al-Tamimi, F. (2011). A study of the laryngeal and pharyngeal consonants in Jordanian Arabic using nasoendoscopy, videofluoroscopy and spectrography. In B. Heselwood, & Z. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (pp. 101–128). Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.319

Jordan, T. R., Almabruk, A. A. A., Gadalla, E. A., McGowan, V. A., White, S. J., Abedipour, L., & Paterson, K. B. (2014). Reading direction and the central perceptual span: Evidence from Arabic and English. *Psychonomic Bulletin & Review*, 21, 505–511. https://doi.org/10.3758/s13423-013-0510-4

Khattab, G. (2007). Lebanese speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 300–312). Clifton Park, NY: Thomson Delmar Learning.

Khattab, G., Al-Tamimi, F., & Heselwood, B. (2006). Acoustic and auditory differences in the /t/-/ṯ/ opposition in male and female speakers of Jordanian Arabic. In S. Boudelaa (Ed.), *Perspectives on Arabic linguistics XVI* (pp. 131–160). Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.266.09kha

Mahfoudhi, A. (2007). Roots and patterns in Arabic lexical processing. In E. Benmamoun (Ed.), *Perspectives on Arabic linguistics XIX* (pp. 123–150). Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.289.11mah

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. de Jong, & N. Meir. (eds.) *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 95–124). Multilingual Matters.

Matar, S., Pylkkänen, L., & Marantz, A. (2019). Left occipital and right frontal involvement in syntactic category prediction: MEG evidence from Standard Arabic. *Neuropsychologia*, 135, 107230. https://doi.org/10.1016/j.neuropsychologia.2019.107230

Moawad, R. (2006). The acquisition of the Arabic gender and number systems. PhD dissertation. University of Wales Bangor.

Mourchid, Y. (1998). The acquisition of case-marking in Moroccan Arabic as a first language. Master Thesis. University of Southern California.

Mustafawi, E., & Mahfoudhi, A. (2002). The development of binding principles: New findings. *Cahiers linguistiques d'Ottawa*, 30, 91–111.

Ntelitheos, D., & Idrissi, A. (2017). Language growth in child Emirati Arabic. In H. Ouali (Ed.), *Perspectives on Arabic linguistics XXIX* (pp. 229–248). Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/sal.5.10nte

Paterson, K., Almabruk, A., McGowan, V., White, S., & Jordan, T. (2015). Effects of word length on eye movement control: The evidence from Arabic. *Psychonomic Bulletin & Review*, 22, 1443–1450. https://doi.org/10.3758/s13423-015-0809-4

Perea, M., Mallouh, R. A., & Carreiras, M. (2010). The search of an input coding scheme: Transposed-letter priming in Arabic. *Psychonomic Bulletin and Review*, 17, 375–380. https://doi.org/10.3758/PBR.17.3.375

Prunet, J.-F., Béland, R., & Idrissi, A. (2000). The mental representation of Semitic words. *Linguistic Inquiry*, 31, 609–648. https://doi.org/10.1162/002438900554497

Ravid, D., & Farah, R. (1999). Learning about noun plurals in early Palestinian Arabic. *First Language*, 19, 187–206. https://doi.org/10.1177/014272379901905603

Ravid, D., & Hayek, L. (2003). Learning about different ways of expressing number in the development of Palestinian Arabic. *First Language*, 23, 41–63. https://doi.org/10.1177/0142723703023001003
Remya, P. G., Anwar, R., & Jeyasekhar, S. (2011). Visual reading patterns on Arabic interfaces: Insights from eye tracking. *Journal of Computing*, 3, 11, 109–114.
Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125–157. https://doi.org/10.1037/0033-295X.105.1.125
Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26, 445–476. https://doi.org/10.1017/S0140525X03000104
Rola, F. (2005). The emergence of noun plurals in Palestinian Arabic: A case study. Paper presented in the *Workshop on the acquisition of morphology in Palestinian Arabic: Universal, particular and contextual factors*. Tel Aviv University, Tel Aviv, Israel, Friday, July 29, 2005.
Roman, G., & Pavard, B. (1987). *A comparative study: How we read Arabic and French*. Amsterdam: Elsevier.
Roman, G., Pavard, B., & Asselah, B. (1985). Traitement perceptive des phrases ambigues en Arabe. *Cahiers de Psychologie Cognitive*, 5, 5–22.
Rvachew, S., Royle, P., Gonnerman, L. M., Stanké, B., Marquis, A., & Herbay, A. (2017). Development of a tool to screen risk of literacy delays in French-speaking children: PHOPHLO. *Canadian Journal of Speech-Language Pathology and Audiology*, 41(3), 321–340.
Safi, M. S. Z. (2006). Language Pathology. In K. Versteegh (Ed.), *Encyclopedia of Arabic language and linguistics* (pp. 687–694). Leiden: Brill.
Saiegh-Haddad, E., & Schiff, R. (2016). The impact of diglossia on voweled and unvoweled word reading in Arabic: A developmental study from childhood to adolescence. *Scientific Studies of Reading* 20, 311–324. https://doi.org/10.1080/10888438.2016.1180526
Salameh, E.-K., Nettelbladt, U., & Norlin, K. (2003). Assessing phonologies in bilingual Swedish-Arabic children with and without language impairment. *Child Language Teaching and Therapy*, 19, 338–364. https://doi.org/10.1191/0265659003ct2580a
Salameh, E.-K., Håkansson, G., & Nettelbladt, U. (2003). Developmental perspectives on bilingual Swedish-Arabic children with and without impairment: A longitudinal study. *International Journal of Language and Communication Disorders*, 39, 65–91. https://doi.org/10.1080/13682820310001595628
Schluter, K. T. (2013). Hearing words without structure: Subliminal speech priming and the organization of the Moroccan Arabic lexicon. PhD dissertation. The University of Arizona, Tucson, AZ.
Shaalan, S. (2010). Investigating grammatical complexity in Gulf-Arabic speaking children with specific language impairment. PhD dissertation. University College London, UK.
Shahin, K. (1995). Child Language evidence on Palestinian Arabic phonology. In E. Clark (Ed.) *Proceedings of the 26th child language research forum* (pp. 104–116). Stanford, CSLI.
Shahin, K. (2002). Remarks on the speech of Arabic-speaking children with cleft palate. *California Linguistics Notes*, 27, 1, 1–10.
Smadi, O. M. (1980). The acquisition of the Jordanian Arabic interrogation and negation by a three-year-old native speaker of Jordanian Arabic. PhD dissertation. The University of Texas at Austin, TX.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords and phonological processes: Some comments on Gathercole, Willis, Emslie and Baddeley. *Applied Psycholinguistics*, 12, 369–373. https://doi.org/10.1017/S0142716400009279

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268 (5217), 1632–1634. https://doi.org/10.1126/science.7777863

Twist, A. E. (2006). A psycholinguistic investigation of the verbal morphology of Maltese. PhD dissertation. University of Arizona, Tucson, AZ.

Ussishkin, A., Dawson, C. R., Wedel, A., & Schluter, K. (2015). Auditory masked priming in Maltese spoken word recognition. *Language, Cognition and Neuroscience* 30(9), 1096–1115. https://doi.org/10.1080/23273798.2015.1005635

Walter, M. A. (2006). Child language. In K. Versteegh (Ed.) *Encyclopedia of Arabic language and linguistics* (pp. 375–378). Leiden: Brill.

Westerlund, M., Kastner, I., Al Kaabi, M., & Pylkkänen, L. (2015). The LATL as locus of composition: MEG evidence from English and Arabic. *Brain and Language*, 141, 124–134. https://doi.org/10.1016/j.bandl.2014.12.003

Yeou, M. and Maeda, S. (2011). Airflow and acoustic modelling of pharyngeal and uvular consonants in Moroccan Arabic. In Z. Hassan and B. Heselwood (Eds.). *Instrumental studies in Arabic Phonetics* (pp. 141–164). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.319.07yeo

Zawaydeh, B. A. (2003). The interaction of the phonetics and phonology of gutturals. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology* (pp. 279–292). Cambridge: Cambridge University Press.

Zeroual, C., Esling, J. H., & Hoole, P. (2011). EMA, endoscopic, ultrasound and acoustic study of two secondary articulations in Moroccan Arabic. In B. Heselwood, & Z. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (pp. 277–297). Amsterdam, John Benjamins. https://doi.org/10.1075/cilt.319

I. Experimental phonetics

Articulatory and acoustic correlates of pharyngealization and pharyngealization spread in Cairene Arabic

A real-time magnetic resonance imaging study

Zainab Hermes[1], Marissa Barlaz[2], Ryan Shosted[2], Zhi-Pei Liang[2] and Brad P. Sutton[2]
[1]University of Chicago, IL / [2]University of Illinois at Urbana-Champaign, IL

Pharyngealized speech sounds in Arabic are articulated with a secondary posterior constriction and a lowered tongue body. This articulatory configuration spreads to adjacent and neighboring segments and has the acoustic consequence of lowering F2 in affected vowels. This study demonstrates that real-time Magnetic Resonance Imaging (rtMRI) can be successfully used to examine the role of (1) vowel length and (2) direction of spread in the extent of the articulatory modifications that occur in the segments to which pharyngealization spreads. Parallel acoustic measurements are also acquired to examine and compare the extent of modifications in formant frequencies. Results from both articulatory and acoustic data demonstrate that the extent of pharyngealization spread significantly varies with respect to these two factors.

Keywords: pharyngealization, pharyngealization spread, Magnetic Resonance Imaging, Cairene Arabic

1. Introduction

Pharyngealization is a phonemic feature of many living Semitic languages, including Arabic, Modern Aramaic, Modern South Arabian, and Ethiopic (Hetzron, 1998). This feature is articulated in Arabic with a secondary constriction in the posterior velopharyngeal and/or pharyngeal region of the vocal tract (Versteegh, 2001). The exact place of this secondary constriction varies according to several factors, including dialect, phonological context, and gender (Khattab, Al-Tamimi, & Heselwood, 2006), and has thus been described as "velarization", "uvularization", and "pharyngealization" (Lehn, 1963). The secondary constriction associated

https://doi.org/10.1075/sal.10.01her

with pharyngealization has the effect of modifying the auditory quality of the pharyngealized speech sound and surrounding vowels and consonants, resulting in an impressionistic auditory quality described as "dark", "heavy", "dull", "thick" (Wahba, 1996), and "intense" (Watson, 2002). The set of pharyngealized speech sounds in Arabic contrasts with their *plain* counterparts, in which the secondary posterior constriction is absent. The set of pharyngealized consonants differs slightly across dialects. Altogether, the various Arabic dialects distinguish five plain-pharyngealized consonant pairs, including stops and fricatives: /t-tˤ, d-dˤ, ð-ðˤ, z-zˤ, s-sˤ/.[1] No dialect has all five pairs.

This study utilizes real-time Magnetic Resonance Imaging (rtMRI) to investigate the articulatory configuration associated with pharyngealization and pharyngealization spread in Cairene Arabic. The real-time Magnetic Resonance Imaging (rtMRI) data in this study provides high-resolution holistic images of the vocal tract in real-time. This method is non-invasive, non-ionizing, and does not drastically inhibit articulatory motion. Additionally, acoustic measures of formant frequencies in this study allow for matching the articulatory configuration with acoustic output.

Pharyngealization is an interesting topic for phoneticians because the secondary articulation occurs in the relatively inaccessible posterior velopharyngeal or pharyngeal region. Instrumental studies have attempted to obtain partial or holistic views of this region during speech production in order to describe the articulation of these sounds precisely. For phonologists, pharyngealization is a suprasegmental phenomenon that spreads according to various patterns that differ across dialects. Differences include the domain and direction of the spread, as well as the presence and behavior of opaque elements (Davis, 1995). Pharyngealization spread has been described as a type of postvelar harmony and has received considerable attention in formal autosegmental approaches (Shahin, 2002, among others). The current study presents a phonetic explanation for some of the phonological patterns that have been observed. Pharyngealization is also of interest to sociolinguists, with studies reporting gender-driven and class-driven differences in the amount of pharyngealization and pharyngealization spread a speaker produces (Royal, 1985; Wahba, 1996). Weaker pharyngealization is reported to mark feminine or effeminate speech and a greater level of social prestige (Royal, 1985). The present study offers insights into how the articulation of pharyngealization may be studied across different demographic groups.

---

1. The IPA symbol for pharyngealization is a superscripted ⟨ˤ⟩. Thus, a pharyngealized voiceless alveolar stop is represented as /tˤ/. Another convention, found in the literature, places a dot beneath the pharyngealized sound. Thus, /tˤ/ is represented as /ṭ/ in that convention. Throughout this text, IPA conventions will be followed, except in citations, in which case the citation will be presented in its original form with the transcription conventions chosen by its author. The IPA symbols different from those in the citation will then be presented in a footnote.

2. Review of literature

2.1 Acoustic correlates of pharyngealization

Arabic pharyngealized consonants are articulated with a secondary constriction in the back velopharyngeal and/or pharyngeal regions of the vocal tract. According to Perturbation Theory (Chiba & Kajiyama, 1941), the velopharyngeal region corresponds to a node (a point of maximum pressure) in the standing wave of the first formant frequency F1, and an anti-node (a point of maximum velocity) in the standing wave of the second formant frequency F2. Perturbation Theory predicts that the resonant frequency will increase if a constriction occurs at a node, and will decrease if the constriction occurs at an anti-node. Thus, in line with these predictions, acoustic studies on Arabic pharyngealized consonants have reported a raised F1 and a lowered F2 in the onsets of vowels following the pharyngealized sound, and in the offsets of vowels preceding it (Jongman, Herd, Al-Masri, Sereno, & Combest, 2011; Khattab et al., 2006; Watson, 2002). These modifications affect surrounding speech sounds, yielding an auditory quality that has been impressionistically described as "dark", "heavy" and "dull" (Wahba, 1996). Another term used by early Arab grammarians, and later by Jakobson (1957), is *mufaxxam* 'intensified' (Watson, 2002). The sounds have also been described as "emphatic", a term that Giannini and Pettorino (1982) suggested was coined by De Sacy (1810) in reference to the "thick" or "heavy" quality of the speech sound. They quote his description of these sounds as being articulated "*avec une sorte d'emphase. Ce que j'appelle emphase ou articulation emphatique est une espèce de renflement qu'il n'est pas aisé de definir mais qui fait en quelque sorte entendre un o sourd après la consonne* (with a sort of emphasis. What I call emphasis or emphatic articulation is a sort of enlarging that is not easy to define, but that leaves one somewhat hearing a hidden 'o' after the consonant)" (De Sacy, 1810, p. 20, as cited in Giannini & Pettorino, 1982, p. 10; translation by Zainab Hermes). The extent of the modification in the formant frequencies of adjacent vowels is determined by the length and quality of the vowel (Yeou, 1997). In his acoustic study, Yeou reported a greater effect of pharyngealization on F1 and F2 in the short vowels adjacent to pharyngealized consonants, relative to their long counterparts, i.e. he reported more raising of F1 and more lowering of F2 in the short vowels. The largest modifications in F1 and F2 occur in the vowel /a/, followed by /i/ and /u/ (Jongman et al., 2011; Yeou, 1997).

2.2 Articulatory correlates of pharyngealization

2.2.1 *Early Arab grammarians*

The pharyngealization feature in Arabic was identified and described as early as the 8th century by Arab grammarians, who referred to the phenomenon as *iṭbāq* (spreading and raising of the back of the tongue), *istiʿlāʾ* (elevation of the dorsum), and *tafkhīm* (thickness or heaviness) (Lehn, 1963). Arab grammarians identified four pharyngealized sounds in Classical Arabic that they termed the *muṭbaqa* sounds, meaning 'covered' or 'enclosed': /tˤ, dˤ, sˤ, ðˤ/. All other speech sounds were termed *ghayr muṭbaqa* (not closed) or *munfatiħa* (opened). Sibawayh (d. 796) described the *muṭbaqa* sounds as having two places of articulation: one involving the front of the tongue rising to form the coronal constriction, and the other involving the back of the tongue rising toward the velum, thus creating a configuration in which the sound is 'enclosed' between these two places. He noted that the secondary articulation is what distinguished the *muṭbaqa* (pharyngealized) sounds from their plain counterparts, and wrote, for example: "without *[iṭbāq], ṣād* [i.e. /sˤ/] would be a *sīn* [i.e. /s/]" (Vollers, 1893, p. 150). Ibn Sina (known in Medieval Europe as Avicenna; d. 1037) related the difference between the two sounds to the shape and volume of the oral cavity associated with each sound. He wrote: "The production of /s/ is similar to (that) of /ṣ/[2] save that, (in /ṣ/) the passage of the air requires (the use) of a larger surface of the tongue both in length and breadth. A sort of hollow is formed in the tongue (surface) to give the rolling of the air a certain resonance." (Semaan, 1963, p. 41). To the *muṭbaqa* set of sounds, three additional sounds were added /q, x, ɣ/ to form a set of sounds having the property of *istiʿlāʾ* (elevation). The articulation of this set was defined as involving the elevation of the tongue toward *al-ħanak al-aʿlā*, literally, *the upper palate*. Interestingly, as Bellem (2007) pointed out, this property was also defined as preventing the presence of another property *imāla* (inclining) which was defined as the fronting and raising of the vowel /a/ toward /i/. Thus, Bellem shows that *imāla* was associated with fronting, whereas *istiʿlāʾ* was associated with backing. The auditory quality of the *istiʿlāʾ* sounds was referred to as *tafkhīm* meaning 'intensification' (Watson, 2002), 'aggrandizing', or 'puffing up' (Bellem, 2007). Thus, the seven *istiʿlāʾ* sounds were also *tafkhīm* sounds. Additionally, /r/ and /l/ in specific contexts were described as having the *tafkhīm* quality (Lehn, 1963). Furthermore, varying degrees of *tafkhīm* were identified based on surrounding vowels, the strongest being when one of the seven sounds is followed by a long /a:/.[3]

---

2. The symbol /ṣ/ corresponds to IPA /sˤ/.

3. Unless otherwise explicitly stated, /a/ represents the front vowel.

Early Arab grammarians understood that the presence of a *tafkhīm* sound influenced its vocalic context. It was particularly well-noted that the vowels immediately following a *tafkhīm* sound were affected. For example, an allophonic variant of /a/ having the *tafkhīm* quality was noted when preceded by a *tafkhīm* sound. Some grammarians also described longer distance pharyngealization spread to other consonants beyond the immediately adjacent vowels, such as al-Suyūṭī (d. 1505) in the following passage in Bellem (2007) that describes a right-to-left spread: "Every *sīn* [/s/] which is followed by *ʿayn* [/ʕ/] or *ghayn* [/ɣ/] or *khāʾ* [/x/] or *qāf* [/q/] or *ṭāʾ* [/tˤ/] may be transformed into *ṣād* [/sˤ/], for instance *yusāqūna* and *yuṣāqūna*[4] … the condition for this is that the *sīn* [/s/] should precede these consonants (i.e. the elevated ones) not follow after them, and that these consonants should be close to it." (Bellem, 2007, p. 25). This right-to-left spreading pattern is observed in many contemporary dialects of Arabic such as Cairene (Youssef, 2014) and Palestinian (Herzallah, 1990).

2.2.2 *More recent literature*

Various experimental methods were used to observe the posterior place of the secondary constriction associated with pharyngealization in different dialects of Arabic. These studies suggested that its exact place differs across dialects (Al-Tamimi & Heselwood, 2011; Norlin, 1987; Obrecht, 1968), and across different phonological environments within the same dialect (Al-Tamimi & Heselwood, 2011; Obrecht, 1968). Thus, in Jordanian Arabic, the secondary articulatory gesture associated with pharyngealization is described as a constriction in the lower oropharynx involving the folding and backward retraction of the epiglottis, in a motion comparable to that implicated in the production of pharyngeal consonants (Al-Tamimi & Heselwood, 2011 and Heselwood & Al-Tamimi, 2011). This is the literal definition of pharyngealization. The oropharynx was defined in Zemlin (1998) as having a superior limit at the level of the soft palate, and a lower boundary at the level of the hyoid bone (Zemlin, 1998, p. 227). In the context of the back vowel /u:/, the articulation of pharyngealized consonants is described as uvularization, involving the approximation of the back of the tongue with the uvula (Heselwood & Al-Tamimi, 2011). A study of Lebanese Arabic described this secondary articulation as velarization (Obrecht, 1968). Additional anterior articulatory gestures are implicated in pharyngealization, as well. These include labialization and the lowering and concavity of the tongue body (Lehn, 1963 and Watson, 2002).

Norlin (1987) summed up the claim that articulatory strategy differs by dialect: "The modern dialects show a wide spectrum of phonemic contrasts in their

4. IPA yusa:qu:na and yusˤa:qu:na respectively.

segments, both among themselves and vis-à-vis Standard Arabic. It should by no means be impossible that [pharyngealization], although certainly a universal phenomenon in all forms of Arabic, might be realised in different ways and degrees in different dialects" (Norlin, 1987, p. 13). In addition to the dialect, and the phonological context, Khattab et al. (2006) suggested that the gender of the speaker may also be a factor in determining the secondary place: "There is no consistent single articulatory exponent of [pharyngealization]. Rather, speakers have a range of articulatory strategies at their disposal, including how high in the pharynx to create a constriction. Which strategy a speaker uses will depend on several factors that may include not only native dialect and phonological context, but also gender and possibly other social variables" (Khattab et al., 2006, p. 140).

Studying the secondary articulation of pharyngealization can be incredibly challenging because of the posterior and relatively inaccessible place in which this articulation occurs. Experimental techniques that were employed previously include nasoendoscopy (Heselwood & Al-Tamimi, 2011), videofluoroscopy (Heselwood & Al-Tamimi, 2011), ultrasound imaging (Lapinskaya, 2013 and Zeroual, Esling, Hoole, & Ridouane, 2011b), electromagnetic articulography (EMA) (Embarki, Ouni, Yeou, Guilleminot, & Al Maqtari, 2011b; Hermes, Wong, Loucks, & Shosted., 2015; Zeroual, Esling, & Hoole., 2011a), and MRI (Shosted, Benmamoun, &. Sutton., 2012).

2.3 MRI in phonetic studies on Arabic

Perhaps the most powerful technique for studying pharyngealization is rtMRI. This technique can provide several simultaneous views of the vocal tract: midsagittal, lateral, coronal, as well as other oblique angles specified by the researcher. In the midsagittal view, MRI provides a full view of the vocal tract, including the posterior velopharyngeal region of interest in the study of pharyngealized consonants. This technique is also non-invasive and does not need to be administered by a medical doctor. The limitations, however, are that it is relatively more costly, and, although it is non-invasive and does not involve subjecting the speakers to ionizing radiation, it is still sometimes intimidating for some speakers. Thus, recruiting participants for an MRI study can be challenging. The audio quality of acoustic data acquired inside an MRI scanner is also deprecated. Special noise-canceling hardware and advanced filtering techniques are required to preserve much of the high-frequency energy associated with phonetic contrasts.

A number of studies on Arabic speech sounds in which the pharyngeal region is implicated employed MRI. Shar and Ingram (2011) compared pharyngeal width and laryngeal height during Arabic gutturals in Assiri Arabic (spoken in Southwest Saudi Arabia). Shosted et al. (2012) used MRI to study the role of the posterior constrictions in pharyngeal and pharyngealized speech sounds in Jordanian and Moroccan Arabic. Israel et al. (2012) used real-time MRI (rtMRI) to study the posterior constriction associated with pharyngealization and the suprasegmental spread of pharyngealization in Lebanese Arabic. Their study implements the algorithm described in Proctor et al. (2010) to extract the lingual contours, and quantitatively compare between the tongue contour displacement as well as the degree of constriction between different speech sounds. Proctor et al. (2010) identified tissue boundaries of the vocal tract in rtMRI data by examining pixel intensities along tract-normal gridlines.

3. Pharyngealization as a phonological feature

3.1 A discussion of coarticulation, coproduction, and spread

Classic phonological theories build on the notion of abstract, mental representations of speech sounds (phonemes) associated with a standard articulation or 'motor plan' (Kühnert & Nolan, 1999). Coarticulation accounts for the observation that these units of sound are not realized identically in all phonological environments, rather they are influenced by surrounding segments. Coarticulation can be attributed, in part, to (1) the physiological constraints of the vocal mechanism, especially in real time, such that it is not possible to instantaneously jump from one articulatory configuration to another, and (2) the setup of our perceptual system and the observation that parallel processing of the acoustic properties of more than one phoneme at once facilitates rapid perception of speech (Kühnert & Nolan, 1999). It is noteworthy that the term coarticulation has also been used in reference to speech sounds that have two simultaneous places of articulation, such as the Arabic pharyngealized consonants that have a primary coronal place and a simultaneous back (velar, uvular, or pharyngeal) place.

The motor constraints of the speech mechanism represent *phonetic* constraints that cause coarticulation. Cross-linguistic studies of coarticulation have further revealed that coarticulatory patterns vary according to language-specific *phonological* constraints. Findings by Manuel and Krakow (1984) showed that languages with considerably larger vowel inventories, such as English, exhibit less coarticulatory effects on vowels. They attribute this to the necessity of maintaining the distinctiveness of vowels in such languages. In contrast, the languages that had relatively fewer distinct vowels in their study (Swahili and Shona) exhibited more coarticulatory effects. They thus argue that coarticulation is determined not only by constraints of motor planning and execution, but also by constraints defined by phonology (here, the size of the phonemic vowel inventory). Thus, it is important to distinguish between coarticulatory effects that are a consequence of the speech producing mechanism in general, and those that are language-specific. The latter correspond to phonological spread patterns. A more extended discussion of the phonological patterns associated with pharyngealization spread will follow.

With regards to vowel inventory, the number of distinct Cairene Arabic vowels (in terms of quality) is five. Like Standard Arabic, Cairene Arabic contrasts /i/, /u/, and /a/, but Cairene Arabic includes two additional close-mid vowels /e/ and /o/ (Gadalla, 2000). In both Standard and Cairene Arabic, vowel length is contrastive for /i/, /u/, and /a/. The vowels /e:/ and /o:/ in Cairene Arabic have no short counterparts (Gadalla, 2000).

Yeou (1997) argued that in Arabic, pharyngealized consonants are the most resistant consonants to coarticulatory effects from adjacent vowels due to the greater constraints on the tongue body required for their articulation. These are the constraints required for achieving (1) the primary (coronal) articulation, and simultaneously (2) the back pharyngeal constriction. He compares this with the English velarized approximant [$l^{ɣ}$] (dark l), the production of which requires the back of the tongue to be raised toward the velum in addition to maintaining the alveolar constriction. This, he argues, makes it more resistant to coarticulatory effects from adjacent vowels. In contrast, the palatal [l^{j}] (light l) is less resistant to coarticulation because of the lesser demands – defined here by the number of gestures – imposed on the tongue to produce this sound.

These observations are in line with findings from Embarki, Ouni, and Salam (2011a) and Embarki et al. (2011b). Both studies investigate the coarticulation patterns triggered by the pharyngealized and contrastive non-pharyngealized consonants across the Yemeni, Kuwaiti, Jordanian, and Moroccan dialects of Arabic. They used locus equations (Yeou, 1997) to measure the amount of coarticulation at consonant-vowel boundaries. Their findings report maximal coarticulatory resistance of the pharyngealized consonants to vowel effects across all dialects. Yeou (1997) remarked that speech sounds that are resistant to coarticulation induce greater coarticulatory effects on adjacent segments. This type of coarticulation would thus correspond to phonological spread. Thus, the coarticulatory effects of pharyngealized consonants are described phonologically as the spread of the pharyngealization feature.

3.2 Phonological studies on the spread of pharyngealization

It has been noted that phonemic inventories differ slightly across Arabic dialects. In addition to the phonemic inventories, Embarki et al. (2011b) also noted that the "gestural and temporal adjustments during larger units" also differ across dialects: "[P]honetic sets […] differ regularly from one Arab country to another. This concerns not only the ways in which the phonemes are produced, but also the gestural and temporal adjustments during larger units such as syllables, phonological words, and other prosodic domains. It is therefore reasonable to assume that coarticulation, which is the result of gestural and temporal adjustments, may differ significantly from one [Arabic dialect] to another." (Embarki et al., 2011b, p. 195–196)

Consequently, the patterns observed for pharyngealization spread differ across dialects. These include differences in the direction of the spread, the domain (distance) of the spread, and the presence and type of opaque elements and their behavior. Moreover, the segments that trigger this spread also vary. Segments that are described as underlyingly or intrinsically pharyngealized, i.e. pharyngealized speech sounds that are phonemically contrastive with plain counterparts, always trigger pharyngealization spread. For Egyptian (Cairene) Arabic, Youssef (2014) argued that the set of segments that trigger pharyngealization spread consists of the coronals /tˤ, dˤ, sˤ, zˤ, rˤ/, and he argues for treating the low back vowel /ɑ/ as an independent phoneme (the opposing view treats it as an allophone of /a/ ([ɑ])), and thus includes it as a segment that triggers pharyngealization spread. In this view then, the pharyngeal quality of /ɑ/ is intrinsic rather than a consequence of spread from surrounding back segments. He also shows that pharyngealization spread in this dialect is bidirectional with no opaque elements. The domain of this spread is the syllable as well as the phonological word including all prefixes and suffixes (Youssef, 2014).

4. Research statement

This study uses rtMRI to describe phonetic correlates of pharyngealization and pharyngealization spread in Cairene Arabic and compares the strength of pharyngealization spread in different directions and across vowels of different lengths. This is done in both the articulatory and acoustic domains. Results from both domains are linked together.

5. Data acquisition

5.1 Articulatory data acquisition

Real-time dynamic Magnetic Resonance Imaging (rtMRI; Fu et al., 2015) is used to investigate the articulatory configuration associated with pharyngealization and pharyngealization spread in Cairene Arabic. The major strength of rtMRI (Fu et al., 2015) is that it provides holistic imaging of the vocal tract with a high spatiotemporal resolution, which allows for capturing the complex dynamics of natural speech

(Lingala, Sutton, Miquel, & Nayak., 2016). High temporal resolution is necessary for capturing the fast-varying dynamics of speech; while high spatial resolution is necessary for capturing the fine features of the articulators implicated in natural speech (Fu et al., 2015). In order to achieve high spatiotemporal resolution, an MR image must be acquired at twice the highest component frequency in the dynamic signal, according to the Nyquist sampling theorem. This requirement is difficult to achieve with limited MR imaging speed. The method developed by Fu et al. (2015), however, makes use of the Partial Separability (PS) model to evade this sampling requirement and allows for decomposing the dynamic signal into partially separable spatial and temporal components. This method meets the requirements of the 2014 Speech MRI Summit (Lingala et al., 2016). The temporal resolution achieved by this method (i.e. the reconstructed sampling frequency) is approximately 100 fps. The spatial resolution achieved for each reconstructed frame is 128 × 128 voxels (volume elements), with each voxel measuring 2.2 mm × 2.2 mm × 8.0 mm (through-plane depth) (Carignan, Shosted, Fu, Liang, & Sutton., 2015).

The articulatory rtMRI data presented in this work were collected from four male native speakers of Cairene Arabic SP1, SP2, SP3, and SP4. The choice of male speakers for this study was made because males typically produce more pharyngealization and stronger pharyngealization than females (Royal, 1985). The effect of sex on the production of pharyngealization would be an interesting question to investigate in further studies.

Acquisition of the rtMRI data was conducted at the Biomedical Imaging Center in the Beckman Institute for Advanced Science and Technology at the University of Illinois in Urbana-Champaign. Prior to the acquisition, the participants were familiarized with the task of repeating the required test material (described in detail in the following section) in their native Cairene dialect. After training, the participant lies supine inside the Magnetom Trio 3T scanner and repeats each test item for approximately 90 seconds (i.e. 1.5 minutes). The reconstruction algorithm identifies the regular changes that occur in the vocal tract, and by focusing on those specific changes, the algorithm can achieve a high resolution. Inside the scanner, the speaker's head is fitted with a head/neck coil and rests on a NoMoCo support system (NoMoCo Pillow, Inc., San Diego, CA). This is in order to reduce movement. As for the effect of posture on the dimensions of the upper airway, a previous study by Jan, Marshall, and Douglas (1994) found that pharyngeal cross-sectional areas slightly decrease when the speaker assumes a supine position. Measurements of the perimeter at the oropharyngeal junction drop from 1.65 ±0.6 cm to 1.31 ±0.07 cm. This suggests that the study of pharyngealization (and vocal dynamics in general) would be better suited to upright MRI technology. Given the limited availability of such scanners, speech imaging researchers have largely accepted this limitation. Furthermore, the speaker lies supine throughout the entire acquisition, and thus, it is reasonable to assume that the consequence of this posture will affect all the speech segments in the same manner. Thus, upon comparing the pharyngeal articulatory configuration in a given target word having a pharyngealized segment, with the pharyngeal articulatory configuration having the contrastive non-pharyngealized (plain) counterpart, the assumption is that the difference represents the effect of the pharyngealization, which is the subject of the present study. This would be the case even if, initially, both articulatory configurations are influenced by the supine posture of the speaker. During scans, the speaker also wears an MR-compatible headset with an attached optical noise-cancelling microphone (Dual Channel-FOMRI, Optoacoustics, Or Yehuda, Israel) to record simultaneous acoustic data. A noise cancellation method is applied to attenuate some of the background noise produced by the MR scanner during acquisition. This acoustic data is used to determine the start and end times of the individual speech sounds of interest in Praat (Boersma & Weenink, 2013).

The rtMRI data acquired for this study were midsagittal. The midsagittal plane was located after acquiring a three-dimensional scan of the head. Using this imaging volume, coronal and axial views were used to interactively place the sagittal slice at the prescribed location. Figure 1 shows an example. This plane provides a holistic view of the vocal tract, including the back pharyngeal and velopharyngeal regions, as well as the lingual contour in the oral cavity. Both the pharyngeal and the lingual configurations are implicated in the study of pharyngealization.

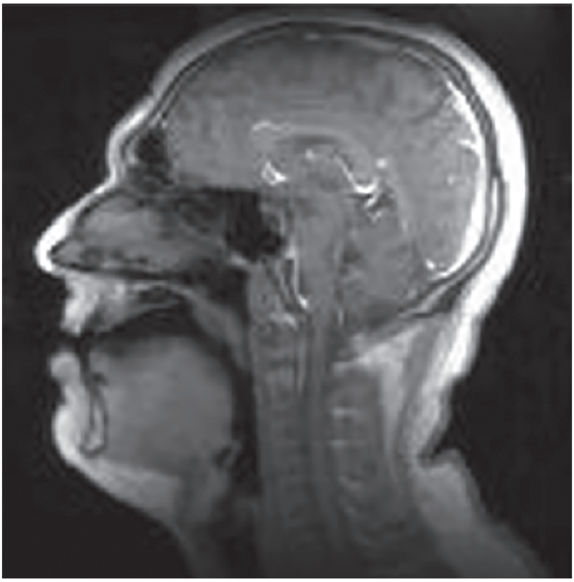


Figure 1. An illustration of a midsagittal view acquired from SP1

5.2 Acoustic data acquisition

Two sets of acoustic data were acquired. The first was acquired while the speaker was inside the MRI scanner and it was simultaneous with the articulatory data acquisition. This acoustic data is synchronized with the rtMRI articulatory data and is used to determine the MR images corresponding to the speech segments of interest. The MR scanner produces considerable background noise during acquisition. A noise cancellation method is applied via software associated with the microphone to attenuate some of this background noise. The resulting acoustics are essential for identifying the start and end times of the speech segments. While formant frequencies are visible in these acoustic data, they remain overall noisy. Therefore, a second set of (cleaner) acoustic data was collected in a sound-attenuated recording facility at the Phonetics and Phonology Lab at Illinois from the same speaker. The speaker lies supine in the booth in a similar fashion as he did in the MRI scanner, to control for effects of gravity on the aperture of the pharynx. An AKG C520 head-worn condenser microphone is positioned at the corner of the speaker's mouth, approximately 3 cm away to reduce the effects of turbulent airflow. Speech is recorded on a Marantz PMD570 recorder outside of the booth. The data in this second cleaner set are used to measure formant frequencies.

6. Test material

To elicit natural speech, the speakers are required to repeat a sentence comprised of target words embedded in the carrier phrase of Cairene Arabic: /ʔallaha X alf marra/ ('He told her X one thousand times' where X is the target word). The carrier phrase is constructed such that the words immediately preceding and following the target item contain no posterior sounds (the vowel /a/ in the words immediately preceding and following the target word is fronted in this context). This is to avoid any potential effect back vowels may have on the target word, minimizing the risk of postvelar harmony due to the tongue root retraction associated with a low back vowel. The target words are selected in order to allow for examining the following:

1. the role of the length of the vowel in the amount of pharyngealization spread
2. the influence of rightward versus leftward spread of pharyngealization

The target words are real-word minimal pairs contrastive in the plain-pharyngealized /sˤ/-/s/, and in which the vowels immediately adjacent to the plain-pharyngealized contrast are /a:/ and /a/. The target words are monosyllabic, with the plain-pharyngealized contrast occurring at the edge of the word (i.e. at the

Table 1. Target words contrastive in plain-pharyngealized /s/-/sˤ/ occurring word-initially or word-finally

Condition	Word-initial		Condition	Word-final	
#Ca:	/sˤa:b/	he hit [a target]	a:C#	/ba:sˤ/	Bus
	/sa:b/	he let go		/ba:s/	he kissed
#Ca	/sˤabb/	he poured	aC#	/basˤsˤ/	he looked
	/sabb/	he insulted		/bass/	enough!

beginning of the word, or at the end of the word). All target words used in this study are listed in Table 1.

The same target words were used during the acquisition of both the articulatory and the acoustic data. For each target word, the speaker was asked to repeat the carrier phrase for 90 seconds. There were, naturally, different speaking rates for each speaker, resulting in a different number of repetitions for each target word by each speaker. Thus, the first 20 valid repetitions from each speaker were included in this study in order to ensure an equal amount of data from each speaker.

7. Data analysis

7.1 Articulatory data analysis

The acoustic data that was acquired simultaneously with the articulatory rtMRI data was segmented in Praat (Boersma & Weenink, 2013) by the first author. The segmentation was based on visual inspection of the waveforms and the wide-band spectrograms associated with them, as well as on auditory inspection of the acoustics. From every repetition of each target word, three segments were identified: C1, V, C2 (the first consonant in the word, the vowel, and the second consonant, respectively). The plain-pharyngealized contrast was either C1 or C2. Accordingly, the start and end times of these segments were extracted. Figure 2 shows an example of this segmentation.

A Matlab (R2015a) script was written to identify the MRI frames corresponding to the segmented speech sounds based on the start and end times. The middle frame within this sequence was selected because it corresponds to the temporal midpoint of the speech sound and can therefore be a good representative of the shape of the vocal tract during that speech sound.

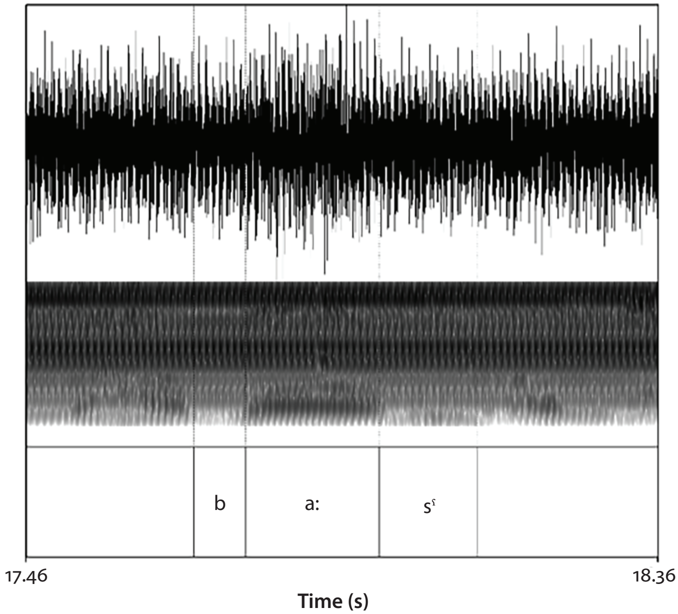


Figure 2. Segmentation of the target word /ba:sˤ/ (bus) extracted from the phrase /ʔallaha ba:sˤ alf marra/ (He told her bus one thousand times). This acoustic data is acquired inside the MR scanner

After identifying the MRI frames of interest, a Matlab interface developed by Proctor et al. (2010) (see also Narayanan et al. (2014), Israel et al. (2012))[5] was used to automatically detect the contours of the vocal tract from the glottis to the lips in those frames. In this application, the interface of which is shown in Figure 3, the user is presented with an MR image and manually clicks on four anatomical landmarks: the glottis, the highest point on the palate, a point on the alveolar ridge, and a point midway (vertically) between both lips.

The xy-coordinates of these landmarks are registered, and tract-normal semi-polar gridlines (orthogonal to the midline of the vocal tract) are generated from the lips to the glottis. Based on changes in pixel intensity, tissue boundaries along the vocal tract are detected. These include the contours of the palate, the back pharyngeal wall, the tongue root (forming the anterior pharyngeal wall), and the contour of the tongue body as shown in Figure 4. These contours provide information about the pharyngeal configuration, as well as the lingual oral configuration.

5. Downloaded from: http://mproctor.net/software.html

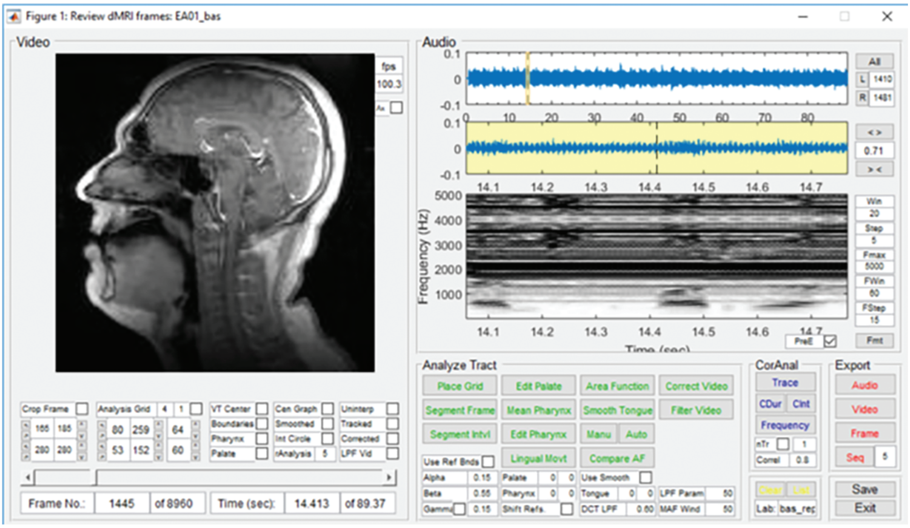


Figure 3. The Matlab interface developed by Proctor et al. (Israel et al., 2012; Narayanan et al., 2014 ; Proctor et al., 2010). The user will manually click on the MRI frame on the left to identify the required anatomical landmarks. The MRI frame is an arbitrarily selected frame from the first speaker (SP1)

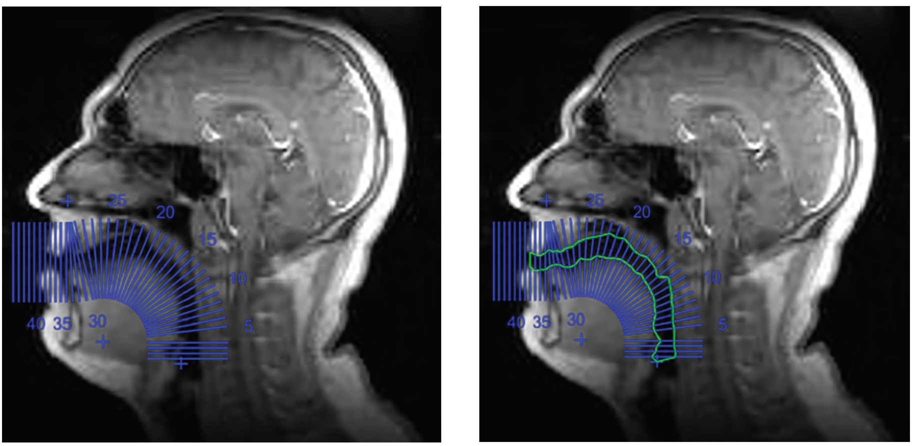


Figure 4. a. (left) Semi-polar gridlines superimposed on an arbitrarily selected MRI frame from SP1 running from the lips to the glottis. b. (right) Automatically-detected vocal tract contours in the same MRI frame in a

Results from the edge detection method implemented are dependent on the visual quality of the MR image. Specifically, it is dependent on the contrast between the vocal contours and the (black) vacuum in the vocal tract. If this contrast is clear and sharp, the algorithm can identify the edge well. Thus, the first step in using this application involves selecting two scalars that determine the brightness and contrast in the MR image. After several trials, the scalar values of 1.0 and 1.5 appeared to yield the best contrast and therefore the best detection of the lingual contours. The application developed by Proctor et al. (2010) allows the user to manually modify the results that the edge-detection method automatically yields, i.e., it is possible to adjust the vocal tract contours that are automatically detected. However, manual adjustments were avoided here for the obvious reasons that manual adjustment is subjective, and hence not reproducible. As there are several repetitions of each segment of interest, the assumption is that the effect of a possible detection error in one repetition will be attenuated, and possibly cancelled out, in averages across all the repetitions.

The lingual and pharyngeal contours for every sound of interest in every repetition of the target words were extracted and plotted. An example of this from SP4 is shown in Figure 5, illustrating the contours in the three segments in /sa:b/ and /sˤa:b/. The contours were then smoothed and averaged as shown in Figure 6. Superimposing the contours of a plain-pharyngealized minimal pair, allows for visual comparison of the vocal configuration during both words.

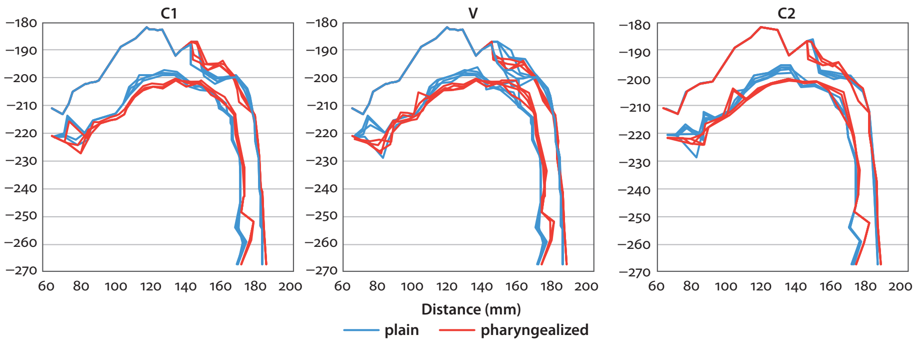


Figure 5. The pharyngeal and lingual contours automatically detected for the 20 repetitions of the target words /sa:b/ in blue ('he left') and /sˤa:b/ in red ('he hit') from SP4

In Figure 5 above, the vocal tract configuration is plotted for both the "pharyngealized word", i.e. the word containing the pharyngealized speech sound, and the "plain word", i.e. its plain contrastive counterpart. The two words are minimal pairs contrastive in the plain-pharyngealized consonant. In this example, the words that are plotted are /sa:b/ and /sˤa:b/. The first panel from the left is a plot of the vocal

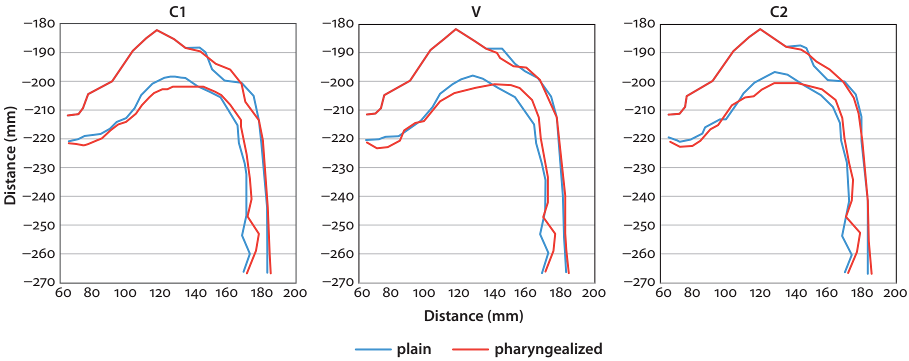


Figure 6. The smoothed pharyngeal and lingual contours averaged from the 20 repetitions of the target words /sa:b/ in blue ('he left') and /sˤa:b/ in red ('he hit') from SP4. Smoothing was performed using a Savitzky-Golay filter

tract configuration corresponding to the midconsonantal /sˤ/ extracted from the pharyngealized word and the midconsonantal /s/ extracted from the plain word. In that panel, the red contours correspond to the vocal tract configuration of the pharyngealized segment, while the blue contours correspond to the vocal tract configuration of the plain segment. Thus, it is possible to compare the vocal tract configurations during both members of the minimal pair. The middle panel is a similar plot of the /a:/ vowel. The vocal tract configuration corresponding to the middle of the /a:/ vowel extracted from the pharyngealized word is plotted in red, while that corresponding to the /a:/ vowel extracted from the plain word is plotted in blue. The third panel on the right is a similar plot of the midconsonantal /b/ extracted from the pharyngealized word in red and the midconsonantal /b/ extracted from the plain word in blue. With these plots:

1. It is possible to observe the change of the vocal tract configuration with respect to time during the articulation of one word. Thus, one can observe pharyngealization spread and the extent and direction of this spread.
2. It is also possible to visually compare the vocal tract configurations of the two words.

In Figure 6 above, these contours are averaged and smoothed in order to correct for any potential errors in the edge-detection or any noise in the image. The smoothing is done using a Savitzky-Golay filter. Using an unweighted linear least-squares regression and a second-degree polynomial model to compute coefficients, this filter produces a moving average for a given contour (Savitzky & Golay, 1964).

For the averaged and smoothed contours in Figure 6, ±95% confidence intervals are calculated and plotted in Figure 7. At the non-overlapping confidence

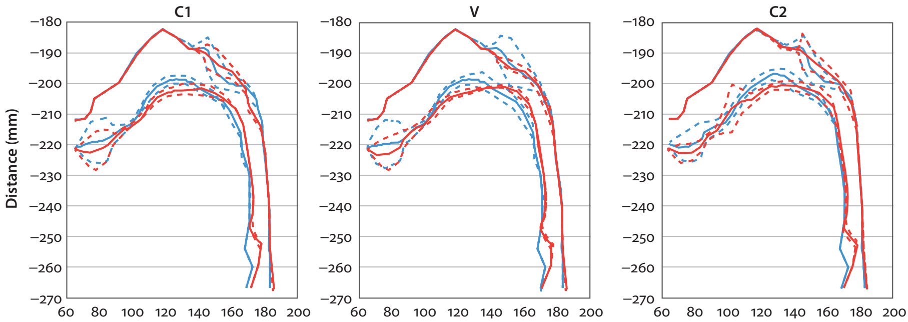


Figure 7. Averaged and smoothed pharyngeal and lingual contours with ±95% confidence intervals. Contours are extracted from the three speech sounds in target words /sa:b/ in blue ('he left') and /sˤa:b/ in red ('he hit') from SP4

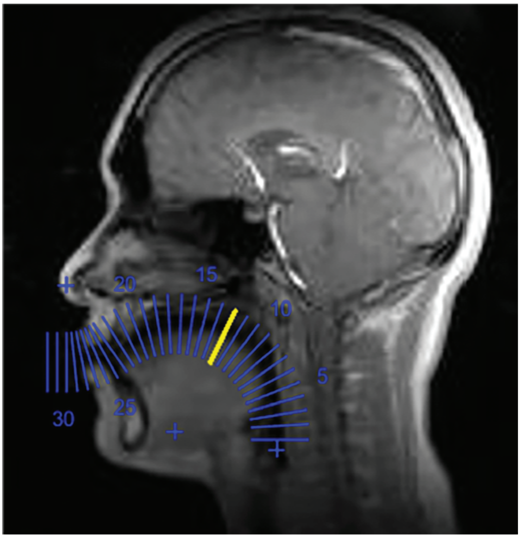
intervals of the contours, it can be asserted with 95% confidence that the differences in vocal tract constrictions are statistically significant.

In order to quantitatively compare the difference between vocal tract configurations, the two-dimensional area of the vocal tract is divided into two areas as follows:

1. The two-dimensional pharyngeal area (2D pharyngeal area) is defined as the two-dimensional area between the glottis and the end of the soft palate. This includes the pharynx.
2. The two-dimensional oral cavity (2D oral area) is defined as the two-dimensional area between the lips and the end of the soft palate. This includes the oral cavity and the velopharyngeal region.

Figure 8 illustrates these two measures. In the MRI frame on the top, the gridline roughly corresponding to the end of the soft palate is demarcated in yellow. In the plot on the bottom, the vocal tract contours are extracted, and the two areas are shown. The 2D oral area is highlighted in magenta and the 2D pharyngeal area is highlighted in green. The dividing line between the two areas corresponds to the yellow gridline in the image on the left.

The 2D pharyngeal area allows for quantifying the pharyngeal constrictions by capturing the 2D change in the configuration of the pharynx, both in the upper regions and lower regions. This measure is useful because it is observed that the constriction resulting from pharyngealization, rather than being localized at a specific point, has the effect of narrowing an extensive part of the pharynx.



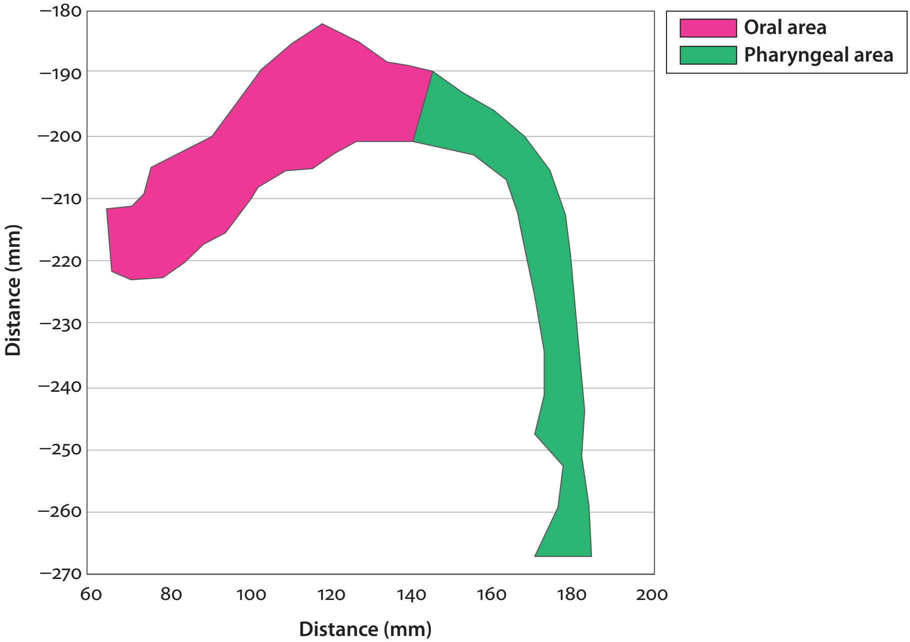


Figure 8. a. (top): The gridline in yellow roughly corresponding to the end of the soft palate. b. (bottom): An illustration of the 2D oral area in magenta and the 2D pharyngeal area in green. The dividing line between the two areas corresponds to the yellow gridline in a

7.2 Acoustic data analysis

Acoustic data analysis is conducted on the "clean" acoustics acquired in the sound-attenuating booth in the Phonetics and Phonology Lab. The speaker produced seven repetitions of the test materials. This acoustic data was also segmented in Praat (Boersma & Weenink, 2013) by the first author. The analysis is conducted on the vowels in the target words, thus only the V segment was identified in the data. In this "clean" data, the start and end of the vowel are marked from the start of the second and third formant frequencies until the end of these formants in the spectrogram. Figure 9 shows an example of the segmentation from a target word produced by SP3. The target word is /ba:sˤ/. The figure shows the segmentation of the vowel /a:/.

The first two formant frequencies F1 and F2 are measured for all vowels using FormantPro,[6] a Praat script developed by Yi Xu (Xu & Gao, 2018). Mean F1 and F2 are defined as the average formant value throughout the duration of a vowel.

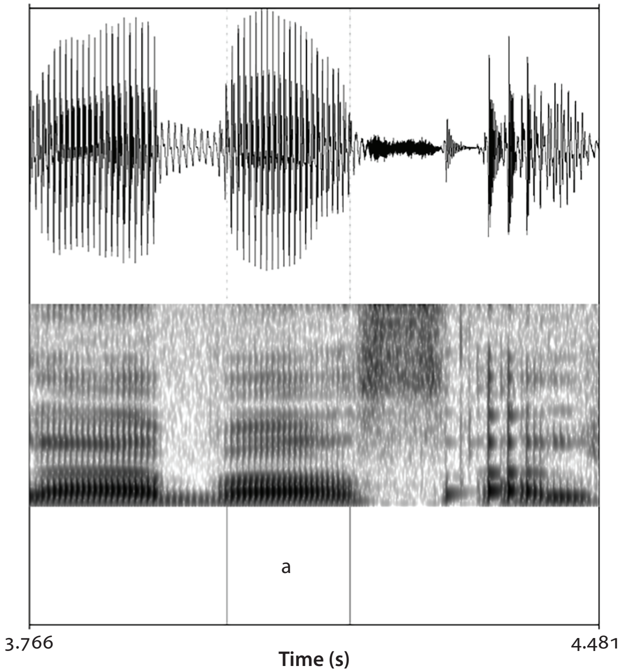


Figure 9. Segmentation of the vowel /a:/ within the target word /ba:sˤ/ ('bus') as produced by SP3

6. Downloaded from: http://www.homepages.ucl.ac.uk/~uclyyix/FormantPro/

8. Results

8.1 Results from articulatory data

The lingual and pharyngeal contours for all four speakers reveal several observations. Firstly, the pharyngealized consonant is produced with a more constricted pharynx than its plain counterpart. This is the case, regardless of the vowel length, and regardless of whether the plain-pharyngealized contrast occurs at the end of the word (in the coda position within a syllable), or at the beginning of the word (in the onset position within a syllable). Furthermore, the constriction of the pharynx is not restricted to one specific point in the pharynx; instead, it extends throughout the pharynx. Additionally, a lowered tongue body is sometimes observed in the pharyngealized consonant when compared to its plain counterpart. This observation is evident visually from tongue contour plots. Quantitatively, the constricted pharynx is captured with the 2D pharyngeal area measure.

Figure 10 shows examples of the articulatory configurations during /s/ and /sˤ/ as produced by SP4. In the plot on the left, the two sounds are extracted from the target words /sa:b/ ('he left') and /sˤa:b/ ('he hit') respectively. In the plot on the right, the two sounds are extracted from the target words /ba:s/ ('he kissed') and /ba:sˤ/ ('bus'). Both plots illustrate the observations above. The contours in red are from the pharyngealized member, and the contours in blue are from the plain member.

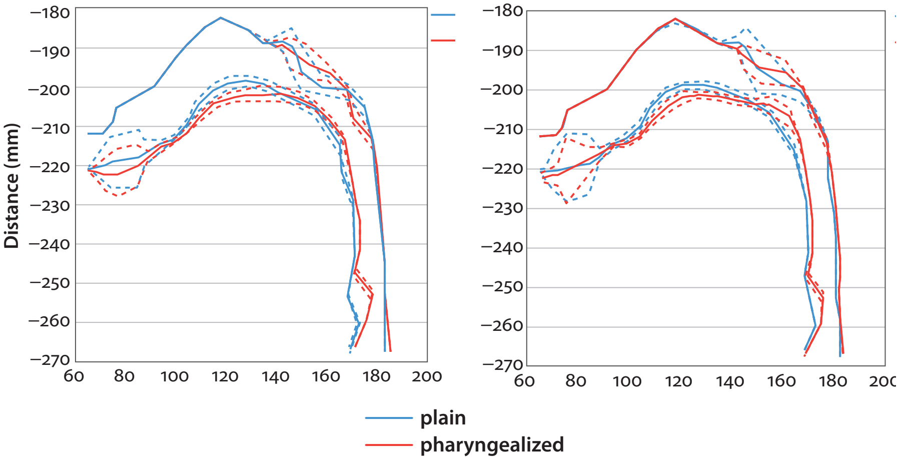


Figure 10. a. (left) Vocal tract contours of pharyngealized /sˤ/ and plain /s/ extracted from /sˤa:b/ ('he hit') and /sa:b/ ('he left') as produced by SP4. b. (right) Vocal tract contours of pharyngealized /sˤ/ and plain /s/ extracted from /ba:sˤ/ ('bus') and /ba:s/ ('he kissed') as produced by SP4

The plot on the left in Figure 11 shows box plots of the 2D pharyngeal areas during plain /s/ and pharyngealized /sˤ/ in /sa:b/ ('he left') and /sˤa:b/ ('he hit') respectively as produced by SP4. The plot on the right in Figure 11 shows the 2D pharyngeal areas when the two sounds occur at the end of the word in ba:s/ ('he kissed') and /ba:sˤ/ ('bus') respectively as produced by the same speaker. Both plots illustrate that the average 2D pharyngeal area is smaller in the pharyngealized member than in the plain counterpart.

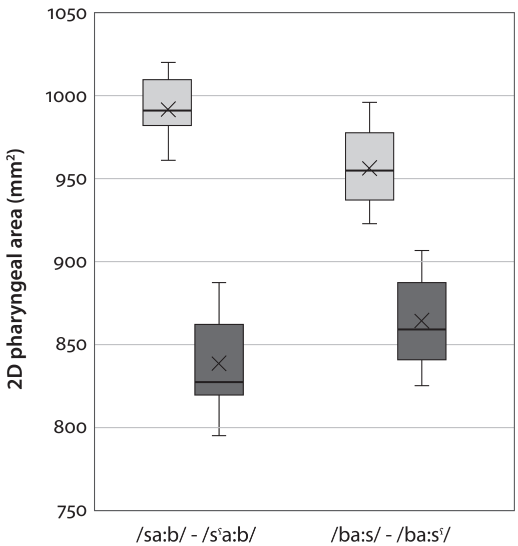


Figure 11. 2D pharyngeal areas in (mm^2) during plain /s/ and pharyngealized /sˤ/ in light gray and dark gray respectively, extracted from /sa:b/ ('he left'), /sˤa:b/ ('he hit'), /ba:s/ ('he kissed'), and /ba:sˤ/ ('bus') as produced by SP4

The articulatory data also shows that within the same word containing the pharyngealized consonant, the other speech segments are produced with a generally more constricted pharynx and a lowered tongue body, compared with their counterparts in the word containing the plain consonant. This is generally the case, regardless of the vowel length, and regardless of where the plain-pharyngealized contrast occurs. This is evident from tongue contour plots and 2D area box plots, and is indicative of pharyngealization spread in both directions. Figure 12 shows vocal tract contours during SP4's production of /sa:b/ and /sˤa:b/. The figure shows a generally more constricted pharynx in all the panels for the segments extracted from the pharyngealized word /sˤa:b/. The figure also shows a lowered tongue in these segments. This thus indicates a rightward spread of pharyngealization.

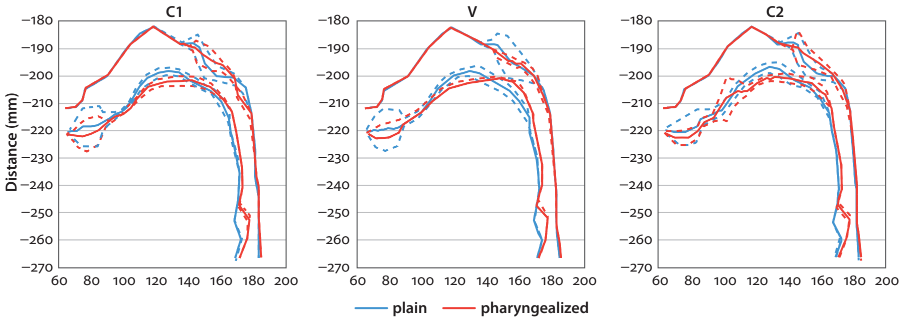


Figure 12. An illustration of rightward pharyngealization spread from SP4. Blue contours represent vocal tract configurations in segments from /sa:b/. Red contours represent vocal tract configuration in segments from /sˤa:b/

Figure 13, in contrast, illustrates leftward spread of pharyngealization. The three panels show the articulatory configuration of the three segments extracted from /ba:s/ in blue, and /ba:sˤ/ in red. Throughout the three panels, a more constricted pharynx and a lowered tongue is evident in the pharyngealized members.

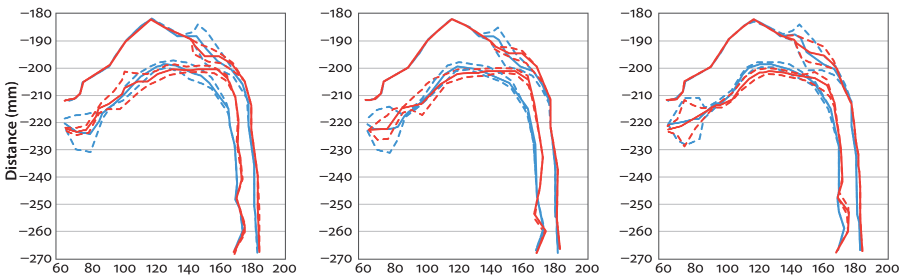


Figure 13. An illustration of leftward pharyngealization spread from SP4. Blue contours represent vocal tract configurations in segments from /ba:s/. Red contours represent vocal tract configuration in segments from /ba:sˤ/

Measures of the 2D pharyngeal areas also corroborate this. As an example, box plots in Figures 14 show the 2D pharyngeal areas measured in the three segments of /sa:b/ and /sˤa:b/ shown in the panels of Figure 12 above. The 2D pharyngeal areas in the pharyngealized member are smaller than in the plain member in all three segments.

In the articulatory data, it is also observed that speech segments preceding the pharyngealized consonants are generally more constricted than those following the pharyngealized consonant across both long and short vowels. This is the case

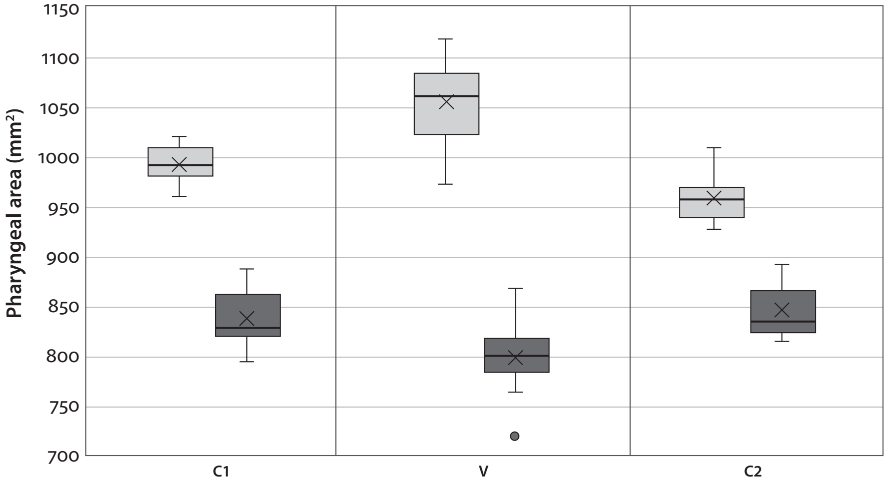


Figure 14. 2D pharyngeal areas in (mm^2) for the C1, V, and C2 segments in /sa:b/ and /sˤa:b/ segments in light gray and dark gray respectively as produced by SP4

in two of the four speakers. In a third speaker, this is the case during a short vowel, but not the long vowel. In the fourth speaker, this trend is not observed. These results may suggest that anticipatory (leftward) spread is often greater in magnitude than perseveratory (rightward) spread. The plots in Figure 15 below demonstrate this from SP4. The plots show the vocal tract contours corresponding to the /b/ segments in /sˤa:b/-/ba:sˤ/ in the panel on the left and the /b/ segments in /sˤabb/-/basˤsˤ/ in the panel on the right.

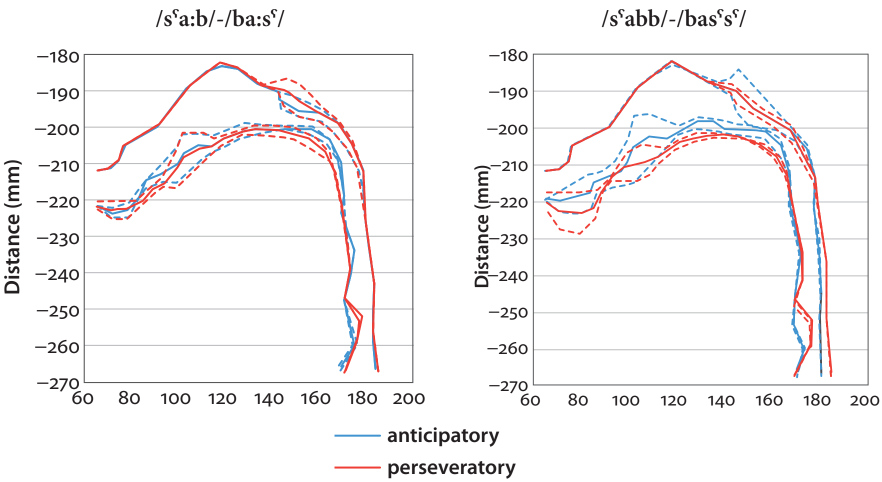


Figure 15. Vocal tract contours of /b/ segments in /sˤa:b/-/ba:sˤ/ (left) and /sˤabb/- /basˤsˤ/ (right) for speaker SP4

In Figure 15, it can be argued that the vocal tract contours of the /b/ in /ba:sˤ/ (i.e., the /b/ modified by anticipatory pharyngealization spread) exhibit a more constricted pharynx, especially in the velopharyngeal region when compared against the vocal tract contours of the /b/ in /sˤa:b/ (i.e. modified by perseveratory pharyngealization spread). This may be interpreted as indicative of a stronger effect of anticipatory spread. To quantitatively compare the effect of anticipatory and perseveratory pharyngealization spread, the 2D pharyngeal areas in the relevant /b/ segments are plotted in Figure 16.

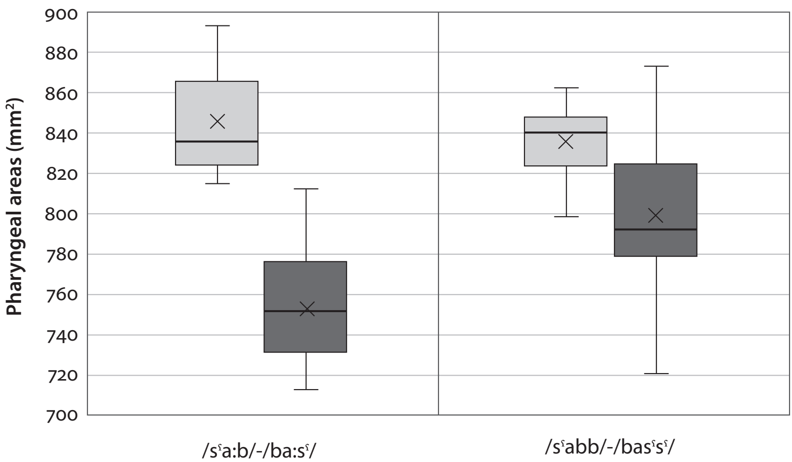


Figure 16. 2D pharyngeal areas (in mm²) of the /b/ segments in /sˤa:b/-/ba:sˤ/ (left) and /sˤabb/- /basˤsˤ/ (right) in light gray and dark gray respectively for SP4

Furthermore, among two minimal pairs having the same pharyngealized consonant but differing in vowel length, the 2D pharyngeal areas are generally smaller in the speech segments of the word containing the longer vowel. This is evident in the vocal tract contour plots, and suggests that the pharynx is more constricted, and thus that the effect of pharyngealization is possibly more substantial when there is a longer vowel. Vocal tract contours in Figure 17 from SP3 illustrate this:

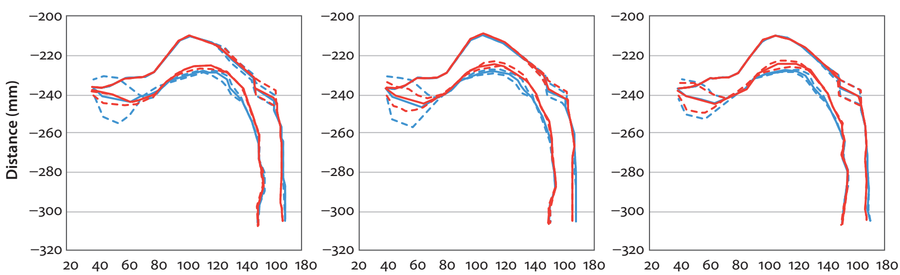


Figure 17. Vocal tract contours during /ba:sˤ/ (‘bus’) in blue and /basˤsˤ/ (‘he looked’) in magenta as produced by SP3

8.2 Results from acoustic data

Analyzing the acoustic data from the four speakers reveals several trends. Firstly, in line with previous studies, the major acoustic consequence of a pharyngealized consonant is the lowering of F2. Thus, F2 in the vowel extracted from the word with the pharyngealized member is lower than F2 in the vowel extracted from the word with the plain counterpart. This is the case, regardless of vowel length and whether the plain-pharyngealized contrast occurs before or after the vowel. It is also observed that F1 in the pharyngealized vowel is higher than F1 in the plain counterpart, though this is not the case in all the data. As an example, mean F1 and F2 values for speaker SP4 are plotted below in Figures 18 and 19, respectively.

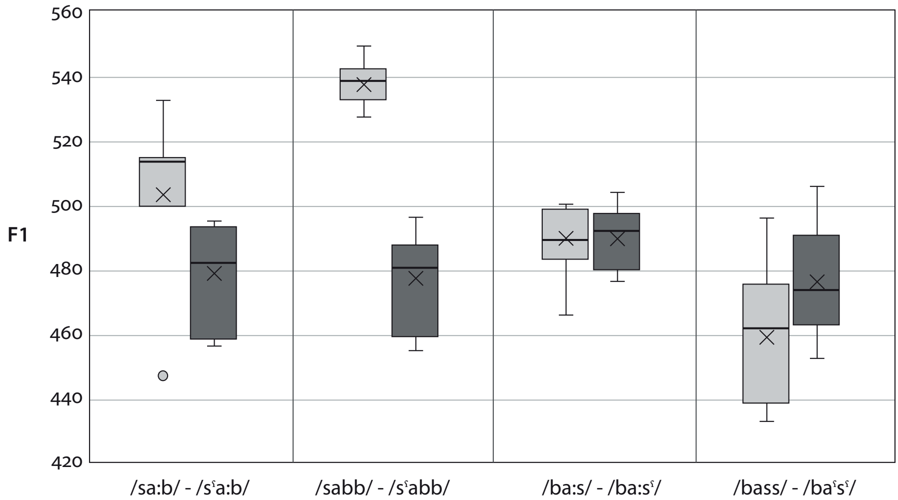


Figure 18. Mean F1 values in /a/ vowels for SP4

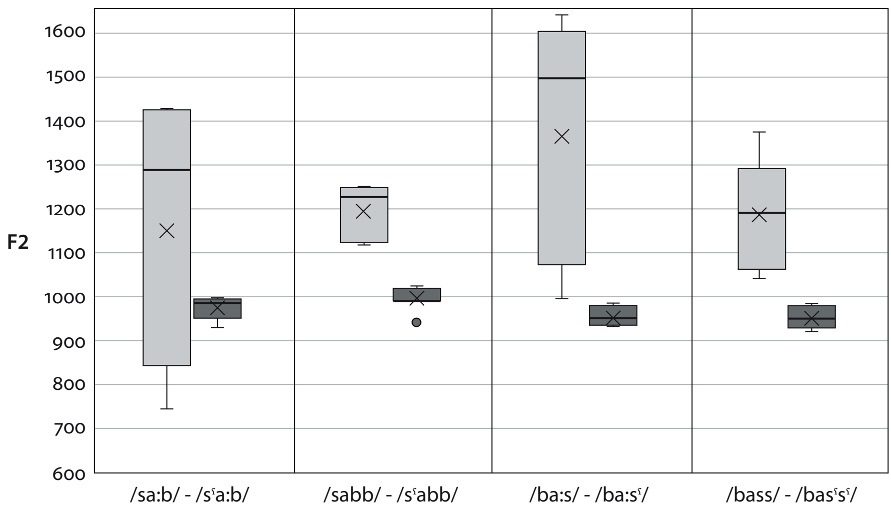


Figure 19. Mean F2 values in /a/ vowels for SP4

From the plots above, it is evident that the difference in F2 modification as a result of pharyngealization is generally greater in longer vowels than in shorter vowels.

9. Discussion

9.1 Discussion of articulatory results

The lingual and pharyngeal contours of the pharyngealized members show an articulatory configuration that generally consists of a decrease in the volume of the pharyngeal cavity and an increase in the oral cavity. This is consistent with articulatory descriptions of pharyngealized consonants presented in Ghazeli (1977) and Watson (2002). In addition to the primary alveolar constriction produced by the approximation of the tongue tip and blade to the dental-alveolar region, Ghazeli's (1977) description identified two principal articulatory correlates for pharyngealized consonants:

1. A secondary back constriction consisting of a "rearward movement of the back of the tongue towards the back wall of the pharynx" (Ghazeli, 1977, p. 72). This is visible in lingual and pharyngeal contour plots.
2. A depression of the tongue's palatine dorsum resulting from the retraction of the back of the tongue to form the secondary back constriction (described in 1). This configuration has also been described in the literature as tongue concavity (Lehn, 1963), which results in a larger oral cavity.

In addition to these correlates, other gestures reported in the literature and associated with pharyngealization include: a retracted primary constriction, lip protrusion (labialization), and tongue concavity (Lehn, 1963) which is also referred to in the literature as sulcalization (Bellem, 2007). It was not possible to examine these gestures in this study. Sulcalization, for example, would not be observed in a midsagittal plane, instead would have to be viewed in a coronal plane. Furthermore, the edge detection method implemented in this study did not detect the lips in such a way as to allow for examining the lip protrusion. Some studies also suggest that the epiglottis plays a role in pharyngealization (Laufer & Baer, 1988). This would be an interesting phenomenon to investigate in further rtMRI studies.

Articulatory results in this study show that the pharyngeal constriction is not formed at one specific point in the pharynx; rather it extends over a broad region, resulting in a generally more constricted pharynx. This is in line with observations in McCarthy (1994), who argued that back sounds are defined not by major articulators, but by place of articulation. Thus, whereas the speech sounds produced in the anterior part of the vocal tract are defined by the active articulators (labials, coronals, dorsals), the speech sounds involving a constriction in the back of the vocal tract are defined by place of articulation. The region of the vocal tract in which the labials, coronals, and dorsals are produced is approximately equal in length to the region reserved for back speech sounds. This leads to an asymmetry in that "finer distinctions of place are made in the front of the vocal tract than in the back" (McCarthy, 1994, p. 199). McCarthy argued that this is linked to the weaker sensory acuity in the posterior region of the vocal tract, when compared with the anterior region – the distribution of the sensory neurons is not uniform throughout the vocal tract. He quoted Grossman (1964): "This review of the reported oral sensory nerve elements reveals a progressive decrease in the frequency of sensory endings from the front to the rear of the mouth in humans … These findings are compatible with the author's initial experimental evidence which indicates that tactile discriminations are most acute in the anterior mucosal surfaces of the mouth. It is probably not coincidental that many important speech articulatory phenomena occur in the same oral regions" (Grossman, 1964, p. 132, quoted in McCarthy, 1994, p. 200). In their MRI study on pharyngeal and pharyngealized Arabic speech sounds, Shosted et al. (2012) divided up the region extending from the velum to the larynx into five areas: laryngeal, hypopharyngeal, epiglottal, oropharyngeal, and nasopharyngeal. Using Principle Component Analysis (PCA) to interpret the variation in the pharyngeal width during the pharyngeal and pharyngealized speech sounds, they conclude that two regions are sufficient for discriminating both types: an upper pharynx, comprising the oropharyngeal and nasopharyngeal regions, and a lower region comprising the remaining laryngeal, hypopharyngeal, and epiglottal regions. The former is the region implicated in the production of pharyngealized speech sounds of interest in this study. Thus, results from that study too, indicate that the large pharyngeal region is relatively poorly differentiated and that the place of articulation in that region is spread out over a wide region. This explains the observation that the back constriction in the pharyngealized consonants in the data in this study is not restricted to a specific point in the pharynx, as well as why the constriction extends over a wide range, and not a single point of contact. A uvular constriction observed in some of the data is in line with what has been reported in the literature (Watson, 2002). Similarly, the depression in the palatine dorsum described is observed to form a concave shape in some data, as reported in the literature (Watson, 2002). The presence of these gestures in some of the data, but not in others, is in line with the observation by Khattab et al. cited above that "there is no consistent single articulatory exponent of [pharyngealization]".

The results also clearly show pharyngealization spread, both to the right and to the left of the pharyngealized consonant, and that the spread is not restricted to vowels adjacent to the pharyngealized consonant, but extends further beyond the adjacent vowels to affect the consonants. Watson (2002) argued that this distinguishes the behavior of pharyngealized consonants from pharyngeals. Unlike pharyngealized consonants, pharyngeals do not have a primary coronal articulation, but only one articulation in the pharynx. According to Watson, both sound types trigger pharyngealization spread in Egyptian Arabic, but the latter type triggers spread that is limited to the adjacent vowels, while the former triggers longer distance spread and targets both vowels and consonants (Watson, 2002, p. 271). Watson argues that "the more contingent the phonetic realization of the non-primary feature on the primary feature, the further the non-primary feature will spread and the less specific the targets of the spread will be". Thus, in the case of pharyngeal consonants, the primary pharyngeal constriction is not restrained by any other stricture in the vocal tract, which produces shorter distance pharyngealization spread to adjacent vowels. In contrast, in the case of the doubly-articulated pharyngealized consonants, the constriction formed by the back of the tongue is highly influenced by the primary coronal one, and thus greater constraints are imposed on the tongue to achieve both the oral and pharyngeal articulations. The primary constriction "adds tension to the tongue dorsum and restricts pharyngeal constriction to the upper pharynx", while the back pharyngeal constriction "retracts the tongue blade from the top of the incisors towards the alveolar ridge" (Watson, 2002, p. 197). Thus the long-distance pharyngealization spread is explained to be a consequence of the longer relaxation time required for such a constrained tongue to return to its neutral state: "the stronger the relationship between a primary and a non-primary feature, the longer the muscle relaxation time, the longer the dorsum and vocal tract take to return to a neutral position, and hence the further pharyngealization will spread within the word" (Watson, 2002, p. 273). As Yeou (1997) argued, this constrained articulatory configuration makes pharyngealized sounds more resistant to coarticulatory effects from adjacent vowels. Pharyngealized sounds, therefore, induce greater coarticulatory effects on surrounding segments (i.e. the pharyngealization spread). This is why pharyngealized consonants are generally considered the locus of pharyngealization (Jongman et al., 2011) modifying surrounding segments, rather than the opposite: i.e. it is generally not understood that pharyngealized consonants are doubly-articulated due to coarticulatory influence of surrounding vowels, for example.

The results suggesting more prominent anticipatory pharyngealization spread can be explained as a consequence of the asymmetry in timing between the occurrence of the primary oral constriction and the secondary pharyngeal constriction described by Watson (2002): "[I]n pharyngealization, the pharynx narrows prior to the hold phase of the primary articulation. Pharyngealization is thus anchored more on the onset of the primary articulation, resulting in the typical anticipatory nature of pharyngealization spread." Davis (1995) also reported that, where bidirectional pharyngealization spread occurs, dominant leftward anticipatory spread is generally more common among most Arabic dialects, and that this explains why opaque phonemes that block spread, are generally opaque to rightward spread and not leftward (anticipatory) spread.

9.2 Discussion of acoustic results

Analysis of the first and second formant frequency measures F1 and F2 of the vowels in the target words reveal that the most consistent acoustic consequence of the presence of a pharyngealized consonant is the lowering of F2. The articulatory interpretation of a lowered F2 can be an increase in the size of the oral cavity due to the depression of the palatine dorsum described above. Watson suggested that the enlargement of the oral cavity can occur at "either extremity of the tract", and that the "enlargement at one end of a tract tend[s] to be enhanced by enlargement at the opposite end" (Watson, 2002, p. 270). She further maintained that, in the case of pharyngealized consonants, the enlargement of the oral cavity at the pharynx may be enhanced by enlargement at the other end of the tract in the form of labialization (lip protrusion). This is analogous to the articulation of labial segments, in which enlargement at the lips is often enhanced by enlargement at the pharynx (Watson, 2002). Indeed, labialization in the form of lip protrusion has been reported for pharyngealized consonants.

The increase in F1 is not as consistent in the results from this study, and an explanation for this can be offered from Watson's analysis (2002): the constriction in the pharynx causes an increase in F1, but the labialization causes a decrease in F1.

10. Conclusions

The aim of this rtMRI and acoustic study was to examine the articulatory configuration of pharyngealized consonants of Cairene Arabic, as well as the phonetic correlates associated with the pharyngealization spread triggered by these consonants. Specifically, this study looked at pharyngealization spread through different vowel lengths (short and long), and in both directions, left-to-right and right-to-left. The articulatory data consisted of rtMRI images. The acoustic data consisted of first and second formant frequency measures. Analysis of both the articulatory and acoustic data suggested stronger and weaker pharyngealization spread across vowels of different lengths. In the acoustic data, this was understood from the variation in the extent of formant frequency modification due to the presence of a pharyngealized consonant. In the articulatory data, this was understood from the smaller 2D pharyngeal area measurements and the more constricted pharyngeal contours. Results from the articulatory data show that speech segments preceding pharyngealized consonants are generally more constricted than those following them, suggesting that anticipatory (leftward) spread is stronger than perseveratory (rightward) spread in this dialect. Furthermore, among two minimal pairs containing the same pharyngealized consonant but differing in vowel length, 2D pharyngeal areas are generally more constricted in the speech segments of the word containing the longer vowel, suggesting that the effect of pharyngealization is stronger in longer vowels.

References

Al-Tamimi, F., & Heselwood, B. (2011). Nasoendoscopic, video-fluoroscopic and acoustic study of plain and emphatic coronals in Jordanian Arabic. In Z. M. Hassan & B. Heselwood (Eds.), *Instrumental studies in Arabic Phonetics* (pp. 165–191). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.319.08tam

Bellem, A. (2007). Towards a comparative typology of emphatics across Semitic and into Arabic dialect phonology. Ph.D. dissertation, University of London.

Boersma, P., & Weenink, D. (2013). *Praat: doing phonetics by computer* [computer program]. Retrieved from http://www.praat.org/

Carignan, C., Shosted, R. K., Fu, M., Liang, Z.-P., & Sutton, B. (2015). A real-time MRI investigation of the role of lingual and pharyngeal articulation in the production of the nasal vowel system of French. *Journal of Phonetics*, 50, 34–51. https://doi.org/10.1016/j.wocn.2015.01.001

Chiba, T., & Kajiyama, M. (1941). *The vowel: Its nature and structure*. Tokyo: Kaiseikan.

Davis, S. (1995). Emphasis spread in Arabic and grounded phonology. *Linguistic Inquiry*, 26, 465–498.

De Sacy, S. (1810). *Grammaire Arabe à l'usage des élèves de l'école spéciale des langues orientales vivantes*. Paris: L'imprimerie Impériale.

Embarki, M., Ouni, S., & Salam, F. (2011a). Speech clarity and coarticulation in Modern Standard Arabic and dialectal Arabic. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 635–638. City University of Hong Kong, Hong Kong.

Embarki, M., Ouni, S., Yeou, M., Guilleminot, C., & Al Maqtari, S. (2011b). Acoustic and electromagnetic articulographic study of pharyngealisation: coarticulatory effects as an index of stylistic and regional variation in Arabic. In Z. M. Hassan & B. Heselwood (Eds), *Instrumental studies in Arabic Phonetics* (pp. 193–216). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.319.09emb

Fu, M., Zhao, B., Carignan, C., Shosted, R. K., Perry, J. L., Kuehn, D. P., Liang, Z.-P., & Sutton, B. (2015). High-resolution dynamic speech imaging with joint low-rank and sparsity constraints. *Magnetic Resonance in Medicine*, 73,1820–1832. https://doi.org/10.1002/mrm.25302

Gadalla, H. (2000). *Comparative morphology of standard and Egyptian Arabic*. Muenchen: Lincom Europa.

Ghazeli, S. (1977). Back consonants and backing coarticulation in Arabic. Ph.D. dissertation, University of Texas at Austin.

Giannini, A., & Pettorino, M. (1982). The emphatic consonants in Arabic. *Speech Laboratory Report 4*. Oriental Institute of Naples.

Grossman, R. C. (1964). Sensory innervation of the oral mucosae. *Journal of the Southern California State Dental Association*, 32, 128–133.

Hermes, Z., Wong, N., Loucks, T., & Shosted, R. (2015). The primary articulation of plain-emphatic /s/-/sˤ/ in Lebanese Arabic: an EMA study. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Paper number 0681, 1–5.

Herzallah, R. (1990). Aspects of Palestinian Arabic Phonology: A non-linear approach. Ph.D. dissertation, Cornell University.

Heselwood, B., & Al-Tamimi, F. (2011). A study of the laryngeal and pharyngeal consonants in Jordanian Arabic using nasoendoscopy, videofluoroscopy and spectrography. In Z. M. Hassan & B. Heselwood (Eds), *Instrumental studies in Arabic Phonetics* (pp. 101–128). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.319.05hes

Hetzron, R. (1998). *The Semitic languages*. New York: Routledge.

Israel, A., Proctor, M., Goldstein, L., Iskarous, K., & Narayanan, S. (2012). Emphatic segments and emphasis spread in Lebanese Arabic: a real-time magnetic resonance imaging study. *13th Annual Conference of the International Speech Communication Association INTERSPEECH*, 3, 2175–2178.

Jakobson, R. (1957). Mufaxxama: the 'emphatic' phonemes in Arabic. In E. Pulgram (Ed), *Studies Presented to Joshua Whatmough on his Sixtieth Birthday* (pp. 105–116). Gravenhage: Mouton & Co.

Jan, M. A., Marshall, I., & Douglas, N. J. (1994). Effect of posture on upper airway dimensions in normal human. *American Journal of Respiratory and Critical Care Medicine*, 149, 145–148. https://doi.org/10.1164/ajrccm.149.1.8111573

Jongman, A., Herd, W., Al-Masri, M., Sereno, J., & Combest, S. (2011). Acoustics and perception of emphasis in Urban Jordanian Arabic. *Journal of Phonetics*, 39, 85–95. https://doi.org/10.1016/j.wocn.2010.11.007

Khattab, G., Al-Tamimi, F., & Heselwood, B. (2006). Acoustic and auditory differences in /t/-/ṭ/ opposition in male and female speakers of Jordanian Arabic. In S. Boudelaa (Ed), *Perspectives on Arabic linguistics XVI* (pp. 131–160). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.266.09kha

Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W. J. Hardcastle & N. Hewlett (Eds), *Coarticulation theory data and techniques* (pp. 7–30). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486395.002

Lapinskaya, N. (2013). An exploratory ultrasound investigation of emphatic articulation in Cairene Arabic. Paper presented in *Ultrafest VI*, Queen Margaret University, Edinburgh. *Poster number* 21, 67–68.

Laufer, A., & Baer, T. (1988). The emphatic and pharyngeal sounds in Hebrew and in Arabic. *Language and Speech*, 31, 181–205. https://doi.org/10.1177/002383098803100205

Lehn, W. (1963). Emphasis in Cairo Arabic. *Language*, 39, 29–39. https://doi.org/10.2307/410760

Lingala, S. G., Sutton, B., Miquel, M., & Nayak, K. (2016). Recommendations for real-time speech MRI. *Journal of Magnetic Resonance Imaging*, 43, 28–44. https://doi.org/10.1002/jmri.24997

Maddieson, I., & Precoda, K. (1991). UPSID-PC: the UCLA phonological segment inventory database. http://web.phonetik.uni-frankfurt.de/upsid_find.html (retrieved May 2016).

Manuel, S. Y., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories: Status Report on Speech Research*, 69–78.

McCarthy, J. J. (1994). The phonetics and phonology of Semitic pharyngeals. In P. A. Keating (Ed.), *Phonological structure and phonetic form: papers in Laboratory Phonology III* (pp. 191–233). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511659461.012

Narayanan, S., Toutios, A., Ramanarayanan, V., Lammert, A., Kim, J., Lee, S., Nayak, K., Kim, Y. C., Zhu, Y., Goldstein, L., Byrd, D., Bresch, E., Ghosh, P., Katsamanis, A., & Proctor, M. (2014). Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC). *The Journal of the Acoustical Society of America*, 136, 1307. https://doi.org/10.1121/1.4890284

Norlin, K. (1987). A phonetic study of emphasis and vowels in Egyptian Arabic. *Lund University Department of Linguistics Working Papers*, 30, 1–119.

Obrecht, D. H. (1968). *Effects of the second formant on the perception of velarisation consonants in Arabic*. The Hague: Mouton. https://doi.org/10.1515/9783111357393

Proctor, M. I., Bone, D., Katsamanis, N. & Narayanan, S. (2010). Rapid semi-automatic segmentation of real-time magnetic resonance images for parametric vocal tract analysis. *Proceedings of the 11th Annual Conference of the International Speech Communication Association INTERSPEECH*, 1576–1579.

Royal, A. (1985). Male/female pharyngealization patterns in Cairo Arabic: a sociolinguistic study of two neighborhoods. Ph.D. dissertation, University of Texas.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627–1639. https://doi.org/10.1021/ac60214a047

Semaan, K. I. (1963). *Arabic phonetics: Ibn Sīnā's Risālah on the points of articulation of the speech-sounds translated from medieval Arabic. Arthur Jeffery memorial monographs*. Lahore: Ashraf Press.

Shahin, K. (2002). *Postvelar harmony*. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.225

Shar, S., & J. Ingram. (2011). MRI investigations of Arabic gutturals. In W.-S. Lee & E. Zee (Eds) *The International Conference of Phonetic Sciences* XVII (pp. 1802–1805). City University of Hong Kong.

Shosted, R., Benmamoun, A., & Sutton, B. (2012). Using magnetic resonance to image the pharynx during Arabic speech: static and dynamic aspects. *13th Annual Conference of the International Speech Communication Association INTERSPEECH*, 3, 2179–2182.

Versteegh, K. (2001). *The Arabic Language*. Edinburgh: Edinburgh University Press.

Vollers, K. (1893). The system of Arabic sounds as based upon Sibaweih and Ibn Yaish. *Transactions of the 9th International Congress of Orientalists*, 2, 130–154.

Wahba, K. M. (1996). Linguistic variation in Alexandrian Arabic: the feature of emphasis. In A. Elgibali (Ed.), *Understanding Arabic: essays in contemporary Arabic linguistics in honor El-Said Badawi* (pp. 103–128). Cairo: The American University in Cairo Press.

Watson, J. E. E. (2002). *The phonology and morphology of Arabic*. New York: Oxford University Press.

Xu, Y., & Gao, H. (2018). FormantPro as a tool for speech analysis and segmentation. *Revista de Estudos da Linguagem*, 26(4), 1435–1454. https://doi.org/10.17851/2237-2083.26.4.1435-1454

Yeou, M. (1997). Locus equations and the degree of coarticulation of Arabic consonants. *Phonetica*, 54, 187–202. https://doi.org/10.1159/000262221

Youssef, I. (2014). Emphasis spread in Cairene Arabic: a reassessment. In O. Durand, A. D. Langone, & G. Mion (Eds.), *Alf lahğa wa lahğa: Proceedings of the 9th AIDA Conference* (pp. 455–464). Lit. Verlag.

Zemlin, W. (1998). *Speech and hearing science anatomy and physiology 4th ed*. Boston: Allyn and Bacon.

Zeroual, C., Esling, J. H., & Hoole, P. (2011a). EMA, endoscopic, ultrasound and acoustic study of two secondary articulations in Moroccan Arabic. In Z. M. Hassan & B. Heselwood (Eds.), *Instrumental studies in Arabic Phonetics* (pp. 277–300). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.319.13zer

Zeroual, C., Esling, J. H., Hoole, P. & Ridouane, R. (2011b). Ultrasound study of Moroccan Arabic labiovelarization. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2272–2275). City University of Hong Kong, Hong Kong.

Affricate variation in Emirati Arabic

An exploratory study

Marta Szreder, Donald Derrick and Chahla Ben-Ammar
United Arab Emirates University / University of Canterbury / University of Toronto

Using a corpus analysis and an elicitation study, we provide evidence regarding the effects of lexical, phonological and socio-phonetic factors on the distribution of affricate variants in Emirati Arabic (EA). The corpus results indicate that the processes are only partially attributable to lexical factors; results reveal previously unreported patterns of vowel effects, and suggest a role of coronal consonants in determining affrication. The experimental results highlight significant within- and across-speaker variability that is not only lexically-determined, but also contradicts traditional hypotheses regarding the geographic distribution of the variants. This exploratory study presents a detailed descriptive analysis of the process in EA, providing suggestions for future investigations.

Keywords: Emirati Arabic, affrication, variation, coronals

1. Introduction

1.1 Arabic in the United Arab Emirates

The United Arab Emirates, located in the southeast end of the Arabian Peninsula and bordered by the Persian Gulf to the north, Oman to the east and Saudi Arabia to the south and west, is a federation composed of seven states, otherwise known as the Emirates, established in 1971. The language context in the UAE is unique in that the society is composed of 88% expatriate communities with different native languages. English is used as the lingua franca of the country, while the official language and the native language of the one million UAE citizens is Arabic. The people of the UAE speak Emirati Arabic (EA), a sub-dialect of Gulf Arabic (GA). GA is a variety of Peninsular Arabic spoken also in Qatar, Kuwait, and parts of Bahrain, Oman, Saudi Arabia and Iraq (Holes, 1995, 2001; Versteegh, 1997). Tables 1 (adapted from Al Ameri, 2009) and 2 (based on Feghali, 2008) present the phonemic inventory of GA.

https://doi.org/10.1075/sal.10.02szr

Table 1. A phonemic chart of Gulf Arabic consonants

		Labial	Dental	Dental-alveolar	Palatal	Velar	Pharyngeal	Glottal
Nasal		m		n				
Plosive	voiceless	p		t, tˁ		k		ʔ
	voiced	b	d			g		
Affricate	voiceless				tʃ			
	voiced				dʒ			
Fricative	voiceless	f	θ	s, sˁ	ʃ	x	ħ	h
	voiced		ð, ðˁ	z		ɣ	ʕ	
Trill				r				
Approximant				l	j	w		

Table 2. Gulf Arabic vowels. Symbols in parentheses represent allophonic variants

	Front		Back	
	Short	Long	Short	Long
Close	i	ii	u (e)	uu (oo)
Close-mid		ee		oo
Open	a (ɑ, e)	aa (ɑɑ, ee)		

According to native informants, three main sub-varieties of EA can be identified: (1) the northern dialect (spoken in the emirates of Dubai, Sharjah, Ajman, Umm Al Quwain and part of Ras Al Khaimah); (2) the Abu Dhabi dialect (spoken in the emirate of Abu Dhabi); and (3) the eastern dialect (spoken in Fujairah, the remainder of Ras Al Khaimah and Khor Fakkan (an exclave of Sharjah). However, the differences across these sub-varieties are slight and primarily thought to involve lexical and morphological aspects.

The current paper looks at the use of affricates in EA, which historically have been affected by two context-sensitive processes: deaffrication (lenition) of [dʒ] to [j], and affrication of [k] to [tʃ]. Since the [tʃ] and [j] variants eventually became the standard features of the dialect, in modern EA the occurrence of [ʤ] and [k] (in specific environments) could be analysed as deaffrication of [tʃ] and affrication of [j] (cf. Al-Wer et al., 2020). However, for the purpose of this study, we follow the traditional labelling and refer to the two patterns of affricate variation as k-affrication and dʒ-deaffrication (lenition). Below, we provide an overview of the previous studies investigating the conditions for the processes of affrication and deaffrication in GA.

1.2 *k*-affrication

The process of *k*-affrication has been extensively studied and is regularly discussed in reference grammars of EA and GA and attributed to both linguistic and extra-linguistic factors. Sociolinguistic studies have investigated the correlation between affrication and extra-linguistic factors, including age, educational level, gender of the speaker and area.

However, linguistic factors were also identified as contributing to the process. The standard accounts (Feghali, 2008; Hoffiz, 1995; Qafiesheh, 1975) observe that alveopalatal variants of the 2nd person feminine clitic, a process traditionally termed *kashkasha*, occur in two cases:

1. If it is the second person feminine singular suffixed pronoun (suffixed to verbs, nouns and particles) or the object suffixed pronoun (suffixed to verbs and participles). Examples:
   a. /ixwaanik/ – /ixwaanitʃ/ ‘your brothers’ (fem. sg.)
   b. /ʃafik/ – /ʃafitʃ/ ‘he saw you’ (fem. sg.)
2. If it is preceded or followed by front vowels. Examples:
   a. /kaððaab/ – /tʃaððaab/ ‘liar’ (m.s.)
   b. /simak/ – /simatʃ/ ‘fish’

Suffix-based affrication is a morphophonological feature of GA that has a syntactic function in that it indicates the gender of the speaker; [-itʃ] for feminine and [-ik] for masculine (Feghali, 2008). In EA these forms do not exhibit variation and as such are not of interest to the current study, which focuses on stem-based affrication.

The main phonological factor associated with stem-based affrication is the vowel context. For example, Johnstone (1963) and Holes (1995) report that in GA dialects, /k/ affricates when it is in contiguity with front vowels /i, ii, ee/ and, conversely, that affrication is often blocked by the presence of back vowels /uu, oo/. This inter-dependence of vowel quality and affrication is further evident in the fact that the vowel /a/ may vary in frontness across dialects together with the affricate. For example, in the word /rakab/ *rode* in the variant with /k/ the vowels may be realised as back /rɑkɑb/, while in the variant with /tʃ/ as front /ratʃab/ (Johnstone, 1963, p. 218). In fact, the vowel effects could have historically been one of the origins of the process of k-affrication (Holes, 1991). Any observed co-occurrence of front vowels and the affricate variant could thus be the result of this historical process, rather than an active phonological rule. Such interpretation is provided e.g. by Youssef (2014) for affrication in Baghdadi Arabic, which no longer appears to be a productive process, but rather a completed historical change, resulting in two separate phonemes: /k/ and /tʃ/. In Baghdadi Arabic, this is evident in the presence of minimal pairs, as well as the variability being mostly restricted to lexical and constrained by sociolinguistic factors.

Other approaches describe k-affrication as a synchronic phonological constraint. Mustafawi (2006) investigates the affrication and de-affrication of /tʃ/ and /dʒ/ (PS *g) in Qatari Arabic, and suggests that affrication in QA is a result of the simple place assimilation process between segments that share the feature [+high], as often observed cross-linguistically. She proposes that the trigger /i, ii/ targets the velar stops and outputs [tʃ] and [dʒ] which are both [+high] segments. The velar stops, which are [dorsal] assimilate to the place feature of /i, ii/ which is [coronal] and therefore, the outputs are coronals. She also suggests that, contrary to previous studies, in the Qatari dialect it is not any front vowel context that triggers affrication. Rather, she claims affrication only occurs when the velar stops are, exclusively, adjacent to the high front vowels /i/ or /ii/. Other segments, including [–high] front vowels block affrication. Al Rasheedi (2015) examines affrication of /k/ to [ts] in Ha'ili Arabic, and concludes that it occurs in the presence of either preceding or following front vowels /a/ and /i/. However, affrication is prohibited if there are other coronal segments in the word, or if the velar stop is geminated. Hence, for example /kaas/ 'glass' cannot be realised as */tsaas/ due to the presence of the coronal /s/, and /rukkaab/ 'passengers' cannot be realised as */rutstsaab/, due to gemination.

Indeed, vowel-conditioned affrication is a process commonly observed in many languages. For example, Quebec French (Béland & Kolinsky, 2005) and Louisiana French spoken in the city of Ville Platte (Russell, 2014) feature the affrication of the stops /t/ and /d/ to their variants [ts] and [dz] preceding the high vowels /i/ and /ʏ/. Vowel-determined affrication is also observed in Western Cretan Greek, wherein the velar obstruents /k, g, x, ɣ/ when preceded by the front vowels /i/ and /e/, are affricated to become /tᶝ, dᶽ, ɕ, ʑ/ (Lengeris & Kappa, 2016). Similarly, Japanese features two processes of affrication (Yoshida, 2001): the first occurs when dental-alveolar stops become palato-alveolar affricates as a result of preceding the high vowel /i/ and the palatal glide /j/. The second occurs when dental-alveolar stops are followed by a high back vowel. Finally, in Tigrinya (Moeng & Carter, 2019), the fricative /s'/ is commonly realised as /ts'/.

In summary, studies agree that k-affrication in GA is related to the presence of front vowels. However, the precise conditions may vary by dialect and may or may not be interpreted as a synchronic constraint. The current study was meant to explore the specific phonological conditioning of the process in detail in EA.

1.3 dʒ-deaffrication

Like affrication, deaffrication of /dʒ/ into /j/ (also termed lenition) has been linked to sociolinguistic and lexical factors. As in the case of affrication, in loanwords, proper nouns, certain classicisms and recently introduced words, the voiced post-alveolar is supposed to remain unchanged. Examples include:

[garaadʒ] *[garaaj] 'garage' (English)
[aldʒazaaʔɪr] *[aljazaaʔɪr] 'Algeria' (proper noun)
[dʒawaaz safar] *[jawaaz safar] 'passport' (literary Arabic)

Unlike affrication, lenition was traditionally often characterised as neither rule-governed nor conditioned by the phonological environment (Feghali, 2008; Hassan, 2017; Hoffiz, 1995; Qafisheh, 1975). However, Mustafawi (2006) considers the process of lenition in Qatari Arabic together with affrication and finds that they are both sensitive to the constraint against a sequence of homorganic segments. Analysing the data from an Optimality Theoretic perspective (Prince & Smolensky, 2004), she proposes that these two processes obey the Obligatory Contour Principle (OCP), which prohibits the co-occurrence of such segments. In this account, the affricate variant does not appear in contiguity with coronal consonants, although some observed variability is attributed to the fact that the OCP is not categorical, and is more strict for immediately neighbouring consonants than for those separated by a vowel. Furthermore, Alaodini (2019) reports that in Dammam there is also an effect of neighboring vowels on the distribution of the variants, such that front vowels pattern with the glide /j/.

1.4 Objectives

There have been few experimental investigations of affricates in EA. Given the different historical origins of the phonological processes affecting affricates, their differential distribution across GA varieties and regions, and the different patterns reported by the multitude of research studies on these sounds in different dialects, the current study was primarily designed as an exploratory descriptive analysis of the production of the two affricates in EA through the means of a corpus analysis and an experimental study.

2. Methodology

2.1 Participants

Thirteen native speakers of Arabic were recruited for the study: 6 females between the ages of 20 to 30 (M_{age} = 22.2 years) and 7 males between the ages of 18 to 24 (M_{age} = 21.4 years). Further four participants were recruited but did not attend the testing session or did not complete it for health reasons. Four participants' data were later excluded, as they reported being citizens of Syria and Yemen, and thus not native speakers of EA. In the end, the analyses thus included data from 9 participants. All participants had been residents of the UAE since birth and were Emirati citizens from Ras Al Khaimah (*N* = 5), Fujairah (*N* = 2), and Abu Dhabi (*N* = 2).

All participants were students at the United Arab Emirates University and as such were all bilingual in Arabic and English. Some participants also reported limited knowledge of German and Korean as L3. All participants reported having no history of speech or hearing impairments and were compensated for their time.

Table 3. Participant information

ID	Age	Sex	Place of origin
S1	22	F	UAE, Fujairah
S2	21	F	UAE, Fujairah
S3	20	F	UAE, Ras Al Khaimah
S6	20	F	UAE, Ras Al Khaimah
S7	20	F	UAE, Ras Al. Khaimah
S8	30	F	UAE, Al Ain
S11	22	M	UAE, Ras Al Khaimah
S13	24	M	UAE, Abu Dhabi
S14	19	M	UAE, Ras Al Khaimah

2.2 Stimuli

Stimuli consisted of 18 Arabic words obtained from EMAC (A Corpus of Emirati Arabic) (Ntelitheos & Idrissi, 2017), with the segments occurring word-initially, word-medially, intervocalically and word-finally. They were selected from a list of words used in other studies on affricate variation (Al Rojaie, 2013; As-Sammer, 2010; Qafiesheh, 1977) and were chosen on the basis of imageability. An effort was made to sample as wide a variety of vowel contexts as possible.

Table 4. Words with /k/

Item	Arabic	Gloss	Word position	V1	V2
kabd	كبد	liver	WI	a	0
kafan	كفن	grave (cloth)	WI	a	a
kiħal	كحل	eyeliner	WI	i	a
baakir	باكر	tomorrow	WM	aa	i
zkaam	زكام	a cold	WM	0	aa
simak	سمك	fish	WF	i	a
keek	كيك	cake	WF	ee	0
misk	مسك	perfume	WF	i	0

Table 5. Words with /dʒ/

Item	Arabic	Gloss	Word position	V1	V2
dʒabal	جبل	mountain	WI	a	a
jaabaan	يابان	Japan	WI	aa	aa
dʒaraada	جرادة	locust	WI	a	aa
dʒumʕa	جمعة	Friday	WI	u	a
radʒdʒaal	رجّال	man	WM	a	aa
ridʒil	رجل	leg	WM	i	i
fadʒir	فجر	dawn	WM	a	i
zudʒaadʒ	زجاج	glass	WF	u	aa
faladʒ	فلج	paralysis	WF	a	a
didʒaadʒ	دجاج	chicken	WF	i	aa

2.3 Experimental data collection and procedure

Acoustic data was collected using a Neumann U87 Ai microphone positioned near the participant's mouth with a USBPre2 microphone preamp and A/D converter. The collected data was captured and automatically synchronized on an ASUS ROG G701VIK laptop running Articulate Assistant Advanced software (AAA) (Articulate Instruments Ltd, 2012).

Participants were seated on a chair located in the Phonetics laboratory at the UAE University and were first given instructions regarding the experiment. The participants then had an initial training run in which they were asked to provide the targeted stimuli based on pictures, so that the experimenter would not model the pronunciation of the target words. If they struggled with elicitation, they were provided with the English equivalent. If that did not result in elicitation, the subsequent trials for the unrecognized word from that participant were excluded from the analysis. Following the training, the participants were presented with the full list of stimuli ten times in a pseudorandomized manner on AAA. The stimulus list was expected to produce 180 tokens per participant and 1,620 tokens in total (18-word items * 10 trials * 9 participants), however, due to errors in participant elicitation, 372 fewer tokens were collected and the total number of tokens was left at 1,248.

2.4 Corpus study

A corpus study was conducted using an adult corpus of Emirati Arabic EMAC (https://faculty.uaeu.ac.ae/dimitrios_n/Research/A-Corpus-of-Emirati-Arabic/a-corpus-of-emirati-arabic.html) and a child Emirati Arabic Language Acquisition Corpus EMALAC (https://faculty.uaeu.ac.ae/dimitrios_n/Research/Emirati-Arabic-Language-Acquis/emirati-arabic-language-acquisition-project.html) (Ntelitheos & Idrissi, 2017), developed by researchers at UAE University, in order to test the existing hypotheses regarding the lexical and phonological constraints on the variation. The above two corpora were selected on account of their representation of the Emirati dialect in both adults and children. Similar to the stimuli included in the elicitation study, the words selected for corpus analysis were also based on previous studies on Arabic (Al Rojaie, 2013; As-Sammer, 2010; Qafiesheh, 1977).

2.5 Corpus study procedure

As the two corpora cannot be searched by means of a search engine (i.e. they can only be downloaded as CLAN and .txt files), they were uploaded on BBEdit, a text editing software, and a multi-file search was conducted to find all instances of stimuli production. The queried stimuli included all possible variations, both affricated and palatalized and not, with their segment environment.

Given the previous body of research on *k*-affrication in Gulf and Emirati Arabic, the experimental study was designed to address the potential evidence for socio-phonetic constraints, as well as the stability of speakers across- and within-words. The corpus study, on the other hand, had the goal of answering two main questions: (1) Does *k*-affrication in Emirati Arabic depend on vowel quality? and (2) Do the patterns of affrication and de-affrication show evidence of lexical constraints?

3. Corpus study results

In order to analyse participant productions and corpora entries, the percentage of affricated and de-affricated segments in relation to their segment environments, participants and individual words was calculated and displayed in stack percentage bar plots scripted in R (R Core Team, 2014).

Table 6. Corpus study target words

Words with k			Words with dʒ		
Item	**Arabic**	**Gloss**	**Item**	**Arabic**	**Gloss**
ʔakil	أكل	*food*	dardʒ	درج	*drawer*
baakir	باكر	*tomorrow*	didʒaadʒ	دجاج	*chicken*
diik	ديك	*rooster*	dʒabal	جبل	*mountain*
kabd	كبد	*liver*	dʒadiid	جديد	*new*
kabiir	كبير	*big*	dʒaahil	جاهل	*child*
kaððaab	كذاب	*liar*	dʒamiil	جميل	*beautiful*
kalb	كلب	*dog*	dʒamʕa	جامعة	*university*
kalma	كلمة	*word*	dʒaraada	جرادة	*grasshopper*
karʃ	كرش	*belly*	dʒariida	جريدة	*newspaper*
kiis	كيس	*bag*	dʒawaaz safar	جواز سفر	*passport*
keek	كيك	*cake*	dʒubin	جبن	*cheese*
kiħal	كحل	*eyeliner*	dʒumʕa	جمعة	*Friday*
makaan	مكان	*place*	dʒuuti	جوتي	*shoes*
misk	مسك	*perfume*	dʒuuʕaan	جوعان	*hungry*
sakkiin	سكين	*knife*	fadʒir	فجر	*dawn*
simak	سمك	*fish*	findʒaan	فنجان	*cup*
zukaam	زكام	*cold*	madʒalla	مجلة	*magazine*
			madʒnuun	مجنون	*crazy*
			masdʒid	مسجد	*mosque*
			radʒdʒaal	رجال	*man*
			ridʒil	رجل	*leg*
			sadʒdʒaada	سجادة	*carpet*
			ʃadʒara	شجرة	*tree*
			taadʒ	تاج	*crown*
			wadʒh	وجه	*face*
			zudʒaadʒ	زجاج	*glass*

3.1 Phonological effects

The first objective addressed in the analysis of the corpus data was to determine the phonological context in which the target segments appeared. All of the figures presented below display affricate variant proportion means and confidence intervals based on interaction with the relevant independent variables. Figures (1a) and (1b) show the breakdown of the produced segments in relation to the preceding (1a) and following (1b) segments for the affrication of /k/ to /tʃ/. The center line (50%) represents an equal count of either variant. The shaded bars indicate the mean bias such that the rightmost extension indicates all the tokens are affricated (/tʃ/), and the leftmost extension means none of the tokens are affricated (/k/). The whiskers represent the 95% binomial proportion confidence intervals (see Dorai-Raj, 2014). The data is sorted by the innermost confidence boundary. The numbers on the right in the boxes represent the token counts.

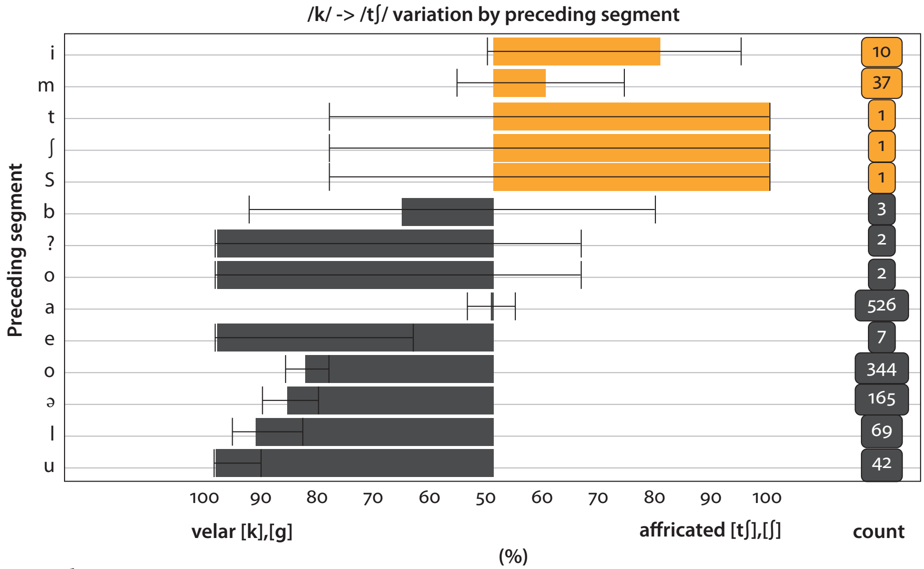
To begin, the context of occurrence of /k/ and /tʃ/ variants was examined with respect to front vowels, to determine whether there was an effect similar to those claimed in previous accounts. Indeed, as evident from Figure 1, while both /k/ and /tʃ/ were observed in the presence of front vowels /i/ and /e/, only the /k/ variant was present in the context of the back vowels /o, u/, confirming the hypothesis that k-affrication is only allowed in the context of front vowels. This finding held for both preceding (Figure 1a) and following (Figure 1b) vowels. Nonetheless, while affrication was possible in the presence of front vowels, it was by no means obligatory: /tʃ/ was more likely when preceded by /i/, but less likely when preceded by /e/, and both were equally likely when followed by these vowels.

Second of all, the effect of consonants was analysed. There were two consonantal contexts that appeared to induce k-affrication (cf. Figure 1a & b). One was the presence of the labial nasal /m/, and the other contiguity with the coronal obstruents /s, ð, ʃ, t/. Although the small token count did not allow for firm conclusions regarding the effect of coronals, the articulatory plausibility of this effect as a possible process of assimilation indicates that it may be of significance (a conclusion further explored and confirmed in a forthcoming study). Apart from the above effects, other environments yielded both /k/ and /tʃ/ tokens.

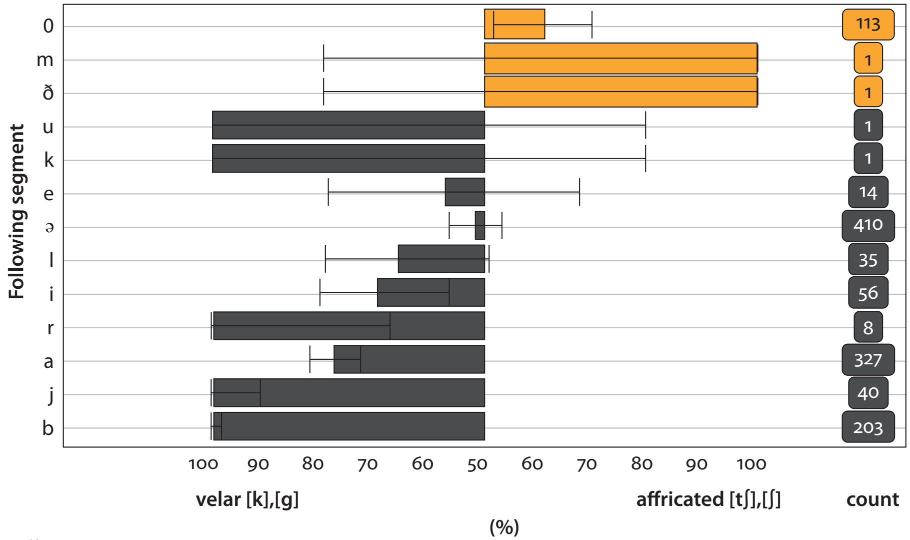
Similarly, the context for dʒ-deaffrication in the target tokens was examined to establish whether any patterns can be observed, despite previous accounts detecting no systematicity in the process. Figure 2 presents the breakdown of the produced phonemes in relation to the preceding and following segments for the deaffrication of /dʒ/ to /j/.

There appeared to be an effect of the vowel /i/, such that it was almost always preceded and followed by /j/. The other front vowel /e/ was more likely following /j/, but equally likely to precede both /j/ and /dʒ/. Deaffrication also did not commonly occur in the presence of the back vowel /u/.

As with k-affrication, there was also a possible effect of coronal segments. However, the effect progressed in the opposite direction. While /k/ was more likely to be affricated in the presence of coronal segments, /dʒ/ was always deaffricated to /j/ in the presence of the consonants /t, d, n, s/ both as the preceding (Figure 2a) and as the following (Figure 2b) segment.

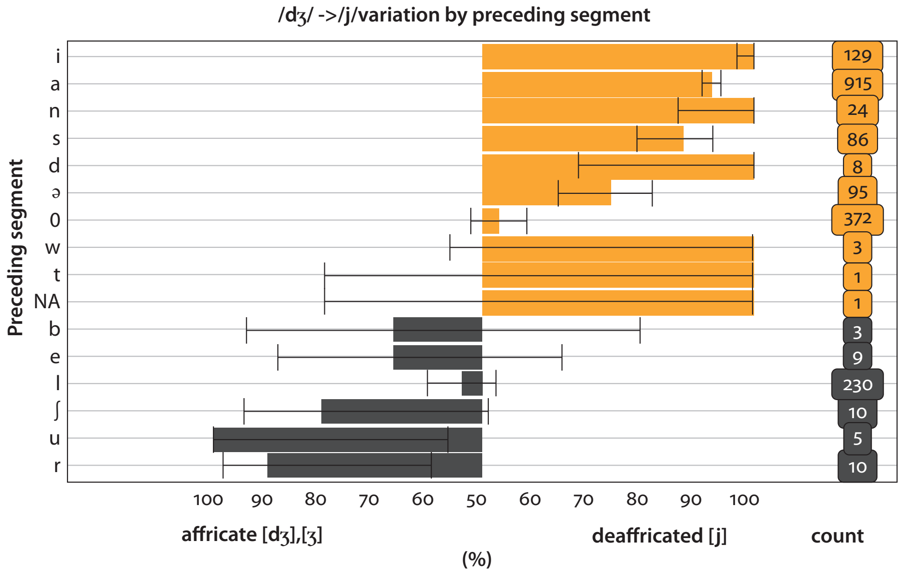


a. Preceding segment

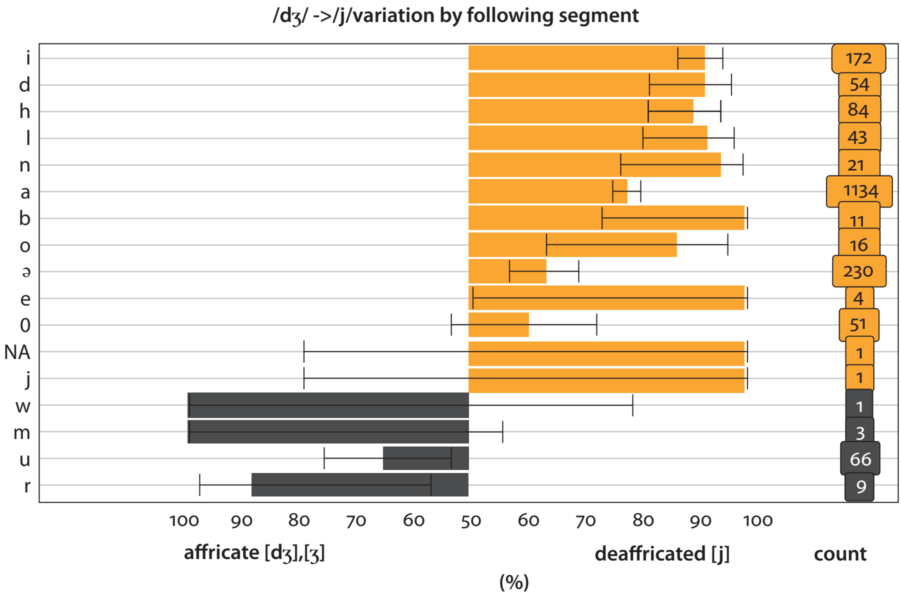


b. Following segment

Figure 1. k-affrication bias based on preceding and following segments.
(Variant /g/ pooled with /k/, and variant/ʃ/ pooled with /tʃ/)



a. Preceding segment



b. Following segment

Figure 2. dʒ-deaffrication bias based on preceding and following segments. (Variant /ʒ/ pooled with /dʒ/.)

3.2 Lexical effects

The second objective of the corpus analysis was to identify the lexical factors relevant to k-affrication and dʒ-deaffrication. Specifically, we aimed to answer the question of whether the different variants are stable within word types, and whether that can be traced back to the word origin, as suggested in previous studies. In order not to overestimate stability, only words with more than fifteen tokens were included in the analysis.

Figure 3 presents the proportion of /k/ and /tʃ/ variants in the target words. It is apparent that most of the words have a preferred form, with half occurring almost exclusively with /k/, and the other half mostly with /tʃ/, although only one word was always realised with /tʃ/. In agreement with previous studies, the only borrowing in this set, 'cake', is systematically realised with /k/. However, as evident in Figure 3, several non-borrowings were also found to never affricate despite having met the theorized conditions of affrication, that is, being in the segment environment of front vowels (e.g. /kiis/ but not /tʃiis/).

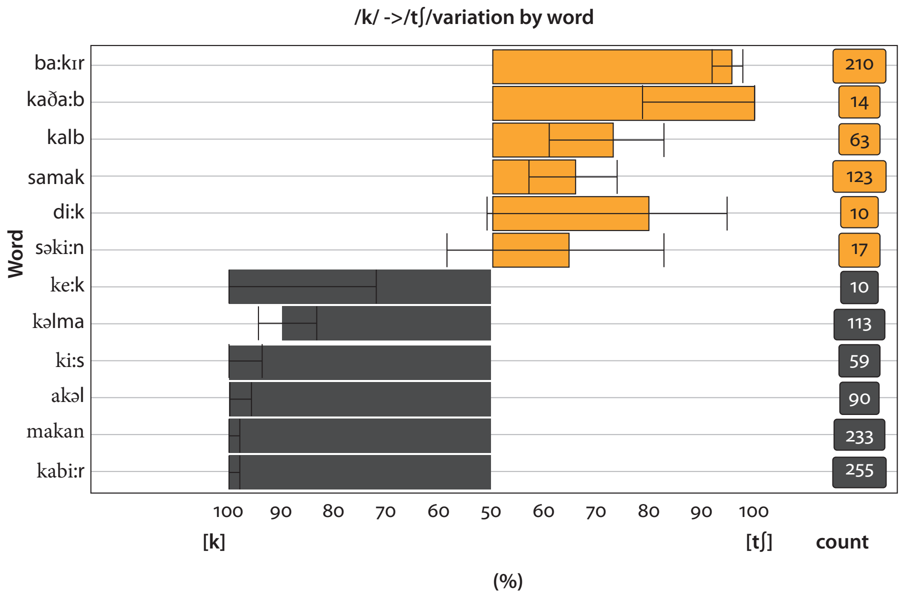


Figure 3. Within-word variability of /k/-/tʃ/

Similar results were obtained for the /dʒ/ and /j/ target words. Figure 4 presents the proportion of /dʒ/ and /j/ variants among the different types. One-third of the target words (7 out of 21) occurred consistently with the affricate variant, while most of the rest (13 of 21) were most often realised with the glide, although with some

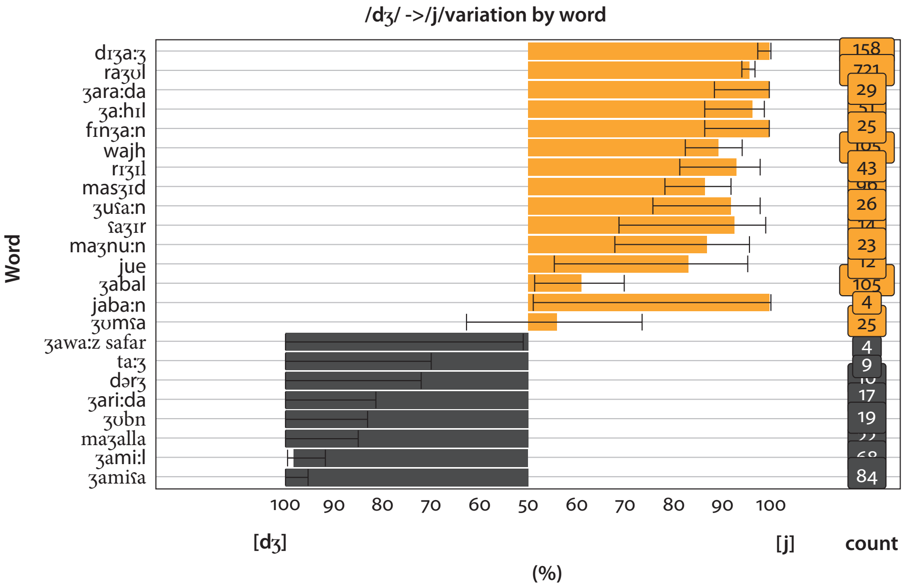


Figure 4. Within-word variability of /dʒ/-/j/

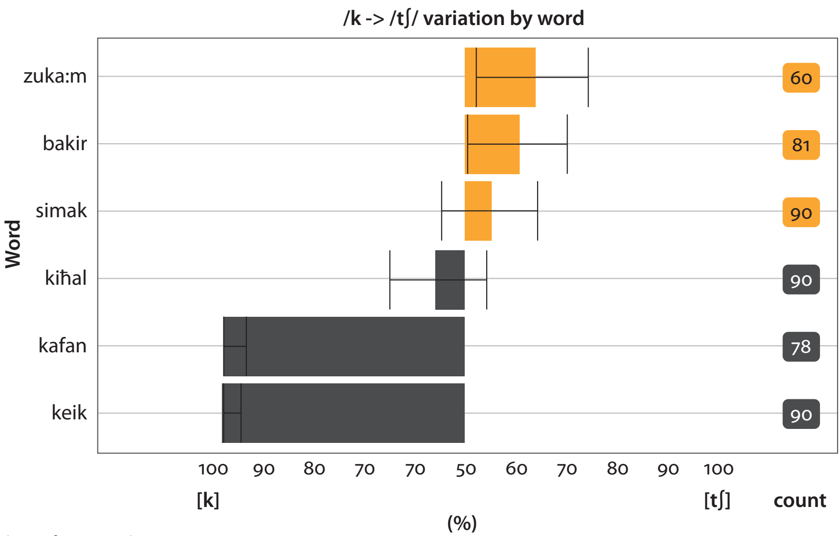
amount of variability. In line with the predictions based on previous accounts, 5 out of the 8 /k/-consistent words were borrowings. Nonetheless, another borrowing from classical Arabic, cup /findʒaan/, was always produced with a glide, as /fɪnjaan/.

In summary, the corpus results confirmed that (1) /k/ is more stable in the dialect than /dʒ/, and (2) the variability of affricates is increased in the presence of front vowels and decreased in borrowings. In addition, the results revealed a possible effect of coronals and the labial /m/.

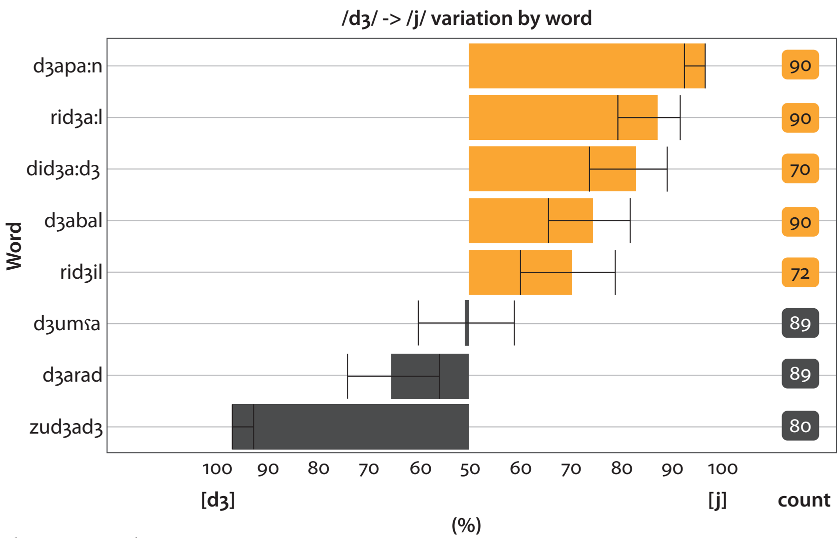
4. Experimental study results

4.1 Lexical effects

The first aim of the experimental study was to test whether the results of the corpus analysis as regards the lexical effects on affricate variability hold (1) with modern speakers (the corpus data dates as far back as the 1970's) and (2) when more than one token is collected from each speaker. The distribution of affricate variants for /k/-/tʃ/ and /dʒ/-/j/ stimuli used in the experimental study is displayed in Figures (5a) and (b). As is apparent, the observed patterns confirm the stability of borrowings (*cake, Japan*), but reveal a much higher amount of within-word variability in non-borrowings than the corpus data.



a. /k/-/tʃ/ stimuli



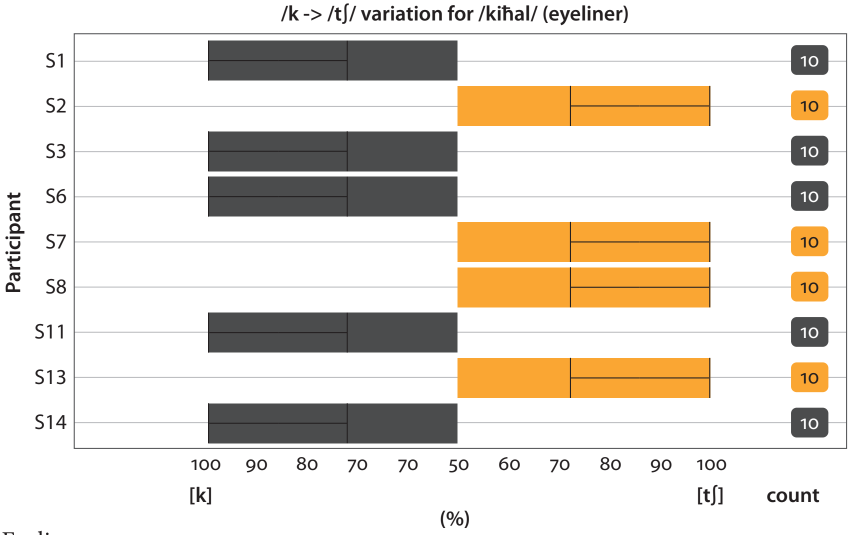
b. /dʒ/-/j/ stimuli

Figure 5. Within-word variability

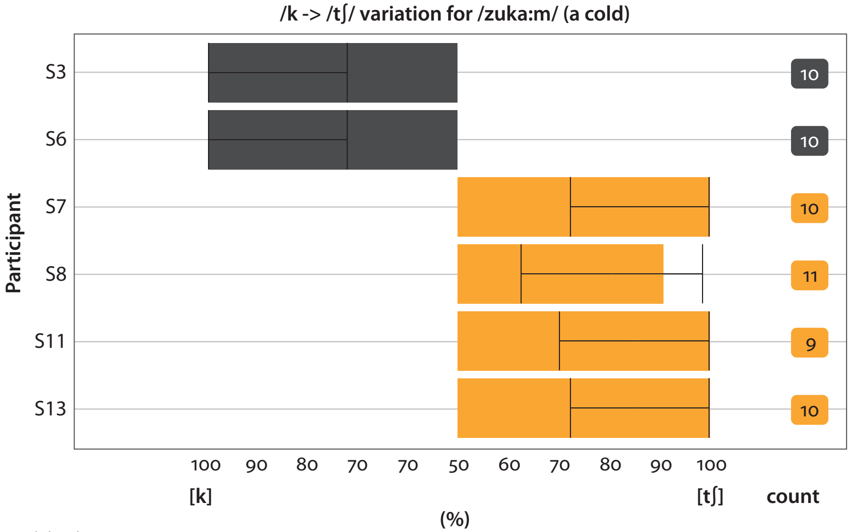
4.2 Speaker effects

The question that could not have been addressed in the corpus study due to the lack of speaker information was whether the variability in affricate realisations within the target words was due to speaker-related differences (i.e. some speakers consistently using one variant and others consistently producing the other) or whether it was the consequence of speakers being variable. However, this question could be answered with the help of the experimental data, as each participant had to produce 10 tokens of each target word.

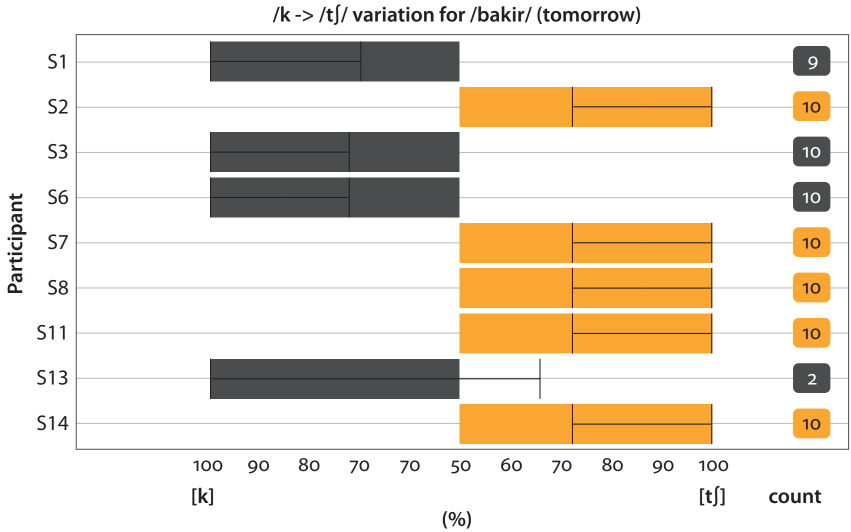
Figure 6 presents the within-word and within-speaker variability for the four /k/ – /tʃ/ stimuli that showed most variability across subject. As far as k-affrication is concerned, the analysis revealed almost across-the-board within-speaker stability, such that each speaker selected their preferred variant for each word and consistently produced it. There was only one token that showed variability (Subject 8 (b) zukaam).



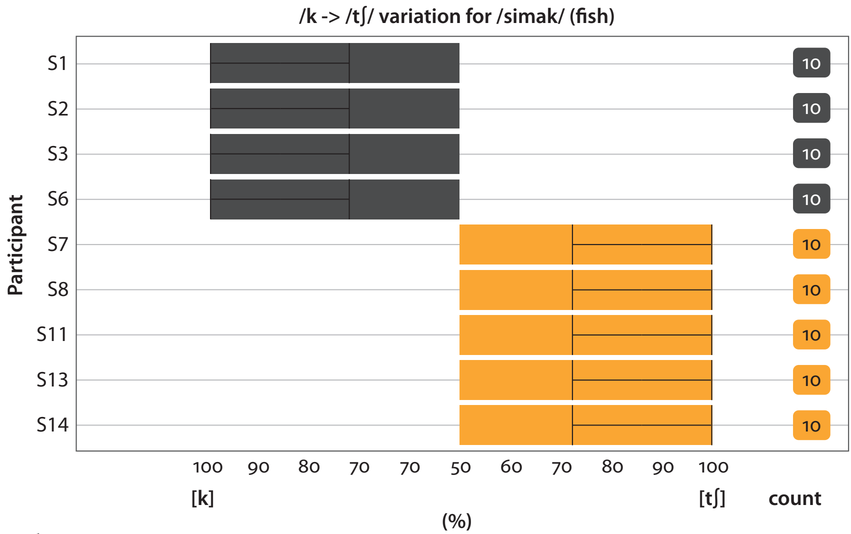
a. Eyeliner



b. Cold/Flu



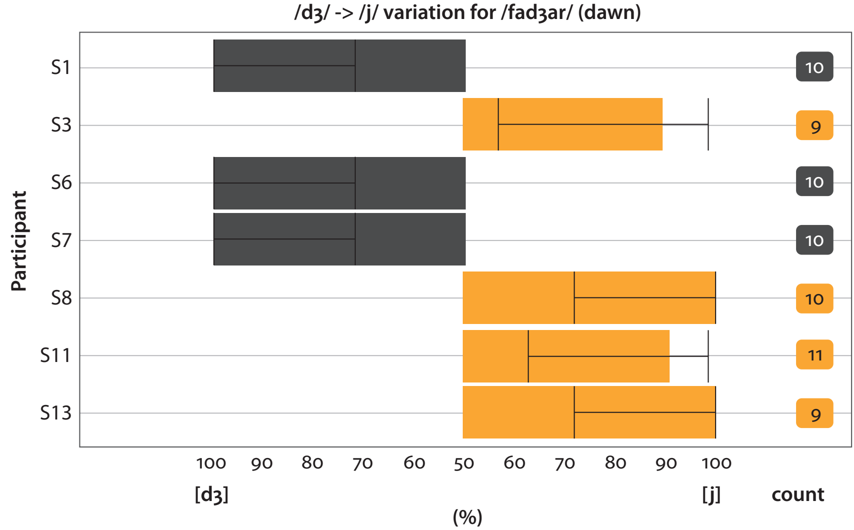
c. Tomorrow



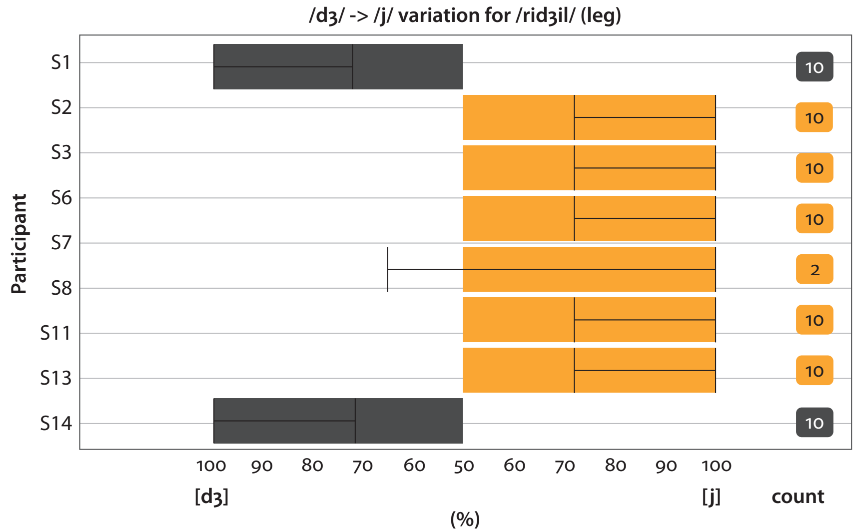
d. Fish

Figure 6. Within-word and within-speaker variability for the most variable /k/-/tʃ/ variants

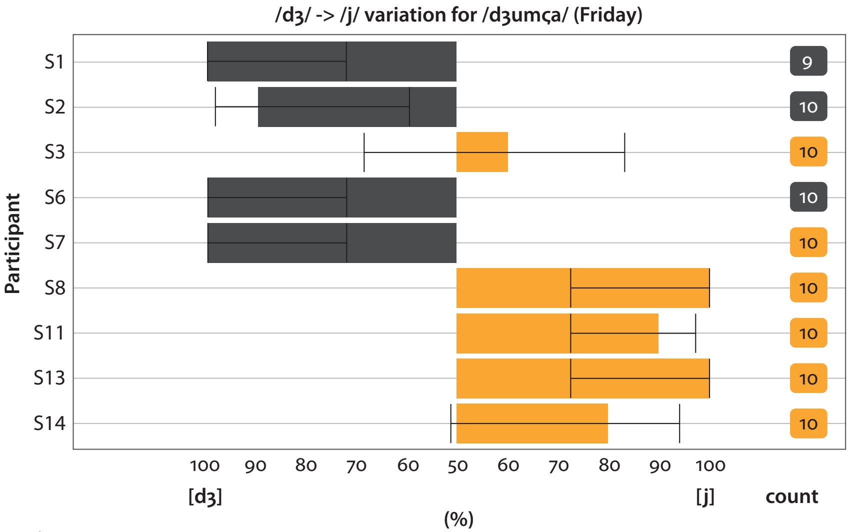
Nonetheless, the results were quite different as regards dʒ-deaffrication. Figure (7a)–(d) present within-word and within-speaker variability rates for the four most variable /dʒ/-/j/ stimuli. While dʒ-words proved to be less stable than k-words in both the corpus and the experimental results, the pattern of variability of dʒ-deaffrication was not confined to across-speaker differences. Rather, several speakers varied their productions within word type. Although the within-speaker within-word variability was not robust (ranging from 10–40% for different items and speakers), it must be kept in mind that all the tokens were collected within one session, using the same stimuli, which presumably would have had the effect of minimizing variability.



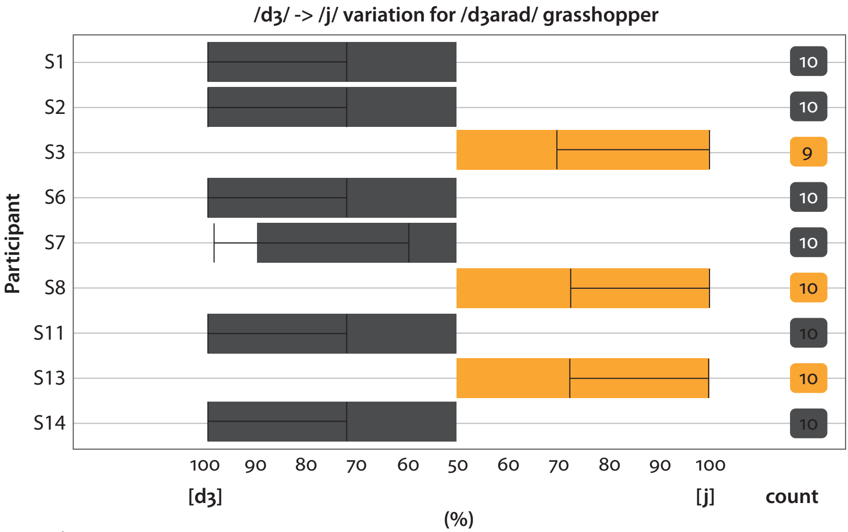
a. Dawn



b. Leg



c. Friday

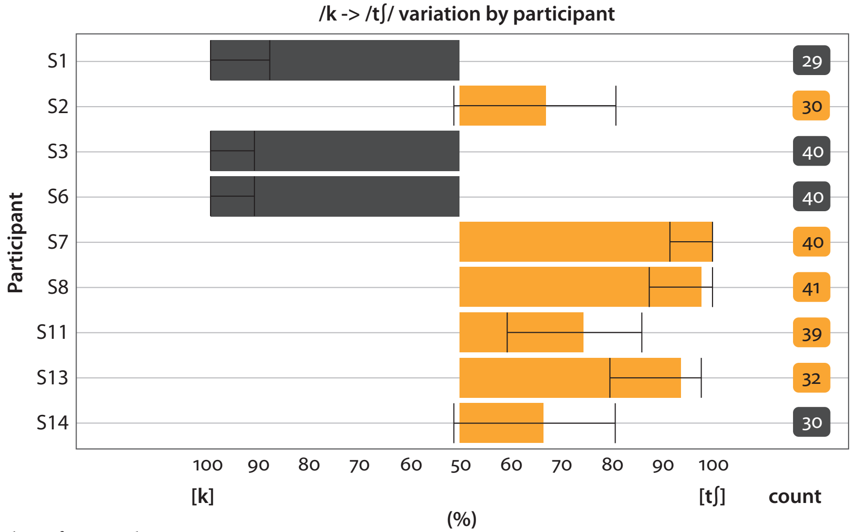


d. Grasshopper

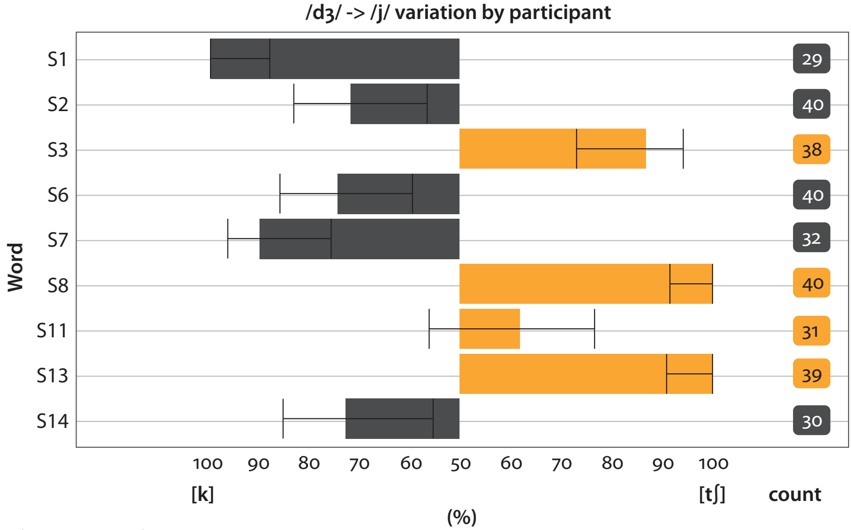
Figure 7. Within-word and within-speaker variability for /dʒ/-/j/ variants

4.3 Sociophonetic factors

The experimental results did not allow for attributing either across- or within-speaker variability to any extra-linguistic factors. Firstly, while speakers showed a tendency to produce items consistently, most did not show a preference for either the affricated or the deaffricated variant. In other words, different speakers generally kept their productions stable across tokens of the same type, but they often varied the types in which they used the phonemes in question. This trend is somewhat more pronounced for dʒ-deaffrication (only 3 out of 9 subjects consistently producing /j/), but also observable for k-affrication (4 out of 9 consistently producing one variant). Figure (8a) and (b) present the distribution of the different phonemes across types and tokens for different subjects.



a. /k/-/tʃ/ stimuli



b. /dʒ/-/j/ stimuli

Figure 8. Within-speaker variability

Due to the small sample size, it was difficult to address the question of the effects of sociolinguistic factors such as area, age and gender. However, the only two participants who were consistent in both k- and dʒ-targets (S8 & S13) showed a trend opposite to predicted: they were both from the southern emirate of Abu Dhabi and they both consistently produced the more 'local' variants.

5. Discussion

This study set out to explore the distribution of affricates and their variants in Emirati Arabic. The analysis of corpus data and experimental results reveals that the pattern of production of affricates by Emirati speakers appears to be potentially conditioned by multiple factors, some phonological, and some lexical and sociophonetic, not always confirming the generalisations proposed by previous accounts of EA, or following the regularities observed in other GA varieties.

5.1 Phonological factors

The results of our corpus analysis suggest that while generalisations based purely on vowel backness are unfounded in the dialect, some effects of fronting are in fact observable for both processes, and there is a potential effect of coronal consonants.

As regards the front vowels /a/ and /e/, neither had a systematic effect on either k-affrication or dʒ-deaffrication in our corpus data. This is perhaps predictable given the inherent variability of the front vowels /a/ and /e/ in GA, which vary with the back /ɑ, u/ depending on the surrounding consonants (Feghali, 2008), and the fact that allophonic variation was not coded for in our corpus. In fact, it can be argued that any predictions of affrication based on these vowels offer little insight into the causes of variation, when the vowel backness depends in turn on the neighboring consonants.

However, there were three other results of our corpus analysis which provide some evidence for the possible effect of fronting on affricate variability in EA. Firstly, both k-affrication and dʒ-deaffrication commonly occurred following the least variable of the GA vowels, the front /ii/. Secondly, k-affrication never occurred in the presence of the back vowel /u/, and dʒ-deaffrication was also less likely in that environment. Finally, k-affrication was also more likely following the bilabial /m/.

Furthermore, the corpus results suggest that there may also be a possible effect of coronal consonants, such that both processes occur more often in their presence. This contradicts Al Rasheedi's (2015) observations in Ha'ili Arabic and Mustafawi's findings in Qatari Arabic. According to these studies, the process is less likely to occur following homorganic consonants, suggesting a constraint against affrication in the presence of homorganic segments in those dialects. However, in our EA data, k-affrication is in fact more likely to occur in the presence of /t, ð, s, ʃ/ immediately before or after the target segment, while /dʒ/ is more likely to be realised as /j/ in the presence of /t, d, n, s/.

Taken together, the above observations contrast with some of the previous findings in other dialects of Arabic. While the vowel /ii/ has been previously associated with increased probability of the voiceless affricate, in our EA data the vowel increases the probability of affrication of /k/, but decreases the probability of a surface /dʒ/. Similarly, /k/ appears to undergo assimilation to coronals, while /dʒ/ – dissimilation.

In summary, the results of the corpus study suggest that the two processes of affrication and deaffrication pattern together (i.e. are conditioned by the same environments) and that there may be a bigger role for phonological factors in the observed patterns of variation than expected, especially for dʒ-deaffrication. Further research including more targeted stimuli should determine whether the patterns are generalisable, and whether they function as active phonological constraints.

Sociolinguistic constraints

While the sample size of our experimental study was too small to allow for generalisations, the observed trends do not provide evidence to support previous sociophonetic accounts of the processes. Previous research of (de-)affrication in GA emphasised the impact of sociolinguistic factors, specifically gender (males more ‘local’ than females), education (lower educated more ‘local’ than higher educated), and area of origin (Northern regions more ‘local’). While these effects were not present in our experimental data, a large amount of intra- and inter-speaker variation was recorded, which was not attributable to any linguistic factors. This may suggest that there exist patterns of sociolinguistic variation in the realisation of affricates which are yet to be discovered.

Many of the test items were consistently produced with the affricate by some speakers and deaffricated by others. However, against our predictions, there appeared to be no clear difference between the amount of affrication in male and female participants. In addition, contrary to our predictions, it was in fact the southern, Abu Dhabi participants who used the local variants /tʃ/ and /j/ most consistently, while the participants from the Northern Emirates were more likely to stick to the more MSA variants /k/ and /dʒ/.

As far as k-affrication is concerned, 6 out of the 9 participants settled on either /k/ or /tʃ/ for all test items. However, as regards dʒ-deaffrication, this regularity was absent: it was not the case that some speakers consistently produced the ‘local’ variants, and others consistently maintained the MSA pronunciation (only two participants did follow that strategy.) Rather, each speaker seemed to mostly favor one pronunciation for each word – although for both processes there was a slight preference for local variants (i.e. [tʃ] and [j]).

Furthermore, even the idiosyncratic speaker consistency was not fully predictable, as there was also within-speaker variability observed for some items containing /dʒ/. This is particularly interesting as the experiment was conducted in one session and in a very constrained environment – conditions that are likely to encourage stable productions. Yet among the 10 tokens produced by participants, often some were pronounced with the affricate /dʒ/, and others with the approximant /j/. This suggests that regardless of any identifiable phonological, lexical, and sociolinguistic factors, there exists an amount of variability that so far escapes explanation.

Although these data are not enough to allow for definitive conclusions, future research should address the question of whether there has been a change in the distribution and/or the social perception of the different affricate variants in the UAE, given the lack of previously reported effects in this study. It is clear that the process of dʒ-affrication tends to occur less consistently within speakers of the EA dialect than k-affrication. The question of whether this instability is a symptom of an ongoing language change is one that would be worth addressing in the future. Finally, an important aspect of the elicitation data was that some amount of variation persisted both within words and within speakers in the experimental data, which escaped a systematic explanation. Identifying the sources of this variation is a challenge that should be addressed by future studies.

5.2 Lexical constraints

The results of our corpus analysis confirmed that lexical borrowings did not appear in variable form. The borrowing status was the factor most commonly cited in previous research as the main and most stable predictor of variability – both k-affrication and dʒ-deaffrication had been proposed to never occur in borrowings. Indeed, in a small number of clear-cut cases (such as e.g. recent borrowings from English) the lexical factors were enough to explain the variability or lack thereof. However, it is worth noting that previous accounts often considered words to be borrowings even if they originated in classical Arabic, and it is likely that many such words are not perceived as borrowings by the community anymore. In the future, a more systematic investigation of lexical factors, taking into account the duration and frequency of use, is needed.

6. Conclusion

In conclusion, the distribution of affricate variants in Emirati Arabic is determined by multiple factors. While a small portion of it appears to be attributable to lexical constraints, such that borrowings remain immune to the variability, a considerable amount may be phonologically determined. The factor that appears most relevant is the presence of front vowels and consonants, which appear to have parallel effects on both k-affrication and dʒ-deaffrication. Nonetheless, the process of k-affrication appears to be more stable than dʒ-deaffrication. Future research should further explore both the phonological constraints and the sociophonetic variability of the affricates and their possible evolution over time.

Acknowledgements

Thanks to the University of Canterbury's New Zealand Institute of Language, Brain, and Behaviour for assistance in modernizing the graphs in this paper. Special thanks to Simon Todd and James Brand.

References

Al Ameri, H. (2009). A Phonological Description of Emirati Arabic. Master's Thesis, Purdue University. Retrieved from https://docs.lib.purdue.edu/dissertations/AAI1475967/

Alaodini, H. (2019). A sociolinguistic study of the Dawāsir dialect in Dammam, Eastern Arabia. Doctoral dissertation, Colchester: University of Essex.

Al Rasheedi, E. S. (2015). Affrication in Ha'ili Arabic: A rule-based approach. *International Journal of Linguistics*, 7, 27–41. https://doi.org/10.5296/ijl.v7i4.8017

Al Rojaie, Y. (2013). Regional dialect levelling in Najdi Arabic: The case of the de-affrication of [k] in the Qasimi dialect. *Language Variation and Change*. 25(1), 43–63. https://doi.org/10.1017/S0954394512000245

Al-Wer, E., Horesh, U., Alammar, D., Alaodini, H., Al-Essa, A., Al-Hawamdeh, A., Al-Qahtani, K., & Ab Hussain, A. (2020). Probing linguistic change in Arabic vernaculars: A sociohistorical perspective. *Language in Society*, 1–22. https://doi.org/10.1017/S0047404520000706

Articulate Instruments Ltd. (2012). *Articulate Assistant Advanced User Guide: Version 2.15*. Edinburgh, UK: Articulate Instruments Ltd.

As-Sammer, M. (2010). Phonological variation in Modern Standard Arabic: The case of the affricate /ʤ/: Oman as a sample. *Journal of Basrah Researches (Humanities Series)*, 36(4), 29–57.

Béland, R., & Kolinsky, R. (2005). One sound heard as two: The perception of affricates in Quebec French by Belgian French speakers. *Journal of Multilingual Communication Disorders*, 3(2), 110–117. https://doi.org/10.1080/14769670500066172

Dorai-Raj, S. (2014). *binom: Binomial Confidence Intervals For Several Parameterizations*. R package version 1.1-1. <https://CRAN.R-project.org/package=binom>

Feghali, H. J. (2008). *Gulf Arabic: The dialects of Kuwait, Bahrain, Qatar, UAE, and Oman*. Hyattsville, MD: Dunwoody Press.

Hassan, H. (2017). The palatalization of /ʤ/ into /j/ in Emirati Arabic (EA): A rule-governed or random alternation? *International Journal of Linguistics*, 9(5), 173–183. https://doi.org/10.5296/ijl.v9i5.11964

Hoffiz, B. (1995). Morphology of United Arab Emirates Arabic, Dubai dialect. Doctoral dissertation, The University of Arizona. Retrieved from https://repository.arizona.edu/handle/10150/187179

Holes, C. (1991). Kashkasha and the fronting and affrication of the velar stops revisited: a contribution to the historical phonology of the peninsular Arabic dialects. In A. S. Kaye (Ed.) *Semitic studies: In honor of Wolf Leslau*. Vol. I (pp. 652–678). Otto Harrassowitz: Wiesbaden.

Holes, C. (1995). *Modern Arabic: Structures, functions and varieties*. London and New York. Longman.

Holes, C. (2001). *Dialect, culture, and society in Eastern Arabia. Vol I: Glossary*. Leiden, The Netherlands: Brill.

Johnstone, T. M. (1963). The affrication of "kaf" and "qaf" in the Arabic dialects of the Arabian Peninsula. *Journal of Semitic Studies*, 8(2), 210–226. https://doi.org/10.1093/jss/8.2.210

Lengeris, A., & Kappa, I. (2016). Palatalization and affrication of velar stops in the (western) Cretan dialect. In M. Mattheoudakis and K. Nikolaidis (Eds.), *Selected Papers of the 21st International Symposium on Theoretical and Applied Linguistics (ISTAL 21)* (pp. 211–223). Thessaloniki: University of Thessaloniki.

Moeng, E. & Carter, W. (2019). Factors in the affrication of the ejective alveolar fricative in Tigrinya. In E. Clem, P. Jenks & H. Sande (eds.), *Theory and description in African Linguistics: Selected papers from the 47th Annual Conference on African Linguistics*, (pp. 269–285). Berlin: Language Science Press.

Mustafawi, E. M. (2006). An optimality theoretic approach to variable consonant alternations in Qatari Arabic. Doctoral dissertation, University of Ottawa, Canada.

Ntelitheos, D., & Idrissi, A. (2017). Language growth in child Emirati Arabic. In H. Ouali (Ed.) *Perspectives on Arabic Linguistics XXIX: Papers from the Annual Symposium on Arabic Linguistics*, Milwaukee, Wisconsin, 2015 (pp. 229–248). Amsterdam: John Benjamins. https://doi.org/10.1075/sal.5.10nte

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Blackwell Publishing. https://doi.org/10.1002/9780470759400

Qafisheh, H. (1975). The affricates /tʃ/ and /dʒ/ in Gulf Arabic. *Journal of the American Association of Teachers of Arabic*. 8(1), 24–40.

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Russell, A. (2014). Preservation of affrication in rural Francophone Louisiana. In J. Boissonneault & A. Reguigui (Eds.), *Langue et territoire Études en sociolinguistique urbaine* (pp. 55–101). Sudbury, Ontario, Canada: Laurentian University.

Versteegh, K. (1997). *The Arabic language*. Columbia University Press.

Yoshida, S. (2001). An element-based analysis of affrication in Japanese. In J. van de Weijer & T. Nishihara (Eds.), Issues in Japanese phonology and morphology (pp. 193–214). New York: Mouton. https://doi.org/10.1515/9783110885989.193

Youssef, I. (2014). Affrication in Baghdadi Arabic: Synchrony and diachrony. *Arabica*, 61, 746–757. https://doi.org/10.1163/15700585-12341323

II. Eye-tracking

Eye movements in Arabic reading

A Review of the current literature

Maryam A. AlJassmi[1,2], Ehab W. Hermena[1] and Kevin B. Paterson[2]
[1]Zayed University / [2]University of Leicester

Measures of eye movements provide a moment-by-moment account of the visual and cognitive processes that underlie normal reading. These measures have been used to develop detailed sophisticated computational models of eye movement control during reading, primarily based on research conducted with European languages such as English and German, which use the Roman orthography. However, relatively little is known about the mechanisms underlying eye movement control during reading for Semitic languages, such as Arabic, which have very different visual and linguistic characteristics. In this chapter, we provide an overview of eye movement research on reading, including research conducted to date on eye movements in Arabic reading. We consider whether current computational models can account for these findings and outline key theoretical questions that remain to be examined empirically.

Keywords: eye movements during reading, Arabic language, word recognition

1. Introduction

Over the past forty years, research on eye movements in reading has been instrumental in developing our understanding of cognitive mechanisms which underpin the ability to read and understand texts in real-time (for a review, see Rayner, 2009). This is because eye movements are intrinsic to reading and allow researchers to examine the process of reading as it occurs naturally. It is also because there exists a close correspondence between eye movements and underlying cognitive processes, so that quantitative measures of eye movements can provide a richly-detailed, moment-by-moment account of the underlying cognitive activity (Rayner, 1998, 2009). This research method contrasts with others which typically require participants to read text in unnatural formats (e.g., in studies using electroencephalography, i.e., EEG / ERP methods, for a recent review, see Bornkessel-Schlesewsky, Staub, &

https://doi.org/10.1075/sal.10.03alj

Schlesewsky, 2016; Swaab, Ledoux, Camblin, & Boudewyn, 2012; see also Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011) or that lack the temporal resolution required to describe underlying neural processes in real-time (e.g., in studies using functional magnetic resonance imaging, i.e., fMRI; e.g., Skeide & Friederici, 2018).

It has been argued that the process of word identification is what drives forward eye movements through text, such that how quickly the reader can identify a word determines *when* and *where* the eyes move (Reichle, Pollatsek, Fisher, & Rayner, 1998). Research investigating variables affecting this dynamic eye movement behaviour has led to the development of sophisticated computational models, which provide the current framework for investigating the cognitive basis of reading (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle et al., 1998; Reichle, Rayner, & Pollatsek, 2003). However, since much of this research has been conducted in Indo-European languages that use the Roman orthography, it is of increasing importance to examine the effects of these variables in other languages, and to establish the extent to which the models apply cross-linguistically. The focus of the present chapter is on the eye movement studies on Arabic reading, a field which is still in its infancy. In terms of its number of readers, Arabic is one of the most popular alphabetic writing systems globally (second only to Latin-scripted writing systems; Olson & Dringer, 2020), yet possesses visual and linguistic characteristics distinct from Latin-scripted languages. Conducting eye movement research on Arabic reading (and the reading of other under-researched languages) is therefore crucial for establishing the extent to which mechanisms that are thought to account for eye movement control when reading Latin-scripted languages might also apply cross-linguistically. In particular, the present chapter will (1) summarise eye movement research in Latinate languages and briefly review those studies conducted in Arabic to date; (2) provide an overview of the Arabic writing system, focusing on characteristics likely to be of interest to eye movement researchers; and (3) outline key research questions that pave the way for a better understanding of the mechanisms of eye movement control in Arabic reading.

2. Mechanisms of eye movement control in reading

When reading, our eyes move by making a series of saccades (rapid eye movements lasting 20–30 milliseconds) separated by brief fixational pauses (see, e.g., Rayner, 1998, 2009). This behaviour is a consequence of limitations in retinal acuity, which is greatest within a narrow region at the centre of gaze (extending no more than about 1°), and declines sharply with increasing eccentricity from this location even within central vision (e.g., Hilz & Cavonius, 1974). As a result, only a very small portion of text (corresponding to no more than about 4–5 letters of normal-sized text) can be processed in high resolution on each fixation, so that readers must move their eyes and make multiple fixations to process a line of text. Crucially, from the perspective of the eye movement researcher, quantitative measures of these eye movements can provide valuable information about how a reader processes text. In particular, while each fixation lasts about 250 milliseconds on average (when reading languages like English or German), there is considerable variability in their actual duration, and precisely how long the reader's eyes dwells on a word has been shown to closely reflect the difficulty experienced in processing that word (see Rayner, 2009). There is also considerable variability in the length of saccades. Typically, these traverse about 7–9 character spaces, roughly equivalent to the distance between the midpoints of adjacent words in English texts. However, saccades can be much shorter and cover as little as one character space, such as when the readers make corrective saccades to re-fixate words optimally (Bowers & Poletti, 2017; Nuthmann, Engbert, & Kliegl, 2005); or much longer, such as during a return sweep, where the reader makes a saccade that carries their gaze from roughly the end of one line of text to the beginning of the next (Parker, Nikolova, Slattery, Liversedge, & Kirkby, 2019; Slattery & Vasilev, 2019). Readers will also sometimes make a forward saccade that skips a word, so that the word is not fixated before the reader looks at the next words along in text (e.g., Brysbaert, & Vitu, 1998; Drieghe, Rayner, & Pollatsek, 2005; Rayner, Slattery, Drieghe, & Liversedge, 2011; see also Rayner, 1998, 2009). Such behaviour is essential as it suggests that saccade planning is sensitive to parafoveal processing (i.e., the pre-processing of upcoming words in a text).

3. The Perceptual span for reading

Given the limitations in retinal acuity, it is important to understand how much information can be acquired on each fixation. This issue has been addressed using gaze-contingent moving window paradigms (e.g., McConkie & Rayner, 1975, 1976; for a review, see Rayner, 2014). In these, text is presented entirely as normal within a small region (window) at the point of fixation, while text outside the window is masked (by, for example, replacing letters in words with X strings). The window is yoked to the reader's eye movements so that when they move their eyes to fixate a new location, text within the window at this new location is shown normally and text at the previous fixation location is masked. This requires high-speed computing and displays with very fast refresh rates to ensure that changes to text displays are made sufficiently rapidly (typically within 10 ms of fixation), so that the reader's phenomenological experience is that the window moves in synchrony with their eyes. In addition, the size of the moving window often is varied across an experiment, following the logic that windows that produce reading rates similar to when text is shown entirely as normal must contain all the information required for normal reading.

The area of text that encompasses this information is described as the reader's *perceptual span*. For skilled young adult readers of languages like English, this extends approximately 3–4 characters to the left of the fixated word (and so includes the beginning of the currently fixated word) and about 14–15 characters (or approximately two words) to the right (e.g., McConkie & Rayner, 1975, 1976; Rayner, Well, & Pollatsek, 1980). The perceptual span (for languages like English) is therefore asymmetric. This asymmetry is not solely a function of acuity, however, but widely believed to reflect greater allocation of attention in the direction of reading to facilitate the pre-processing of words and to guide decisions about where next to move the eyes (Balota, Pollatsek, & Rayner, 1985; Briihl & Inhoff, 1995; Drieghe et al., 2005; McConkie & Rayner, 1975; Rayner, 1975a; Rayner, Well, Pollatsek, & Bertera, 1982; White, Johnson, Liversedge, & Rayner, 2008; for a review, see Schotter, Angele, & Rayner, 2012).

As this asymmetry is believed to be a function of reading direction, several studies have investigated whether it changes for languages, such as Arabic, Hebrew and Urdu that are read from right to left compared to languages like English that are read from left to right. Such studies have investigated the perceptual span of bilingual readers who are native readers of a right to left language and proficient in English. A seminal study by Pollatsek, Bolozky, Well, and Rayner (1981) examined the perceptual span for a small sample ($N = 6$) of Hebrew-English bilinguals. They used a paradigm similar to that used in previous perceptual span studies, and compared effects for moving windows that extended asymmetrically to the right or left of fixation (where letters in words outside the windows were replaced with 'x's) with text shown entirely as normal. The results revealed that the perceptual span for English extended asymmetrically to the right, as observed previously for native English readers. However, for Hebrew, this asymmetry was reversed, so that the perceptual span extended asymmetrically to the left of fixation, consistent with greater allocation of attention in the direction of reading.

A subsequent study by Jordan et al. (2014) addressed this same question but with a larger sample size ($N = 12$) of readers who were bilingual in Arabic and English (and a follow-up study by Paterson et al., 2014, investigated the same issue with Urdu-English bilinguals). In this study, native Arabic-speaking participants with good English abilities read sentences in Arabic or English (in separate sessions). The sentences were displayed either entirely normally or text was presented as normal only within gaze-contingent moving windows. By contrast with previous studies, however, Jordan et al. used a blurring technique rather than letter substitutions to mask words outside the moving windows. They took this approach because it is standard for Arabic to be written in a proportional font as cursive text (i.e., where letters join together in words) and substituting letters in words would disrupt the normal visual appearance of text. The blurring, by comparison, allowed them to effectively mask the identities of words while preserving the approximate orthographic properties of words and their layout in text, thereby preserving the normal visual appearance of text. The findings from this study broadly replicated those reported by Pollatsek et al. (1981), showing that the perceptual span extends asymmetrically to the right in English, and asymmetrically to the left in Arabic, for this bilingual population, providing further evidence that perceptual span asymmetries are a function of reading direction.

4. Word factors affecting the spatial and temporal characteristics of eye movements

Research on the spatial and temporal characteristics of eye movements in reading has been crucial for revealing the influence of the visual and linguistic characteristics of a text on *when* and *where* the eyes move (Rayner, 1998, 2009), and central to the development of models of eye movement control (e.g., Engbert et al., 2005; Pollatsek, Reichle, & Rayner, 2006; Reichle et al., 1998, 2003). This research provides compelling evidence that eye movements are under cognitive control, such that processing difficulty associated with the characteristics of words directly influences eye movement behaviour (see, e.g., Rayner, Sereno, & Raney, 1996). The central claim is that three variables (a word's length, familiarity, operationalised as the frequency of its written usage, and its contextual predictability) primarily guide eye movements during reading, by influencing the time spent fixating words and the likelihood of skipping these words.

Numerous studies show the influence of these factors when reading languages that use the Roman orthography (like English). First, short words tend to be read more quickly and skipped more often than longer words (e.g., Joseph, Liversedge, Blythe, White, & Rayner, 2009; Juhasz & Rayner, 2003; Kliegl, Grabner, Rolfs, & Engbert, 2004; Paterson, McGowan, & Jordan, 2013; Rayner & McConkie, 1976; Rayner, Sereno, & Raney, 1996; Rayner et al., 2011; Vitu, O'Regan, Inhoff, & Topolski, 1995). Moreover, words that have a high lexical frequency are read more quickly and skipped more often than lower frequency words (e.g., Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Rayner, Sereno, & Raney, 1996). Finally, words that are more predictable from the preceding sentence context are read faster and skipped more often than words that are less predictable (Balota et al., 1985; Brothers, Swaab, & Traxler, 2015; Brothers, Swaab, & Traxler, 2017; Ehrlich & Rayner, 1981; Rayner & Well, 1996; Staub, 2015). When taken together, such findings show that the movements of the eyes during reading are strongly influenced by the visual, lexical and contextual characteristics of words.

Other studies point to word-based influences on where saccades tend to land during reading. Most end in fixations on words rather than interword spaces, suggesting that the oculomotor system aims saccades towards words using low spatial frequency cues to word boundaries (i.e., word-shaped blobs separated by spaces). Saccades towards upcoming words in text also exhibit a systematic tendency to land at a so-called preferred viewing location (PVL) between the word's beginning and middle letters, certainly in languages like English (Rayner, 1979). This further suggests that low spatial frequency cues are used to target saccades towards the centre of words but that these often fall short of this location. This is attributed to random error in oculomotor control and also the range effect, which is a tendency to overshoot close targets and undershoot more distant targets (McConkie, Kerr, Reddix, & Zola, 1988). Consequently, not all saccades land at the PVL, and saccades often will land further to the left in longer words and sometimes overshoot the centre of short words (Joseph et al., 2009; McConkie et al., 1988; Paterson et al., 2013). Other evidence that readers benefit from saccades that land at the PVL comes from the finding that readers tend to make longer fixation and are less likely to re-fixate a word (i.e., make more than one fixation on it) when the landing position is at the PVL than the beginning or end of words (e.g., Nuthmann, Engbert & Kliegl, 2005; Rayner et al., 1996; Vitu, McConkie, Kerr & O'Regan, 2001).

These findings suggest that words are recognised most efficiently when fixated at the PVL and that readers will make a corrective eye movement when saccades land at sub-optimal locations towards the beginning or ends of words. This has also led to the widespread belief that there must be some correspondence between the PVL and the *optimal viewing position* (OVP) effect reported in studies of isolated word recognition (e.g., Hyönä & Bertram, 2011; Jordan, Paterson, Kurtev, & Xu, 2010; Li et al., 2017; O'Regan, 1980; O'Regan & Jacobs, 1992; O'Regan, Lévy-Schoen, Pynte, & Brugaillère, 1984; Vitu, O'Regan, & Mittau, 1990). In these studies, words are displayed to participants at different horizontal offsets relative to a specific screen location that participants are instructed to fixate, so that the participant initially will fixate a specific inter-word position (e.g., the first, middle, or end letter). The common finding from studies conducted in languages that use the Roman orthography is that participants are quicker to recognise words (measured by word naming or lexical decision latencies, or the duration of fixations on the words) when these are fixated at locations between their beginning and middle letters. Based on the similarity between the OVP and the PVL, it often is assumed that this shows that saccades are targeted towards the PVL in textual reading to optimize word recognition.

This view is also relevant when considering OVP and PVL effects for Semitic languages like Arabic and Hebrew. Studies of OVP effects shows that, in contrast with effects for languages that use the Roman orthography, words in these languages generally are recognised most efficiently when fixated at their centre rather than between the beginning and middle letters of words (Deutsch & Rayner, 1999; Farid & Grainger, 1996; Jordan, Almabruk, McGowan, & Paterson, 2011). Various explanations of this difference in OVP effects have been proposed, including that it reflects morphological structure, and therefore the location of information about the core meaning of a word, which differs for Semitic languages like Arabic and Hebrew compared to Latinate languages like English (see Stevens & Grainger, 2003). We will return to this issue when we discuss the characteristics of the Arabic writing system.

5. Parafoveal processing in reading

As already noted, readers acquire parafoveal information from upcoming words to pre-process these words and guide saccade-targeting. This raises the question of which information is processed parafoveally, and whether words are processed serially or in a parallel fashion. Useful clues to these inquiries are generally drawn from research using gaze-contingent boundary paradigms (Rayner, 1975a; for reviews, see Hyönä, 2011; Schotter et al., 2012). To set up such a boundary paradigm, an invisible boundary is placed immediately in front of the target word in a sentence. Before the reader saccades across such boundary to reach the target word, the target word is shown entirely as normal (a valid preview) or changed (an invalid preview). The invalid preview can be created by changing all the letters in the word, typically while maintaining word length. When the participant makes a saccade that crosses the boundary, a display change is initiated so that the preview word is replaced by the target word. This display change requires high-performance computing and a display screen with a fast refresh-rate to ensure it is implemented quickly, ideally before the saccade ends (and so within 20–30 ms), and without the participant's awareness in most cases. To be certain about this, at the end of the experiment participants are usually asked if they noticed any display changes, and their entire dataset may be excluded if they noticed too many changes.

The principal consideration in these experiments is whether a preview benefit is observed. There is a processing advantage for the target word (i.e., shorter reading time) if the preview is valid. Such a finding demonstrates that the readers acquired parafoveal preview information prior to the boundary change and that their reading performance was influenced by this preview information. It is also possible, by manipulating the relationship between the preview and the target word, to investigate which information is processed parafoveally. For instance, numerous studies show a preview benefit when the target word (e.g., *song*) and preview (e.g., *sorp*) share the same beginning letters and are similar in shape (Rayner, 1975a, 1978; Rayner, Balota, & Pollatsek, 1986; Rayner, McConkie, & Zola, 1980). These findings suggest that readers acquire orthographic information parafoveally and that previewing the first couple of letters in words can facilitate target word recognition, presumably because this information helps initiate lexical access (Balota et al., 1985; Briihl & Inhoff, 1995; Drieghe et al., 2005; Inhoff, 1987, 1989a, 1989b, 1990; Inhoff & Tousman, 1990; Lima & Inhoff, 1985; Rayner, 1975b; White et al., 2008). Other studies show that phonological information can also be acquired parafoveally, so that there is a preview benefit when the preview and target word sound similar but are composed of different letters (i.e., using either homophones such as *bear* and *bare*, or pseudohomophones such as *brain* and *brane*; e.g., Pollatsek, Lesch, Morris, & Rayner, 1992; see Schotter et al., 2012, for a review). Whether semantic information is processed parafoveally is more controversial. In English, no preview benefit is observed when the preview and target words are semantically related (e.g., *tune* & *song*; Rayner et al., 1986). However, such effects have been reported in German (Hohenstein, Laubrock, & Kliegl, 2010) and Chinese (Yan, Richter, Shu, & Kliegl, 2009). Similarly, whereas evidence seems to suggest that Hebrew readers extract morphological information from upcoming words in the parafovea (Deutsch, Frost, Pelleg, Pollatsek, & Rayner, 2003; Deutsch, Frost, Pollatsek, & Rayner, 2000, 2005), investigations in English have failed to replicate the same findings (Inhoff, 1989b; Juhasz, White, Liversedge, & Rayner, 2008; Kambe, 2004). This raises the possibility that the extent to which parafoveal processing is observed may vary cross-linguistically depending on the nature of the orthography and other characteristics of the writing system.

6. Models of eye movement control in reading

Research discussed in the previous sections has been the basis for the development of computational models of eye movement control. These provide an explicit formalisation of theoretical accounts that have used to simulate key findings in the literature and to generate and test theoretical predictions. Two currently dominant models, the E-Z Reader model (Pollatsek, Reichle, & Rayner, 2006; Reichle et al., 1998; Reichle et al, 2003) and the SWIFT model (Engbert & Kliegl, 2012; Engbert, Longtin, & Kliegl, 2002; Engbert et al., 2005), build on an earlier generation of models that aimed to explain perceptual and attentional processes in reading, including the Morrison model (Morrison, 1984) and the Mr. Chips ideal-observer model (Legge, Klitz, & Tjan, 1997). We briefly outline these two models, along with the more recent OB1 model (Snell & Grainger, 2019; Snell, van Leipsig, Grainger, & Meeter, 2018).

All three models start from the assumption that readers acquire linguistic information from within a narrow region (the perceptual span) across which attention is distributed and that, due to acuity limits, words nearer fixation are seen most clearly. An important distinction between these models, however, relates to the extent to which words within the perceptual span are processed serially or in parallel. A key assumption of the E-Z Reader model is that words are identified serially, in the order in which they appear in text. According to this model, an initial stage of pre-attentive visual processing informs a two-stage process of lexical identification, inspired by activation-verification models of word recognition (e.g., Paap, Newsome, McDonald, & Schvaneveldt, 1982). During the first stage, called the familiarity check (called L1), the system assesses if the currently fixated word is likely to be recognised imminently. This decision is influenced by information about word length, frequency and predictability, and is completed faster if the word is short, high frequency or highly predictable. Full lexical identification is slower, however, and only achieved following completion of a second stage of lexical identification (called L2), during which information about the word's orthography, phonology and meaning is accessed. Once the familiarity check is complete, the system begins programming a saccade and shifts attention to the next word. As soon as attention is switched to the next word, this initiates a familiarity check for this word. If this is completed before saccade-programming finishes, the saccade is terminated and a new saccade program initiated that will skip the next word and target a word following it. The likelihood of this occurring is a function of the length, frequency, and predictability of the next word, so that the model uses sequential shifts in attention to account for parafoveal processing and word-skipping.

Compared with E-Z Reader, SWIFT allows for multiple words to be processed in parallel and assumes that saccade programming is not directly informed by lexical identification. In this model, attention is allocated as a distributed gradient across words. This allows parallel lexical processing of several words at a time, although as the attention gradient allocates most attention to the fixated word and less to words further from fixation, words will be processed at different rates. Decisions about when to make a saccade are determined by a random timer, and saccades directed towards the word that is most active in the attention gradient but not yet lexically identified. Although these decisions are not triggered by word identification in the model, difficulty with lexical processing can interrupt this process to delay the system making a saccade. Finally, word-skipping occurs because of parallel lexical processing and takes place when the next word along has already been identified prior to the completion of a saccade programme so that the subsequent saccade is directed towards a later word in the sentence.

OB1 is a more recent example of a parallel-graded attention model that also incorporates the assumption that multiple words can be processed in parallel. Similarly to SWIFT, it also assumes that saccades are generated at a random pace but can be inhibited when word recognition is difficult, and that saccades are targeted towards words that are most active in the attention gradient but not yet identified. A distinctive feature of OB1 is that it includes an account of lexical and sub-lexical processing (at the level of letters and bigrams) while existing models do not incorporate specific mechanisms for word recognition.

Despite their differences, both E-Z Reader and SWIFT models can successfully predict visual and lexical influences on eye movements during reading (and see Reichle, Warren, & McConnell, 2009, for a version of the E-Z Reader model that tackles higher-level language processing). An important concern for the present chapter is whether these models can, in principle, account for these influences on eye movement control during Arabic reading. This is important because it provides evidence for the cross-linguistic generalizability of the different models and creates a framework for future research.

7. The Arabic writing system

Arabic is a Semitic language that uses a phonemic alphabet with 28 letters (Abu-Rabia, 1997; Daniels, 2013). Some letters are visually similar, in the sense they share the same grapheme, and are distinguishable by the number and location of small dots (Alotaibi, 2007; Perea, Abu Mallouh, Mohammed, Khalifa, & Carreiras, 2016). For example, the letters /b/ (ب), /t/ (ت), /θ/ or soft th (ث) share the same basic grapheme but differ in the number of dots above or below it. Text in this language is read horizontally from right to left and printed in cursive script in which most letters in words link with the letters to each side using short horizontal lines (called ligatures). There are six letters (و، ز، ر، ذ، د، ا) that are exceptions to this. These link only to the preceding letter and so create spaces that divide words into separate parts. Some letter shapes change because of this linking between letters and their location in words (see Carreiras, Perea, & Abu Mallouh, 2012; Daniels, 2013; see also Boudelaa, Norris, Mahfoudhi, & Kinoshita, 2019). For instance, the letter /ain/ can be written as عـ,ـعـ, or ـع depending on whether it occupies a beginning, medial or end location in words, and as ع when it is not followed by another letter and follows one of the six exception letters that do not link with other letters. Finally, all letters in Arabic are phonemic consonants except for three (ا، و، ي) that can additionally serve as long vowels. Short vowels are occasionally represented using additional small diacritic mark above or below a consonant. For example, adding different diacritics to the consonants /ktb/(كتب) changes the pronunciation and meaning to /kataba/ (كَتَبَ) meaning "he wrote", /kutiba/ (كُتِبَ) meaning "it's written" and /kutubun/ (كُتُبٌ) meaning "books". Therefore, the presence of diacritics increases the transparency of the language and facilitates word recognition (Abu-Rabia & Taha, 2006). However, other than in religious texts or children's books, including school books, diacritics are only rarely used in written Arabic and usually only to disambiguate words (Hermena, Drieghe, Hellmuth, & Liversedge, 2015; Lallier et al., 2018).

In languages like English, morphology is mainly inflectional, with prefixes and suffixes added linearly to expand the basic meaning of the word stem. For instance, the word "print" can be inflected using suffixes to produce "printed" or "printer" and a prefix to produce "reprint". By comparison, Arabic uses a non-concatenative morphology in which the root and inflectional parts of a word intermingle (Abu-Rabia, 2002; Boudelaa & Marslen-Wilson, 2001, 2004). The root in Arabic usually corresponds to three consonants, while the word pattern is a sequence of letters that provide inflectional meaning about word class, gender and number. These two units intermingle such that, while root letters retain their order, they are interspersed with the word pattern. For example, the verb /jaktubu/(يكتب) meaning "he writes", the noun /kita:ba/ (كتابة) meaning "writing", and the noun /ka:tib/ (كاتب) meaning "writer", are all derived from the three-letter root /ktb/(كتب) by interweaving the root with different word patterns. Because of this interweaving, root letters, which convey the core meaning of a word are much less likely to appear at the beginning of a word in non- concatenative languages like Arabic compared to concatenative languages like English. This is potentially important as these beginning letters can provide valuable parafoveal information when reading Latinate languages like English (e.g., Blanchard, Pollatsek, & Rayner, 1989; Rayner, McConkie, & Zola, 1980; for a review, see Schotter, Angele, & Rayner, 2012), and so the absence of this information in Arabic words may limit parafoveal processing when reading this language.

More generally, because of their morphology, Arabic words tend to be more linguistically dense compared to words in Latinate languages like English. They are generally short, with 90% of Arabic words between 4 and 8 letters long (Boudelaa & Marslen-Wilson, 2010), while often expressing complex meanings. For instance, the five-letter word /katabata:/ (كتبتا) translates as the four-word phrase "the two females wrote" in English. This greater complexity of orthography and morphology might increase the word decoding time, potentially resulting in slower reading speeds for Arabic compared to other languages (Brysbaert, 2019; Hermena et al., 2019; Roman & Pavard, 1987; see also Liversedge, Drieghe, Li, Yan, Bai & Hyönä, 2016).

8. Eye movement research on Arabic reading

There is a growing body of research on Arabic word recognition. However, very few studies investigate the mechanisms of eye movement control in Arabic reading. In an earlier section, we outlined one of the few such studies, which reported an eye movement experiment on perceptual span effects in Arabic reading (Jordan et al., 2014). This showed that, as with Hebrew, the perceptual span is asymmetrically longer to the left of fixation in Arabic reading. Several other eye movement studies have investigated other aspects of word and sentence processing during Arabic reading, including factors, such as word length and lexical frequency, that have a central role in eye guidance in current models of eye movement control in reading.

One study that investigated word length effects examined eye movements while reading sentences containing either a 3-, 5-, or 7-letter target word (Paterson, Almabruk, McGowan, White, & Jordan, 2015). These words were matched for contextual predictability and lexical frequency (using the Aralex database of Arabic word frequencies; Boudelaa & Marslen-Wilson, 2010), across the word lengths. Words in Arabic with the same number of letters can vary quite substantially in spatial length because individual letters vary in width. In this study, the words were carefully selected, so that letter and spatial lengths were closely matched. Reading times were much longer than for the same length of words in Latinate languages like English (see Rayner, 2009), which may be a consequence of the greater information density of Arabic words (e.g., Roman & Pavard, 1987). However, as with languages like English, longer letter length was also associated with increased reading times and higher re-fixation probabilities for words. Moreover, this effect was observed during the initial processing of the words (prior to the reader fixating words to the left and so later in the sentence), and so the study provided clear evidence that, as with languages that use the Roman orthography, word length influences decisions about when to move the eyes in Arabic reading. Word length effect on word-skipping, by contrast, were restricted to the 3-letter words (with no difference between 5-letter and 7-letter words) and even then the skipping rates were much lower than reported for languages like English and German (only 10% for 3-letter words). Consequently, while word length also influenced decisions about when to move the eyes, skipping rates were generally very low and largely insensitive to word length. The study also looked at the landing positions of saccades on these words and found that these tended to land between the beginning and middle letters of the longer words (and so slightly to the right of word centre) and to slightly overshoot the centre of short words. This was essentially the same pattern as reported for Hebrew (Deutsch & Rayner, 1999), and opposite to that reported for languages that use the Roman orthography, and so suggests that word length information is used for saccade-targeting similarly across these languages.

Hermena, Liversedge and Drieghe (2017) expanded upon the findings of Paterson et al. (2015) in a study that aimed to disentangle the effects of letter and spatial length on reading times and saccade-targeting. To do so, they exploited the natural variability in the spatial length of words containing the same number of letters when Arabic words are printed using a conventional proportional font in which letters are not equally wide. Hermena et al. selected a set of 5-letter and 7-letter target words that were matched for spatial length so that these sets of words occupied the same amount of horizontal space that was classified as either narrow or wide (wider 5- and 7-letter words were about 10 pixels wider than narrow 5- and 7-letter words). The researchers presented these words in sentences using a proportional Arabic font. Hermena et al. also found that reading rates were high and word-skipping rates were low compared to those of languages like English and German. However, their results also showed that number of letters rather than spatial length determined the reading times for words. However, saccadic control was influenced exclusively by the spatial length of words, so that word-skipping and saccade landing positions were influenced by spatial length and not the number of letters in words. Hermena et al. also examined the influence of initial bigram characteristics on saccade-targeting. Participants read sentences containing 6- or 7-letter target words with initial bigrams that were (a) very high frequency (الـ, /al/, which is the article *the*), (b) relatively high frequency (للـ, /lil/, meaning *for the*), or (c) low frequency, where initial bigrams belonged to the word stem. They found that landing positions were insensitive to the initial bigram frequency and that saccades always landed towards the centre of words, providing evidence that saccadic control is insensitive to word orthography in Arabic reading.

A further study by Hermena and his colleagues examined frequency effects on eye movements in Arabic reading (Hermena, Liversedge, Bouamama, & Drieghe, 2019). They did this by embedding in sentences words that were of high or low frequency but matched for length (i.e., number of letters and spatial length) and contextual predictability. Again, reading times were high and word-skipping rates were low compared to languages like English and German. However, readers made fewer fixations and had shorter reading times for the higher frequency words, as observed for languages that use the Roman orthography. However, skipping rates for these words were low (less than 4%) and did not differ for high- compared to low-frequency words. This suggested that, in Arabic reading, readers are less likely to use information about the lexical frequency to skip words. With a follow-up experiment, Hermena et al. examined the influence of frequency of a word's root (i.e., the consonantal trigram that conveys the core meaning of the word). They found no effect on reading times when words were matched for overall word frequency (as well as length and predictability) but had either a high or low root frequency, suggesting that familiarity with the core meaning of the word did not influence the process of lexical identification during reading. Studies conducted to date therefore provide evidence that two of the three principal factors that guide the lexical processing of words during reading in languages that use the Roman orthography also influence reading times for words but have a much weaker influence of word-skipping in Arabic reading. It will also be important to establish whether contextual predictability influences eye movement control during Arabic reading and, in particular, whether it can facilitate parafoveal word recognition and so potentially increase skipping rates for words.

Several other eye movement studies have investigated the processing of diacritics during Arabic reading. These, as we described earlier, are small marks used to indicate vowel information that are placed above or below the consonants in the word. These are not widely used in modern writing, except in religious texts and children's books, or occasionally in other texts to disambiguate words. One study, by Hermena, Drieghe, Hellmuth, and Liversedge (2015) investigated the processing of this information in normal reading by presenting participants with sentences that either contained no diacritics, were fully vowelised by including diacritics in every word, or where diacritics were used only to disambiguate a specific homographic verb (as passive) in the sentence. Compared to sentences containing no diacritics, the fully vowelised sentences took slightly (but reliably) longer time to read, as they had fewer fixations that were longer in duration. Hermena et al. explained this in terms of increased visual crowding due to the added diacritical marks. In addition, reading times were shorter when homographic words were disambiguated using diacritics compared to when not, indicating that skilled Arabic readers can make rapid use of diacritic information to disambiguate words. However, this benefit of diacritics was not observed when sentences were fully vowelised, indicating that the selective use of diacritics in most modern Arabic writing serves to cue readers in the presence of ambiguity. By comparison, this information may have been redundant in fully vowelised sentence and so no longer served as a useful cue for ambiguity resolution.

A follow-up study by Hermena, Liversedge, and Drieghe (2016) investigated whether diacritics are processed parafoveally (that is, preprocessed before the word is fixated or skipped). They used the boundary paradigm described earlier. In the experiment, an invisible boundary was placed immediately before the target word, which was always a heterophonic homographic noun that had a dominant or subordinate meaning (based on the form of diacritisation/ pronunciation). Three types of preview were created for words that were disambiguated in favour of the dominant meaning by the use of diacritics (e.g., as in قَدَرٌ /qadarun/, meaning 'destiny'). In one, the preview provided the accurate diacritic information. In another, the preview provided inaccurate diacritics that disambiguated the word in favour of an alternative meaning (e.g., قِدْرٌ /qidru/, meaning pot). A third preview provided no diacritic marks so that the preview was ambiguous between the two meanings (e.g., قدر /qadaru/ or /qidru/, could mean destiny or pot). Compared to accurate previews, inaccurate previews produced longer reading times on these words, and the parafoveal processing effects of accurate or inaccurate previews were heavily modulated by the launch distance. Launch distance refers to the distance between the location of the fixation on the pretarget word, and the location of the initial fixation on the target word. In line with previous research findings, the obtained effects were greater for closer launch sites. The results therefore indicated that, subject to the quality of the parafoveal preview (as determined by launch distance), diacritic information can be processed parafoveally to facilitate word disambiguation.

9. Final comments and future directions

In the present chapter, we outlined the approach taken to investigating eye movements during reading in languages that use the Roman orthography and described the relatively few studies that have investigated eye movement control during Arabic reading. These show that, like readers of languages that use the Roman orthography, Arabic readers make similar use of letter length and lexical frequency to recognise words as readers of languages like English and German. Moreover, studies examining the effects of word length on saccade-targeting show that Arabic readers use information about spatial length but not letter length to target their saccades towards the centre, which is due to the fact that oculomotor errors often land towards the right of the centre for longer words and near the centre for shorter words. Such findings suggest that factors shown to be fundamental to eye movement control in reading languages that use the Roman orthography are also likely to have a key influence on eye movement control during Arabic reading.

Other aspects of the findings obtained to date also reveal interesting differences in eye movement behaviour when reading Arabic rather than languages like English and German. First, reading times tend to be longer in Arabic compared to languages that use the Roman orthography, indicating that readers have greater difficulty identifying these words, potentially because of their greater informational density (e.g., Roman & Pavard, 1987). At the same time, whereas skipping rates for words in English can be 15% or higher depending on the difficulty of text and expertise of the reader, skipping rates appear much lower in Arabic (typically less than 10%). Moreover, word-skipping is only weakly affected by word length (Paterson et al., 2015) and mostly unaffected by the lexical frequency of words (Hermena et al., 2019), while saccade-targeting appears to be unaffected by the presence of highly familiar orthographic information at word boundaries (Hermena et al., 2017). Such findings point to more limited parafoveal processing during reading for Arabic compared to languages like English and German. It will be important for subsequent research to investigate why parafoveal processing is more limited in this language and to establish the boundary conditions under which parafoveal processing effects might be observed in Arabic reading.

Studies of the influence of diacritics on word recognition might provide the first insights into this. These are used occasionally in everyday text to disambiguate words and research by Hermena et al. (2016) shows that Arabic readers are sensitive to this information and have faster reading times when the presence of diacritics disambiguate words. However, there is little evidence that even this information is used extensively in parafoveal processing as the benefit of receiving accurate rather than inaccurate or no diacritic parafoveal previews of words was observed only when the prior fixation was very close to the word. This may be because these important cues to word identity are only visible to readers when the prior fixation is close to the word. Other aspects of the Arabic orthography also mitigate against observing clear parafoveal processing effects. In particular, because Arabic uses a non-concatenative morphology, those letters that convey the core meaning of a word are less likely to appear near the beginning of words in Arabic compared to in languages like English and German, and so also may be less visible to readers during parafoveal processing. Alternatively, reduced parafoveal processing and lower skipping rates in Arabic reading may reflect both the greater complexity and ambiguity of Arabic words. The greater complexity is also a consequence of the morphological structure of the language. In particular, unlike in English, grammatical information about gender and number is included within the word rather than as a separate function word, with the consequence that Arabic words convey complex grammatical information that is unlikely to be processed parafoveally and may require that words are fixated more frequently during reading. Moreover, because vowels are not marked orthographically, and readers must use context to determine the meaning and pronunciation of words, this ambiguity in the orthographic form of a word may also require readers to fixate words more frequently. The characteristics of the Arabic script may also limit parafoveal processing during reading, resulting in lower skipping rates and longer reading times for words. Evidence from studies using cursive text suggests the use of cursive script slows the recognition of words (Roldán, Marcet, & Perea, 2018; see also Yakup, Abliz, Sereno, & Perea, 2015). This may slow the recognition of fixated words and impede parafoveal processing during Arabic reading, while such effects are likely to be aggravated by the variability in letter shape depending on the location of letters in words and the characteristics of adjacent letters.

An obvious first step in addressing this question will involve examining the effects of contextual predictability on word recognition. This will be valuable for establishing the extent to which parafoveal processing is limited by other factors influencing the pre-processing of upcoming words. In particular, if the findings show that highly predictable words are skipped only when these are very short and grammatically very simple, this will provide good evidence that the grammatical complexity of Arabic words mitigates against these being easily recognised during parafoveal processing. It will also be important, however, for future eye movement research to consider the consequences of the characteristics of the Arabic writing system for different populations of readers, including developing child readers, older adult readers, and groups with specific visual or reading impairments.

Finally, it will be important to investigate whether current models of eye movement control during reading can account for Arabic reading behaviour. Research to date suggests that key variables implemented in these models to account for reading in languages that use the Roman orthography are also important for modelling Arabic reading behaviour. However, it also seems likely that parameters in the models governing lexical processing and saccade-targeting will be required to account for longer reading times, reduced word-skipping and decreased parafoveal processing in this language. Such work nevertheless will be valuable for extending the reach of these models and understanding mechanisms of eye movement control for reading across different writing systems.

References

Abu-Rabia, S. (1997). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers. *Reading and Writing*, 9, 65–78. https://doi.org/10.1023/A:1007962408827

Abu-Rabia, S. (2002). Reading in a root-based-morphology language: the case of Arabic. *Journal of Research in Reading*, 25, 299–309. https://doi.org/10.1111/1467-9817.00177

Abu-Rabia, S., & Taha, H. (2006). Reading in Arabic orthography: Characteristics, research findings, and assessment. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 321–338). Mahwah, NJ: Erlbaum.

Alotaibi, A. Z. (2007). The effect of font size and type on reading performance with Arabic words in normally sighted and simulated cataract subjects. *Clinical and Experimental Optometry*, 90, 203–206. https://doi.org/10.1111/j.1444-0938.2007.00123.x

Balota, D., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17, 364–390. https://doi.org/10.1016/0010-0285(85)90013-1

Blanchard, H. E., Pollatsek, A., & Rayner, K. (1989). The acquisition of parafoveal word information in reading. *Perception & Psychophysics*, 46, 85–94. https://doi.org/10.3758/BF03208078

Bornkessel-Schlesewsky, I., Staub, A., & Schlesewsky, M. (2016). The time course of sentence processing in the brain. In G. Hickok & S. L. Small (Eds.), *The neurobiology of language* (pp. 607–620). North Holland: Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00049-3

Boudelaa, S., & Marslen-Wilson, W. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, 81, 65–92. https://doi.org/10.1016/S0010-0277(01)00119-6

Boudelaa, S., & Marslen-Wilson, W. (2004). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language*, 90, 106–116. https://doi.org/10.1016/S0093-934X(03)00424-3

Boudelaa, S., & Marslen-Wilson, W. D. (2010). Aralex: a lexical database for Modern Standard Arabic. *Behavior Research Methods*, 42, 481–487. https://doi.org/10.3758/BRM.42.2.481

Boudelaa, S., Norris, D., Mahfoudhi, A., & Kinoshita, S. (2019). Transposed letter priming effects and allographic variation in Arabic: Insights from lexical decision and the same–different task. *Journal of Experimental Psychology: Human Perception and Performance*, 45, 729–757. https://doi.org/10.1037/xhp0000621

Bowers, N. R., & Poletti, M. (2017). Microsaccades during reading. *PLoS ONE*, 12, e0185180. https://doi.org/10.1371/journal.pone.0185180

Briihl, D., & Inhoff, A. W. (1995). Integrating information across fixations during reading: The use of orthographic bodies and of exterior letters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 55–67. https://doi.org/10.3758/BF03202439

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: prediction takes precedence. *Cognition*, 136, 135–149. https://doi.org/10.1016/j.cognition.2014.10.017

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, 93, 203–216. https://doi.org/10.1016/j.jml.2016.10.002

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language* 109, 1–30. https://doi.org/10.1016/j.jml.2019.104047

Brysbaert, M., & Vitu, F. (1998). Word skipping: implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception I* (pp. 125–147), Elsevier Science Ltd; North Holland. https://doi.org/10.1016/B978-008043361-5/50007-9

Carreiras, M., Perea, M., & Abu Mallouh, R. (2012). Priming of abstract letter representations may be universal: The case of Arabic. *Psychonomic Bulletin & Review*, 19, 685–690. https://doi.org/10.3758/s13423-012-0260-8

Daniels, P. T. (2013). The Arabic writing system. In J. Owens (Ed.), *The Oxford handbook of Arabic linguistics* (pp. 422–431). New York, NY: Oxford University Press.

Deutsch, A., Frost, R., Pelleg, S., Pollatsek, A., & Rayner, K. (2003). Early morphological effects in reading: Evidence from parafoveal preview benefit in Hebrew. *Psychonomic Bulletin & Review*, 10, 415–422. https://doi.org/10.3758/BF03196500

Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2000). Early morphological effects in word recognition in Hebrew: Evidence from parafoveal preview benefit. *Language and Cognitive Processes*, 15, 487–506. https://doi.org/10.1080/01690960050119670

Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2005). Morphological parafoveal preview benefit effects in reading: Evidence from Hebrew. *Language and Cognitive Processes*, 20, 341–371. https://doi.org/10.1080/01690960444000115

Deutsch, A., & Rayner, K. (1999). Initial fixation location effects in reading Hebrew words. *Language and Cognitive Processes*, 14, 393–421. https://doi.org/10.1080/016909699386284

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: analyses and review. *Journal of Experimental Psychology: General*, 140, 552–572. https://doi.org/10.1037/a0023885

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 954–969. https://doi.org/10.1037/0096-1523.31.5.954

Ehrlich, S., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20, 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

Engbert, R., & Kliegl, R. (2012). Parallel graded attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 787–800). Oxford, New York: Oxford University Press.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42, 621–636. https://doi.org/10.1016/S0042-6989(01)00301-7

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112, 777–813. https://doi.org/10.1037/0033-295X.112.4.777

Farid, M., & Grainger, J. (1996). How initial fixation position influences visual word recognition: A comparison of French and Arabic. *Brain and Language*, 53, 351–368. https://doi.org/10.1006/brln.1996.0053

Hermena, E. W., Liversedge, S. P., Bouamama, S., & Drieghe, D. (2019). Orthographic and root frequency effects in Arabic: Evidence from eye movements and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45, 934–954. https://doi.org/10.1037/xlm0000626

Hermena, E., Drieghe, D., Hellmuth, S., & Liversedge, S. (2015). Processing of Arabic diacritical marks: Phonological–syntactic disambiguation of homographic verbs and visual crowding effects. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 494–507. https://doi.org/10.1037/xhp0000032

Hermena, E., Liversedge, S., & Drieghe, D. (2016). Parafoveal processing of Arabic diacritical marks. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 2021–2038. https://doi.org/10.1037/xhp0000294

Hermena, E., Liversedge, S., & Drieghe, D. (2017). The influence of a word's number of letters, spatial extent, and initial bigram characteristics on eye movement control during reading: Evidence from Arabic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 451–471. https://doi.org/10.1037/xlm0000319

Hilz, R., & Cavonius, C. R. (1974). Functional organization of the peripheral retina: Sensitivity to periodic stimuli. *Vision Research*, 14, 1333–1337. https://doi.org/10.1016/0042-6989(74)90006-6

Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1150–1170. https://doi.org/10.1037/a0020233

Hyönä, J. (2011). Foveal and parafoveal processing during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 819–838). New York, NY, US: Oxford University Press.

Hyönä, J., & Bertram, R. (2011). Optimal viewing position effects in reading Finnish. *Vision Research*, 51, 1279–1287. https://doi.org/10.1016/j.visres.2011.04.004

Inhoff, A. W. (1987). Parafoveal word perception during eye fixations in reading: Effects of visual salience and word structure. In M. Coltheart (Ed.), *Attention and performance* (Vol. 12, pp. 403–420). London: Erlbaum.

Inhoff, A. W. (1989a). Lexical access during eye fixations in reading: Are word access codes used to integrate lexical information across interword fixations? *Journal of Memory and Language*, 28, 444–461. https://doi.org/10.1016/0749-596X(89)90021-1

Inhoff, A. W. (1989b). Parafoveal processing of words and saccade computation during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 544–555. https://doi.org/10.1037/0096-1523.15.3.544

Inhoff, A. W. (1990). Integrating information across eye fixations in reading: The role of letter and word units. *Acta Psychologica*, 73, 281–297. https://doi.org/10.1016/0001-6918(90)90027-D

Inhoff, A. W., & Tousman, S. (1990). Lexical integration across saccades in reading. *Psychological Research*, 52, 330–337. https://doi.org/10.1007/BF00868065

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40, 431–439. https://doi.org/10.3758/BF03208203

Jordan, T. R., Almabruk, A. A., McGowan, V. A., & Paterson, K. B. (2011). Evaluating hemispheric divisions in processing fixated words: The evidence from Arabic. *Cortex*, 47, 992–997. https://doi.org/10.1016/j.cortex.2011.02.012

Jordan, T. R., Paterson, K. B., Kurtev, S., & Xu, M. (2010). Re-evaluating split-fovea processing in word recognition: Effects of fixation location within words. *Cortex*, 46, 298–309. https://doi.org/10.1016/j.cortex.2009.01.008

Jordan, T., Almabruk, A., Gadalla, E., McGowan, V., White, S., Abedipour, L., & Paterson, K. (2014). Reading direction and the central perceptual span: Evidence from Arabic and English. *Psychonomic Bulletin & Review*, 21, 505–511. https://doi.org/10.3758/s13423-013-0510-4

Joseph, H. S., Liversedge, S. P., Blythe, H. I., White, S. J., & Rayner, K. (2009). Word length and landing position effects during reading in children and adults. *Vision Research*, 49, 2078–2086. https://doi.org/10.1016/j.visres.2009.05.015

Juhasz, B. J., White, S. J., Liversedge, S. P., & Rayner, K. (2008). Eye movements and the use of parafoveal word length information in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1560–1579. https://doi.org/10.1037/a0012319

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1312–1318. https://doi.org/10.1037/0278-7393.29.6.1312

Kambe, G. (2004). Parafoveal processing of prefixed words during eye fixations in reading: Evidence against morphological influences on parafoveal pre-processing. *Perception & Psychophysics*, 66, 279–292. https://doi.org/10.3758/BF03194879

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 262–284. https://doi.org/10.1080/09541440340000213

Lallier, M., Abu Mallouh, R., Mohammed, A., Khalifa, B., Perea, M., & Carreiras, M. (2018). Does the visual attention span play a role in reading in Arabic? *Scientific Studies of Reading*, 22, 181–190. https://doi.org/10.1080/10888438.2017.1421958

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: an ideal-observer model of reading. *Psychological Review*, 104, 524–553. https://doi.org/10.1037/0033-295X.104.3.524

Li, L., Li, S., Wang, J., McGowan, V. A., Liu, P., Jordan, T. R., & Paterson, K. B. (2017). Aging and the optimal viewing position effect in visual word recognition: Evidence from English. *Psychology and Aging*, 32, 367–376. https://doi.org/10.1037/pag0000163

Lima, S. D., & Inhoff, A. W. (1985). Lexical access during eye fixations in reading: Effects of word-initial letter sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 272–285. https://doi.org/10.1037/0096-1523.11.3.272

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, 147, 1–20. https://doi.org/10.1016/j.cognition.2015.10.013

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, 28, 1107–1118. https://doi.org/10.1016/0042-6989(88)90137-X

McConkie, G., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17, 578–586. https://doi.org/10.3758/BF03203972

McConkie, G., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of The Psychonomic Society*, 8, 365–368. https://doi.org/10.3758/BF03335168

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 667–682. https://doi.org/10.1037//0096-1523.10.5.667

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research*, 45, 2201–2217. https://doi.org/10.1016/j.visres.2005.02.014

Olson, D. R., & Dringer, D. (2020). Alphabet. *Encyclopædia Britannica*. Encyclopædia Britannica, inc. Retrieved on 27/8/20 from https://www.britannica.com/topic/alphabet-writing

O'Regan, J. K., & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 185–197. https://doi.org/10.1037/0096-1523.18.1.185

O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugaillère, B. É. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 250–257. https://doi.org/10.1037//0096-1523.10.2.250

O'Regan, J. K. (1980). The control of saccade size and fixation duration in reading: The limits of linguistic control. *Perception & Psychophysics*, 28, 112–117. https://doi.org/10.3758/BF03204335

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition: The word-superiority effect. *Psychological Review*, 89, 573–594. https://doi.org/10.1037/0033-295X.89.5.573

Parker, A. J., Nikolova, M., Slattery, T. J., Liversedge, S. P., & Kirkby, J. A. (2019). Binocular coordination and return-sweep saccades among skilled adult readers, *Journal of Vision*, 19, 1–19. https://doi.org/10.1167/19.6.10

Paterson, K. B., McGowan, V. A., & Jordan, T. R. (2013). Aging and the control of binocular fixations during reading. *Psychology and Aging*, 28, 789–795. https://doi.org/10.1037/a0033328

Paterson, K. B., McGowan, V. A., White, S. J., Malik, S., Abedipour, L., & Jordan, T. R. (2014). Reading direction and the central perceptual span in Urdu and English. *PloS ONE*, 9, e88358. https://doi.org/10.1371/journal.pone.0088358

Paterson, K., Almabruk, A., McGowan, V., White, S., & Jordan, T. (2015). Effects of word length on eye movement control: The evidence from Arabic. *Psychonomic Bulletin & Review*, 22, 1443–1450. https://doi.org/10.3758/s13423-015-0809-4

Perea, M., Abu Mallouh, R., Mohammed, A., Khalifa, B., & Carreiras, M. (2016). Do diacritical marks play a role at the early stages of word recognition in Arabic? *Frontiers in Psychology*, 7, 1255. https://doi.org/10.3389/fpsyg.2016.01255

Pollatsek, A., Bolozky, S., Well, A., & Rayner, K. (1981). Asymmetries in the perceptual span for Israeli readers. *Brain and Language*, 14, 174–180. https://doi.org/10.1016/0093-934X(81)90073-0
Pollatsek, A., Lesch, M., Morris, R., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 148–162. https://doi.org/10.1037//0096-1523.18.1.148
Pollatsek, A., Reichle, E., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, 52, 1–56. https://doi.org/10.1016/j.cogpsych.2005.06.001
Rayner, K. (1975a). Parafoveal identification during a fixation in reading. *Acta Psychologica*, 39, 271–281. https://doi.org/10.1016/0001-6918(75)90011-6
Rayner, K. (1975b). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7, 65–81. https://doi.org/10.1016/0010-0285(75)90005-5
Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, 85, 618–660. https://doi.org/10.1037/0033-2909.85.3.618
Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, 8, 21–30. https://doi.org/10.1068/p080021
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. https://doi.org/10.1037/0033-2909.124.3.372
Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457–1506. https://doi.org/10.1080/17470210902816461
Rayner, K. (2014). The gaze-contingent moving window in reading: Development and review. *Visual Cognition*, 22, 242–258. https://doi.org/10.1080/13506285.2013.879084
Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191–201. https://doi.org/10.3758/BF03197692
Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, 16, 829–837. https://doi.org/10.1016/0042-6989(76)90143-7
Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504–509. https://doi.org/10.3758/BF03214555
Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 40, 473–483. https://doi.org/10.1037/h0080111
Rayner, K., McConkie, G., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology*, 12, 206–226. https://doi.org/10.1016/0010-0285(80)90009-2
Rayner, K., Sereno, S., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of model. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1188–1200. https://doi.org/10.1037/0096-1523.22.5.1188
Rayner, K., Slattery, T., Drieghe, D., & Liversedge, S. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 514–528. https://doi.org/10.1037/a0020990
Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, 27, 537–544. https://doi.org/10.3758/BF03198682

Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Perception & Psychophysics*, 31, 537–550. https://doi.org/10.3758/BF03204186

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125–157. https://doi.org/10.1037/0033-295X.105.1.125

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26, 445–476. https://doi.org/10.1017/S0140525X03000104

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16, 1–21. https://doi.org/10.3758/PBR.16.1.1

Roldán, M., Marcet, A., & Perea, M. (2018). Is there a cost at encoding words with joined letters during visual word recognition? *Psicológica Journal*, 39, 279–291. https://doi.org/10.2478/psicolj-2018-0012

Roman, G., & Pavard, B. (1987). A comparative study: How we read in Arabic and French. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements from physiology to cognition* (pp. 431–440). Amsterdam, the Netherlands: Elsevier. https://doi.org/10.1016/B978-0-444-70113-8.50064-3

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception & Psychophysics*, 74, 5–35. https://doi.org/10.3758/s13414-011-0219-2

Skeide, M. A., & Friederici, A. D. (2018). Neurolinguistic studies of sentence comprehension. In E. Fernandez, & H. Cairns (Eds.), *The Handbook of Psycholinguistics* (pp. 438–456). Hoboken: Wiley-Blackwell. https://doi.org/10.1002/9781118829516.ch19

Slattery, T. J., & Vasilev, M. R. (2019). An eye-movement exploration into return-sweep targeting during reading. *Attention, Perception, & Psychophysics*, 1–7. https://doi.org/10.3758/s13414-019-01742-3

Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences*, 23(7), 537–546. https://doi.org/10.1016/j.tics.2019.04.006

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological Review*, 125, 969–984. https://doi.org/10.1037/rev0000119

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, 9, 311–327. https://doi.org/10.1111/lnc3.12151

Stevens, M., & Grainger, J. (2003). Letter visibility and the viewing position effect in visual word recognition. *Perception & Psychophysics*, 65, 133–151. https://doi.org/10.3758/BF03194790

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In E. S. Kappenman and S. J. Luck (Eds.), *Oxford handbook of event-related potential components* (pp. 397–440). Oxford: Oxford University Press.

Vitu, F., McConkie, G., Kerr, P., & O'Regan, J. (2001). Fixation location effects on fixation durations during reading: an inverted optimal viewing position effect. *Vision Research*, 41, 3513–3533. https://doi.org/10.1016/S0042-6989(01)00166-3

Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception & Psychophysics*, 47, 583–600. https://doi.org/10.3758/BF03203111

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception & Psychophysics*, 57, 352–364. https://doi.org/10.3758/BF03213060

White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: The importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1261–1276. https://doi.org/10.1037/0096-1523.34.5.1261

Yakup, M., Abliz, W., Sereno, J., & Perea, M. (2015). Extending models of visual-word recognition to semicursive scripts: Evidence from masked priming in Uyghur. *Journal of Experimental Psuychology: Human Perception and Performance*, 41, 1553–1562. https://doi.org/10.1037/xhp0000143

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, 16, 561–566. https://doi.org/10.3758/PBR.16.3.561

An eye-tracking study of phonological awareness in Emirati Arabic

Alexandra Marquis, Meera Al Kaabi, Tommi Leung and Fatima Boush
United Arab Emirates University

This paper studies phonological awareness in Emirati Arabic by tracking eye movements. Thirty-eight Arabic participants, divided into feedback and no feedback groups, were evaluated on three phonological conditions: (1) explicit instructions for *onset consonant matching*, (2) implicit instructions for *segmentation of initial consonant*, and (3) *rhyme matching*. We hypothesized that (1) fixations would differ according to the phonological conditions, (2) subjects would perform better on explicit instructions compared to implicit instructions, (3) subjects would perform better in the two consonant conditions compared to the rhyme condition, and (4) feedback would improve performance. The experiment shows that accuracy on the consonant conditions was higher than in the rhyme condition. Fixation patterns differed between the consonant conditions and the rhyme condition, and this effect was dependent on the administration of feedback, suggesting that eye tracking can be used to evaluate phonological awareness in Emirati Arabic.

Keywords: Emirati Arabic (EA), eye-tracking, phonological awareness, word recognition, neurolinguistics, Arabic language, adults

1. Introduction

Phonological awareness is the ability to perceive and manipulate the sounds of spoken words (e.g., Alamri, 2017; Goswami & Bryant, 1990; Holm, Farrier, & Dodd, 2008). It is considered a good predictor of reading and spelling abilities in children (e.g., Holm et al., 2008; Melby-Lervåg, Lyster, & Hulme, 2012; Schneider, Roth, & Ennemoser, 2000). For instance, Melby-Lervåg et al. (2012) conducted a meta-analytic review of 235 studies and concluded that phonemic awareness was the strongest predictor of individual reading abilities. Another study investigated early linguistic abilities and their effect on learning to spell and read in Arabic, finding that

https://doi.org/10.1075/sal.10.04mar

children with deficits in reading accuracy faced early difficulties in morphological and phonological awareness skills, which confirmed phonological awareness skills as the strongest predictor for reading (Saiegh-Haddad & Taha, 2017). Moreover, phonological skills in kindergarten children have been shown to be associated with spelling skills a year later (e.g., Speece, Roth, Cooper, & De La Paz, 1999).

The linguistic situation in the Arab world is complex given the diglossic state of Arabic, which is exhibited in two distinct forms: (1) Modern Standard Arabic (MSA): the literary form, also known as /fusħa/ which is used by educated people in reading and writing, as well as in formal communications, interviews, newspapers and other formal settings, and (2) the spoken form or dialect, also known as /ʕaammijja/ which is used in daily conversation and informal settings. Previous studies proposed that these two forms are couched in two separate cognitive systems. For instance, Saiegh-Haddad (2008) and Asaad and Eviatar (2014) conducted studies on letter learning and suggested that letters corresponding to sounds that did not occur in the speaker's dialect were more difficult to learn and to identify than the ones that existed in the spoken Arabic dialect. Moreover, Tibi and Kirby (2018) pointed out that the diglossic situation in Arabic could affect reading accuracy and speed. That is, MSA words that do not exist in Arabic dialects would be less likely to be recognized, especially if the reading context activates the dialectal counterparts of these words.

MSA comprises three short vowels, /a/, /u/, and /i/, and their long counterparts, /aa/, /uu/, and /ii/ and the diphthongs /aw/ and /aj/. Emirati Arabic has the same vowels found in MSA, with the addition of the short vowel /ə/, and the long vowels /oo/ and /ee/. While syllables in MSA always begin with a consonant followed by a vowel (e.g., CV, CVC, CVV, CVVC or CVCC), Emirati Arabic allows for the clustering of two onset consonants (e.g., /ktaab/ 'book', Hoffiz, 1995; Qafisheh, 1977).

2. Eye-tracking

Measuring eye movements has an advantage over traditional measurements of attention such as reaction time (RT), as eye movements are governed by an automatic process that does not impose additional operational requirements alongside the primary task goals. In most cases, eye movements allow us to gain a naturalistic, yet indirect, understanding of the cognitive processes associated with moment-by-moment task completion. Therefore, we consider eye-tracking measurements as an additional index for attention, and perhaps even a superior one, as it eliminates potential confounds of recording responses mechanically.

Eye-tracking research relies heavily on what is known as the 'eye-mind' hypothesis (Just & Carpenter, 1976a, 1976b) which states that the direction at which the eyes are fixated at any given time is where attention, and therefore information processing, is actively taking place. Objects falling within the immediate visual field are the subject of direct attention and are therefore under active cognitive processing (e.g., Conklin & Pellicer-Sánchez, 2016; Pickering, Frisson, McElree, & Traxler, 2004). Moreover, the fixation duration also reflects the associated processing demands on that object, such that as soon as an object is no longer fixated, it no longer requires further processing. Therefore, it is assumed that different task demands will result in differential fixation patterns depending on the cognitive load associated with the task. In general, average fixations during free viewing of visual scenes tend to last approximately 330 milliseconds (Conklin & Pellicer-Sánchez, 2016). However, this number varies according to the task being completed. Items requiring reduced cognitive processing receive shorter fixations than those that require higher cognitive processing loads (Conklin & Pellicer-Sánchez, 2016).

Apart from reading studies (see Rayner, 2009 for a review), eye tracking has been increasingly implicated in behavioral experiments including choice tasks (e.g., Orquin & Loose, 2013 for a review), data visualization and perception of various displays (e.g., Bylinskii, Borkin, Kim, Pfister, & Oliva, 2017; Lahrache, El Ouazzani, & El Qadi, 2018), as well as speech perception and comprehension (e.g., Tanenhaus & Trueswell, 2006). The latter area of research, and most relevant to the current study, is associated with what is known as the *Visual World Paradigm* (VWP), where "on each trial, participants hear an utterance while looking at an experimental display. Participants' eye movements are recorded for later analyses" (Huettig, Rommers, & Meyer, 2011). In such experiments, a visual display is coupled with spoken utterances. Eye movements are recorded, as they are thought to be influenced by participants' goal to completing a particular task (e.g., Pyykkönen-Klauck & Crocker, 2016; Salverda, Brown, & Tanenhaus, 2011; Salverda & Tanenhaus, 2017), integrating a spoken utterance with a visual display.

3. The visual world paradigm

The current study of phonological awareness in Emirati Arabic was conducted using the VWP as the experimental method for spoken word recognition (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Altmann, 2004; Altmann & Kamide, 1999; Cooper, 1974; Ferreira, Apel, & Henderson, 2008; Huettig, Rommers, & Meyer, 2011; Knoeferle & Crocker, 2007; Richardson, Altmann, Spivey, & Hoover, 2009; Richardson & Spivey, 2000; Spivey & Geng, 2001; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The VWP is now widely considered as a standard protocol to investigate the interplay between linguistic and visual information processing. Since the seminal work by Cooper (1974), it has been established that there exists an interactive relation between the locus of eye fixation and spoken language in the sense that the latter somehow controls the former in a time-logged fashion. A series of work (e.g., Allopenna et al., 1998; Tanenhaus et al., 1995) further detailed a particular version of the VWP commonly used by many subsequent studies. In such a version, line drawings of four objects (usually arranged on a 2 × 2 grid) are shown on a computer screen. Upon receiving audio instructions, participants have to move the displayed object by clicking on the computer mouse and dragging it to the location of a fixed geometric shape (e.g., after hearing the sentence "Pick up the beaker. Now, put it below the diamond."). The objects displayed usually include the target (also known as the 'referent') and three distractors. Among the distractors, some are competitive in that they share some linguistic properties with the target (e.g., the same onset syllable as in 'candle' and 'candy'). The primary objective of this particular configuration is not only to test how fast subjects move their eye gaze to the correct target, but also to see how distractors may shed light on the language processing system vis-à-vis word recognition/identification. Distractors must be chosen depending on the type of language processing task targeted (e.g., phonological, semantic, pragmatic, etc.)

In an influential study conducted by Allopenna et al. (1998), which focused on the real-time phonetic processing within the VWP given a particular target (e.g., 'beaker'). The first distractor was an 'onset competitor' (e.g., 'beetle'), the second one was a 'rhyme competitor' (e.g., 'speaker'), and the third one was unrelated (e.g., 'carriage'). As the spoken linguistic message started to unfold, participants demonstrated comparable fixations on the objects that shared the same onset (e.g., 'beaker' and 'beetle'). As more phonetic cues were presented to the participants, fixations on the onset competitor started to decline, and fixations on the rhyme competitor (e.g., 'speaker') started to increase. This dynamic shift of fixations provides insights to the word recognition model such as the Cohort model (Marslen-Wilson, 1987), the TRACE model (McClelland & Elman, 1986), and the Shortlist model (Norris, 1994).

4. Current study

The purpose of previous research on phonological awareness was to establish a link between phonological awareness and literacy (e.g., Holm et al., 2008; Landerl et al., 2019; Rvachew et al., 2017). The current research aims at exploring the feasibility of using eye tracking to assess phonological awareness of Emirati Arabic speakers using a variant of the VWP. The current phonological awareness assessment is an extension of the phonological awareness task used in the Language Acquisition Test for Arabic (LATFA, Marquis, 2021 this volume). The LATFA phonological awareness assessment comprises three phonological conditions which are expanded on below.

1. Explicit instructions for *onset consonant matching* (OCM): participants are introduced to a cartoon character preceding the following example statement. “This animal likes things that start with the same sound. The sound that it likes is /s/. Which of these things does it like?

/ʔarnab/	**/səkkiin/**	/ʕəriiʃ/	/nəmər/”
‘rabbit’	‘knife’	‘palm house’	‘tiger’

2. Implicit instructions for *segmentation of initial phoneme* (SIC): participants are introduced to a cartoon character preceding the following example statement. “This is Dana. Dana likes things that begin with the same sound as her name. Which of these things does Dana like?

/ħajwaan/	/xaðˤra/	/**dallah**/	/ʕətˤər/”
‘animal’	‘vegetables’	‘flask’	‘perfume’

3. *Rhyme matching* (RM): participants are introduced to a cartoon character preceding the following example statement.
   “This is Lulu. Lulu likes things that sound like her name, which of these things does Lulu like?

/**ʔubuu**/	/mətˤar/	/dənya/	/mənaz/”
‘father’	‘rain’	‘world’	‘cradle’

The current study adopts the aforementioned task conditions, using the eye-tracker as an alternative methodology. The use of eye-tracking can allow us to gain further insight into the implicit linguistic knowledge associated with phonological awareness. Successful validation of the use of eye movements on adults will allow us to use this task with children in order to evaluate their phonological awareness skills.

We hypothesized that fixation indices would vary across the experimental conditions. We expected explicit instructions to facilitate participants’ performance on the task. Thus, we predicted that eye movements would reflect more efficient fixation patterns in the OCM condition in comparison to the SIC condition. Moreover, since Arabic is argued to consist of consonantal roots as abstract units (e.g., Idrissi, Prunet, & Béland, 2008), we hypothesized that participants would perform better in the consonant conditions (i.e., OCM and SIC) in comparison to the rhyme condition (i.e., RM). Finally, we expected that providing feedback during practice trials would facilitate participants’ performance overall.

5. Method

5.1 Participants

Forty-one Arabic-speaking female university students were evaluated. Data from three participants were excluded from the analyses because they spoke a dialect other than Emirati Arabic. Therefore, we report the data from thirty-eight Emirati Arabic speakers (Age_{Mean} = *19.8*, Range = *17–24*). Participants were randomly assigned to receive feedback during practice trials (feedback group) or assigned to the no feedback group.

5.2 Stimuli

Images used for the current experiment were adapted from the stimuli set used for the LATFA phonological awareness assessment (Marquis, 2016–2018), with minor substitutions for items that became too heavily pixelated due to re-scaling on the monitor screen used for eye tracking. In cases where images were replaced, they were either identical or as visually similar to the originals as possible. The Emirati Arabic Language Acquisition Corpus (EMALAC, Ntelitheos & Idrissi, 2017) was used as a database for the selection of highly familiar nouns to ensure participants' familiarity with the items and to avoid saliency effects. For each trial, an auditory question was introduced by a character whose name served as a prime for the target phoneme, followed by a display showing the four possible choices arranged in a grid combination (2 × 2) containing the target noun and three distractors (see Figure 1). In total, there were 192 different noun-image pairs. No image was repeated more than twice throughout the whole experiment, and never for the same target phoneme or phonological task. All auditory stimuli were recorded by a native female Emirati Arabic adult, and consisted of verbal task instructions containing the target phoneme in Emirati Arabic.

For each of the three phonological conditions, there were four different target phonemes, where each phoneme was tested on three separate trials with three different targets. Therefore, each phonological condition included a total of 12 trials. Each trial began with a display showing a cartoon character that introduces the target phoneme using an auditory script associated with that character (see Figure 1, left side). There was a different script for each target phoneme, and each script introduced the nature of the task (i.e., whether it was *segmentation of initial consonant, onset consonant matching*, or *rhyme matching*). Next, the trial continues with the presentation of a display containing four images: the target and three unrelated distractors pseudorandomized in a 2 × 2 grid arrangement (see Figure 1, right side).

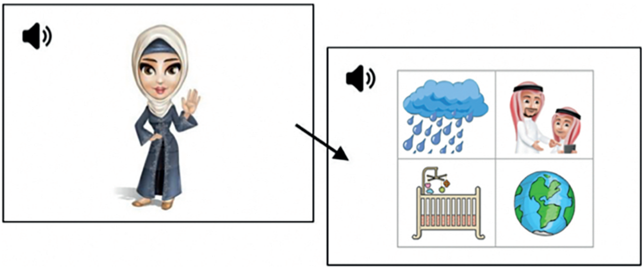


Figure 1. Trial example

Screen 1 (Left):

هاذي لولو. لؤلؤ تحب الأشياء الي تشابه اسمها. شو من هاي الأشياء لؤلؤ تحب
haaðii luuluu. luuluu t-ħəbb ʔəl-ʔaʃjaaʔ ʔilli t-ʃaabəh
this-SG.F Lulu Lulu 3SG.F-likes the-things.PL that 3SG.F-similar
ʔəsim-ha. ʃuu mən hal-ʔaʃjaaʔ luuluu t-ħəbb?
name-her what of these-things Lulu 3SG.F-like?
'This is Lulu. Lulu likes things that sound like her name. Which of these things does Lulu likes?'

Screen 2 (Right):

/ʔubuu/ /mətˤar/ /dənja/ /mənaz/
father rain world cradle
*Read from top-right to left.

5.3 Design

We adapted the phonological awareness task used in LATFA (Marquis, 2016–2018; Marquis, 2021 this volume) to an audio-coupled VWP. The design of the paradigm was a mixed factorial, four-alternative forced-choice (4-AFC) *task-based* VWP (see Salverda et al., 2011 for more on task-based approaches to visual research). There were two independent variables: (1) phonological condition with three levels (OCM, SIC, and RM) and (2) feedback group (feedback vs no feedback).

5.4 Procedure

Eye movements were recorded using an EyeLink 1000Plus tracker fitted on a desktop mount, and a head-and-chin rest was used to stabilize the participants' head position. While viewing was binocular, the tracker was set to record only from the right eye at a sampling rate of 1000 Hz. Participants were seated 85 cm from the screen of a 24-inch BenQ gaming monitor. Calibration and validation were performed using the default nine-point calibration grid.

At the beginning of each phonological condition block, four practice trials (excluded from the analyses) were presented in order to familiarize participants with each particular task. Trials within each block were pseudorandomized. Participants were tested individually and were asked to sit comfortably in front of the computer monitor, adjust the height of their seat if needed, with their chin and head on the table mounted support. A nine-point grid calibration was performed and validated for each participant before the experiment started. Participants then put on noise-cancelling headphones and were instructed to listen carefully to the instructions and recordings presented via the headphones and to select, using the mouse, between the four different images appearing on the display. Each trial ended when participants recorded their choice by left-clicking on the mouse and the next trial automatically began. Participants sat through 12 trials per condition, for a total of three condition blocks. Participants were allowed breaks in between blocks if needed. In addition, a central drift correction was performed before each block. Participants were briefed about the study by the experimenter at the end of the experiment and were allowed to ask questions. On average, each session lasted about 20 minutes.

5.5 Measures

For all participants, the following measures were collected and analyzed:

- *Accuracy.* The accuracy measure is the percentage of correct responses (i.e., correctly selecting the target for each trial), automatically calculated for each participant by the Data Viewer software and averaged across participants.
- *Average Fixation Count.* The average fixation count is the average number of fixations made to the target in each trial, automatically calculated for each participant by the Data Viewer software and averaged across participants. Individual fixations were defined as periods where the eyes were maintained stable and the saccade amplitude of eye movements did not exceed 0.0 degree per second. Fixations spanning 1 degree or less were automatically merged by the Data Viewer software.
- *Average Fixation Duration.* The average fixation duration is the average length in milliseconds (*ms*) of individual fixations made to the target in each trial, automatically calculated for each participant by the Data Viewer software and averaged across participants.

6. Results

A multivariate analysis of variance (MANOVA) was performed on the three levels of phonological condition (OCM, SIC, and RM) as within-subjects variable and feedback group (Feedback vs No Feedback) as the between-subjects variable on three eye movement measures: accuracy, average fixation count, and average fixation duration. Results show significant effects of phonological condition, $F(6, 17) = 6.68$, $\lambda = 0.298$, $p < 0.001$, $\eta^2 = 0.70$, and feedback group, $F(3, 20) = 99.6$, $\lambda = 0.063$, $p < 0.001$, $\eta^2 = 0.94$, and a significant interaction between phonological condition and feedback group $F(6, 17) = 5.59$, $\lambda = 0.336$, $p = 0.002$, $\eta^2 = 0.66$. Results are summarized in Tables 1 and 2.

Univariate tests show a significant main effect of phonological condition on accuracy, $F(1, 22) = 38.4$, $p < .001$, $\eta^2 = 0.64$, and a significant interaction of average fixation duration with feedback group, $F(1, 22) = 27.9$, $p < .001$, $\eta^2 = 0.56$. There were significant main effects of feedback group for average fixation duration, $F(1, 22) = 18.1$, $p < .001$, $\eta^2 = 0.42$, and average fixation count, $F(1, 22) = 220.6$, $p < .001$, $\eta^2 = 0.91$. Pairwise comparisons on phonological condition revealed that accuracy on RM scores ($M = 75.5\%$, $SD = 0.18$) were significantly lower than both OCM ($M = 98.7\%$, $SD = 0.02$; $p < 0.001$) and SIC ($M = 96.9\%$, $SD = 0.04$; $p < 0.001$) consonant conditions. The difference between the OCM and SIC consonant conditions was non-significant ($p = 0.073$).

The data were split by feedback group, and a repeated measures ANOVA was performed on fixation duration. Within-subjects effects showed significant results of Feedback, $F(2, 22) = 13.5$, $p = .002$, $\eta^2 = 0.55$, and No Feedback, $F(2, 22) = 4.86$, $p = .029$, $\eta^2 = 0.31$, on phonological condition. Pairwise comparisons revealed significant differences between fixation durations for RM ($M = 325$ *ms*, $SD = 52.4$) and the SIC ($M = 277$ *ms*, $SD = 23.8$) at the no feedback level ($p = 0.01$), and between RM ($M = 312$ *ms*, $SD = 24.9$) and both SIC ($M = 344$ *ms*, $SD = 35.4$, $p = 0.001$) and OCM ($M = 376$, $SD = 51.2$, $p = 0.006$) consonant conditions at the feedback level. No further significant differences were found.

Table 1. Summary of results for the feedback group

	Measure		
Factor	***Means (SD)***		
	Accuracy (%)	**Fixation count**	**Fixation duration (*ms*)**
Phonological condition			
OCM	99.1 (0.02)	8.2 (1.81)	376 (51.2)
SIC	99.1 (0.02)	8.3 (1.60)	344 (23.8)
RM	77.6 (0.18)	8.2 (2.58)	312 (24.9)

Table 2. Summary of results for the no feedback group

	Measure		
Factor	*Means (SD)*		
	Accuracy (%)	**Fixation count**	**Fixation duration (*ms*)**
Phonological condition			
OCM	98.3 (0.02)	1.6 (0.18)	292 (34.2)
SIC	94.6 (0.03)	1.9 (0.46)	277 (24.0)
RM	73.3 (0.18)	2.7 (1.31)	325 (52.4)

7. Discussion

Eye movements are goal-based, i.e. the immediate relevance of visual information to goal execution is key to the interpretation of results. The results we obtained confirm our hypotheses that: (1) fixation indices vary across the experimental conditions, (2) participants perform better in the two consonant conditions (i.e., OCM and SIC) in comparison to the rhyme condition (i.e., RM), and (3) that providing feedback during practice trials facilitate participants' performance overall. However, our fixation metrics could not distinguish phonological awareness skills between our explicit (OCM) and implicit (SIC) consonant conditions.

Additionally, participants responded to the tasks in the consonant conditions with higher accuracy than in the rhyme condition. These results indicate that participants are better able to perform with consonants in comparison to rhymes. They also dovetail with the root-and-pattern theory which is dominant in the comparative Semitic literature (Ratcliffe, 2013) and concur with some psycholinguistic experiments verifying the independent status of consonantal roots in Arabic (e.g., Boudelaa & Marslen-Wilson, 2001).

Participants who received feedback during practice made more fixations and longer average durations to targets before responding (for consonants). In sum, our eye-tracking results reveal that awareness of rhymes (i.e., vowels) is less accurate than awareness of consonants in adult Emirati Arabic speakers. Furthermore, our data shed light on feedback effects for phonological awareness in Arabic. The goal of the current study was to explore the feasibility of using eye tracking to assess phonological awareness of Emirati Arabic speakers using a variant of the VWP. Based on our results and the assumption that participants' gaze behaviors reflect their on-line attentional processes, we conclude that eye tracking is an efficient methodology to measure phonological awareness. In addition, administering feedback during practice appears to enhance participants' efficiency in their phonological awareness abilities. Since phonological awareness is considered to be a good predictor of reading and spelling abilities in children (e.g., Holm et al., 2008; Melby-Lervåg et al., 2012; Schneider et al., 2000), outcomes of phonological awareness research using eye tracking could inform us on how to improve educational targets suited to the particularities of Arabic. Our results, particularly those on rhymes, are surprising given that previous eye-tracking research failed to find differences between onset and rhymes, or indicated a performance advantage for rhymes over onsets (e.g., Brouwer & Bradlow, 2016; Desroches, Joanisse, & Robertson, 2006) while behavioral research suggest that the onset-rhyme distinction is important in children's phonological development (e.g., Kirtley, Bryant, MacLean, & Bradley, 1989) or that rhyme awareness plays a major role in the development of literacy (e.g., Goswami, 1999; Goswami & Bryant, 1990). In Arabic, previous research (e.g., al Mannai & Everatt, 2005; Saiegh-Haddad & Geva, 2008) revealed that phonological awareness might be used to predict literacy abilities. In the future, our research should target Arabic children's phonological awareness using the eye-tracking methodology in order to predict their literacy skills.

Acknowledgements

We are thankful to Akeela Fatheen Abdul Gafoor, Al Yazia Mohammed Al Mansouri and Wadha Ayed Muaid Asi Al Otaibi for research assistance.

Funding

This project was supported by the United Arab Emirates University Research Start-up Grant Award No. 31H128.

References

al Mannai, H., & Everatt, J. (2005). Phonological processing skills as predictors of literacy amongst Arabic speaking Bahraini children. *Dyslexia*, 11, 269–291. https://doi.org/10.1002/dys.303

Alamri, A. (2017). Phonological, semantic and root activation in spoken word recognition in Arabic: Evidence from eye movements. Ph.D. dissertation, University of Ottawa, Canada.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439. https://doi.org/10.1006/jmla.1997.2558

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, 93, 79–87. https://doi.org/10.1016/j.cognition.2004.02.005

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Asaad, H., & Eviatar, Z. (2014). Learning to read in Arabic: The long and winding road. *Reading and Writing*, 27(4), 649–664. https://doi.org/10.1007/s11145-013-9469-9

Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, 81(1), 65–92. https://doi.org/10.1016/S0010-0277(01)00119-6

Brouwer, S., & Bradlow, A. R. (2016). The temporal dynamics of spoken word recognition in adverse listening conditions. *Journal of Psycholinguistic Research*, 45(5), 1151–1160. https://doi.org/10.1007/s10936-015-9396-9

Bylinskii, Z., Borkin, M. A., Kim, N. W., Pfister, H., & Oliva, A. (2017). Eye fixation metrics for large scale evaluation and comparison of information visualizations. In M. Burch, L. Chuang, B. Fisher, A. Schmidt, & D. Weiskopf (Eds.), *ETVIS: Workshop on Eye Tracking and Visualization. Mathematics and Visualization*. (pp. 235–255). Chicago: Springer. https://doi.org/10.1007/978-3-319-47024-5_14

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, 32(3), 453–467. https://doi.org/10.1177/0267658316637401

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition*, 100, 32–42. https://doi.org/10.1016/j.cognition.2005.09.001

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, 12(11), 405–410. https://doi.org/10.1016/j.tics.2008.07.007

Goswami, U. (1999). Causal connections in beginning reading: The importance of rhyme. *Journal of Research in Reading*, 22(3), 217–240. https://doi.org/10.1111/1467-9817.00087

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. New York: Routledge.

Hoffiz, B. T., III (1995). Morphology of United Arab Emirates Arabic, Dubai dialect. Ph.D. Dissertation, The University of Arizona. http://hdl.handle.net/10150/187179

Holm, A., Farrier, F., & Dodd, B. (2008). Phonological awareness, reading accuracy and spelling ability of children with inconsistent phonological disorder. *International Journal of Language & Communication Disorders*, 43(3), 300–322. https://doi.org/10.1080/13682820701445032

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Idrissi, A., Prunet, J.-F., & Béland, R. (2008). On the mental representation of Arabic roots. *Linguistic Inquiry*, 39(2), 221–259. https://doi.org/10.1162/ling.2008.39.2.221

Just, M. A., & Carpenter, P. A. (1976a). Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Just, M. A., & Carpenter, P. A. (1976b). The role of eye-fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation*, 8(2), 139–143. https://doi.org/10.3758/BF03201761

Kirtley, C., Bryant, P., MacLean, M., & Bradley, L. (1989). Rhyme, rime and the onset of reading. *Journal of Experimental Child Psychology*, 48, 224–245. https://doi.org/10.1016/0022-0965(89)90004-0

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, 57(4), 519–543. https://doi.org/10.1016/j.jml.2007.01.003

Lahrache, S., El Ouazzani, R., & El Qadi, A. (2018). Visualizations memorability through visual attention and image features. *Procedia Computer Science*, 127, 328–335. https://doi.org/10.1016/j.procs.2018.01.129

Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G., Parrila, R., & Georgiou, G. K. (2019). Phonological awareness and rapid automatized naming as longitudinal predictors of reading in five alphabetic orthographies with varying degrees of consistency. *Scientific Studies of Reading*, 23(3), 220–234. https://doi.org/10.1080/10888438.2018.1510936

Marquis, A. (2016–2018). *Development of an assessment tool for oral skills in Emirati Arabic speaking children*. Al Ain, United Arab Emirates: United Arab Emirates University.

Marquis, A. (2021). LATFA: An assessment tool for Emirati Arabic-speaking children. In D. Ntelitheos and T. Leung (Eds.), *Experimental Arabic Linguistics* (pp. 125–142). John Benjamins Publishing Company. https://doi.org/10.1075/sal.10.05mar

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychological Bulletin*, 138(2), 322–52. https://doi.org/10.1037/a0026744

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Ntelitheos, D., & Idrissi, A. (2017). Language growth in child Emirati Arabic. In H. Ouali (Ed.), *Perspectives on Arabic linguistics XXIX: Papers from the Annual Symposium on Arabic Linguistics, Milwaukee, Wisconsin, 2015* (pp. 229–248). Amsterdam: John Benjamins. https://doi.org/10.1075/sal.5

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, 144(1), 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In M. Carreiras and C. Clifton , Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 33–50). New York: Psychology Press. https://doi.org/10.4324/9780203509050

Pyykkönen-Klauck, P., & Crocker, M. W. (2016). Attention and eye movement metrics in visual world eye tracking. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually situated language comprehension* (pp. 67–82). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.18452/19861

Qafisheh, H. A. (1977). *A short reference grammar of Gulf Arabic*. Tucson, AZ: University of Arizona Press.

Ratcliffe, R. R. (2013). Morphology. In J. Owens (ed.), *The Oxford handbook of Arabic linguistics* (pp. 71–91). Oxford: Oxford University Press.

Rayner, K. (2009). The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Richardson, D. C., Altmann, G. T., Spivey, M. J., & Hoover, M. A. (2009). Much ado about eye movements to nothing: A response to Ferreira et al.: taking a new look at looking at nothing. *Trends in Cognitive Sciences*, 13(6), 235–236. https://doi.org/10.1016/j.tics.2009.02.006

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, 76(3), 269–295. https://doi.org/10.1016/S0010-0277(00)00084-6

Rvachew, S., Royle, P., Gonnerman, L. M., Stanké, B., Marquis, A., & Herbay, A. (2017). Development of a tool to screen risk of literacy delays in French-speaking children: PHOPHLO. *Canadian Journal of Speech-Language Pathology and Audiology*, 41, 321–340.

Saiegh-Haddad, E. (2008). On the challenges posed by diglossia to the acquisition of basic reading processes in Arabic. *Language and Literacy*, 1, 101–126.

Saiegh-Haddad, E., & Geva, E. (2008). Morphological awareness, phonological awareness, and reading in English–Arabic bilingual children. *Reading and Writing*, 21, 481–504. https://doi.org/10.1007/s11145-007-9074-x

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. https://doi.org/10.1002/dys.1572

Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*, 137(2), 172–180. https://doi.org/10.1016/j.actpsy.2010.09.010

Salverda, A. P., & Tanenhaus, M. K. (2017). The visual world paradigm. In A. M. B. de Groot, and P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide*, (pp. 89–110). New York: John Wiley & Sons.

Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology*, 92(2), 284. https://doi.org/10.1037/0022-0663.92.2.284

Speece, D. L., Roth, F. P., Cooper, D. H., & De La Paz, S. (1999). The relevance of oral language skills to early literacy: A multivariate analysis. *Applied Psycholinguistics*, 20(2), 167–190. https://doi.org/10.1017/S0142716499002015

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, 65, 235–241. https://doi.org/10.1007/s004260100059

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In M. J. Traxler & M. A. Gernsbacher (eds.), *Handbook of Psycholinguistics*, (2nd ed., pp. 863–900). New York: Elsevier. https://doi.org/10.1016/B978-012369374-7/50023-7

Tibi, S., & Kirby, J. R. (2018). Investigating phonological awareness and naming speed as predictors of reading in Arabic. *Scientific Studies of Reading*, 22(1), 70–84. https://doi.org/10.1080/10888438.2017.1340948

III. Typical and atypical language development

LATFA

An assessment tool for Emirati Arabic-speaking children

Alexandra Marquis
United Arab Emirates University

Research demonstrates the existing links between oral language and literacy. The present chapter introduces the Language Acquisition Test for Arabic (LATFA), an assessment tool based on oral language skills to predict future literacy difficulties in Emirati Arabic (EA) speaking children. A total of forty-three EA speaking children were evaluated in EA using the LATFA tool on their language perception (i.e., phonological awareness & speech perception) and language production (i.e., morphological awareness & phonological production) at the end of their first grade in school. A year later, the same children were assessed with three literacy tasks. Results show that the LATFA tool can help in predicting and identifying literacy difficulties faced by EA speaking children.

Keywords: Emirati Arabic (EA), LATFA, assessment tool, literacy, oral language, children

1. Introduction

In this day and age, literacy abilities may be considered paramount to evolve in the society. Literacy includes a number of skills that require knowledge of letters, letter-sound pairing, decoding (reading) words, spelling, writing of sentences and passages (grammatically), reading fluency, reading comprehension. These literacy skills must be learned explicitly, through teaching at school. Language, contrary to literacy, is learned implicitly from the first year of life, prior to entering school. What we mean by language includes the acquisition of language-specific phonemes (i.e., what sounds exist in the language), phonotactic rules (i.e., what phoneme sequences are permitted), morphological rules (i.e., how to create words and how words might be related to one another), language-specific syntactic rules (i.e., how to create sentences and utterances), that all link with usage and meaning.

https://doi.org/10.1075/sal.10.05mar

It has been demonstrated that literacy abilities are linked to oral language skills (e.g., Boets, Wouters, van Wieringen, De Smedt, & Ghesquière, 2008; Holm, Farrier, & Dodd, 2008; Pacton & Deacon, 2008; Pennala et al., 2010; van Viersen et al., 2018). Language may be divided into two main components, language perception and language production, both comprising measurable skills that can be associated with literacy proficiency. For instance, skills in the perception of speech have been found to correlate with literacy capacities (e.g., Boets et al., 2008; Pennala et al., 2010; Vanvooren, Poelmans, De Vos, Ghesquière, & Wouters, 2017; White-Schwoch et al., 2017). The evaluation of speech perception capacities may involve different methods including speech-in-noise (e.g., Boets et al., 2008; Vanvooren et al., 2017; White-Schwoch et al., 2017), length discrimination (e.g., Pennala et al., 2010) or error perception (e.g., Rvachew & Grawburg, 2006; Rvachew et al., 2017). Phonological awareness, the ability to perceive and manipulate the sounds of language, is a good predictor of literacy abilities (e.g., Goswami, 2000; Goswami & Bryant, 1990; Holm et al., 2008; Pan et al., 2015; see Melby-Lervåg, Solveig-Alma Lyster, & Hulme, 2012 for a review). In addition, morphological awareness may also influence literacy (e.g., Bourassa & Treiman, 2001; Bourassa, Treiman, & Kessler, 2006; Pan et al., 2015; Speece, Roth, Cooper, & De La Paz, 1999; Wolter, Wood, & D'zatko, 2009; see Pacton & Deacon, 2008 for a review). Techniques used to assess morphological awareness include tasks of writing inflectional endings (e.g., Bourassa et al., 2006; Wolter et al., 2009), morphological construction (e.g., Pan et al., 2015), oral morphological production (e.g., Marquis, Royle, Gonnerman, & Rvachew, 2012; Pan et al., 2015; Speece et al., 1999; Wolter et al., 2009), and morphological comprehension (e.g., Speece et al., 1999). The production of speech is another element of language that may affect children's successful literacy development (e.g., Hayiou-Thomas, Carroll, Leavett, Hulme, & Snowling, 2017; Masso, Baker, McLeod, & Wang, 2017; Overby, Masterson, & Preston, 2015). Children who suffer from speech or language impairments may face difficulties when acquiring literacy skills (e.g., Hayiou-Thomas et al., 2017; McLeod, Harrison, & Wang, 2019; Puranik, Petcher, Al Otaiba, Catts, & Lonigan, 2008; van Viersen et al., 2018). Specifically, children with speech sound disorder perform poorly on literacy measures including print knowledge (e.g., Masso et al., 2017), spelling (e.g., Hayiou-Thomas et al., 2017; Lewis, Freebairn, & Taylor, 2000; Lewis, Freebairn, & Taylor, 2002; Overby et al., 2015), and reading (e.g., Hayiou-Thomas et al., 2017; Lewis et al., 2000; Lewis et al., 2002). In short, based on the literature, it appears that it is possible to identify children at risk of literacy difficulties by assessing their oral language skills.

In the present paper, we introduce the Language Acquisition Test for Arabic (LATFA), an assessment tool based on oral language skills aiming to predict future literacy difficulties in EA speaking children. The tool is an adaptation of a similar tool in Québec French (Rvachew et al., 2017). We hypothesize that the tool could be to a great extent successful in identifying children at risk of presenting literacy difficulties a year later.

2. Methodology

All testing protocols were approved by the United Arab Emirates University Research Ethics Committee, the Abu Dhabi Department of Education and Knowledge, and the United Arab Emirates Ministry of Education. Children participants were recruited from their school by sending an information package (a letter describing the project and the consent forms redacted in Arabic) to the parents who were asked to return the signed consent form if they agreed to the participation of their child in the project. All participants were enrolled in the primary school in Al Ain, Emirate of Abu Dhabi, United Arab Emirates. All children present in the classrooms were considered eligible for participation as long as the parents consented, and the child agreed to participate and was able to cooperate with the testing procedures. A telephone interview was also conducted in Arabic with each participant's parent to obtain demographic, literacy, health, and language information via questionnaires.

Children were tested (individually for Phase 1 and in small groups for Phase 2) in a suitable quiet room inside the school at the end of the school year. A trained native Arabic-speaking research assistant administered both phases of the assessment protocol and subsequently completed all scoring, transcription under the supervision of the main researcher. Reliability coding was later performed by other trained native Arabic-speaking research assistants.

3. Phase 1: Oral language assessment

3.1 Participants

Forty-three typically developing EA speaking children participants attending first grade (*mean age* = 6;7; *SD* = .35; *Range* = 6;2–7;5) were evaluated (10 girls and 33 boys). Three additional children were evaluated, but their data were excluded from the analyses because they failed to complete one or all tasks. According to parents' reports, all children spoke EA at home, were educated in Arabic, and none of them had been diagnosed with any language or hearing deficit.

3.2 Procedure

The Language Acquisition Test for Arabic (LATFA) is an evaluation of oral language skills that comprises four different language assessment tasks evaluating language perception (i.e., phonological awareness & speech perception) and language production (i.e., morphological awareness & phonological production) in EA. The four language tasks were administered over two separate sessions, each of which included one perception task and one production task. During one session, the speech perception task was tested with the morphological awareness task. During another session, the phonological awareness task was tested with the phonological production task. The order of the sessions was counterbalanced across participants. Each session lasted about 20 minutes. Evaluation sessions occurred within a maximum of two weeks from each other. Sessions were recorded with a Sony ICD-UX560F stereo digital recorder at a sampling frequency of 44 kHz and a quantization rate of 24 bits. Responses were annotated on response sheets by the experimenter. Participants' oral answers to the production tasks were transcribed from audio recordings and verified for transcription reliability.

3.3 Speech perception task

The Speech Perception task was modeled on the speech perception test developed by Rvachew (2009) and adapted to EA. The EA stimuli were selected from the Emirati Arabic Language Acquisition Corpus (EMALAC, Ntelitheos & Idrissi, 2017). To evaluate speech perception, the task includes a two-alternative, forced-choice word identification task. The participant hears natural speech recorded from native EA speaking adults and children, with and without speech articulation problems. The words are presented in blocks of 10 items, five target (correct) productions (from typical native EA speaking adults and children) and five misarticulated (incorrect) versions of the target word (from native EA speaking adults and children with speech articulation problems). The participant listens to each word and must choose the target picture when a correct pronunciation is heard and an X when a misarticulation is heard. An iPad was used to run the task that presented the pseudo-randomized stimuli for each block. Participants listened to the stimuli over headphones, presented within a comfortable level. The examiner would let the participant click on the next button after each trial was completed. The participant's responses were recorded on a response sheet. Before each block, the experimenter would read the appropriate script (e.g., 'You will hear people saying the word (____). You must indicate whether you heard the word (____) or not'). Regardless of whether the participant's correctly or incorrectly responded, the experimenter would smile to encourage the participant after each response. The Arabic adaptation included three-word stimuli presented in two separate blocks. Two blocks of the target word /flu:s/ 'money', two blocks of the target word /warda/ 'flower', and two blocks of the target word /θaldʒ/ 'ice' (see Figure 1). Misarticulated versions of the words consist of common misarticulations, (e.g., /fluus/ → /fluθ/; /warda/ → /wadda/; /θaldʒ/ → /θalʒ/) and do not represent acceptable allophonic variations in EA (see Al Ameri, 2009 for details). The test comprised 60 items, 20 for each target word. The test was scored as percentage of items correct for each target word and overall.

فلوس /flu:s/ 'money' **وردة /warda/ 'flower'** **ثلج /θaldʒ/ 'ice'**

Figure 1. Images and target words used for the *Speech Perception* task

3.4 Phonological awareness task

The Phonological Awareness task is an adaptation of the French Test de Conscience Phonologique Préscolaire (Brosseau-Lapré & Rvachew, 2008; modeled on the phonological awareness test by Bird, Bishop, & Freeman, 1995), adapted to EA. The EA stimuli were selected from the EMALAC Corpus (Ntelitheos & Idrissi, 2017). The task, presented on an iPad, includes three conditions: Rhyme Matching (RM), Onset Consonant Matching (OCM) and Segmentation of Initial Consonant (SIC). The child is presented with a character and its name, and told that the character "likes things that sound like its name" or that the character "likes things that start with the same sound as its name". For each trial, the child is presented with four image choices, the target and three distractors (see Figure 2). In total, there were 12 trials for each condition, 3 trials per target, and four practice trials. Instructions were pre-recorded by a native EA female speaker to ensure consistency. Participants were asked to listen to each script, look at the given images and record their answers by touching the image of their choice. A trained research assistant noted all the participant's choices in a form and encouraged the child for her/his effort. Responses were then compiled for analyses.

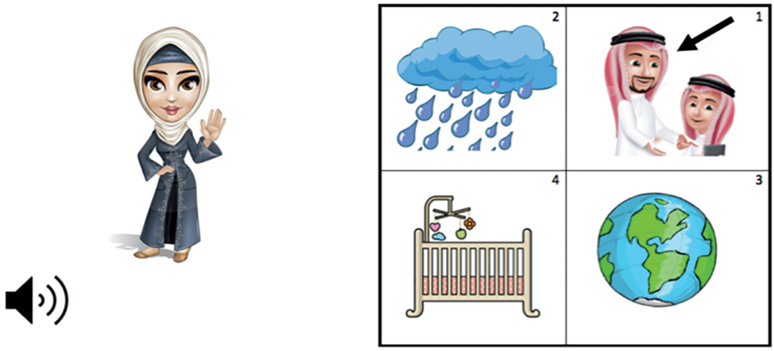


Figure 2. Example of a trial used in the *Phonological Awareness* task: Rhyme Matching (RM) Condition. Screen 1 (left): هذي لؤلؤ. لؤلؤ تحب الأشياء الي تشابه اسمها. شو من هاي الأشياء لؤلؤ تحب؟ '*This is Lulu. Lulu likes things that sound like her name. Which of these things is the one that Lulu likes*?' Screen 2 (right): 1. أب [ʔubuu] 'father' (target), and distractors: 2. مطر [mətˤar] 'rain', 3. دنيا [dənya] 'world' & 4. منز [mənaz] 'cradle'

3.5 Morphological awareness task

To assess Morphological Awareness in EA, a verb elicitation task was developed (see Marquis, 2019). The task was adapted from the French task Jeu de Verbes (Marquis et al., 2012). The target stimuli used in this task were twenty-four perfect tense verb forms in Arabic, ten with biconsonantal template root forms (e.g., طاح /tˤaaħ/ 'to fall') and fourteen with triconsonantal template root forms (e.g., كتب /kataba/ 'to write') (see e.g., Al Kaabi, 2015). The number of verbs for each category is representative of the total number of verbs found in EA children productions from the EMALAC corpus (Ntelitheos & Idrissi, 2017). In addition, four verbs were presented as practice items to introduce the task to the children. Half of the verbs, including the practice items, had to agree with a feminine subject (e.g., شربت /ʃarbət/ 'she drank') and the other half with a masculine subject (e.g., ساق /saag/ 'he drove'). All verbs were frequent and known by younger children (i.e., by 3 to 4 years of age) based on the EMALAC corpus (Ntelitheos & Idrissi, 2017). Verbs were presented along with an image depicting the actions. The images were created by a professional artist to represent the local culture (see Figure 3). The elicitation task included a testing script in which, for each stimulus, two verb tokens were presented as exemplars, followed by a question aiming at eliciting the target perfect tense (see Figure 3). To ensure consistency of testing across all participants, stimuli were audio recorded by a native female EA speaker. A trained research assistant noted all the participant's responses in a form and encouraged the child for her/his effort. Responses were then compiled for analyses.

1. فاطمة بتشرب الماي

   /fatˤma bətaʃrəb alma:j/

   'Fatma will drink water.'

2. فاطمة دايما تشرب الماي

   /fatˤma da:jiman taʃrəb əlma:j/

   'Fatma always drinks water.'

3. فاطمة دايما تشرب الماي

   /ʃu: sawwət fatˤma ʔams/

   'what did Fatma do yesterday?'

Expected target response:

شربت الماي

/ʃarəbət alma:j/

'Fatma/she drank water'

Figure 3. Example of a trial used in the *Morphological Awareness* task

3.6 Phonological production task

The Phonological Production task evaluates accurate phonological production in EA. It was modeled on the French Test de Dépistage Francophone de Phonologie by Rvachew et al. (2013). The task included thirty words which contained the EA phonemes and are commonly used, based on the EMALAC corpus (Ntelitheos & Idrissi, 2017). Colorful images were used to represent the target words. The task included four possible ways to elicit children's production of the target words, recorded by a native female EA speaker. Two versions were made for this task, one for female participants and one for male participants because verbs in Arabic must agree in number and gender. The first attempt is the spontaneous production where the child is asked /haaði/ or /haaða/ 'This is ___?' with an interrogative tone. If the child produces no response or does not produce the target word, s/he is provided with a semantic hint (see Figure 4 for an example). If the child still produces no response or produces a word other than the target one, delayed imitation of the target word is attempted where s/he hears the target followed by a semantic hint (see

Figure 4). Finally, after these three attempts, if the child still has not produced the target word, s/he is asked to say the target word (see Figure 4). A trained research assistant compiled participants' productions for correct consonant production, error types[1] and total number of errors per participant.

1. هذا ____________؟

   /ha:ða/__________?

   'This is___________?'

2. اللي يغرّد سيف سيف سيف ____________؟

   /illi jəɣarrid si:f si:f si:f/_____________?

   'It sings sif sif sif___________?'

3. عصفور. اللي يغرّد سيف سيف سيف ____________؟

   /ʕasˤfu:r. illi jəɣarrid si:f si:f si:f/_____________?

   'Bird. It sings sif sif sif___________?'

4. قولي عصفور

   /guli ʕasˤfu:r/

   'Say bird.'

Figure 4. Example of a trial used in the *Phonological Production* task

4. Phase 2: Literacy assessment

4.1 Participants

For Phase 2, the same forty-three typically developing EA speaking children participants assessed in Phase 1 were evaluated a year later, when they were attending the second grade (*mean age* = 7;7; *SD* = .35; *Range* = 7;2–8;5) (10 girls and 33 boys). According to parents' reports, all children spoke EA at home, were educated in Arabic and none of them had been diagnosed with any language or hearing deficit.

1. Consonant errors included deletion, substitution, metathesis, and insertion. Vowel errors, insertion or substitution, were disregarded from the analyses.

4.2 Procedure

The literacy assessment included three different written assessment tasks in Arabic: (1) the Arabic alphabet task, (2) a dictation task, and (3) the Arabic plurals task. The three different written assessment tasks were administered in a single session including four to ten participants according to room availability. Each session lasted nearly about 20 minutes. Responses to each task were then compiled for analyses.

4.3 Arabic alphabet task

The development of literacy skills (i.e., reading, spelling, and writing) begins with the acquisition of the language's graphemes and letters (at least in alphabetic language systems). The Arabic alphabet task is a dictation task of the 28 Arabic letters (see Table 1). Children were given a pencil and a paper and were instructed, in Arabic, to write in Arabic as many letters of the Arabic alphabet as they could in one minute. Participants' responses were compiled for total number of correct letters per participant.

Table 1. The Arabic letters with their names

Arabic letter	Name	Arabic letter	Name
أ	alif	ض	daad
ب	baa	ط	tˤaa
ت	taa	ظ	dhaa
ث	thaa	ع	ajn
ج	jiim	غ	ghajn
ح	ħaa	ف	faa
خ	khaa	ق	qaaf
د	daal	ك	kaaf
ذ	thaal	ل	lam
ر	raa	م	mim
ز	zaaj	ن	nuun
س	siin	ه	haa
ش	shiin	و	waw
ص	saad	ي	jaa

4.4 Dictation task

The dictation task evaluates the writing abilities of children on their writing skills. The task is an Arabic adaptation of the dictation task found in the Batterie d'Évaluation de Lecture et d'Orthographee (BELO; George & Pech-Georgel, 2006). The task is divided into three parts. The first part concerns the writing of ten mono- or bi-syllables (non-words) that respect the phonotactics of Arabic (e.g., تاش /taaʃ/). The second part includes twenty-one existing words in Arabic of one to four syllables in length (e.g., بَابْ /baab/ door; تُفَّاحَة /təffaaħa/ 'apple'). The third part is a dictation of four separate sentences containing between five and eight words each in MSA (e.g., عِنْدِي بَيْتٌ جَميلٌ مَعَ حَدِيْقَةٍ /ʕindii bajtun dʒamiilun maʕa ħadiiqa/[2] 'I have a pretty house with a garden.'). Words for the second and third part of the dictation task are all expected to be well known by EA speaking children of this age according to the EMALAC corpus (Ntelitheos & Idrissi, 2017). Children were given a pencil and a paper and were instructed, in Arabic, to write in Arabic what they heard. Participants' responses were compiled for correct spelling, error types and the total number of errors.

4.5 Arabic plurals task

The Arabic plurals task is an Arabic adaptation of the Wug task developed by Berko (1958). The seminal work by Berko (1958) used two items for the plural, but since Arabic has dual forms, we used three items to evaluate children's knowledge of the Arabic plural. The forced-choice task included a total of fifteen 'sound' and fifteen 'broken' feminine and masculine plurals (for descriptions and discussions, see Boudelaa & Gaskell, 2002; McCarthy & Prince, 1990). The script for the task is that a character or an object is introduced, e.g., 'This is a man', then 'Now there are three. These are three ___?', followed by four choices, the target and three distractors. The distractors include a non-word phonologically similar to the target (e.g., /radʒlu:n/ for the target /ridʒaal/ 'men'), an Arabic word that is phonologically similar to the target (e.g., /ridʒəl/ 'foot' for the target /ridʒaal/ 'men'), and an ungrammatical plural form phonologically similar to the target (e.g., /radʒulaat/ *man.F.Pl for the target /ridʒaal/ 'men') (see Figure 5). All target words for this task are expected to be known by EA speaking children of this age according to the EMALAC corpus

2. Thanks to a reviewer for pointing out that this sounds a bit stilted. A more natural way to say this would be *ʕindi bajtun ʤamilun wa maʕahu ħadiqa*, which translates to "I have a pretty house and with it a garden". However, this was the original material used in the interactions and is reported here as such.

(Ntelitheos & Idrissi, 2017). Children were given a pencil with the paper test. In addition to the written script in Arabic, the research assistant would read the script to the participants to ensure they understood each target and choice. Participants had to circle the word they thought was appropriate. Participants' responses were compiled for error types and total number of errors per participant.

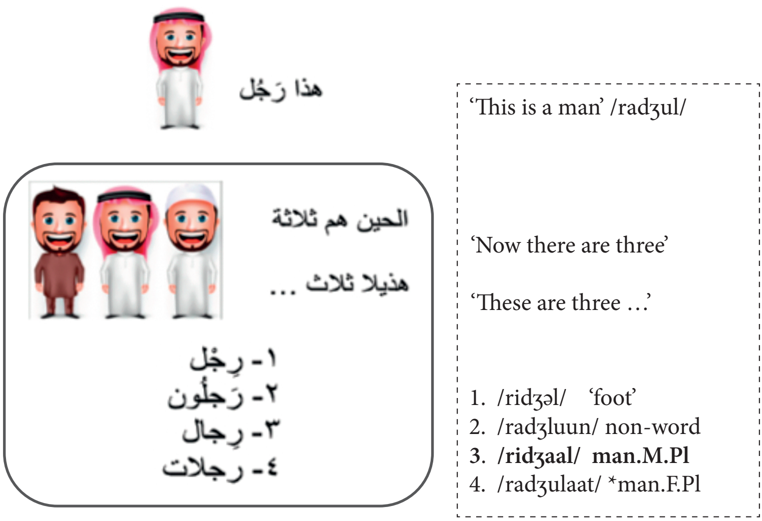


Figure 5. Example of a trial for the *Arabic Plurals* task on the left with the translation and IPA transcriptions of the target (in the singular and plural) and distractors on the right

5. Results

Participants' results were compiled for each task at each phase. Table 2 reveals the results on the LATFA tool regarding the participants' assessment of oral language abilities in the first grade. For the speech perception task, participants' performance ranged from 30 to 51 correct items on a total of 60. As for the phonological awareness task, children obtained from 10 to 29 correct answers on a possible maximum score of 36. The morphological awareness task shows the most variation, ranging from 0 correct morphological production to a perfect score of 24 on 24. Finally, the phonological production task exhibits the highest success rate, where participants reached scores ranging between 80 and 100 percent correct production.

Table 2. Participants' scores on the four tasks of the LATFA tool in the first grade, including minimum scores (Min), maximum scores (Max), mean scores (*M*), standard deviations of scores (SD), and Cut-off scores for Speech Perception (number correct over 60 items), Phonological Awareness (number correct over 36 items), Morphological Awareness (number correct over 24 items), and Phonological Production (percentage consonant correct over 30 words and 101 consonants). In line with Rvachew et al. (2017), the cut-off scores are approximately 1.25 standard deviations below the mean, with rounding and some adjustments for skewed distributions.

Task	Min	Max	*M*	*SD*	Cut-off score
Speech Perception	30	51	42.67	7.10	34
Phonological Awareness	10	29	22.00	4.97	16
Morphological Awareness	0	24	13.44	8.51	3
Phonological Production	80	100	98.73	3.41	94

In Table 3, participants' results on the literacy assessment in the second grade are presented. First, on the Arabic alphabet task, children obtained between 2 and 25 correct responses on a maximum score of 28. The dictation task reveals the most variation of success with participants' correct responses ranging from 1 to 55 out of 56. The Arabic plural task shows participants' scores ranging between 10 and a perfect score of 30 on 30 items.

Table 3. Participants' scores on Phase 2: Literacy assessment in the second grade, including minimum scores (Min), maximum scores (Max), mean scores (*M*), standard deviations of scores (SD), and Cut-off scores for Arabic alphabet (number correct over 28 items), Dictation (number correct over 56 items: 10 non-words, 21 words, 25 words in sentences), and Arabic Plurals (number correct over 30 items). In line with Rvachew et al. (2017), the cut-off scores are approximately 1.25 standard deviations below the mean, with rounding and some adjustments for skewed distributions.

Task	Min	Max	*M*	*SD*	Cut-off score
Arabic Alphabet	2	25	9.28	5.60	3
Dictation	1	55	30.16	15.05	11
Arabic Plurals	10	30	24	4.56	18

Table 4 presents the number of participants who passed or failed on each task of the four tasks of the LATFA tool (Phase 1: Oral language assessment) in the first grade considering the cut-off scores of 1.25 standard deviations below the mean (following Rvachew et al., 2017). Thus, children who obtained a score below the cut-off score failed a particular test. The ratio of failure for the speech perception task was 21% (9 children out of 43), 14% (6 children out of 43) for the phonological awareness task, 16% (7 children out of 43) for the morphological awareness task, and only 5% (2 children out of 43) for the phonological production task. Taken together, 37 children passed the LATFA tool (i.e., passed at least three of the four tasks), resulting in a total of 6 children who failed the LATFA tool (i.e., failed two or more of the four tasks). An in-depth investigation reveals that only a total of 16 children (out of the 43) failed one or more of the four tasks.

Table 4. Mean scores (*M*), standard deviations of scores (SD), Cut-off scores (approximately 1.25 standard deviations below the mean), number and percentage of pass, number and percentage of fail on the four tasks of the LATFA tool in the first grade.

Task	Score *M*	Score *SD*	Cut-off score	Pass *n* %	Fail *n* %
Speech Perception	42.67	7.10	34	34 79%	9 21%
Phonological Awareness	22.00	4.97	16	37 86%	6 14%
Morphological Awareness	13.44	8.51	3	36 84%	7 16%
Phonological Production	98.73	3.41	94	41 95%	2 5%

Table 5 presents the number of participants who passed or failed on each task of the literacy assessment in the second grade considering the cut-off scores of 1.25 standard deviations (following Rvachew et al., 2017) below the mean (i.e., children who obtained a score below the cut-off score failed a particular test). The ratio of failure for the Arabic alphabet was only a low 2% (1 child out of 43). The ratio of failure was the highest for the dictation task, with 14% of failure (6 children out of 43). While only 7% of children failed the task on Arabic plurals (3 out of 43). In-depth investigation reveals that a total of 7 children (out of the 43) failed at least one of the literacy assessment tasks in the second grade (2 of them failed two of the three literacy tasks). Amongst these 7 children, 4 of them had also failed at least 2 tasks of the LATFA tool in the first grade (2 of these 4 failed two of the three literacy tasks). In sum, of the 6 children who failed the LATFA tool (i.e., failed two or more of the four oral language tasks in the first grade), 4 of them also failed, a year later, at least one of our three literacy assessment tasks.

Table 5. Mean scores (*M*), standard deviations of scores (SD), Cut-off scores (approximately 1.25 standard deviations below the mean), number and percentage of pass, number and percentage of fail on the three tasks of Phase 2: Literacy assessment in the second grade.

Task	Score *M*	Score *SD*	Cut-off score	Pass *n* %	Fail *n* %
Arabic Alphabet	9.28	5.60	3	42 98%	1 2%
Dictation	30.16	15.05	11	37 86%	6 14%
Arabic Plurals	24	4.56	18	40 93%	3 7%

6. Discussion

A number of studies have established the connection between oral language and literacy skills (e.g., Boets et al., 2008; Holm et al., 2008; Pacton & Deacon, 2008; Pennala et al., 2010; van Viersen et al., 2018). In particular, phonological awareness and perception of speech correlate with literacy capacities (e.g., Boets et al., 2008; Goswami, 2000; Goswami & Bryant, 1990; Holm et al., 2008; Pan et al., 2015; Pennala et al., 2010; Rvachew & Grawburg, 2006; Rvachew et al., 2017; Vanvooren et al., 2017; White-Schwoch et al., 2017). In production, both morphological and speech production have been proven to impact literacy development (e.g., Hayiou-Thomas et al., 2017; Lewis et al., 2000; Lewis et al., 2002; Marquis et al., 2012; Masso et al., 2017; McLeod et al., 2019; Overby et al., 2015; Pan et al., 2015; Puranik et al., 2008; Speece et al., 1999; van Viersen et al., 2018; Wolter et al., 2009). The current study thus adds to the existing body of literature by including Arabic data.

Research on linguistic abilities in Arabic speaking children is sparse (but see Abu-Rabia & Siegel, 2002; Saiegh-Haddad & Geva, 2008; Taibah & Haynes, 2011) while assessment tools in the Arabic language are scarce (see Elbeheri, Everatt, Reid, & al Mannai, 2006; as well as Shaalan, 2009 for a discussion). One particularity of Arabic, under the root-based approach, is that Arabic has consonant roots, vocalic melodies, and templates (which combine the melodies and consonantal roots) (e.g., Idrissi, Prunet, & Béland, 2008; McCarthy, 1979, 1981; Prunet, Béland, & Idrissi, 2000). Other particularities of Arabic are that Arabic script is written from right to left, is cursive, in addition to having a transparent orthography (presenting or lacking diacritics for short vowels). To paint a detailed portrait of language acquisition in children, we must consider a variety of linguistic factors, present in various language families, in order to highlight characteristics that may be otherwise overlooked in other languages.

The goal of the current study was to introduce an assessment tool based on oral language skills that could be used to identify EA children at risk of presenting literacy difficulties. The oral language skills that were evaluated in the first grade included two tasks of language perception (i.e., phonological awareness & speech perception) and two tasks of language production (i.e., morphological awareness & phonological production) in EA. The literacy skills that were used included three different written assessment tasks in Arabic (i.e., writing the Arabic alphabet, a dictation task, and a task evaluating knowledge of Arabic plurals). Our study reveals that we have developed an assessment tool based on oral language skills in Arabic that determined a near 70% of successful prediction of literacy difficulties in the children who failed two or more oral tasks the year before.

7. Future directions

In order to ascertain the predictive validity of our assessment tool, a larger sample size of participants with multiple schools in various Emirates should be reached. The next step would be to determine which particular items of the assessment tool are the best predictors of literacy skills in order to develop a screening test with fewer items that could easily be used by speech-language therapists, as well as school teachers to quickly determine, early on, children facing potential writing difficulties.

Acknowledgements

Special thanks to the school, the parents and their children for participating in this project. We are thankful to Akeela Fatheen Abdul Gafoor, Al Yazia Mohammed Al Mansouri and Wadha Ayed Muaid Asi Al Otaibi for research assistance, and Edith Lebel for creating the images for the Morphological Awareness task. In addition, many thanks to Susan Rvachew for consulting on this project.

Funding

Finally, this project was supported by the United Arab Emirates University Research Start-up Grant Award No. 31H107.

References

Abu-Rabia, S., & Siegel, L. S. (2002). Reading, syntactic, orthographic, and working memory skills of bilingual Arabic-English speaking Canadian children. *Journal of Psycholinguistic Research*, 31(6), 661–678. https://doi.org/10.1023/A:1021221206119

Al Ameri, H. (2009). A phonological description of Emirati Arabic. Master thesis. Purdue University, West Lafayette.

Al Kaabi, M. (2015). The anatomy of Arabic words: The role of the root in representations and processing. Ph.D. dissertation. New York University, New York.

Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2–3), 150–177. https://doi.org/10.1080/00437956.1958.11659661

Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech, Language, and Hearing Research*, 38, 446–462. https://doi.org/10.1044/jshr.3802.446

Boets, B., Wouters, J., van Wieringen, A., De Smedt, B., & Ghesquière, P. (2008). Modelling relations between sensory processing, speech perception, orthographic and phonological ability, and literacy achievement. *Brain and Language*, 106(1), 29–40. https://doi.org/10.1016/j.bandl.2007.12.004

Boudelaa, S., & Gaskell, M. G. (2002). A re-examination of the default system for Arabic plurals. *Language and Cognitive Processes*, 17(3), 321–343. https://doi.org/10.1080/01690960143000245

Bourassa, D. C., & Treiman, R. (2001). Spelling development and disability: The importance of linguistic factors. *Language, Speech, and Hearing Services in Schools*, 32, 172–181. https://doi.org/10.1044/0161-1461(2001/016)

Bourassa, D. C., Treiman, R., & Kessler, B. (2006). Use of morphology in spelling by children with dyslexia and typically developing children. *Memory and Cognition*, 34(3), 703–714. https://doi.org/10.3758/BF03193589

Brosseau-Lapré, F., & Rvachew, S. (2008). *Test de Conscience Phonologique Préscolaire* [Computer Software].

Elbeheri, G., Everatt, J., Reid, G., & al Mannai, H. (2006). Dyslexia assessment in Arabic. *Journal of Research in Special Educational Needs*, 6(3),143–152. https://doi.org/10.1111/j.1471-3802.2006.00072.x

George, F., & Pech-Georgel, C. (2006). *BELO: Batterie d'évaluation de lecture et d'orthographe*. Dardilly: Éditions Solal.

Goswami, U. (2000). Phonological representations, reading development and dyslexia: Towards a cross-linguistic theoretical framework. *Dyslexia*, 6, 133–151. https://doi.org/10.1002/(SICI)1099-0909(200004/06)6:2<133::AID-DYS160>3.0.CO;2-A

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Hillsdale: Lawrence Erlbaum Associates.

Hayiou-Thomas, M. E., Carroll, J. M., Leavett, R., Hulme, C., & Snowling, M. J. (2017). When does speech sound disorder matter for literacy? The role of disordered speech errors, co-occurring language impairment and family risk of dyslexia. *The Journal of Child Psychology and Psychiatry*, 58(2), 197–205. https://doi.org/10.1111/jcpp.12648

Holm, A., Farrier, F., & Dodd, B. (2008). Phonological awareness, reading accuracy and spelling ability of children with inconsistent phonological disorder. *International Journal of Language and Communication Disorders*, 43(3), 300–322. https://doi.org/10.1080/13682820701445032

Idrissi, A., Prunet, J.-F., & Béland, R. (2008). On the mental representations of Arabic roots. *Linguistic Inquiry*, 39(2), 221–259. https://doi.org/10.1162/ling.2008.39.2.221

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2000). Academic outcomes in children with histories of speech sound disorders. *Journal of Communication Disorders*, 33(1), 11–30. https://doi.org/10.1016/S0021-9924(99)00023-4

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2002). Correlates of spelling abilities in children with early speech sound disorders. *Reading and Writing: An Interdisciplinary Journal*, 15(3–4), 389–407. https://doi.org/10.1023/A:1015237202592

Marquis, A. (2019). Verb production assessment test for Emirati Arabic children. In W. Zoghbor, S. Al Alami, & T. Alexiou (Eds.), *Proceedings of the 1st Applied Linguistics & Language Teaching (ALLT) Conference* (pp. 87–99). Dubai: Zayed University Press.

Marquis, A., Royle, P., Gonnerman, L. M., & Rvachew, S. (2012). La conjugaison du verbe en début de scolarisation. *TIPA-Travaux Interdisciplinaires sur la Parole et le Langage*, 28, 1–16. https://doi.org/10.4000/tipa.201

Masso, S., Baker, E., McLeod, S., & Wang, C. (2017). Polysyllable speech accuracy and predictors of later literacy development in preschool children with speech sound disorders. *Journal of Speech, Language, and Hearing Research*, 60, 1877–1890. https://doi.org/10.1044/2017_JSLHR-S-16-0171

McCarthy, J. J. (1979). On stress and syllabification. *Linguistic Inquiry*, 10(3), 443–465. http://www.jstor.org/stable/4178121

McCarthy, J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12(3), 373–418.

McCarthy, J. J., & Prince, A. (1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory*, 8(2), 209–283. https://doi.org/10.1007/BF00208524

McLeod, S., Harrison, L. J., & Wang, C. (2019). A longitudinal population study of literacy and numeracy outcomes for children identified with speech, language, and communication needs in early childhood. *Early Childhood Research Quarterly*, 47, 507–517. https://doi.org/10.1016/j.ecresq.2018.07.004

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychological Bulletin*, 138(2), 322–52. https://doi.org/10.1037/a0026744

Ntelitheos, D., & Idrissi, A. (2017). Language growth in child Emirati Arabic. In H. Ouali (Ed.), *Perspectives on Arabic Linguistics XXIX* (pp. 229–248). Amsterdam: John Benjamins. https://doi.org/10.1075/sal.5

Overby, M. S., Masterson, J. J., & Preston, J. L. (2015). Preliteracy speech sound production skill and linguistic characteristics of grade 3 spellings: A study using the Templin archive. *Journal of Speech, Language, and Hearing Research*, 58(6), 1–16. https://doi.org/10.1044/2015_JSLHR-S-14-0276

Pacton, S., & Deacon, S. H. (2008). The timing and mechanisms of children's use of morphological information in spelling: A review of evidence from French and English. *Cognitive Development*, 23, 339–359. https://doi.org/10.1016/j.cogdev.2007.09.004

Pan, J., Song, S., Su, M., McBride, C., Liu, H., Zhang, Y., Li, H., & Shu, H. (2015). On the relationship between phonological awareness, morphological awareness and Chinese literacy skills: evidence from an 8-year longitudinal study. *Developmental Science*, 19(6), 1–10. https://doi.org/10.1111/desc.12356

Pennala, R., Eklund, K., Hämäläinen, J., Richardson, U., Martin, M., Leiwo, M., Leppänen, P. H., & Lyytinen, H. (2010). Perception of phonemic length and its relation to reading and spelling skills in children with family risk for dyslexia in the first three grades of school. *Journal of Speech Language, and Hearing Research*, 53, 710–724. https://doi.org/10.1044/1092-4388(2009/08-0133)

Prunet, J. F., Béland, R., & Idrissi, A. (2000). The mental representation of Semitic words. *Linguistic inquiry*, 31(4), 609–648. https://doi.org/10.1162/002438900554497

Puranik, C., Petcher, Y., Al Otaiba, S., Catts, H. W., & Lonigan, C. J. (2008). Development of oral reading fluency in children with speech or language impairments: A growth curve analysis. *Journal of Learning Disabilities*, 41, 545–560. https://doi.org/10.1177/0022219408317858

Rvachew, S. (2009). *Speech Assessment and Interactive Learning System* (Version 2) [Software].

Rvachew, S., & Grawburg, M. (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research*, 49, 74–87. https://doi.org/10.1044/1092-4388(2006/006)

Rvachew, S., Marquis, A., Brosseau-Lapré, F., Paul, M., Royle, P., & Gonnerman, L. M. (2013). Speech articulation performance of francophone children in the early school years: Norming of the Test de Dépistage Francophone de Phonologie. *Clinical Linguistics & Phonetics*, 27(12), 950–968. https://doi.org/10.3109/02699206.2013.830149

Rvachew, S., Royle, P., Gonnerman, L. M., Stanké, B., Marquis, A., & Herbay, A. (2017). Development of a tool to screen risk of literacy delays in French-speaking children: PHOPHLO. *Canadian Journal of Speech-Language Pathology and Audiology*, 41(3), 321–340.

Saiegh-Haddad, E., & Geva, E. (2008). Morphological awareness, phonological awareness, and reading in English–Arabic bilingual children. *Reading and Writing*, 21(5), 481–504. https://doi.org/10.1007/s11145-007-9074-x

Shaalan, S. (2009). Considerations for developing and adapting language and literacy assessments in Arabic-speaking countries. In E. L. Grigorenko (Ed.), *Multicultural psychoeducational assessment* (pp. 287–314). New York: Springer Publishing Company.

Speece, D. L., Roth, F. P., Cooper, D. H., & De La Paz, S. (1999). The relevance of oral language skills to early literacy: A multivariate analysis. *Applied Psycholinguistics* 20(2), 167–90. https://doi.org/10.1017/S0142716499002015

Taibah, N. J., & Haynes, C. W. (2011). Contributions of phonological processing skills to reading skills in Arabic speaking children. *Reading and Writing*, 24(9), 1019–1042. https://doi.org/10.1007/s11145-010-9273-8

van Viersen, S., de Bree, E. H., Zee, M., Maassen, B., van der Leij, A., & de Jong, P. F. (2018). Pathways into literacy: The role of early oral language abilities and family risk for dyslexia. *Psychological Science*, 29(3), 418–428. https://doi.org/10.1177/0956797617736886

Vanvooren, S., Poelmans, H., De Vos, A., Ghesquière, P., & Wouters, J. (2017). Do prereaders' auditory processing and speech perception predict later literacy? *Research in Developmental Disabilities*, 70, 138–151. https://doi.org/10.1016/j.ridd.2017.09.005

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., Zecker, S. G., & Kraus, N. (2017). Auditory processing in noise: A preschool biomarker for literacy. *PLoS Biology*, 13(7), e1002196. https://doi.org/10.1371/journal.pbio.1002196

Wolter, J. A., Wood, A., & D'zatko, K. W. (2009). The influence of morphological awareness on the literacy development of first-grade children. *Language, Speech and Hearing Services in Schools*, 40, 286–298. https://doi.org/10.1044/0161-1461(2009/08-0001)

Sentence repetition in children with autism spectrum disorder in Saudi Arabia

An investigation of morphosyntactic abilities

Mada Abdullah Al-Hassan and Theodoros Marinis
University of Reading / University of Konstanz

This study investigated the morphosyntactic abilities of Saudi Arabic-speaking children with Autism Spectrum Disorder (ASD) using a Saudi Arabic Sentence Repetition Task (Saudi-SRT) developed for this study. The first objective of the study was to determine whether Saudi Arabic-speaking children with autism living in Saudi have difficulties in morphosyntax as a function of their verbal abilities. For this reason, the participating children with autism were divided into subgroups of children with autism with normal language (ALN) and children with autism with language impairment (ALI) based on their verbal abilities. The second objective of this study was to identify whether the children's performance differs as a function of the scoring scheme used. Towards that aim, two scoring schemes were used based on whether children repeated sentences verbatim versus whether they used the targeted structure accurately. The third objective was to address differences between the structures used in the Saudi-SRT. The study involved 62 five- to seven-year-old children who speak Saudi Arabic as their first language: 20 children with autism (n = 10 ALN and n = 10 ALI) and 42 typically developing (TD) control children. The children with autism showed difficulties in morphosyntax as a function of their verbal abilities, but the scoring scheme modulated this. In nonverbal abilities and receptive vocabulary, ALI scored lower than TD and ALN, and there was no difference between TD and ALN. As for the grammatical abilities, the less sensitive verbatim scheme showed a similar pattern in the ALN and ALI children who performed less well than the TD children; in contrast, the more sensitive structural scheme showed that in the most syntactically complex structures, the ALN children performed similarly to the TD children and better than the ALI children.

Keywords: Autism Spectrum Disorder, Specific Language Impairment, morpho-syntax, sentence repetition, Saudi-Arabic

https://doi.org/10.1075/sal.10.06alh

1. Introduction

Autism spectrum disorder (ASD) is a communicative and behavioral disorder, described as a "developmental disorder" due to early development of symptoms at the first years of life, despite the fact that its diagnosis can be throughout the lifespan (Sundar Raj, Prasath, & Adalarasu, 2015). Impaired capacities mark children with autism for reciprocal socio-communicative interaction and a restricted, stereotyped repetitive repertoire of interests and activities (Solari, Grimm, McIntyre, Zajic, & Mundy, 2019; World Health Organization, 2013). The vast majority of studies on the language abilities of children with autism focused on pragmatics and atypical prosody (McCann & Peppé, 2003; Tager-Flusberg, 1999). The grammatical capabilities of children with autism have received less attention, with limited studies addressing the relation and input of elements of grammar to communication and social interaction (Colle, Baron-Cohen, Wheelwright, & Van Der Lely, 2008; Hobson, Lee, & Hobson, 2010). Within the domain of grammar, relatively few studies have investigated the acquisition of morphosyntax (Durrleman, Delage, Prévost, & Tuller, 2017; Eigsti & Bennetto, 2009; Roberts, Rice, & Tager-Flusberg, 2004; Terzi, Marinis, Kotsopoulou, & Francis, 2014; Terzi, Marinis, Zafeiri, & Francis, 2019). The majority of existing studies that investigate the grammatical abilities of children with autism have focused on monolingual English-speaking children. Research in the grammatical abilities of Saudi Arabic-speaking children with autism has not been done before. The present study fills this gap by investigating the morphosyntactic abilities of children with autism in Saudi Arabia compared to typically developing (TD) children.

Some previous studies on children with autism, for example, Tager-Flusberg and Calkins (1990), showed that children with autism do not differ from TD children in their grammatical abilities. Tager-Flusberg and Calkins (1990) investigated the development of grammatical complexity as reflected by Mean Length of Utterance (MLU) in a group of children with autism compared to a group of children with Down Syndrome (DS) and a group of TD controls. The DS children were chosen to match the children with autism in chronological age (CA) and language level, but they were not matched on IQ or nonverbal mental age (MA) levels. A sample of spontaneous utterances was collected during bimonthly visits to the children's homes. The researchers found that children with autism follow the same general developmental path as TD children in the acquisition of grammatical and lexical aspects of language, as there were no significant qualitative differences between the groups. However, children with autism showed a delayed onset of development.

In a more recent study, Eigsti, Bennetto, and Dadlani (2007) investigated syntactic and higher-level discourse abilities in three groups: children with autism, children with non-specific Developmental Delays (DD) matched on nonverbal IQ

(NVIQ), gender, and CA, and TD children matched on NVIQ and gender. Both the DD children and the children with autism were verbal and relatively high functioning, with mean NVIQ scores in the low – average range. Children participated in a free play session that took place during their visit to the lab. Eigsti et al. (2007) found that children with autism produced syntactically less complex language than the TD controls. They suggested that children with autism exhibit a higher degree of developmental scatter; that is, they produce grammatical structures that are less predictable based on previous productions. This result expanded upon the 1982 finding by Bartolucci and colleagues that children with autism produce fewer grammatical morphemes than controls. The results indicated that the autism group reaches its syntactic abilities through an atypical developmental pathway.

The heterogeneity among children with autism at the language levels and its development has not been integrated into the investigations of language skills, thus potentially obscuring essential aspects of language disorder in this population. As a result, difficulties in morphosyntax for children with autism remained undetected (Eigsti et al., 2007; Tager-Flusberg & Calkins, 1990). Mapping out a detailed linguistic profile for the population with autism is crucial to establishing reliable differences between autism and other developmental disorders, such as Specific Language Impairment (SLI). SLI is "*a developmental disorder that is diagnosed based on language levels that fall significantly below age expectations, but in the absence of other conditions (e.g., hearing loss, intellectual disabilities, or evidence of organic pathology)*" (Kjelgaard & Tager-Flusberg, 2001, p. 3).

Studies of children with SLI in various languages reveal varied symptoms; nearly all of them seem to reflect deficiencies in grammatical computation and phonological short-term memory (Leonard, 2014). *"The symptoms of the disorder are sensitive to the type of language being acquired, with extraordinary weaknesses seen in those areas of language that are relatively challenging for younger TD children"* (Leonard, 2014, p. 1). Cross-linguistic research showed that the characteristics of the ambient language influenced the relative strengths and weaknesses of children with SLI. For example, deficits with tense and agreement were attenuated in specific languages, replaced by other weaknesses not seen in English (Leonard, 2014, p. 2). In studies of SLI in children, proposed clinical markers have been theoretically based on limitations in cognitive processing mechanisms (e.g., phonological memory) and linguistic knowledge (i.e., finiteness marking; Conti-Ramsden, 2003). *"The term clinical marker was introduced to the language disorders literature by Rice and Wexler (1996) to refer to highly accurate diagnostic tasks that are based on characteristics or behaviours that are indicative of SLI"* (Poll et al., 2010, p. 415). Three types of clinical markers have been explored in children with SLI: (a) non-word repetition, (b) sentence repetition, and (c) grammaticality judgments of finiteness marking. Non-word repetition tasks, which tap both short-term phonological memory and phonological processing, have been shown to result in a much higher likelihood of correct identification of children with SLI when compared with a traditional standardised test (Poll, Betz, & Miller, 2010).

SLI is generally defined in children whose NVIQ scores are within the normal range, but whose language performance on language tests falls below mean. In addition, there is a considerable heterogeneity in their pattern of language skills (e.g., Tomblin & Zhang, 1999). To understand the overlap between ASD and SLI, ASD groups have been divided into two subgroups based on their scores on a verbal assessment, such as the Peabody Picture Vocabulary Test (PPVT), and an assessment of nonverbal abilities, such as the Raven's Coloured Progressive Matrices (CPM). One of these studies is the study by Kjelgaard and Tager-Flusberg (2001).

Kjelgaard and Tager-Flusberg (2001) were interested in revisiting the findings from previous studies by Bartak, Rutter, and Cox (1977), which suggested some similarities between ASD and SLI. Towards that end, they explored the range of language abilities among a large group of children with autism varying in age and IQ level using a battery of standardised language tests, to test the following abilities: (1) phonological abilities, (2) lexical and higher-order semantics, and (3) grammatical language abilities (Kjelgaard and Tager-Flusberg, 2001). Kjelgaard and Tager-Flusberg (2001) looked into subgroups that were defined on the basis of the children's performance on one of the major language measures. They identified subgroups of children with autism and found that some children with autism demonstrated normal language skills (ALN); for other children, their language skills are significantly below CA expectations (ALI). The profile of performance across the standardised measures for the ALI children was similar to the profile that had been found among children with SLI, displaying impairment in syntax, as reported by Tomblin and Zhang (1999).

Kjelgaard and Tager-Flusberg (2001) found that there was a very wide range of language abilities across the IQ spectrum within children with autism, explaining previous well-known clinical literature (Lord & Paul, 1997). They also found a significant relationship between IQ and language abilities, especially as measured by vocabulary tests. Furthermore, they found that among lower IQ children some had language skills within the normal range, and among high IQ children some had language skills in the impaired range. Kjelgaard and Tager-Flusberg (2001) argued only the children with higher IQ scores were able to complete the full range of language tests, and that children with higher IQ scores had better language abilities overall. Also, they argued that IQ itself accounts for some of the heterogeneity found in language skills among children with autism. In addition, language skills can be independent of IQ in autism, and can be more important in understanding both current functioning and long-term prognosis for children with this disorder (Venter, Lord, & Schopler, 1992).

The first complete and well-informed study of morphosyntactic development in children with autism was conducted by Roberts et al. (2004) that investigated third-person singular and past tense usage to refine the language phenotypes of verbal children with autism, with specific emphasis on identifying subgroups within the population. The participants were divided into three groups based on their performance on the PPVT: ASD, SLI, and ALI. Roberts et al.'s (2004) main findings were that the ASD and the SLI groups showed high rates of omission on both the present and past tense probes, and the children with autism and ALI children performed worse than the children with SLI in the same language range. They argued the possibility of an SLI subgroup within their sample of children with autism based on a similarity of symptoms across the two conditions. They also highlighted the importance of matching the groups in nonverbal abilities, even if the groups were matched for their verbal abilities. They found that the performance of the children with autism was weaker than expected for their general level of both linguistic and cognitive development, indicating a specific morphosyntactic deficit.

Tager-Flusberg (2015) also investigated whether children with ALI show the same pattern of performance, including both correct and error responses, as do children with SLI on tasks tapping both non-word discrimination and repetition. Four groups participated in this study: ALN children, ALI children, children with SLI, and TD children as a control group. The majority of those participants were verbal but impaired in language relative to age-matched peers. Tager-Flusberg's (2015) hypothesis was that the ALI subgroup has ASD and co-morbid SLI and compared children with autism to children with SLI and TD controls on aspects of language processing that are impaired in children with SLI: repetition of non-words. The results showed a pattern of performance among the children with autism and language impairment that was similar to those with SLI and contrasted with the children with autism and no language impairment and typical controls, providing further evidence for the hypothesis that a subgroup of children with autism has co-morbid SLI.

A further study investigating morphosyntactic development in children with autism was the one by Park, Yelland, Taffe, and Gray (2012). Park et al. (2012) explored three possibilities of morphosyntactic development in children with autism: whether children with autism have atypical morphological and syntactic development; whether children with autism use rote learning as an atypical grammatical process; and whether children with autism acquire grammatical morphemes in atypical order (Park et al., 2012: p. 98). The study included three groups of children aged from 3–6 years: (1) children with autism, (2) children with DD without autism, and (3) TD children. The demographic differences between the groups, including age and non-verbal cognitive functioning, were statistically controlled. Park et al. (2012) assessed the children's cognitive functioning, receptive and expressive language, expressive morphological and syntactic skills. Park et al. (2012) found that children with autism have some skills which may be atypical (the use of verb phrases), some skills which are delayed due to overall developmental delay (the use of regular plurals, past tense and third person singular, the use of noun phrases and sentence structures and length of utterance), and some skills which are intact (the use of *-ing*, *in/on*, articles, and questions). Park et al. (2012) argued that the difficulties for the children with autism in some sub-skills are not unique to autism but likely to be due to a developmental delay.

Two additional recent studies showed similar outcomes in English grammatical development. Perovic, Modyanova, and Wexler (2013a) investigated syntactic binding in children with autism who were matched with two groups of TD children, one on language abilities and the other one on NVIQ. Perovic and colleagues investigated reflexive and personal pronoun usage using a two-choice picture selection task adopted from Wexler and Chien (1985). Perovic et al. (2013a) found that children with autism showed more difficulties interpreting personal pronouns than TD controls, although the two groups performed similarly in terms of their verbal and nonverbal abilities. This was in line with the reported language delays and pragmatic deficits in this population. However, they exhibited difficulties interpreting reflexives, performing significantly worse than either of the two groups of controls matched.

In a follow-up study, Perovic, Modyanova, Wexler, and Naigles (2013b) investigated binding in children with autism and children with Williams syndrome (WS). They divided the children with autism into two subgroups according to the presence or absence of language impairment: ALI and ALN, respectively. The performance of the ALI and ALN groups was compared to that of two TD control groups matched to each ASD group on gender and nonverbal MA, and to a group of WS children matched on CA. Perovic et al. (2013b) found a remarkably different pattern of comprehension of reflexives and personal pronouns in the ALI children than in the ALN, WS, and TD control children. The results indicated that all five groups had an equal delay in their comprehension of personal pronouns. This was in line with delays widely reported in the TD literature, argued to be due to delayed pragmatic abilities (Perovic et al., 2013b). However, and most strikingly, the ALI group also showed a pronounced difficulty in the comprehension of reflexive pronouns. This was argued to be due to the existence of a particular impairment concerning reflexives, which is unrelated to either general language delays or cognitive deficits generally present in many individuals with ASD and WS, or to general pragmatic deficits, known to be particularly prevalent in the population with autism.

A study by Janke and Perovic (2015) also involved the comprehension of binding in high functioning children with autism (HFA). The performance of HFA children was compared with that of two groups of younger TD children: one matched on nonverbal MA and the other one on verbal MA. They found that the HFA group demonstrated satisfactory performance on reflexives on a par with the TD matched children, but their comprehension of personal pronouns was somewhat weaker, and there was no significant difference between the groups which confirms the existing literature (Janke and Perovic, 2015). The weak comprehension could be related to difficulties in Theory of Mind (ToM), although this possibility was not discussed by Janke and Perovic. This finding contrasts with the results of studies by Perovic et al. (2013a; b) who investigated HFA children and discovered severe difficulties with the binding of reflexives. Janke and Perovic (2015) argued that performance on binding in ASD is mixed. The picture emerging is that children classified as Low Functioning Autism (LFA) do exhibit problems in this area of grammar. However, HFA children perform on par with their nonverbal MA-matched peers.

Modyanova, Perovic, and Wexler (2017) investigated the nature of finiteness abilities and compared tense marking abilities in children with ALI and with ALN, and that of two TD control groups matched on nonverbal MA and verbal MA. The authors divided the children with autism into two groups (ALN, ALI) in line with the classifications in the literature (Kjelgaard & Tager-Flusberg, 2001). The authors assessed the participants' NVIQ, expressive vocabulary, receptive vocabulary and receptive grammar. Modyanova et al. (2017) found large deficits in the ALI group, in marking tense, where they performed significantly worse than their younger TD controls', and significantly worse than that of the ALN group, which in contrast performed similarly to their TD controls. Moreover, both ASD groups showed distributions of null or overt subjects with non-finite and finite verbs in agreement with those found in young TD children. A key difference, however, was that the ALI group used (rather than simply omitted) the wrong tense in some sentences, a feature not reported in the Optional Infinitive (OI) stage of TD or SLI children. Modyanova et al. (2017) argued that there was a clear distinction in the morphosyntactic abilities of the two subgroups of children with autism. The language system responsible for finiteness in the ALN group seems to be functioning comparably to that of the TD children, whereas the ALI group seems to experience an extensive grammatical deficit concerning finiteness which does not seem to improve with age. Crucially, the ALI group seemed to have worse grammatical abilities even than those reported for SLI.

A further study investigating the grammatical abilities of children with ASD is the study by Wittke, Mastergeorge, Ozonoff, Rogers, and Naigles (2017) that recruited a heterogeneous sample of children with autism. The participants completed extensive behavioural testing, including language assessments as part of their participation in the project. The participants were divided into three groups based on their language and NVIQ: (1) high Verbal children scored in the normal range (standard scores above 85) for both non-verbal and vocabulary language testing, (2) low Verbal children whose NVIQ was below 85, and standardised vocabulary testing was commensurate with their NVIQ; and (3) minimally Verbal children whose NVIQ and vocabulary performance was significantly below average (i.e., standard scores below 70). In addition, Wittke et al. (2017) divided the verbal children into three language sub-groups: (1) two children from the Minimally Verbal group were combined with the children in the Low Verbal group to form the "ALI" group, (2) The children from the High Verbal group were assigned into one of two groups based on their frequency of grammatical errors: (a) children who produced grammatical errors in more than 10% of their total utterance" were placed in the "Grammatical Impairment" (GI) group, and (b) children who produced grammatical errors in fewer than 10% of their total utterance were placed in the "ALN" group.

Wittke et al. (2017) found that ALN children scored similarly to what would be expected from TD children; they made few grammatical morpheme errors in nouns, verbs, and pronouns, and their lexical abilities also presented as intact and they were producing a variety of word types. This group had a similar rate as children in the sample from Kjelgaard and Tager-Flusberg (2001) for children with autism who have normal cognitive and language abilities. However, the ALI group produced much less speech than their other verbal peers with ASD; they produced fewer grammatical morphemes, were significantly more likely to use echolalia, and presented lower NVIQ scores than the other two groups. Wittke et al. (2017) found that both ALI and ALN groups presented with language patterns that were mostly commensurate with their non-verbal abilities and autism severity. As for the GI group, Wittke et al. (2017) found significant weaknesses in their morphosyntactic production, including more frequent verb, noun, and pronoun morphology errors, as well as more overall ungrammatical language. However, the GI group was more advanced than the ALI group on NVIQ and vocabulary and more frequent usage of grammatical rates. Wittke et al. (2017) argued that deficits in NVIQ coincided with language impairments that included more atypical language use and less frequent grammatical marker use and that the findings aligned with some of the previous research that claimed that some children with autism met the general criteria for SLI, evidenced by impaired grammatical skills with relative strength in vocabulary (Durrleman & Zufferey, 2009; Kjelgaard & Tager-Flusberg, 2001; Roberts et al., 2004). Wittke et al. (2017) also argued that their findings lent support to theories that suggest that the acquisition of grammar depends, at least somewhat, on factors external to general cognition (e.g., Lewis & Landau, 2015; Tuller et al., 2017; Valian, 2015). That is, while this sample of children with autism included two subgroups whose language was generally commensurate with their NVIQ (i.e., the ALI and ALN groups), it also included one subgroup whose NVIQ and vocabulary abilities were high, yet whose grammatical abilities were markedly impaired.

All studies mentioned above investigated the grammatical abilities of English-speaking children with autism. The number of studies investigating grammatical abilities of children with autism in languages other than English is even smaller and have been conducted in Greek, French, Mandarin, Danish, Egyptian Arabic, Saudi Arabic, and Hebrew (Abdalla & Crago, 2008; Brynskov et al., 2017; Durrleman & Delage, 2016; Meir & Novogrodsky, 2019; Morsi, 2009; Prévost et al., 2018; Terzi et al., 2014; Terzi, Marinis, & Francis, 2016a; b; Zhou, Crain, Gao, Tang, & Jia, 2015).

Terzi and colleagues investigated the grammatical abilities of Greek-speaking children with autism. Terzi et al. (2014) investigated pronoun reference and verbs with nonactive morphology (morphology that marks passive and reflexive constructions, anti-causatives, reciprocals, and middles) in HFA children and matched HFA children with TD children on CA, PPVT scores, and Raven's scores. Terzi et al. (2014) found that children with autism were less accurate in the comprehension and production of clitics but did not differ from the TD children in interpreting reflexives and nonactive morphology. Terzi et al. (2014) findings are unlike the findings in Perovic et al. (2013a; b) which did not include HFA children in their study. Terzi et al. suggested that the difficulties found in their study could be a result of difficulties in the syntax-pragmatics or the syntax-phonology interface.

In a follow-up study, Terzi et al. (2016a) investigated the usage of reflexive pronouns, pronominal object clitics, and their counterpart determiner phrases by HFA children individually matched with TD children on age, PPVT scores, and Raven's scores, similarly to Terzi et al. (2014). However, in this study, they narrowed the age range to avoid confounds due to different developmental trends in the two groups. Terzi et al. (2016a) found that the children with autism performed well on reflexives but performed significantly worse than the TD children on clitic pronouns. In the production and comprehension of clitics, the children with autism fell behind their language matched TD controls. Terzi et al. (2016a) argued that this behaviour did not reflect problems with syntax proper but was a consequence of pragmatic shortcomings with consequences for the syntax-pragmatics interface.

In a further follow-up study, Terzi et al. (2016b) investigated whether the difficulties with clitics had a purely morphosyntactic source or whether they were the consequence of difficulties at the interface of morphosyntax with discourse or with discourse and prosody, i.e., clitics based on a more complex syntax. Their objective was to replicate the study by Terzi et al. (2014) in a new group of HFA children of similar age. The participating HFA children were individually matched with TD children in age and PPVT scores. The results from the simple clitics replicated the findings by Terzi et al. (2014), with HFA children falling behind their language matched controls on both the comprehension and production of object clitics, with the gap being wider for production. For comprehension, children with autism committed the same errors as in the previous study by Terzi et al. (2014). However, the children with autism did not differ from the TD children in the comprehension and production of clitics in clitic left dislocation structures.

Grammatical abilities of French-speaking children with autism were investigated by Durrleman and Delage (2016) and Prévost et al. (2018). Durrleman and Delage (2016) investigated the first- and third-person accusative clitic production. Their participants included ASD and SLI individuals as well as age-matched and younger TD controls. They assessed grammar using a sentence completion task and a ToM task. Durrleman and Delage (2016) found similar deficits for both ASD and SLI participants on the production of accusative clitics, which they argued to be unrelated to their nonverbal abilities. Their results are relevant for the debate around the existence of language disorders in those individuals with ASD who would either resemble SLI (Roberts et al., 2004) or slightly resemble non-specific language impairment resulting from their general cognitive ability (Tager-Flusberg & Calkins, 1990).

Prévost et al. (2018) included children with autism divided into two subgroups: ALI and ALN. Also included were a group of age-matched children with SLI and a TD control group. The performance of the children in the ALI group and the SLI group was similar in third person accusative clitics that were produced at lower rates than first-person pronouns and in contrast to the children in the ALN group who performed similarly to the TD children (Durrleman & Delage, 2016; Terzi et al., 2014). The researchers argued that ALI children's performance was not related to the NVIQ of the children with autism and struck a parallel to SLI.

The grammatical abilities of Mandarin-speaking children were addressed in Zhou et al. (2015) that investigated the production of grammatical morphemes by Mandarin-speaking children with high functioning autism. Zhou et al. (2015) found that Mandarin-speaking children with autism scored significantly lower in grammatical morphemes than the age-matched and IQ-matched TD group and the MLU-matched TD group. They argued that the difficulty in the use of grammatical morphemes could not be attributed to articulation difficulties because all children passed the articulation test conducted after the experimental session. Zhou et al. (2015) argued that their findings have important implications for understanding the grammatical abilities of children with autism because previous research found that a subgroup of English-speaking children with autism exhibited deficits in using grammatical morphemes to mark tense (Pierce & Bartolucci, 1977; Roberts et al., 2004).

The language competence of Danish children with autism were investigated by Brynskov et al. (2017) who compared the language abilities of young Danish children with autism with and without early language delay to TD children matched on nonverbal mental age. The children completed three language tasks: (1) the PPVT that assesses receptive vocabulary, (2) the CELF that assesses language structure, and (3) the Test of Danish Receptive Morphology (DIM) that assesses receptive morphology. Brynskov et al. (2017) found that the children with ASD had significant delays in both syntax and morphology and vocabulary measures with significant within-group heterogeneity; furthermore, syntactic and morphological measures were impaired even for subgroups matched on vocabulary. Also, children in the ASD group without early language delay showed syntactic and morphological impairment, with intact performance on vocabulary and sentence repetition. Brynskov et al. (2017) argued that the findings indicated that syntactic and morphological impairments were a significant concern for high functioning children with autism and might be overlooked if language evaluation focused exclusively on vocabulary.

An even smaller number of studies have investigated the language abilities of bilingual children with ASD. Towards that end, Meir and Novogrodsky (2019) investigated the production of third-person subject and object pronouns in monolingual and bilingual children with HFA and TD children who were growing up in Israel. The bilingual children spoke Hebrew as a majority language and Russian as their Heritage Language. Furthermore, it evaluated the underlying linguistic and non-linguistic prerequisites of pronoun use by assessing the role of morphosyntactic skills, mentalizing skills (e.g., ToM abilities), and executive functioning (e.g., working memory and inhibition) on pronoun use. The participants were divided into four groups: HFA monolingual children, HFA bilingual children, monolingual TD children, and bilingual TD children. Their results yielded a robust effect of HFA on the use of pronouns and children with HFA scored lower in pronoun production compared to the TD groups. Morphosyntactic abilities of the children with HFA primarily predicted both third-person subject and object pronouns. Subject pronoun use was predicted by ToM skills and working memory, confirming that pronoun use is a complex phenomenon, which requires the integration of multiple linguistic and non-linguistic components. Meir and Novogrodsky (2019) concluded that morphosyntactic development was a prerequisite for third-person subject and object pronoun use in children with HFA, and ToM and working memory were involved in third-person subject pronoun use.

To date no studies have been conducted to investigate the grammatical abilities of Arabic-speaking children with autism. Given the overlap between the profile of children with ASD and children with SLI, to provide some background on the language difficulties Arabic children with ASD may have, we include two studies that investigated the grammatical abilities of Arabic children with SLI: Abdalla and Crago (2008) and Morsi (2009). Abdalla and Crago (2008) examined the ability of Arabic-speaking children with SLI compared to TD controls to produce finite verb inflections and explored the error patterns in the children's use of verbs in finite verb contexts (e.g., gender, number, person, and tense). The children were monolingual and spoke a dialect of Arabic known as Urban Hijazi Arabic (UHA) and consisted of three groups: children with SLI, TD MLU-equivalent controls and age-matched controls. The children completed four speech and language diagnostic measures that were adopted for Arabic because no standardized tests for language exist in Saudi Arabia: (a) a battery of language tests derived from the Test of Auditory Comprehension of Language Revised and the Preschool Language Scale; (b) an informal articulation test to assess various Arabic phonemes in three word positions using picture stimuli of familiar items; (c) cognitive screening tasks (association, matching, categorization, and problem solving); and (d) analysis of a spontaneous language sample to examine the types and variety of grammatical structures used (morphology, utterance length) as well as conversational discourse/pragmatics and lexical diversity. Abdalla and Crago (2008) found that the Arabic-speaking children with SLI who are acquiring UHA performed significantly different on the percentage of correct use of tense and subject-verb agreement forms from the TD MLU-equivalent controls and age-matched controls (Abdalla and Crago, 2008). They argued that the grammatical morphology is indeed an area that is vulnerable for the Arabic-speaking children with SLI, much as it is for children with SLI who speak a variety of other languages (Bedore & Leonard, 1998; Crago & Allen, 2001; de Jong, 1999; Fletcher, Leonard, Stokes, & Wong, 2005; Hansson, Nettelbladt, & Leonard, 2000; Leonard, 2000; Leonard, Camarata, Brown, & Camarata, 2004; Paradis & Crago, 2001; Rice, Noll, & Grimm, 1997; Rice, Wexler, & Cleave, 1995; Tager-Flusberg & Cooper, 1999).

Morsi (2009) investigated the tense, subject-verb agreement, noun-adjective agreement, sentence repetition and digit recall in monolingual Egyptian Arabic-speaking children with SLI. Morsi (2009) found that present tense verbs and sentence repetition were the areas of most difficulty for SLI children. The results showed that the predominant error type was using a default type form (imperative or imperative-like error) which substituted the present tense form. The percentage of errors of verb production showed that in general, for present tense verbs, tense was impaired, but agreement was not significantly affected, while for past tense verbs, neither tense nor agreement was greatly affected.

The literature reviewed above shows a lack of consensus on whether or not grammar is affected in children with autism, if so, which domains of grammar and whether NVIQ and general language abilities contribute to their grammatical profile. This is coupled with the scarcity of studies in languages other than English and a lack of implementing experimental methods that investigate grammatical development in more than one linguistic structure.

1.1 Present study

The present study addresses this gap by investigating the morphosyntactic abilities of Saudi Arabic-speaking children with autism using a new Saudi Sentence Repetition Task (Saudi-SRT) that was developed for this study. Sentence Repetition tasks involve listening to sentences and repeating them verbatim. To accurately repeat sentences, participants have to be able to process the sentences in terms of all levels of representation (phonological, morphosyntactic, semantic); then they have to extract its meaning and use their production system to regenerate the meaning of the sentence (Lombardi & Potter, 1992; Potter & Lombardi, 1990, 1998). There is a close relationship between accuracy in verbatim repetition and levels of representation related to the following abilities: (1) comprehension, (2) production, (3) store and retrieve language material from memory.

The task was developed in the Saudi spoken vernacular because this is the Arabic variety spoken within the home in Saudi families. Investigating Saudi Arabic in children with ASD is important not only because of its novelty since this has not been done before, but also because of the issue of diglossia and the specific structures of the Saudi Arabic language (see Methodology).

The first objective of the study was to determine whether Saudi Arabic-speaking children with autism have difficulties in morphosyntax as a function of their verbal abilities. For this reason, the participating children with autism were divided into subgroups of ALN children and ALI children in line with the classifications in the literature (Kjelgaard & Tager-Flusberg, 2001). This approach allows us to characterise precisely the severity of the deficit in morphosyntax found in children with autism relative to TD peers. The second objective was to identify whether the children's performance differs as a function of the scoring scheme used. Towards that end, two scoring schemes were used based on whether the children repeated sentences in a verbatim way versus whether they used the targeted structure accurately. The third objective was to address differences between the structures used in the Saudi-SRT. This will enable us to identify if specific structures are affected differently in children with ASD. These objectives gave rise to the following research questions:

1. Are there differences in performance on the Saudi-SRT between the ALN, ALI, and TD children?
2. Are the same differences attested in scorings based on verbatim repetition versus accuracy in using the target structure?
3. Are there differences between the structures used in the Saudi-SRT?

The methodology section below describes the participants, the tasks used, including the Saudi-Arabic Sentence Repetition task, as well as the study procedure.

2. Methodology

2.1 Participants

The study included 62 five- to seven-year-old children with Saudi Arabic as their first language and English as their second language. The sample comprised 20 children with autism and 42 TD control children. Children with autism were further divided into two equally sized groups based on their verbal abilities (ALN and ALI), which were measured using baseline tasks (see results section). This gave rise to three groups: 10 ALN children aged 5;1–6.4 ($M = 5.83$, $SD = .43$); 10 ALI children aged 5;6–6.2 ($M = 6.04$, $SD = .19$); and 42 TD control children aged 5;0–6.7 ($M = 5.80$, $SD = .72$). A one-way ANOVA with the group as a between factor (TD, ALI, ALN) showed no significant differences between them in age ($F(2, 59) = 1.688$, $p = .194$).

The children with autism were recruited from research centres and schools in Riyadh using the database of children with autism that was organised and administrated by the Charitable Society of Autism Families and enabled a search for children with autism who met the criteria. The TD children were recruited from mainstream schools in Riyadh. All parents gave consent before the commencement of the study. The study was reviewed by the Research Ethics Committee of the School of Psychology & Clinical Language Sciences, University of Reading, and was given a favourable ethical opinion for conduct.

2.2 Selection criteria

The selection criteria for the TD children were: (1) The child should not have a history of speech/language delay or disorder; (2) The child's age should fall within the age range of 5–7 years; (3) The child should not meet criteria for ASD at any point in their development, confirmation through parent questionnaire 'GARS'; (4) The child should not have a first-degree relative with an ASD diagnosis. Four selection criteria were used for the recruitment of children with ASD. The children: (1) The child should have had a documented ASD diagnosis made by a physician or psychologist specializing in autism before the age of 5, confirmation through parent questionnaire 'GARS'; (2) The child should meet criteria for ASD on the ADOS (both Social and Communication domains and total score) and according to best estimate clinical judgment; (3) The child should have not been diagnosed with any other disorder, and (4) The child should fall within the age range of 5–7 years. Only children fulfilling all four criteria were included in the study.

Within the autism group, children were divided into two subgroups based on their performance on verbal measures (PPVT) and a cut-off point of one standard deviation below the mean on the standard score. If the children scored more than one standard deviation below the mean on the PPVT they were included in the ALI group, whereas if they scored within the normal range, they were included in the ALN group.

3. Baseline tasks

A battery of baseline tasks was used to ascertain the children's verbal and nonverbal abilities. The Raven's CPM (Raven, Raven, & Court, 2000) was used to assess the children's nonverbal abilities. The adapted Arabic version of the PPVT (Dunn & Dunn, 2007) (adapted by Abu-Allam & Hadi, 1998) was used to measure their receptive vocabulary. The Arabic PPVT followed the design and rules of the original task but had been adapted for the Kuwaiti Arabic language and culture and standardized with Kuwaiti Arabic speaking children.

4. Experimental task: The Saudi-Arabic Sentence Repetition Task

The Saudi-SRT was developed based on the LITMUS SRep guidelines (Marinis & Armon-Lotem, 2015). These include the following two principles: (1) Include in all SRep tasks a set of syntactically complex structures that are difficult for children with SLI across languages and involve embedding and/or syntactic movement along with a set of syntactically simple structures as control structures (language-independent structures), and (2) Include a set of structures for each language that are difficult for children with SLI in the specific language (language-specific structures) (Marinis & Armon-Lotem, 2015).

The task comprised 48 sentences that contained a range of eight different grammatical structures: object wh-questions, topicalisations, subject relative clauses, and object relative clauses (language-independent structures), past tense, present tense, accusative clitic pronouns, and sentential complements (language-specific structures). The syntactic features of the task were morphologically balanced in gender and tense, phonologically balanced in sentence length on the basis of the number of words and syllables, and syntactically in the word order verb-subject. See examples of the sentences in (1) to (8) below.

(1) Past tense
ʔakal-at al-batˤtˤa xubz fii al-buħajra
Eat.PFV-1SG.F DET-duck bread at DET-lake
'The duck ate bread at the lake.'

(2) Present tense
ta-qraʔ atˤ-tˤaalib-a al-qisˤsˤah fii al-lajl
3SG.F-read.PFV DET-student-SG.F DET-story. at DET-night
'The student is reading the story at night.'

(3) Object Wh-questions
maaðaa rasamat al-ʔamiir-a aʃ-ʃahr al-maadˤii
what draw.PFV-3SG.F DET-princess-SG.F DET-month DET-last
'What did the princess draw last month?'

(4) Topicalisations
dʒazarun ʔaxaða al-ʔarnab fii asˤ-sˤaba:ħ
carrot take PFV-1SG.F DET-rabbit in DET-morning
'A carrot, the rabbit took (it) in the morning.'

(5) Object clitic pronouns
aʃ-ʃantˤata dʒadda-tii ʔadˤaaʕ-at fii al-madʒlis
DET-bag grandmother-1SG.F.POSS lost-1SG.F in DET-living room
'The bag, my grandmother lost (it) in the living room.'

(6) Sentential complements
ðˤannat al-ʔum anna al-ʔarnab ʔakala al-mawzah
think.PFV-3SG.F DET-mother that DET-rabbit eat DET-banana
'The mother thought that the rabbit ate the banana.'

(7) Subject relative clauses
raʔaa ar-radʒul allaðii sˤawwar ad-duktuur
see.PFV.3SG.M DET-man who photograph. PFV-3SG.M DET-doctor
'He saw the man who photographed the doctor.'

(8) Object relative clauses
raʔat ad-dumj-a allatii ʔa-ħdˤarat-ha al-fataat
3SG.F-see.PFV DET-doll-F that 3SG.N-bring-3PL.N DET-girl
'She saw the doll that the girl brought.'

For past tense, the response had to contain the same tense as the target sentence and the verb-subject agreement for number and gender; items in the target sentence also had to contain the same basic order (VSO) and had to contain a main verb, a subject and an object in the right positions in order to be scored as correct; the same applied for present tense. For wh-object questions, the response had to contain a fronted wh-question word followed by verb-subject inversion; the response needed to contain the appropriate wh-word and a subordinate clause in order to be scored as correct in terms of the target structure. In topicalizations, the response had to contain a fronted object followed by the verb and the subject. For clitic accusative pronouns, the response had to contain the same structure as the target OSV followed by the pronoun. In sentences targeting sentential complements, for variable tense the response sentence had to contain one main clause with the conjunction (i.e., that) and had to be accompanied by a main and a subordinate clause which had to be correctly marked for tense; for invariable tense, the response sentence had to contain the main clause after the first main verb. Sentences with relative clauses as the target structure were required to contain a relative pronoun and a subordinate clause that modified the target antecedent (i.e., subject vs. object relative clause). For subject relative clauses, a relative clause had to be present modifying a subject. For object relative clauses, a relative clause had to be present modifying an object. The relative pronoun could be dropped from the sentences.

In this task, children listened to a series of sentences in spoken Saudi Arabic. A native speaker of Saudi Arabic pre-recorded the sentences. The sentences were incorporated into a PowerPoint presentation to ensure that all children listened to the sentences in the same way. The sentences were embedded in a helping game to make the task more engaging. A man was shown in a PowerPoint presentation going through several roads to reach home (see Figure 1). The administration time was 30 minutes.

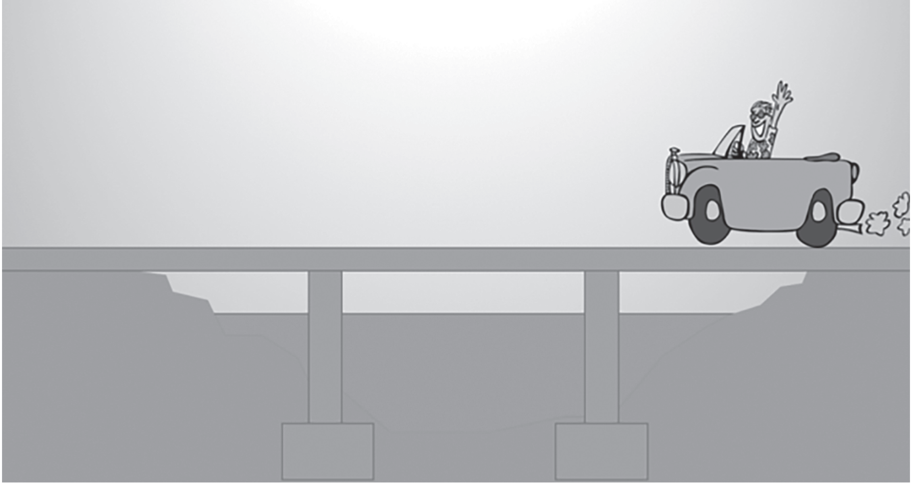


Figure 1. Example from Saudi-SRT PowerPoint with the man driving along his path

In this task, the car moved from one point to another. Children heard each time a sentence and had to repeat it verbatim to enable the car to move. That proved to be very motivating for children who were eager to repeat the sentences to find out where the car would go and how it would reach home. The sentences were presented through headphones to ensure excellent and consistent quality of input.

External speakers were used for presentation for the few children who objected to using the headphones. The children's responses were recorded with a high-quality microphone to allow detailed scoring after testing.

The children were seated in front of a computer screen showing the PowerPoint presentation. They were instructed to listen carefully to the sentences and repeat what they heard. A practice session presented two sentences to demonstrate the task to the children and to ensure that the children understood what they had to do. If the children did not repeat the sentences spontaneously, the experimenter explained the task again and encouraged them to repeat what they heard. The practice sentences could be repeated several times until the children understood what they had to do. Most children understood the task quickly and easily.

For the experimental sentences, children listened to each sentence once. The experimenter would repeat the sentence if there were a loud noise in the room or another interruption. Verbal praise was used irrespective of the children's performance to motivate them to continue. If the child self-corrected, her final response was scored (whether correct or incorrect). The children's responses were audio-recorded and marked on an answer sheet during the testing. They were further transcribed and scored off-line based on the recording.

4.1 Scoring

The transcribed responses were scored using several scoring schemes. In this paper, we present the results from the scoring based on the test of language development (TOLD) (Newcomer & Hammill, 2008) and the structural scoring that assesses accuracy in the particular structure targeted (Marinis & Armon-Lotem, 2015). The TOLD scoring allocates a score of 1 if the sentence was repeated entirely verbatim and a score of 0 if there were one or more changes from the original in the child's response (Newcomer & Hammill, 2008). The structural scoring allocated a score of 1 if the child used the target structure accurately, irrespective of whether or not there were changes in other parts of the sentence. A score of 0 was given if the child made an error in the target sentence structure, e.g., changing the tense of a verb in sentences targeting past tense (see Table 1).

Table 1. Examples of the two scoring schemes for the Saudi-SRT task

Scoring scheme	***Response***	***Analysis***	***Score***
TOLD	asir, ʃarab al-ləban fi l-kaas	repeated verbatim	1
	asir fi al-d͡ʒubin al-kaas	repeated non-verbatim	0
Structural Scoring	asir, ʃarab al-ləban fi l-kaas	correct structure	1
	ʃarab asir al-ləban fi l-kaas	incorrect structure	0

4.2 Procedure

Children were assessed individually in a quiet room in their schools or centres. The data were collected by the researcher, who is a native speaker of Saudi Arabic and a fluent speaker of English. The data collection for the ALI and ALN children was divided into four sessions of around 20 minutes each. The test battery for the TD children lasted approximately 60 minutes and was divided into two sessions of around 30 minutes each. The first session consisted of measuring the children's NVIQ and Arabic receptive vocabulary. The second session consisted of only Saudi Arabic grammar. The participating children were allowed short breaks between tasks if desired. The children were rewarded for their participation with a badge and a certificate after finishing the sessions.

5. Results

5.1 Baseline tasks

Table 2 shows the descriptive statistics for the baseline tasks.

Table 2. Descriptive statistics for the baseline tasks

Group		*Raven's raw score*	*Arabic PPVT raw score*	*Arabic PPVT standard score*	*Arabic PPVT z-score*
TD	Mean	15.86	20.86	103.90	.255
	SD	4.88	15.12	23.6	1.58
	Range	18	72	92	6.1
	Min-Max	8–26	5–77	68–160	−2.1–4.0
ALN	Mean	14.60	15.30	95.50	−.293
	SD	1.83	4.24	7.5	.4964
	Range	6	14	25	1.7
	Min-Max	11–17	8–22	85–100	−1.0–.7
ALI	Mean	11.40	8.10	70.70	−1.940
	SD	2.17	5.15	5.71	.3816
	Range	6	16	14	.9
	Min-Max	8–14	1–17	64–78	−2.4––1.5

Note. TD = Typically Developing, ALN = Children with autism with normal language, and ALI = Children with autism with language impairment.

A one-way ANOVA with the children's score on the Raven's as the dependent variable and Group (TD, ALN, and ALI) as the independent variable showed a significant difference between the groups in their nonverbal abilities ($F(2, 59) = 4.56$, *p* = .014), see Table 3. A Tukey post hoc test revealed that the children with ALI scored significantly lower than the TD children ($p < .001$) and the children with ALN ($p = .006$). However, there was no significant difference in performance between the TD children and the ALN children ($p = .391$).

Table 3. One-way analysis of variance of the Raven's raw score

Source	*df*	*SS*	*MS*	*F*	*P*
Between Groups	2	161.799	80.900	4.563	.014
Within Groups	59	1045.943	17.728		
Total	61	1207.742			

The same pattern emerged in the children's receptive vocabulary. A separate one-way ANOVA with each score of the Arabic PPVT as the dependent variable and Group (TD, ALN, and ALI) as the independent variable showed a significant difference between the groups in all three scores: raw score ($F(2, 59) = 4.21$, $p = .020$), standard score ($F(2, 59) = 11.16$, $p < .001$), and *z*-score ($F(2, 59) = 10.98$, $p < .001$), see Table 4. A Tukey post hoc test for each one of the three scores showed that the children with ALI scored significantly lower than the TD children (all $p < .02$) and the children with ALN (raw: $p = .009$; standard: $p = .020$; *z*-score: $p = .021$), but there was no significant difference between the TD children and the ALN children (raw: $p = .108$; standard: $p = .462$; *z*-score: $p = .476$).

Table 4. One-way analysis of variance of the Arabic peabody picture vocabulary test

Scoring	Source	*df*	*SS*	*MS*	*F*	*P*
Raw Score	Between Groups	2	1395.279	697.638	4.211	.020
	Within Groups	59	9774.143	165.663		
	Total	61	11169.419			
Standard Score	Between Groups	2	8939.458	4469.729	11.165	.000
	Within Groups	59	23620.219	400.343		
	Total	61	32559.677			
Zscore	Between Groups	2	39.040	19.520	10.986	.000
	Within Groups	59	104.832	1.777		
	Total	61	143.872			

5.2 The Saudi-Arabic sentence repetition task

Table 5 shows the descriptive statistics for the two scorings of the Saudi SRT.
A one-way ANOVA with each scoring scheme as the dependent variable and Group (TD, ALN, and ALI) as the independent variable showed a significant difference

Table 5. Descriptive statistics for Saudi-SRT

Group		*TOLD scoring*	*Structural scoring*
TD	Mean	.362	.602
	SD	.2284	.2076
	Range	.8	.7
ALN	Mean	.230	.490
	SD	.0823	.1287
	Range	.3	.4
ALI	Mean	.098	.310
	SD	.0948	.1101
	Range	.3	.3

Note. Saudi-SRT = Saudi Sentence Repetition Task.

between the groups in both scoring schemes: TOLD scoring ($F(2, 59) = 7.99$, $p = .001$; structural scoring ($F(2, 59) = 6.26$, $p = .003$), see Table 6. A Tukey post hoc for the TOLD scoring revealed higher accuracy for the TD children compared to the ALN ($p = .012$) and the ALI children ($p < .001$). The ALN children had a higher accuracy than the ALI children ($p = .010$). In contrast, a Tukey post hoc for structural scoring revealed higher accuracy for the TD children compared to the ALI children ($p = .013$) but no significant difference between the TD children and the ALN children ($p > .05$). The ALN children had a higher scoring than the ALI children ($p = .04$).

Table 6. One-way analysis of variance of the TOLD and structural scoring for the Saudi-SRT

Scoring	Source	*df*	*SS*	*MS*	*F*	*P*
TOLD	Between Groups	2	.618	.309	7.989	.001
	Within Groups	59	2.281	.039		
	Total	61	2.899			
Structural Scoring	Between Groups	2	.122	.061	6.262	.003
	Within Groups	59	.575	.010		
	Total	61	.697			

Note. TOLD = the test of language development.
Saudi-SRT = Saudi Sentence Repetition Task.

To investigate differences between the three groups and the eight structures, the data were analysed for each structure and each scoring scheme separately. Figure 2 shows the results of the eight structures based on the TOLD scoring, and Figure 3 shows the results of the eight structures based on the structural scoring.

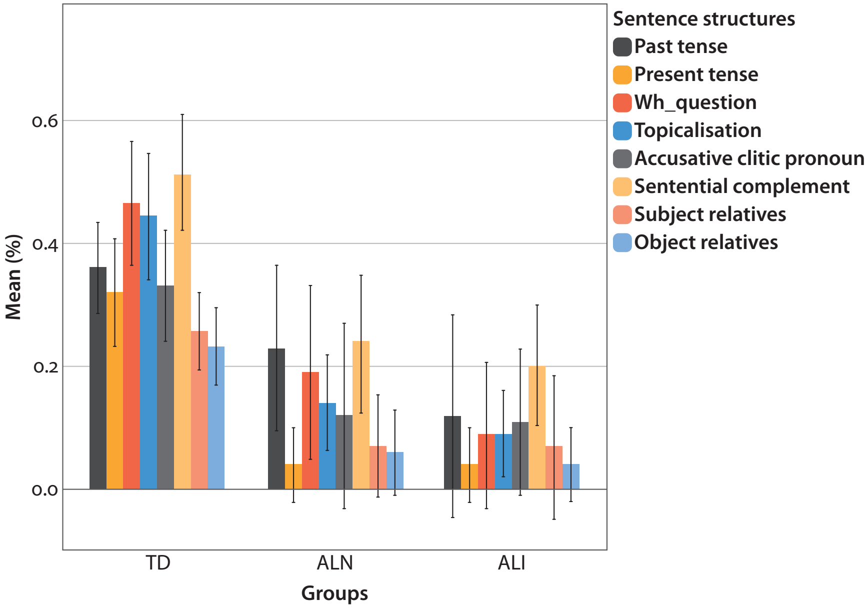


Figure 2. Performance by structure on the TOLD scoring

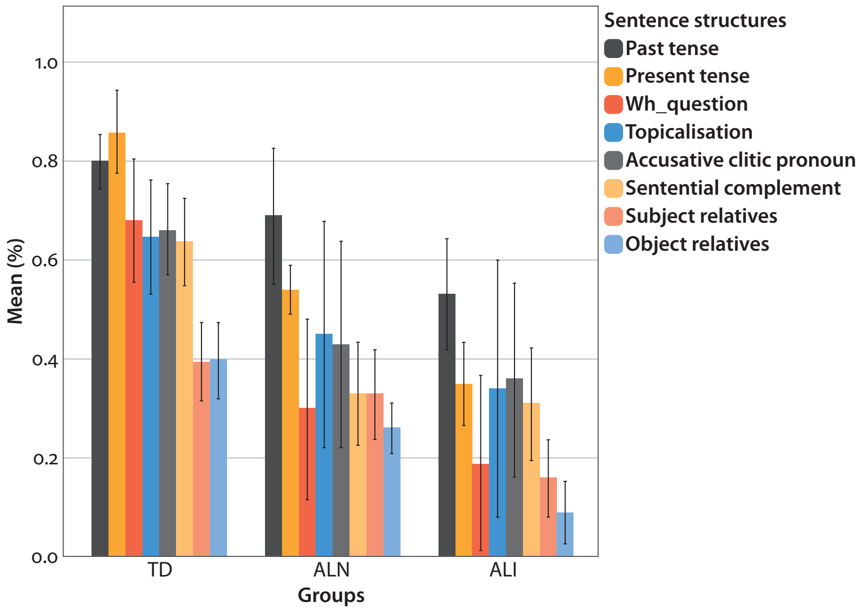


Figure 3. Performance by structure on the structural scoring

To identify differences between the groups and structures in the TOLD scoring, a repeated measure ANOVA was conducted with the TOLD scoring as the dependent variable, the group (TD, ALN, and ALI) as the between-subjects variable, and sentence structure (past tense, present tense, wh-questions, topicalisations, accusative clitic pronouns, sentential complements, subject relative clauses, and object relative clauses) as the within-subjects variable. The Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2(27) = 41.84$, $p = .035$; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .82$). This showed a significant difference between the groups ($F(2, 59) = 11.32$, $p = .000$, $\eta^2 = .28$) and a significant difference between the sentence structures ($F(5.730, 338.072) = 7.76$, $p = .000$, $\eta^2 = .116$), but no significant interaction between group and structures ($F(11.460, 338.072) = 1.238$, $p = .258$, $\eta^2 = .040$), see Table 7 and 8.

Table 7. Repeated measure analysis of variance of the TOLD between groups

Source	***df***	***SS***	***MS***	***F***	***P***
Intercept	1	1.586	1.586	42.368	.000
Three Groups	2	.848	.424	11.328	.000
Error	59	2.209	.037		

Note. TOLD = the test of language development.

Table 8. Repeated measure analysis of variance of the TOLD

Source	***df***	***SS***	***MS***	***F***	***P***
Sentence Structures	5.730	1.406	.245	7.757	.000
	338.072	10.693	.032		
Groups * Structures	11.460	.449	.039	1.238	.258
	338.072	10.693	.032		

Note. TOLD = the test of language development.

To identify differences between the sentence structures, a Tukey post hoc test was conducted for all three groups together because there was a lack of interaction between group and structure. This revealed that across the three groups children were more accurate in past tense compared to subject relative clauses and object relative clauses (both $p < .003$) and in wh-questions compared to present tense, accusative clitic pronouns, subject relative clauses and object relative clauses (all $p < .02$). They were also more accurate in topicalisations compared to present tense, subject relatives and object relative clauses (all $p < .02$), and in sentential complements compared to past tense, present tense, accusative clitic pronouns, subject relative clauses and object relative clauses (all $p < .006$).

A similar analysis was conducted for the structural scoring. The Mauchly's test indicated that this assumption of sphericity had also been violated, χ^2 (27) = 56.85, p = .000; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = .792). The analysis showed significant differences between the performance of the groups (F(2, 59) = 10.89, p = .000, η^2 = .270) as well as between the sentence structures (F(5.54, 327.25) = 17.64, p = .000, η^2 = .230). There was also a significant interaction between group and structures (F(11.09, 327.25) = 2.30, p = .010, η^2 = .072), see Table 9 and 10.

Table 9. Repeated measure analysis of variance of the structural scoring between groups

Source	*df*	*SS*	*MS*	*F*	*P*
Intercept	1	8.140	8.140	223.530	.000
Three Groups	2	.793	.397	10.892	.000
Error	59	2.149	.036		

Table 10. Repeated measure analysis of variance of the structural scoring

Source	*df*	*SS*	*MS*	*F*	*P*
Sentence Structures	5.54	6.745	1.216	17.64	.000
	327.25	22.551	.069		
Groups * Structures	11.09	1.754	.158	2.294	.010
	327.25	22.551	.069		

To identify the source of the interaction, one-way ANOVAs with the group as an independent variable were conducted for each structure separately. This showed no significant differences between the three groups on past tense and accusative clitic pronouns (both p >.05). However, there were differences in present tense, wh-questions, topicalisations, sentential complements, subject relative clauses, and object relative clauses (all p < .002). In the present tense, wh-questions, and sentential complements, the ALI and ALN children had lower accuracy than the TD children (all p < .05). In subject relative clauses and object relative clauses, the ALI children were less accurate compared to the ALN and the TD children (both p < .05), while there was no difference between the ALN and the TD children (both p > .05). In topicalisations, the ALI children performed less well compared to the TD children (p = .047).

A comparison between the structures of the sentences for each group separately revealed that the TD children performed equally well in past tense and present tense, and they were more accurate in these structures compared to wh-questions, topicalisations, accusative clitic pronouns, sentential complements, subject relative clauses, and object relative clauses (all p < .05). In contrast, they performed equally well in subject relative clauses and object relative clauses. The ALN children were more accurate in past tense and present tense compared to topicalisations and accusative clitic pronouns (both $p < .05$). Moreover, they performed better in sentential complements and subject relative clauses compared to wh-questions (both $p < .001$). The ALI children were more accurate in past tense compared to wh-questions, subject relative clauses, and object relative clauses (all $p < .04$). Moreover, they were less accurate in subject relative clauses and object relative clauses compared to all other sentence structures ($p < .05$).

6. Discussion

This is the first study to investigate the grammatical abilities of children with autism in Saudi Arabia. The overarching aim was to document the strengths and weaknesses of the morphosyntactic abilities in Saudi-speaking children with autism. The first objective was to determine whether Saudi-speaking children with autism have difficulties in morphosyntax as a function of their verbal abilities. For this reason, children with autism were divided into subgroups of ALN and ALI children. The second objective was to identify whether the children's performance differs as a function of the scoring scheme, i.e., a verbatim scoring scheme and a structural scoring scheme. The third objective was to address differences between the structures used in the Saudi-SRT.

In order to determine whether Saudi-speaking children with autism have difficulties in morphosyntax as a function of their verbal abilities, we divided the children with ASD into two subgroups based on their performance on the PPVT. Children whose standard score on the PPVT was more than one standard deviation below the mean were included in the ALI group and those who scored within the normal range were included in the ALN group. As a result, the ALI children scored lower than the ALN children and the TD children on the PPVT and there was no difference between the ALN children and the TD children on this task.

The main findings of this study can be summarised as follows (see also Table 11). Firstly, the ASD children had difficulties in morphosyntax as a function of their verbal abilities, but their scores were modulated via the scoring scheme. In terms of verbatim repetition, both the ALN and ALI children were less accurate in the Saudi-SRT compared to the TD children and no significant difference was attested between the ALN and the ALI group; however, on the structural scoring scheme, the ALN children performed similarly to the TD children and better than the ALI children. Secondly, the two scoring schemes showed a different pattern of performance between the groups and structures. One first difference is that the verbatim scoring did not show an interaction between group and structure, indicating that the pattern of performance in the structures was similar in the three groups, whereas the structural scoring showed a group by structure interaction, indicating that the groups showed a different pattern of performance for the different structures. This difference between the two scoring schemes suggests that the verbatim scoring scheme is less sensitive in identifying differences between the groups than the structural scoring scheme.

The verbatim scoring scheme showed that across the three groups children were more accurate in past tense compared to subject relative and object relative clauses and in wh-questions compared to present tense, accusative clitic pronouns, subject and object relative clauses. They were also more accurate in topicalisations compared to present tense, subject and object relative clauses, and in sentential complements compared to past tense, present tense, accusative clitic pronouns, subject and object relative clauses.

In contrast to the verbatim scheme, the structural scoring scheme showed that the three groups had a different pattern of performance and importantly, also the two groups of children with ASD performed differently from each other. In terms of similarities, the three groups performed similarly on past tense and accusative clitic pronouns. In all other structures, the ALI children performed less well than the TD children. The ALN children performed similarly to the ALI children and less well than the TD children in present tense, wh-questions, topicalisations, and sentential complements. However, in subject and object relative clauses the ALN children were more accurate than the ALI children and on a par with the TD children. When we looked at each group separately, the TD children were more accurate in past tense and present tense compared to wh-questions, topicalisations, accusative clitic pronouns, sentential complements, subject, and object relative clauses. They performed equally well in subject and object relative clauses. The ALN children were more accurate in past tense and present tense than they were on topicalisations and accusative clitic pronouns. They also performed better on sentential complements and subject relative clauses compared to wh-questions. The ALI children were more accurate in past tense compared to wh-questions, subject, and object relative clauses. They were also less accurate in subject and object relative clauses compared to all sentence structures.

These are novel and significant findings both for the understanding of the way ASD manifests itself in Saudi Arabic speaking children in Saudi Arabia as well as for the understanding of the grammatical abilities of children with autism irrespective of the language they acquire. The results indicate that subgroups of children with autism vary in their performance compared to TD children, depending on the structures tested and the way the structures were scored. The Saudi-SRT was designed with eight structures, starting from morphologically complex but syntactically simple structures (past tense and present tense) to structures that are both

Table 11. Summary of main findings

1	**Analyses of all data, TOLD Scoring:** ALI = ALN < TD children **Analyses of all data, Structural scoring:** ALI < ALN = TD children
2	**Analyses by sentence type: TOLD Scoring, all groups:** – Past tense > subject and object relative clauses – Wh-questions > present tense, accusative clitic pronouns, subject and relative clauses. – Topicalisations > present tense, subject and object relative clauses. – Sentential complements > past tense, present tense, accusative clitic pronouns, subject and object relative clauses
3	**Analyses by sentence type: Structural Scoring, differences between groups:** – ***ALI = ALN = TD:*** past tense and accusative clitic pronouns. – ***ALI = ALN < TD:*** present tense, wh-questions, topicalisations, and sentential complements. – ***ALI < ALN = TD:*** subject and object relative clauses.
4	**Analyses by sentence type: Structural scoring, differences within groups:** – ***TD children:*** past tense and present tense > wh-questions, topicalisations, accusative clitic pronouns, sentential complements, subject = object relative clauses. – ***ALN children:*** past tense and present tense > topicalisations and accusative clitic pronouns; sentential complements and subject relative clauses > wh-questions. – ***ALI children:*** past tense > wh-questions, subject and object relative clauses; subject and object relative clauses < all other sentence structures.

morphologically and syntactically complex (subject relatives and object relatives). This enabled us to address differences between the groups based on morphological complexity, syntactic complexity or a combination of the two.

Morphological complexity on its own did not seem to pose significant difficulties in any of the groups but this could relate to the morphological manipulations in the specific task. Children performed better in the structures eliciting past and present tense compared to all other sentences and children with autism seemed to perform less well than TD children only in present tense forms. This was in contrast to syntactic complexity that impacted all three groups of children and more severely children with ALI. Subject and object relative clauses were the most challenging structures for the ALI children and were the structures that differentiated the ALI from the ALN children.

The relationship between NVIQ and language has long been highlighted in previous studies. Kjelgaard and Tager-Flusberg (2001) stated that 'IQ itself accounts for some of the heterogeneity found in language among the children with autism' and Roberts et al. (2004) highlighted the importance of matching groups in terms of nonverbal abilities, even when the groups were matched for their verbal abilities. Deficits in morphosyntax are not universal in autism in contrast to the universal impairments found in communication skills in this population that are among the defining features of autism, such as echolalia (Tager-Flusberg & Calkins, 1990). Our findings demonstrate that there are fundamental insights when children with autism are grouped based on their verbal abilities and when more sensitive scoring procedures are applied that are based on scoring only the grammatical abilities independently from the overall ability to repeat a sentence in a verbatim way. We argue that due to the significant heterogeneity among children with autism, the detailed linguistic profile in the population with autism is crucial to detect morphosyntax difficulties for that population.

Many previous studies on the grammatical abilities of children with autism investigated a single domain of grammar, e.g. third-person singular and past tense morphology in English (Roberts et al., 2004), binding in English (Janke & Perovic, 2015; Perovic et al., 2013a, 2013b), accusative clitics in French (Durrleman & Delage, 2016; Prévost et al., 2018). Terzi et al. (2014, 2016a, 2016b) included more than one structures in Greek, for example, binding and accusative clitics as well as passives and interfaces between syntax and discourse, but also in those studies the number and range of structures tested is limited compared to the structures included in the present study. Therefore, in previous studies, it is difficult to compare and identify structures that are particularly vulnerable in groups of ALN and ALI children within the same study. In the remainder of this section, we will compare our findings to findings from previous studies addressing similar structures in other languages.

The acquisition of tense was addressed in Roberts et al. (2004) who found that children with autism showed high rates of omission on both third-person singular and past tense, and children with autism and ALI performed worse than children with SLI in the same verbal IQ range. This indicates difficulties in morphosyntactic complexity for English-speaking children with autism. The verbatim and structural scoring schemes in our study also showed some difficulties in present tense morphology in children with ALN and ALI. However, these difficulties were less severe than difficulties with structures that have increased syntactic complexity.

The acquisition of clitics was investigated in Durrleman and Delage (2016); Prévost et al. (2018), and Terzi at al. (2014, 2016a, 2016b). Durrleman and Delage (2016) found similar deficits for both ASD and SLI participants on the production of accusative clitics, which they argued to be unrelated to their nonverbal abilities. Their results are relevant for the debate around the existence of language disorders in those with ASD that would either resemble SLI (Roberts et al., 2004) or slightly resemble nonspecific language impairment resulting from the general cognitive ability (Tager-Flusberg & Calkins, 1990). Prévost et al. (2018, pp. 5–13) found similarities in the production of accusative clitics between the children with ALI and the children with SLI in contrast to the children in the ALN group, who performed similarly to the TD children. Interestingly, although the structural scoring did not show differences between the groups on the production of accusative clitics, the ALN children had lower performance in accusative clitics compared to present and past tense, and the ALI children had better performance on accusative clitics compared to subject and object relative clauses. This reflects the nature of syntactic complexity of accusative clitics compared to past and present tense as well as subject and object relative clauses. Past and present tense are only morphologically complex, whereas accusative clitics are both morphologically and syntactically complex. Subject and object relative clauses have higher syntactic complexity than accusative clitics.

The previous two studies that addressed the grammatical abilities of Arabic children with SLI investigated the finite verb inflections and explored the error patterns in the children's use of verbs in finite verb contexts (e.g., gender, number, person, and tense) (Abdalla & Crago, 2008), and the tense, subject-verb agreement, noun-adjective agreement, sentence repetition and digit recall (Morsi, 2009). Abdalla and Crago (2008) found that children with SLI had problems with tense and subject-verb agreement and indicates that these are vulnerable domains in Arabic-speaking children with SLI. Morsi (2009) also found that children with SLI showed areas of difficulties in present tense verbs, in which tense was impaired, but agreement was not significantly affected. That indicates difficulties in present tense for the Arabic-speaking children with SLI but not in past tense. Our findings demonstrate similarities between the three groups on past tense, a structure that does not seem to be affected in Arabic-speaking children with ASD. In contrast, present tense inflection seems to be vulnerable in children with ASD because both the ALN and ALI children were less accurate in the production of present tense forms, especially in the structural scoring. This is in line with the study by Morsi (2009) and demonstrates that present tense in Arabic is vulnerable not only for children with SLI but also for children with autism. This is expected for children with ALI but somewhat surprising for children with ALN because their language abilities do not differ from those of TD children. The question of why there is a vulnerability in present and not past tense in children with SLI as well as children with ASD cannot be answered on the basis of the data presented in this study, and therefore, it is open for future research.

The findings from the present study shed light on the language abilities of children with autism in Saudi Arabia within the domain of morphosyntax. Further research should identify the error patterns in the structures investigated in the present study and relate these findings to a broader range of syntactic and pragmatic measures as well as ToM and verbal and nonverbal executive functions among children with autism. Such research will help to advance our understanding of the nature of the language deficits attested across a range of children with autism.

Acknowledgements

We would like to thank all the participating children, centres, schools, families, and speech and language therapists for allowing us to recruit the TD and children with autism. Earlier parts of this work were presented at the 3rd meeting of the Experimental Arabic Linguistics Conference (EXAL) at the United Arab Emirates University (October 2018), at the Bilingualism and Language Impairment (Bi-SLI) conference at the University of Reading (June 2018), and the Child Language Symposium (CLS) conference in Reading (June 2018). We would like to thank the audiences from these events for their interest, comments, and suggestions. At the time this study was conducted, TM was supported by the Language Abilities in Children with Autism (LACA) project, funded by the NWO Scheme Internationalisation in the Humanities.

References

Abdalla, F., & Crago, M. (2008). Verb morphology deficits in Arabic-speaking children with specific language impairment. *Applied Psycholinguistics*, 29(2), 315–340. https://doi.org/10.1017/S0142716408080156

Abu-Allam, R., & Hadi, F. (1998). *Picture vocabulary test*. Centre for Childhood, Evaluation, and Teaching, Kuwait.

Bartak, L., Rutter, M., & Cox, A. (1977). A comparative study of infantile autism and specific developmental receptive language disorders III. Discriminant function analysis. *Journal of Autism and Childhood Schizophrenia*, 7(4), 383–396. https://doi.org/10.1007/BF01540396

Bartolucci, G. (1982). Formal aspects of language in childhood autism. *Advances in Child Behavior Analysis and Therapy*, 2, 159–185.

Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, 41, 1185–1192. https://doi.org/10.1044/jslhr.4105.1185

Brynskov, C., Eigsti, I. M., Jørgensen, M., Lemcke, S., Bohn, O. S., & Krøjgaard, P. (2017). Syntax and morphology in danish-speaking children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 47(2), 373–383. https://doi.org/10.1007/s10803-016-2962-7

Colle, L., Baron-Cohen, S., Wheelwright, S., & Van Der Lely, H. K. J. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*, 38(1), 28–40. https://doi.org/10.1007/s10803-007-0357-5

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language Impairment (SLI). *Journal of Speech, Language, and Hearing Research*, 46(5), 1029–1037. https://doi.org/10.1044/1092-4388(2003/082)

Crago, M., & Allen, S. (2001). Early finiteness in Inuktitut: The role of language structure and input. *Language Acquisition*, 9, 59–111. https://doi.org/10.1207/S15327817LA0901_02

de Jong, J. (1999). *Specific language impairment in Dutch: Inflectional morphology and argument structure*. Enschede, The Netherlands: Print Partners Ipskamp.

Dunn, L., & Dunn, M. (2007). *PPVT-4: Peabody picture vocabulary test*. Minneapolis, MN: Pearson Assessments.

Durrleman, S., & Zufferey, S. (2009). The nature of syntactic impairment in autism. *Rivista di Grammatica Generativa*, 34, 57–86.

Durrleman, S., & Delage, H. (2016). Autism spectrum disorder and specific language impairment: Overlaps in syntactic profiles. *Language Acquisition*, 23(4), 361–386. https://doi.org/10.1080/10489223.2016.1179741

Durrleman, S., Delage, H., Prévost, P., & Tuller, L. (2017). The comprehension of passives in autism spectrum disorder. *Glossa: A Journal of General Linguistics*, 2(1), 88. https://doi.org/10.5334/gjgl.205

Eigsti, I. M., Bennetto, L., & Dadlani, M. B. (2007). Beyond pragmatics: Morphosyntactic development in autism. *Journal of Autism and Developmental Disorders*, 37(6), 1007–1023. https://doi.org/10.1007/s10803-006-0239-2

Eigsti, I. M., & Bennetto, L. (2009). Grammaticality judgments in autism: Deviance or delay. *Journal of Child Language*, 36(05), 999–1021. https://doi.org/10.1017/S0305000909009362

Fletcher, P., Leonard, L., Stokes, S., & Wong, A. (2005). The expression of aspect in Cantonese-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 44, 621–634. https://doi.org/10.1044/1092-4388(2005/043)

Hansson, K., Nettelbladt, U., & Leonard, L. (2000). Specific language impairment in Swedish: The status of verb morphology and word order. *Journal of Speech, Language, and Hearing Research*, 43, 84–86. https://doi.org/10.1044/jslhr.4304.848

Hobson, R. P., Lee, A., & Hobson, J. A. (2010). Personal pronouns and communicative engagement in autism. *Journal of Autism and Developmental Disorders*, 40(6), 653–664. https://doi.org/10.1007/s10803-009-0910-5

Janke, V., & Perovic, A. (2015). Intact grammar in HFA? Evidence from control and binding. *Lingua*, 164, 68–86. https://doi.org/10.1016/j.lingua.2015.06.009

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes*, 16 (2–3), 287–308. https://doi.org/10.1080/01690960042000058

Leonard, L. (2000). Specific language impairment across languages. In D. V. M. Bishop & L. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome* (pp. 115–129). Hove: Psychology Press.

Leonard, L. (2014). Specific language impairment across languages. *Child Development Perspectives*, 8(1), 1–5. https://doi.org/10.1111/cdep.12053

Leonard, L., Camarata, S., Brown, B., & Camarata, M. (2004). Tense and agreement in the speech of children with specific language impairment: Patterns of generalization through intervention. *Journal of Speech, Language, and Hearing Research*, 47, 1363–1379. https://doi.org/10.1044/1092-4388(2004/102)

Lewis, S., & Landau, B. (2015). Language development in children with Williams syndrome: genes, modularity, and the importance of development. In E. Bavin and L. Naigles (Eds.), *Cambridge handbook of child language*, 2nd Edn. (pp. 609–636). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316095829.028

Lombardi, L., & Potter, M. (1992). The regeneration of syntax in short term memory. *Journal of Memory And Language*, 31(6), 713–733. https://doi.org/10.1016/0749-596X(92)90036-W

Lord, C., & Paul, R. (1997). Language and communication in autism. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive development disorders*, 2nd ed. (pp. 195–225). New York: John Wiley.

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. *Multilingual Matters*, 1–32.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders*, 38(4), 325–350. https://doi.org/10.1080/1368282031000154204

Meir, N., & Novogrodsky, R. (2019). Prerequisites of third-person pronoun use in monolingual and bilingual children with autism and typical language development. *Frontiers in Psychology*, 10, 1–14. https://doi.org/10.3389/fpsyg.2019.02289

Modyanova, N., Perovic, A., & Wexler, K. (2017). Grammar is differentially impaired in subgroups of autism spectrum disorders: Evidence from an investigation of tense marking and morphosyntax. *Frontiers in Psychology*, 8, 1–23. https://doi.org/10.3389/fpsyg.2017.00320

Morsi, R. (2009). Specific language impairment in Egyptian Arabic A preliminary investigation. Ph.D. dissertation. University of Reading, UK.

Newcomer, P., & Hammill, D. D. (2008). *Test of language development: Primary, 4th edition*. Austin, TX: Pro-ed.

Paradis, J., & Crago, M. (2001). The morphosyntax of specific language impairment in French: An extended optional default account. *Language Acquisition*, 9, 269–300. https://doi.org/10.1207/S15327817LA0904_01

Park, C. J., Yelland, G. W., Taffe, J. R., & Gray, K. M. (2012). Morphological and syntactic skills in language samples of pre-school aged children with autism: atypical development? *International Journal of Speech-language Pathology*, 14(2), 95–108. https://doi.org/10.3109/17549507.2011.645555

Perovic, A., Modyanova, N., & Wexler, K. (2013a). Comprehension of reflexive and personal pronouns in children with autism: A syntactic or pragmatic deficit? *Applied Psycholinguistics*, 34(4), 813–835. https://doi.org/10.1017/S0142716412000033

Perovic, A., Modyanova, N., Wexler, K., & Naigles, L. R. (2013b). Comparison of grammar in neurodevelopmental disorders: The case of binding in Williams Syndrome and autism with and without language impairment. *Language Acquisition*, 20(2), 133–154. https://doi.org/10.1080/10489223.2013.766742

Pierce, S., & Bartolucci, G. (1977). A syntactic investigation of verbal autistic, mentally retarded, and normal children. *Journal of Autism and Childhood Schizophrenia*, 7(2), 121–134. https://doi.org/10.1007/BF01537724

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of clinical markers of specific language impairment in adults. *Journal of Speech, Language, and Hearing Research*, 53(2), 414–429. https://doi.org/10.1044/1092-4388(2009/08-0016)

Potter, M., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal Of Memory And Language*, 29(6), 633–654. https://doi.org/10.1016/0749-596X(90)90042-X

Potter, M., & Lombardi, L. (1998). Syntactic Priming in Immediate Recall of Sentences. *Journal Of Memory And Language*, 38(3), 265–282. https://doi.org/10.1006/jmla.1997.2546

Prévost, P., Tuller, L., Zebib, R., Barthez, M. A., Malvy, J., & Bonnet-Brilhault, F. (2018). Pragmatic versus structural difficulties in the production of pronominal clitics in French-speaking children with autism spectrum disorder. *Autism and Developmental Language Impairments*, 3, 5–14. https://doi.org/10.1177/2396941518799643

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for raven's progressive matrices and vocabulary scales. Section 3: The standard progressive matrices*. Oxford: Oxford Psychologists Press.

Rice, M., Wexler, K., & Cleave, P. (1995). Specific language impairments as a period of extended optional infinitive. *Journal of Speech and Hearing Research*, 38, 850–863. https://doi.org/10.1044/jshr.3804.850

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in english-speaking children. *Journal of Speech, Language, and Hearing Research*, 39(6), 1239–1257. https://doi.org/10.1044/jshr.3906.1239

Rice, M., Noll, K., & Grimm, H. (1997). An extended optional infinitive stage in German speaking children with specific language impairment. *Language Acquisition*, 6, 255–295. https://doi.org/10.1207/s15327817la0604_1

Roberts, J. A., Rice, M. L., & Tager-Flusberg, H. (2004). Tense marking in children with autism. *Applied Psycholinguistics*, 25(3), 429–448. https://doi.org/10.1017/S0142716404001201

Solari, E. J., Grimm, R. P., McIntyre, N. S., Zajic, M., & Mundy, P. C. (2019). Longitudinal stability of reading profiles in individuals with higher functioning autism. *Autism*, 23(8). 1911–1926. https://doi.org/10.1177/1362361318812423

Sundar Raj, M., Prasath, M., & Adalarasu, K. (2015). Autism spectrum disorder. *International Journal of Pharmacy and Technology*, 7(2), 8638–8647. https://doi.org/10.1007/978-1-4419-1698-3_100159

Tager-Flusberg, H., & Calkins, S. (1990). Does imitation facilitate the acquisition of grammar? Evidence from a study of autistic, Down's syndrome and normal children. *Journal of Child Language*, 17(3), 591–606. https://doi.org/10.1017/S0305000900010898

Tager-Flusberg, H. (1999). A psychological approach to understanding the social and language impairments in autism. *International Review of Psychiatry*, 11(4), 325–334. https://doi.org/10.1080/09540269974203

Tager-Flusberg, H. (2015). Defining language impairments in a subgroup of children with autism spectrum disorder. *Science China Life Sciences*, 58(10), 1044–1052. https://doi.org/10.1007/s11427-012-4297-8

Tager-Flusberg, H., & Cooper, J. (1999). Present and future possibilities for defining a phenotype for specific language impairment. *Journal of Speech, Language, and Hearing Research*, 42, 1261–1274. https://doi.org/10.1044/jslhr.4205.1275

Terzi, A., Marinis, T., Kotsopoulou, A., & Francis, K. (2014). Grammatical abilities of greek speaking children with autism. *Language Acquisition*, 21(1), 4–44. https://doi.org/10.1080/10489223.2013.855216

Terzi, A., Marinis, T., & Francis, K. (2016a). Syntax and its interfaces at the low and high ends of the autism spectrum. In A. M. Di Sciullo, (Ed.), *Biolinguistic investigations on the language faculty* (pp. 195–212). Amsterdam: John Benjamins. https://doi.org/10.1075/la.235.09ter

Terzi, A., Marinis, T., & Francis, K. (2016b). The interface of syntax with pragmatics and prosody in children with autism spectrum disorders. *Journal of autism and developmental disorders*, 46(8), 2692–2706. https://doi.org/10.1007/s10803-016-2811-8

Terzi, A., Marinis, T., Zafeiri, A., & Francis, K. (2019). Subject and object pronouns in high functioning children with ASD of a null-subject language. *Frontiers in Psychology*, 10, 1–8. https://doi.org/10.3389/fpsyg.2019.01301

Tomblin, J. B., & Zhang, X. (1999). Language patterns and etiology in children with specific language impairment. In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders* (pp. 361–382). Cambridge, MA: MIT Press/Bradford Books.

Tuller, L., Ferre, S., Prevost, P., Barthez, M., Malvy, J., & Bonnet-Brihault, F. (2017). The effect of comuputational complexity on the acquisition of French by children with ASD. In L. R. Naigles (Ed.), *Innovative investigations of language in autism spectrum disorder* (pp. 115–140). New York, NY: American Psychological Association. https://doi.org/10.1037/15964-007

Valian, V. (2015). Innateness and learnability. In E. Bavin and L. Naigles (Eds.), *Cambridge handbook of child language*, 2nd Edn. (pp. 13–14). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316095829.002

Venter, A., Lord, C., & Schopler, E. (1992). A follow-up study of high-functioning autistic children. *Journal of Child Psychology and Psychiatry*, 33(3), 489–597. https://doi.org/10.1111/j.1469-7610.1992.tb00887.x

Wexler, K., & Chien, Y.-C. (1985). The developemnt of lexical anaphor and pronouns. *Paper and Report on Child Language Development*, 24, 139–149.

Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., & Naigles, L. R. (2017). Grammatical language impairment in autism spectrum disorder: Exploring language phenotypes beyond standardized testing. *Frontiers in Psychology*, 8, 1–12. https://doi.org/10.3389/fpsyg.2017.00532

World Health Orgnaization. (2013). Autism spectrum disorders and other developmental disorders: From raising awareness to building capacity. *World Health Organization, Geneva, Switzerland*, 1, 1–36. Retrieved from https://www.who.int/mental_health/maternal-child/autism_report/en/

Zhou, P., Crain, S., Gao, L., Tang, Y., & Jia, M. (2015). The use of grammatical morphemesby mandarin-speaking children with high functioning autism. *Journal of Autism and Developmental Disorders*, 45(5), 1428–1436. https://doi.org/10.1007/s10803-014-2304-6

The relationship between word and nonword repetition and receptive and expressive vocabulary skills in Gulf Arabic speaking children

Mariam Khater
Mohammed Bin Rashid Center for Special Education – Operated by The New England Center for Children©

This study investigates early phonological skills, as represented by nonword repetition (NWR), in typically developing (TD) Gulf Arabic speaking children and children with language impairment, and tries to examine findings in relation to two important NWR hypotheses: the phonological short term memory account (PSTM, Gathercole & Baddeley, 1990a) and the linguistic account of Snowling, Chiat, and Hulme (1991). We developed a new Arabic word and nonword repetition test (WNRep) and tested 44 TD children and a clinical group (CL) of 15 children with language impairment, between two and four years old. Scores were measured using the percent phonemes correct (PPC) and whole word correct (WWC) methods. The results show that the TD group scored significantly higher than the CL group on the WNRep and across one, two and three-syllable words/nonwords, and that NWR scores correlated significantly with receptive and expressive vocabulary tests. Apart from its ability to differentiate between TD and children with language impairment, NWR results revealed significant differences in groups' performance even on one-syllable word and nonwords, which differs from findings in other languages. The two different scoring methods did not make any significant contribution to the groups' results.

Keywords: vocabulary, acquisition, nonword repetition, language impairment

1. Introduction

Children in the early stages of language development try to mimic adult words. These words are initially novel for children, however, with time and development children will be able to repeat more difficult words and learn their meanings.

https://doi.org/10.1075/sal.10.07kha

Baddeley, Gathercole, and Papagno (1998) proposed that children's ability to repeat novel multisyllabic words is a predictor of their overall language learning ability. Therefore, there has been an increased interest in investigating children's ability to repeat unfamiliar words in order to know more about the underlying processes that may predict language abilities in typical and atypical language development. Nonword repetition has become a common tool to measure the ability of repeating novel words, as it was used with different populations: children and adults, typically developing individuals and those with language impairments, and across different languages. Nonword repetition stimuli were manipulated in different ways to develop different tasks in order to measure different aspects of the language, such as phonology, morphology, semantics.

In this study, we investigate whether real word repetition and/or NWR have the potential to be used as diagnostic assessment tools to distinguish typically developing children from children with language impairment in Gulf Arabic speaking children. This will provide some essential information regarding the nature of the relationship between real word repetition and nonword repetition and receptive and expressive vocabulary skills in Arabic speaking children with typical and atypical language development. In addition, we examine the predictive value of real word repetition and/or nonword repetition with receptive and expressive vocabulary. We investigate if using two different scoring methods (PCC and WWC) will have any impact on these three areas of investigation. We additionally investigate the effects of item length on children's repetition accuracy. Furthermore, the present study will also add evidence from Arabic to the existing literature on the role of real and nonword repetition in typical and atypical language development. Some findings of non-Indo-European languages (e.g., Cantonese) did not find nonword repetition as a clinical marker in children with SLI (Stokes, Wong, Fletcher, & Leonard, 2006), while a previous study in Gulf Arabic that used phonologically complex nonwords with school aged children with SLI found a significant effect for NWR (Shaalan, 2010). Our study uses less phonologically complex stimuli with younger children as we avoided the use of clusters. Finally, the current study will involve developing some new assessment materials that are necessary to evaluate phonological and vocabulary skills of children with and without language impairment in Arabic. Therefore, a battery of tests was developed, including: a word and nonword repetition test (WNRep), a receptive vocabulary test, and an expressive vocabulary test. These tests will be described in the following section.

1.1 Relationship between NWR and vocabulary in TD children and children with SLI

Studies of NWR in typically developing (TD) children have found significant correlations between NWR and receptive vocabulary (Briscoe, Bishop, & Norbury, 2001; Coady & Evans, 2008; Gathercole, Willis, Emslie, & Baddeley, 1991; Gathercole, Willis, Emslie, & Baddeley, 1992). These studies explained this correlation differently, based on two main accounts: the phonological short term memory (PSTM) account of nonword repetition (Gathercole & Baddeley, 1989, 1990a; Gathercole, Willis, Baddeley, & Emslie, 1994) and the phonological processing account (Snowling, Chiat, & Hulme, 1991). The significant correlation between NWR and vocabulary is not surprising as it is argued that NWR mimics child's task when learning new words, as learning a new word involves attending to novel acoustic information that is used to create phonological representations. Therefore, TD children with better vocabulary tend to score better on NWR tasks when compared with children with lower vocabulary scores (Bowey, 1996, 2001; Gathercole & Baddeley, 1989; Metsala, 1999). Though this correlation was found to be significant at age of four (Gathercole & Baddeley, 1989; Gathercole et al., 1991), it was no longer significant by the age of five (Gathercole et al., 1992). Therefore, PSTM stops influencing vocabulary growth by this age and it is vocabulary that seems to affect NWR after this age. This is because children employ richer vocabularies to facilitate NWR by using lexical and sublexical information. Between the age of four and five years NWR is influenced by vocabulary development, while before the age of four it is NWR that has stronger influence on vocabulary development (Gathercole et al., 1992). However, it is not clear how the direction of influence between nonword repetition and vocabulary reverses with age. Gathercole (2006) and Edwards, Beckman, & Munson (2004) acknowledge other effects that may influence NWR, such as phonological and frequency of consonant sequences. These all can be included under the general umbrella of phonotactics.

The link between NWR and vocabulary seems to hold for receptive vocabulary, but not for expressive vocabulary. Briscoe et al. (2001) found that receptive, rather than expressive, vocabulary tended to correlate more strongly with NWR in the SNHL and TD groups. No explanation was provided about why receptive, but not expressive, vocabulary accounted for this correlation between NWR and vocabulary development in TD children.

Gathercole and Baddeley (1990a) reported that children with Specific Language Impairment (SLI),[1] repeated significantly fewer nonwords correctly when compared with TD children of a similar age who were matched on nonverbal intelligence, as well as a younger language-matched group. The NWR skills of children with SLI (aged 8;6) were compared with those of typically developing age-matched children (TDAM) and typically developing younger (TDY) children. Children with SLI performed significantly more poorly than their TDAM and TDY peers on the three- and four-syllable nonwords. The mean performance of children with SLI was approximately 4 years below their chronological age. Gathercole and Baddeley (1990a) attributed this deficit to limitations in the phonological short term memory of children with SLI. Children with SLI demonstrated proportionally more difficultly in repeating longer nonwords than shorter ones, indicating that they had limited phonological capacity and that SLI is essentially a disorder of phonological short-term memory (Baddeley et al.,1998; Gathercole & Baddeley, 1990a).

Many studies showed more evidence supporting Gathercole and Baddeley's (1990b) claim that lower performance on tests of working memory in children with SLI is mostly captured by an explanation of poor storage and processing of phonological information (Bishop, North, & Donlan, 1996; Conti-Ramsden & Hesketh, 2003; Dollaghan & Campbell, 1998; Weismer et al., 2000). Conti-Ramsden and Hesketh (2003) compared the performance of 5 year old children with SLI to that of typically developing language matched (TDLM) peers (aged 3;0) on four possible clinical markers: (a) a past-tense task, (b) a noun plural task, (c) a NWR task, and (d) digit span. Children with SLI performed significantly below the TDLM children on digit span and NWR. Therefore, Bishop et al. (1996) proposed that nonword repetition could be considered a primary behavioral marker of SLI and that deficits in these children's ability to retain phonological representations over time could be the underlying cause of some of the syntactic deficits in children with SLI.

1.2 Theories of NWR

The original Baddeley and Hitch (1974) model of working memory (WM) consisted of three components: A central control system called the central executive and two subsidiary systems called the visuo-spatial sketchpad and the phonological loop.

1. SLI is a set of significant receptive and or expressive language impairments in the absence of cognitive, sensorimotor, social-emotional and environmental deficits (Bishop, 1997). Currently, there is a debate about the nature of SLI and whether it really exists in isolation of any other deficits and more terms are introduced to replace SLI, such as primary language impairment (Bishop, 2014; Kohnert, Windsor, & Ebert, 2009), Developmental Language Disorder (Bishop, Snowling, Thomspon, Greenhalgh, & The CATALISE Consortium, 2016) or Language Learning Impairment (LLI).

As proposed by Baddeley and Hitch (1974), the central executive is responsible for the control of working memory. It is a pool with limited capacity of general processing resources, and it controls the interaction between the other components. The visuo-spatial sketchpad is the second component and it is responsible for integrating and processing visual, spatial and nonverbal information.

The third component is the phonological loop, which is specialized for the storage of verbal material. It consists of two components, phonological storage which holds speech-based information for 1–2 seconds, and the articulatory rehearsal component, which converts words into articulatory or spoken words before entering phonological storage. The phonological loop is responsible for sorting and processing sound combinations, so any impairment in this part of working memory will cause deficits in the phonological representation, affecting the process of learning new words (Archibald & Gathercole, 2006; Baddeley et al., 1998; Gathercole & Baddeley, 1990a; Gathercole et al., 1999b). Baddeley et al. (1998) proposed that 'the function of the phonological loop is to provide temporary storage of unfamiliar phonological forms while more permanent memory representations are being constructed' (1998, p. 159).

The working memory model (Baddeley, 2003), might succeed in explaining the different factors that are involved in working memory, however it is not clear how these factors work together. For example, there is evidence that the phonological loop is responsible for sorting and processing sound combinations and any deficit in this part would affect word learning (Archibald & Gathercole, 2006; Baddeley et al., 1998; Gathercole & Baddeley, 1990a; Gathercole et al., 1999a). However, it is not clear yet if the deficit will affect the phonological loop only or if it will combine with other components. Furthermore, the WM model does not explain how semantic, syntactic and lexical components interact in NWR (Baddeley et al.1998). The link between verbal short term memory (STM), which is another term for the PSTM, and long term memory (LTM) is also questioned by Cowan (2001), who found that Baddeley's (2003) WM model failed to explain adequately how verbal STM is linked to LTM.

There are two different accounts that try to investigate the WM model in relation to NWR by exploring the mechanisms underlying nonword repetition in the light of the working memory model and examining if the verbal short term memory represented by NWR is assisted by long-term memory (represented by linguistic knowledge) or by phonological memory only. The first account is the *phonological short-term memory* (PSTM) account by Gathercole and Baddeley (1989, 1990a) and Gathercole et al. (1994). The second account is *the linguistic account or the phonological processing* account by Snowling et al. (1991).

The PSTM account of nonword repetition proposed that a person's ability to repeat what they hear helps in learning new words. Gathercole (2006) suggested that there is a high correlation between nonword repetition and novel word learning because both are constrained by phonological storage. That is, children's ability to repeat an unfamiliar sequence of phonemes can determine their ability to store and learn a sequence of phonemes pertaining to a novel word. Since every word is a novel word when it is first introduced to the child, both nonword repetition and vocabulary acquisition may have shared cognitive and neural mechanisms (Gupta & MacWhinney, 1997). The proponents of PSTM have claimed that storage of phonological information in the phonological loop is critical for learning words (Gathercole & Baddeley, 1990a; Gathercole et al., 1994).

The phonological loop is responsible for storing temporary phonological information which influences nonword repetition. Therefore, NWR is considered a pure measure of PSTM, as it presents non-lexical material that allows the elimination of any familiarity effect. Furthermore, the phonological structure of nonwords does not require a long-term lexical memory; the accurate repetition of the nonwords requires only a temporary storage for sound sequences in the phonological loop (Baddeley, Gathercole & Papagno, 1998; Dollaghan & Campbell, 1998). The PSTM account explains that word learning (long term lexical) difficulty in children with SLI is due to difficulties with phonological short-term memory (phonological loop). Although the PSTM account acknowledges the contributions of the different phonological processes that link nonword repetition to vocabulary knowledge (Snowling et al. 1991), Gathercole et al. (1991) claim that PSTM is the most significant contributor to children's subsequent vocabulary development, and that the phonological processes offer little in predicting future vocabulary size.

Finally, the linguistic or phonological processing account attempts to provide an additional explanation for the nature of the relationship between NWR and vocabulary acquisition (Snowling et al. 1991). It considers the phonological loop as only the initial process in nonword repetition, with various other processes participating in the same task, such as speech perception, phonological awareness, and output processes. Beside these processes there are other factors that play a role in nonword repetition, such as lexical knowledge and phonotactic probability. Gathercole et al. (1991) proposed that lexical knowledge can be used to support nonword repetition as the similarity to the sublexical units within nonwords increases, which in turn will influence accurate nonword repetition. The linguistic account, challenges the PSTM account and argues that nonword repetition is not a pure measure of phonological short term memory, as there are several other factors involved in the processing of nonwords, influenced by PSTM or not. These various factors are discussed in detail in the following section with reference to the two competing accounts of NWR.

1.3 Factors affecting nonword accuracy

Many studies have found an increase in nonword repetition accuracy with age (Edwards et al., 2004; Gathercole & Baddeley, 1989; Gathercole et al., 1991). Both the competence of temporary storage and articulation skills increase with age and support the subvocal rehearsal function of the phonological loop (Hoff, Core, & Bridges, 2008). Subvocal rehearsal helps in the maintenance of phonological memory traces in the storage (Gathercole & Baddeley, 1989).

Another variable influencing nonword repetition is vocabulary size, or the estimate of how many words a child has in his/her mental lexicon. Children with typical language development who obtain high scores on vocabulary measures are more likely to have better nonword repetition performance in contrast to children with low vocabulary scores (Bowey, 1996; Edwards et al., 2004; Gathercole & Baddeley, 1989, 1990a; Gathercole et al., 1991). The connection between phonological working memory and vocabulary size changes throughout development. Gathercole and Baddeley (1990a) proposed that children who performed better on nonword repetition exhibited faster learning for new words than the children who were less skillful at nonword repetition. Gathercole (1995) and Gathercole et al., (1991) found that children with low repetition accuracy showed low scores on receptive vocabulary knowledge. Correspondingly, children who showed better performance on nonword repetition achieved high vocabulary scores on standardized vocabulary tests (Gathercole, 1995; Gathercole, et al., 1991). These results indicate the bidirectional nature of the relationship between working memory and long term knowledge of words and word parts.

On the other hand, Horohov and Oetting (2004) argued that the amount of exposure to new words influences word learning more than lexical knowledge; the more the child is exposed to a new word the more he/she will be able to maintain it as a mental lexicon entry. Furthermore, Coady and Evans (2008) stated that children with language impairment succeeded in learning new words but they needed more exposure over time to the words.

Studies have found a strong effect of nonword length (number of syllables) on repetition performance (Gathercole & Baddeley, 1989, 1990a; Gathercole et al., 1994), which supports the PSTM account. In these studies, longer nonwords typically resulted in more repetition errors than shorter nonwords. The longer the word is the more likely it will be forgotten because repeating longer words imposes more demands on memory, articulation, and duration (Baddeley, 2003). Santos, Bueno, & Gathercole (2006) explored nonword repetition skills in typically developing children from four to ten years old and found that errors increased with longer nonwords (of four and five syllables) and that accuracy increased with age.

Furthermore, children with language impairment have consistently displayed significant deficits when repeating long nonwords of three or more syllables (Bishop et al., 1996; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990a).

Wordlikeness is defined as the degree of likeness between a nonword and the phonological form of words stored in an individual's lexicon and it has been shown to influence NWR results (Dollaghan, Biber, & Campbell, 1995; Edwards et al., 2004; Gathercole, 1995). Many studies (e.g., Gathercole, 1995; Gathercole & Adams, 1994; Gathercole et al., 1999a) suggested that the more wordlike the nonword, the more likely it will be repeated accurately. According to Baddeley (2003), judging wordlikeness is based on two main factors: either through phonological similarity where the nonword consists of a known or familiar phoneme structure, or lexico-semantic similarity. Therefore, (Gathercole & Baddeley, 1996) developed a test of phonological working memory 'The Children's Test of Nonword Repetition (CNRep) where the nonword contains a real word, such as "under" as in "*under*brantuand" or a real morpheme, such as "ing" as in "blonterstap*ing*". Several studies reporting on wordlikeness manipulation effects (Gathercole et al, 1991; Archibald and Gathercole, 2006; Roy and Chiat, 2004) found that repetition of high wordlikeness items is facilitated by long-term lexical knowledge and is therefore less sensitive to phonological memory constraints, whereas nonword repetition for low wordlikeness nonwords is dependent on phonological memory. Furthermore, typically developing children benefit from wordlikeness effects while children with language impairment were less sensitive to these effects.

Several studies investigated the effects of stress on nonword repetition, which was not discussed in the PSTM account. Dollaghan et al. (1995) investigated children's performance on nonword repetition by manipulating effects of stress using two patterns. The first pattern consisted of a nonsense syllable stress, i.e., a stress pattern that does not exist or very atypical of English words, where nonwords did not contain any weak stress and only tense vowels were used. The other stress pattern was typical of English stress pattern (a weak syllable and strong syllable stress pattern). Results showed that children were more likely to accurately repeat nonwords with stressed syllables corresponding to real words than nonwords with nonsense syllable stress. Furthermore, Roy and Chiat (2004) investigated stress effects on the type of nonword repetition errors, and found that children aged 2 to 4 years were sensitive to stress when repeating words and nonwords. Children were more likely to omit the unstressed syllables than the stressed ones.

More recent studies have shown increasing interest in using real word repetition along with NWR to examine the contributions of the different processes involved in NWR while controlling for articulatory processes, since there are many young TD children who fail to complete NWR due to their limited articulation skills (e.g., Chiat & Roy, 2007; Stokes & Klee, 2009). Hoff et al. (2008) used the individualized scoring method of controlling articulation performance in nonword repetition in a different direction, as they used real word repetition along with nonwords. They argued that they used real words to control for articulation because they place less demand on phonological memory than nonwords. Therefore, if the child was able to repeat a real word correctly, but not the nonword that consisted of the same phonemes, this indicated an error in phonological memory. However, if the child failed to repeat both the nonword and the real word, this indicated the child had limited articulation skills. Word length was found to affect the accuracy of real word repetition in typically developing children and children with language impairment and in a differential manner: typically developing children were less affected by word length in real word repetition than NWR, as the word familiarity helped in improving their repetition accuracy (Chiat & Roy, 2007; Dispaldro, Benelli, Marcolini, & Stella, 2009; Dispaldro, Deevy, Altoé, Benelli, & Leonard, 2011; Roy & Chiat, 2004). However, Chiat and Roy (2007) found that children with language impairment were less affected by word familiarity, and that the interaction between length and word status (word vs. nonword) was not significant. The authors attributed this to the idea that although real word repetition puts less demands on phonological short-term memory compared with nonword repetition, children with language delay could not benefit from this advantage in their performance on real word repetition due to their language deficits. Dispaldro, Leonard, & Deevy (2013) investigated the utility of real word/nonword repetition as a clinical marker in Italian children with SLI aged between 3;11–5;8. Both tasks succeeded in distinguishing the SLI group from the TD children.

The two main scoring methods that are used in most word and nonword repetition studies are the whole word correct (WWC) method where each item was scored as a whole, either entirely correct or incorrect (CNRep test, Gathercole & Baddeley, 1996), and the percentage of phonemes correct (PPC) scoring the number of phonemes produced correctly in each item (Nonword Repetition Test, Dollaghan & Campbell, 1998). Graf Estes, Evans, and Else-Quest (2007) compared these two scoring methods across 23 different English studies with TD children and children with SLI for word and NWR. They hypothesized that the WWC method could penalize children with SLI as they might have more phonological errors compared to TD groups, while the PPC scoring method could provide a more appropriate evaluation of the two groups. The results showed that these two different scoring methods did not influence the groups' results on the NWR.

1.4 The Arabic word and nonword repetition (WNRep) test

The main objective in developing the Arabic Word-Nonword repetition test (WNRep) was to assess two to four year old Qatari Arabic speaking children's early phonological skills as measured by word length in syllables; to examine its utility in distinguishing children with language impairment from typically developing children; and to examine if word repetition or/and nonword repetition can predict receptive and expressive vocabulary size in both groups. The WNRep was modeled on the Preschool Repetition Test (PSRep; Roy & Chiat, 2004), which consists of 36 test items; 18 words followed by 18 nonwords. Words and nonwords were made up of equal numbers of 1–3 syllable items, with systematic manipulation of stress, so that half the words have strong/weak stress (SW), while the other half have weak/strong (WS) stress. The words and nonwords were phonologically matched, with nonwords created by altering the vowel in single syllable words and reversing two consonants in each word to create a corresponding nonword (Roy & Chiat, 2004).

To develop the WNRep test many factors were taken into consideration to control for variables that might influence this task. The variables that were considered were: articulatory complexity, word length, wordlikeness, language specific phonotactic rules, and word familiarity.

Gulf Arabic has 30 consonants, however only 10 consonants were used in the WNRep, and most of them were early acquired sounds. The sounds used were: /b/, /d/, /t/, /k/, /f/, /s/, /m/, /n/, /l/ and /r/. The selection of these sounds was based on Ayyad (2011); Amayreh (2003), and Amayreh and Dyson (2000) showing that younger children (aged 45–54 months) have mostly acquired these sounds. Although /s/ and /r/ are not acquired early, they were included as it was difficult to find a range of common words that do not contain one of these consonants. According to Bukshaisha (1985), all these consonants can occur in any position in Gulf Arabic words. Clusters were also avoided in this study as we tested young children who may not have mastered the production of clusters yet (Ragheb and Davis, 2010; Ayyad, 2011). On the other hand, gemination which is defined as a cluster of two identical consonants (Delattre, 1971) was included given that that 16% of words in Gulf Arabic have geminate consonants (Alqattan, 2015 for the related Kuwaiti Arabic variety).

All Arabic long and short vowels were included in the WNRep test: /a/, /aa/, /u/, /uu/, /i/, /ii/, /ee/, /oo/. No diphthongs were included in the test as Modern Standard Arabic (MSA) diphthongs (e.g., /aj/ and /aw/) are almost always realized as long vowels /ɛɛ/ and /oo/ (Ayyad; 2011). Furthermore, there are no studies that have shown the age of acquisition of diphthongs in Arabic.

The WNRep test consists of 48 items equally divided into one, two and three syllable items. Both the list of words and nonwords consisted of 8 one syllable items, 8 two syllables items and 8 three syllable items. The items on each list were ordered in a semi-random fashion.

Vowels were changed in all nonword items but vowel length, shape of syllables and stress of words were kept the same for words and nonwords. The word structures used in both lists were as follow: for one-syllable items the syllable structures used were CVC; for two-syllable CV.CVC and CVC.CVC; and for three-syllable CVC. CV.CV, CV.CV.CVC, CV.CV.CV, and CV.CVC.CVC. Four items in the WNRep test with the structures CVC.CVC and CV.CVC.CVC included a geminate.

Nonwords were phonotactically legal respecting Arabic phonological constraints, such as the Obligatory Contour Principle on place of articulation (OCP-Place), where consonants produced in the same place of articulation are disfavoured within the same root (Frisch, Pierrehumbert, & Broe, 2004; Frisch & Zawaydeh, 2001). For example, there are no words in Arabic that have the root √f.b.m, and therefore no nonwords were formed in this experiment based on these illegal combinations. Examples of consonantal roots that are phonotactically legal but do not exist in Gulf Arabic are √k.d.f or √s.b.d (Shaalan, 2010).

The word list in the WNRep test has words that were selected from speech samples that were collected in a previous study with 56 Gulf Arabic speaking children aged from two to six years old (Khater & Shaalan, 2007), therefore the words were common and familiar to the children. In contrast, the converted nonword list did not consist of any familiar roots and/or patterns, therefore the nonword list was designed to avoid any wordlikeness.

The original words used to form the WNRep were all nouns; no verbs, adjectives, or adverbs were used to create this test. The words and nonwords were phonologically matched, and nonwords were created by alternating the vowel in all syllables of the words, avoiding any real vocalic pattern that is common Gulf Arabic. The vocalic patterns that were used with nonwords were as follow:

1. For the two syllabic nonwords: (i-u) (e.g., /bituk/), (u-u) (e.g., /sukkub/), (a-u) (e.g., /lakus/), (o-i) (e.g, /moril/) and (u-a) (e.g., /nujam/.
2. For the three syllable nonwords the vocalic patterns used were as follow: (a-a-i), (a- i) (e.g., /jasari/), (u-i), (e.g.,/kusimɑ/), (u-a-ɔ) (e.g., /lufanɔ/), (u-a-u) (e.g., /nufatul/) and (u-a-ə) (e.g.,/fumajjək/).

Furthermore, none of the nonwords contain any existing roots in order to avoid any morphological similarity; all the nonwords consist of non-roots. The nonexistence of these nonwords was also checked in an Arabic dictionary *Mu'jam Al-Waseet* (Mustafa, AlZayaat, Hamed, & AlNajaar, 2004) to ensure that no real root was used as a non-root item. Though the one syllable nonwords were not real words in Gulf Arabic on their own, it is possible to find those syllables as parts of real multisyllabic words. Moreover, there was no attempt to control for stress patterns in WNRep items, as items were randomly given strong/weak (SW) and weak/strong (WS) stress patterns. Some examples for one syllable /kis/ "bag" →/saak/, two syllables /see.kəl/ "bicycle"→/la.kus/ and three syllables /laimuun/ "lemon"→ /nul.jaam/

1.5 Developing the Arabic expressive vocabulary test (AEVT)

The Arabic Expressive Vocabulary test (AEVT) was developed to be used in this current study with Qatari Arabic speaking children, as there is currently no Arabic expressive vocabulary test available to be used in Gulf Arabic. This test was designed to be administered with young children aged from two to four years old. It was developed following the methodology used in the Expressive One-Word Picture Vocabulary Test (EOWPVT; Martin & Brownell, 2000), a standardized test that provides assessment for the verbal expressive vocabulary for English individuals aged 2;0 to over 80 years old. The Arabic version followed the same principles of the picture display, scoring and organizing the stimuli in groups according to age bands. The bands were 11 months for each group and there were 8 groups. Each group consisted of 8 pictures (with a total of 64 pictures for the whole test). There were also four trial items to enable the children to understand the test procedures and to be familiar with the examiner and the test itself. A booklet was made that consisted of 68 pages (4 pages for practice items and 64 for test items). Each page contained one coloured picture taken from non-copyrighted materials (Windows clip art) from the same source to make sure that all pictures had the same characteristics.

AEVT items were selected based on item type or category and item difficulty. Thus, the items belonged to different groups, (e.g., actions or verbs, single and plurals nouns that belonged to different categories (households, animals, food, clothes, toys). The items were arranged according to their difficulty following the normal expressive vocabulary development in Gulf Arabic speaking children. Two methods were used to determine the words' difficulty level to children. The first method was using representative speech samples that were collected in a previous study by Khater and Shaalan (2007), from 56 Gulf Arabic speaking children aged 2;0–6;0 years old. The items used in the test for each age band were selected from the speech samples at the same age group. The second method involved a familiarity rating collected from 24 Gulf Arabic speaking adults for 600 words (Shaalan, 2010) that was used to develop the Arabic Picture Vocabulary test (APVT) also used in this current study. The researcher in this current study modified the order of the APVT items according to the item analysis done by Shaalan (2010) following testing 107 children (4;6 to 9;6 years old). Furthermore, two pictures were modified or changed by the current author as they were found to be difficult to recognize according to Shaalan (2010). The criteria for choosing these words were similar to those used in the British Picture Vocabulary test (BPVT; Dunn & Dunn, 1997).

The WNRep test and AEVT tests were administered to a group of typically developing children (TD) and a clinical group of children with language impairment (CL) to confirm the following predictions:

1. The TD group will have significantly higher scores than the CL group on all measures (repetition of 1, 2 and 3-syllable words and nonwords, as measured by percent phonemes correct (PPC) and whole words correct (WWC); receptive vocabulary and expressive vocabulary.
2. Item length will affect repetition performance in both groups, with a significant decrease in scores as length increases.
3. Word type will affect repetition performance in both groups, with significantly higher scores for words than nonwords.
4. Scores for WNRep will be significantly correlated with scores on the APVT and AEVT in both groups and using both scoring methods.

As indicated in the above predictions, it was furthermore hypothesised that using two different scoring methods, PPC and WWC, would not affect outcomes of the current study, i.e. the two scoring methods would yield the same effects of group and item factors.

2. Methods

2.1 Participants

A summary of participants' characteristics is shown in Table 1:

Table 1. Summary of the characteristics of participants

Participants	TD group	Clinical group
Number of participants (Male: Female)	44 (21:23)	15 (10:5)
Mean age in months (years)	38 (3;2)	43 (3;7)
Range in months (years)	27–47 (2;3- 3;11)	33–57 (2;9–4;9)

All typically developing children included in this study had no history of speech or language impairment or neurological, developmental and cognitive impairments, such as congenital malformation, hearing loss, or autism. The children in the clinical group (CL) were referred to the clinic with impaired language not combined with any history of congenital abnormalities, cognitive disabilities, hearing loss, oral – motor difficulties or autism were included in this study. The criteria of selecting the CL group in this experiment are similar to the diagnostic criteria of SLI according to definitions by Bishop (1997) and Leonard (1998). However, it was not possible to label the CL group in this study as an SLI group, due to the insufficient investigations by the referral sources (e.g., no IQ tests were conducted in the speech clinic). In addition, the use of term SLI has been controversial (Ebbels, 2014; Reilly, Tomblin, et al., 2014; Gallagher, 2014; Rice, 2014; and Taylor, 2014).

2.2 Procedures and scores

All children completed the battery of tests (APVT, EVT, and WNRep) in the same session, which ranged between 45 and 90 minutes depending on the child's ability to tolerate the tasks. All children received an articulation screener and a developmental verbal dyspraxia screener (Shaalan, 2010) to rule out any severe phonological disorders. Children who needed a break were given 10 to 15 minutes to play with some toys before the WNRep, the word repetition, and the NWR tasks were administered.

2.3 Reliability

Reliability (the consistency of performing a test, DeVellis, 2012) is measured to ensure that changes in test scores are only due to changes in the target variable. Inter-rater reliability is a method used to assess reliability and it measures the correlation between the scores of two different examiners or raters. A second examiner, a Gulf Arabic speaking speech-language therapist with experience in scoring and administrating repetition tasks, was asked to evaluate 10% of the children's scores using the PPC scoring method in WNRep and Arabic Expressive Vocabulary test (EAVT). Inter-rater agreement between the examiners was at a high level of ($\alpha = 1.0$) for EAVT and ($\alpha = .90$) for WNRep test.

3. Results and analysis

The following section presents the descriptive and inferential statistics of the PPC scoring method only. The WWC scoring methods results are not presented in this current study. However, the WWC method results do not differ from the PPC and have no bearings on the findings of this study.

3.1 Descriptive statistics of WNRep

The descriptive statistics of the children's performance on word and nonword repetition for both CL and TD groups, using the PPC scoring method, are presented in Table :

Table 2. Scores of for word and nonword repetition tests for both typically developing children (TD) and clinical children (CL) Percentage of phonemes correct (PPC) method

	Group	Age (months)	Word repetition	Nonword repetition
TD	M	38.39 (3:3)	96.7	79.2
N = 44	SD	6	6.45	11.62
	Range	26–47	91–122	73–122
CL	M	43.3	85.36	69.9
N = 15	SD	8.09	15.93	20.13
	Range	33–57	66–122	55–110

The descriptive statistics of the TD and CL group scores on different word and nonword lengths (one, two and three syllable) and for each word type are presented in Table 3:

Table 3. Children's mean scores (and standard deviations, SDs) on the different word lengths for the words and nonword repetition test, based on the (PPC) scoring method. TD = typically developing children, CL = clinical group

Repetition type	Group type	Mean	SD	N
1 syllable words	TD	99.71	5.78	44
	CL	93.88	7.53	15
2 syllable words	TD	97.09	5.33	44
	CL	87.28	12.11	15
3 syllable words	TD	95.12	8.32	44
	CL	80.12	18.66	15
1 syllable nonwords	TD	95.64	6.72	44
	CL	86.66	14.45	15
2 syllable nonwords	TD	88.68	11.24	44
	CL	81.08	14.03	15
3 syllable nonwords	TD	85.78	14.29	44
	CL	69.21	23.17	15

The first question in this experiment was whether the performance of the TD and CL group differed across different word lengths and word types. It was predicted that the TD group would score higher than the CL group across different word types and lengths. Furthermore, it was expected that both groups would perform better on words vs nonwords and shorter words/nonwords vs longer words/nonwords. To answer these questions, a repeated measures ANOVA was conducted to investigate the effects of word length (3: one, two and three syllables' length), and word type (2: words and nonwords) as within subject factors, and group (2: TD vs. CL) as a between-subject factor. There was a significant main effect of group (F (1, 57) = 20.59, p = <.001, η^2 = .265) as the TD group had significantly higher PPC scores on word/nonword repetition (M = 87.95, SD = 9.035) than the CL group (M = 77.63, SD = 18.03). There was also a significant main effect of word type (F (1, 57) = 42.4, p < .001, η^2 = .4.27) as children's performance was better for words (M = 93.8, SD = 9.3) than nonwords (M = 76.9, SD = 7.6). A significant main effect of word length was found (with Greenhouse-Geisser correction which was conducted due to violation of specificity) ε = 0.785, F(1.51, 86.2) = 4.61, p < .05. Children's PPC scores decreased as the word length increased (1 syllable: M = 85.19, SD = 10.24; 2 syllables: M = 69.17, SD = 14.78, 3 syllables: M = 59.75, SD = 20.79). The interaction between word length and word type was significant (ε = 0.869), F(1.73, 99.0) = 2.54, p < .05.

In addition to these significant main effects, there were also significant interactions between group and word length (F (1, 57) = 5.15, p = .027, η^2 = .083), and between word type and length (F (1, 57) = 1.37, p < .001, η^2 = .002). The interaction between group and word type, on the other hand, was not significant (F = (1,57) = .126, p = .723, η^2 = .002), nor was the three-way interaction between group, word type and length (F (1, 57) = .137, p = .713, η^2 = .002).

A Bonferroni corrected post hoc test revealed a significant difference in children's performance from both groups on different word and nonword lengths (one, two and three syllables). The longer the word and the nonword was, the less accurate the repetition was. Children's scores were significantly higher on one syllabic words/nonwords when compared to two syllable words and nonwords (p < .05). They also scored higher on one syllable words and nonwords when compared to three syllable words and nonwords (p < .05). Similarly, their scores on the two syllabic words/nonwords were significantly higher than on the three syllable words and nonwords (p < .05). An independent t-test was conducted to compare each group (TD and CL) at the interaction between word length and group. Results from the statistical comparisons (see Table 3 above) showed that there was a significant difference between the groups at one, two and three syllables for words and nonwords. The CL group performed significantly less accurately than the TD group at every length for means and SDs for words and nonwords (see Figure 1).

In addition, a Paired Sample t-test showed that one-syllable vs two-syllable words effects are significant (t (43) = 3.5, p = .001), as re for two-syllable vs. three-syllable words (t (43) = 4.14, p < .001) and one-syllable vs. three-syllable words (t (43) = 2.52, p = .015). Likewise, in the CL group there was a significant difference between

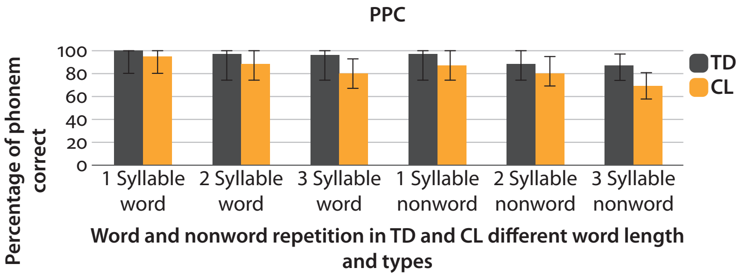


Figure 1. PPC scores for word and nonword repetition in TD and CL in different word length and type

the different word lengths, with significant effects in one-syllable vs two-syllable (t (14) = 2.36, p = .033), two-syllable vs. three-syllable (t (14) = 4.17, p = .001) and one-syllable vs. three-syllable (t (14) = 3.5, p = .004).

Therefore, the interaction between group and length must have arisen from the magnitude of the difference between groups at different lengths. Looking at the effect sizes in the graphs in Figure 1, it is evident that the magnitude of the difference between groups increases with length, with the most marked difference between groups occurring in the three syllable items. Hence, the CL group were more affected by word/nonword length than the TD group.

Table 4. Summary of the independent samples t-test results comparing TD and CL groups at each length for word and nonword repetition using PPC scoring method

Word length	t-value	Significance	Effect size
One syllable	$t(57) = 3.24$	$p < .001$	$\eta^2 = 0.304$
Two syllable	$t(57) = 3.36$	$p = .001$	$\eta_2 = .0.431$
Three syllable	$t(57) = 3.27$	$p < .001$	$\eta^2 = 0.400$

Next, to follow up the significant interaction between word length and word type, paired-samples t-tests were conducted to compare the combined groups' performance on word and nonword repetition at each length. The results showed significant differences between words and nonwords at every word length: one-syllable (t (58) = 4.3, p < .001), two-syllables (t (58) = 5.9, p < .001), and three-syllables (t (58) = 5.4, p = .001). There were also significant differences when the one-syllable nonwords were compared with the two-syllable nonwords (t (58) = 3.2, p = .002), and the two-syllable nonwords with the three-syllable nonwords (t (58) = 3,32, p < .001). Thus, the accuracy of nonword repetition increased when the nonword length decreased. Similarly, one-syllable words were repeated significantly better than two-syllable words (t (58) = 3.2, p = .002), one-syllable words than three-syllable words (t (58) = 4.1, p < .001) and two-syllable words than three-syllable words (t (58) = 3.3, p = .002. Therefore, the interaction between word type and word length must have arisen, again, from the magnitude of the difference between the word type at different lengths.

The results of the main ANOVA and the follow up analysis are consistent with our hypothesis. The TD group scored higher than the CL group across different word types and lengths. Moreover, both groups had superior scores in repeating words vs nonword and higher scores in repeating shorter words/nonwords vs longer words/nonwords.

3.2 Groups' performance on receptive and expressive vocabulary tests

To investigate the children's performance on the receptive vocabulary test (APVT), an independent sample t-test- was conducted for the TD group and CL group. It was hypothesised that the CL group will score significantly less than the TD group in APVT. Results showed that the difference was not significant between the TD group (M = 23.4, SD = 7.2) and CL group (M = 20.8, SD = 6.9) in APVT (t (57) = 1.21, p = .228). This result was contrary to what was hypothesised. Figure 2 shows the scores of the TD and CL group on the APVT:

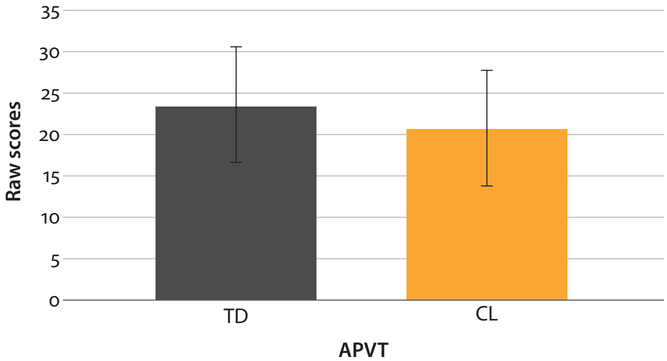


Figure 2. TD and CL performance in Arabic picture vocabulary test (APVT)

Another independent sample t-test was conducted for the TD and CL groups to investigate children's performance on the expressive vocabulary test (AEVT). It was also hypothesised that the CL group will score significantly less than the TD group on AEVT. Results showed that the TD group (M = 18.84, SD = 5.63) scored marginally higher than the CL group (M = 15.73, SD = 5.42), however the difference between the two groups just failed statistical significance (t (57) = 1.85, p = .069). This result was contrary to what was hypothesised. Figure 3 shows the group's performance on AEVT.

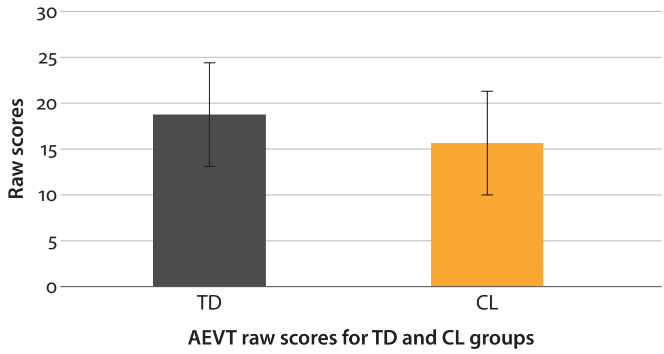


Figure 3. TD and CL groups' performance on Expressive Arabic Vocabulary test (AEVT)

The lack of significant difference between the CL and the TD group could be due to the small CL group sample ($N = 15$), in addition to three outlier children from the CL group who performed well in all tests despite meeting the criteria for the CL group selection. The children's scores in both vocabulary tests were recalculated while excluding the three outlier children ($n = 12$). The results of the revised independent sample T-test showed that the TD group performed significantly better than the CL group on both receptive and expressive vocabulary tests. On the receptive vocabulary test (APVT), the TD group performed significantly better ($M = 23.4$, $SD = 7.2$) than the CL group ($M = 18.17$, $SD = 4.8$), $t\ (54) = 2.36$, $p = .022$). On the expressive vocabulary test (AEVT), the TD group had significantly better raw scores ($M = 18.8$, $SD = 5.6$) than the CL group ($M = 13.92$, $SD = 4.3$), $t\ (54) = 1.85$, $p = .007$).

3.3 Correlation between word and nonword repetition and receptive and expressive vocabulary in TD children

A Pearson product-moment correlation coefficient was calculated to measure the correlation between word and nonword repetition and the Arabic Picture Vocabulary Test (APVT) and the Expressive Arabic Vocabulary Test (AEVT) skills in TD children ($n = 44$), using the PPC scoring method. It was expected that word and nonword repetition scores would significantly correlate with receptive and expressive vocabulary test on one hand and the word repetition scores would also significantly correlate with nonword repetition on the other hand. The reference of labelling the correlation strength in this section is based on the following r: 0–0.19 is regarded as very weak, 0.2–0.39 as weak, 0.40–0.59 as moderate, 0.6–0.79 as strong and 0.8–1 as very strong correlation

The Pearson product-moment correlations coefficients were also calculated to measure the correlation between word and nonword repetition and receptive and expressive vocabulary in typically developing children (TD) ($n = 44$). Results showed that the word repetition test significantly correlated with age ($r = .453$, $p < .01$), the APVT ($r = .331$, $p < .05$), the AEVT ($r = .560$, $p < .001$), and the nonword repetition test ($r = .602$, $p < .001$). The nonword repetition test significantly correlated with age in months ($r = .322$, $p < .05$), AEVT ($r = .376$, $p < .05$) and the Arabic Picture Vocabulary test APVT ($r = .331$, $p < .05$). The APVT and AEVT were strongly correlated with each other ($r = .612$, $p < .001$). Both vocabulary tests were correlated strongly with age in months (age and APVT, $r = .491$, $p < .001$ and age and the AEVT, $r = .806$, $p < .001$). The results of correlation using the PPC scoring method were consistent with what was predicted.

The Pearson product-moment correlation coefficient was also used to measure the correlation between word and nonword repetition and receptive APVT and expressive AEVT vocabulary skills in the clinical group (CL) ($n = 15$). Results of the various correlations showed that the word repetition test significantly correlated with the nonword repetition test ($r = .786$, $p < .001$). However, the word repetition test did not correlate significantly with other variables e.g., age ($r = .113$, $p < .0688$), the APVT ($r = .110$, $p < .697$), or the AEVT ($r = .155$, $p < .582$). Nonword repetition correlated significantly with word repetition, however the correlation with the other factors (age in months, APVT, and the AEVT) was not significant. The APVT and AEVT were strongly correlated ($r = .692$, $p < .000$). Neither vocabulary tests nor word and nonword repetition tests correlated with age in months. The results of the correlation using the WWC scoring method were partially against predictions as the word/nonword repetition scores did not correlate with the receptive or expressive scores. On the other hand, the significant correlation between word and nonword repetition was consistent with what was hypothesised.

For the PPC scoring method, the correlation between word and nonword repetition and receptive and expressive vocabulary in the clinical group (CL) ($n = 15$), showed that the nonword repetition test significantly correlated only with the word repetition test ($r = .733$, $p < .002$) but did not correlate significantly with age ($r = .032$, $p < .910$), the APVT ($r = .131$, $p < .634$), or the AEVT ($r = .215$, $p < .443$) Furthermore, the nonword repetition test did not correlate with the other variables (age ($r = -.160$, $p = .569$), APVT ($r = -.020$, p.943) and AEVT ($r = .411$, $p = 128$). The APVT and AEVT were strongly correlated ($r = .692$, $p < .000$). None of the vocabulary tests correlated significantly with age in months, or the word and nonword repetition test. One possible reason for this nonsignificant correlation was the inclusion of three children who scored within normal range in both vocabulary tests who had received speech therapy for more than three months while the other participants in the CL group had either no speech therapy or had just been referred to the speech therapy unit at the time of the study. This information about length of speech therapy services was available after the individual results for each participant were analysed. Reanalysis for the correlation was conducted after the three participants were excluded and the results showed significant correlation between vocabulary tests and word nonword repetition. Figure 4 show sscatterplots of the correlation between (APVT), (AEVT), word and nonword repetition tests for all children (TD and CL) using PPC scoring method.

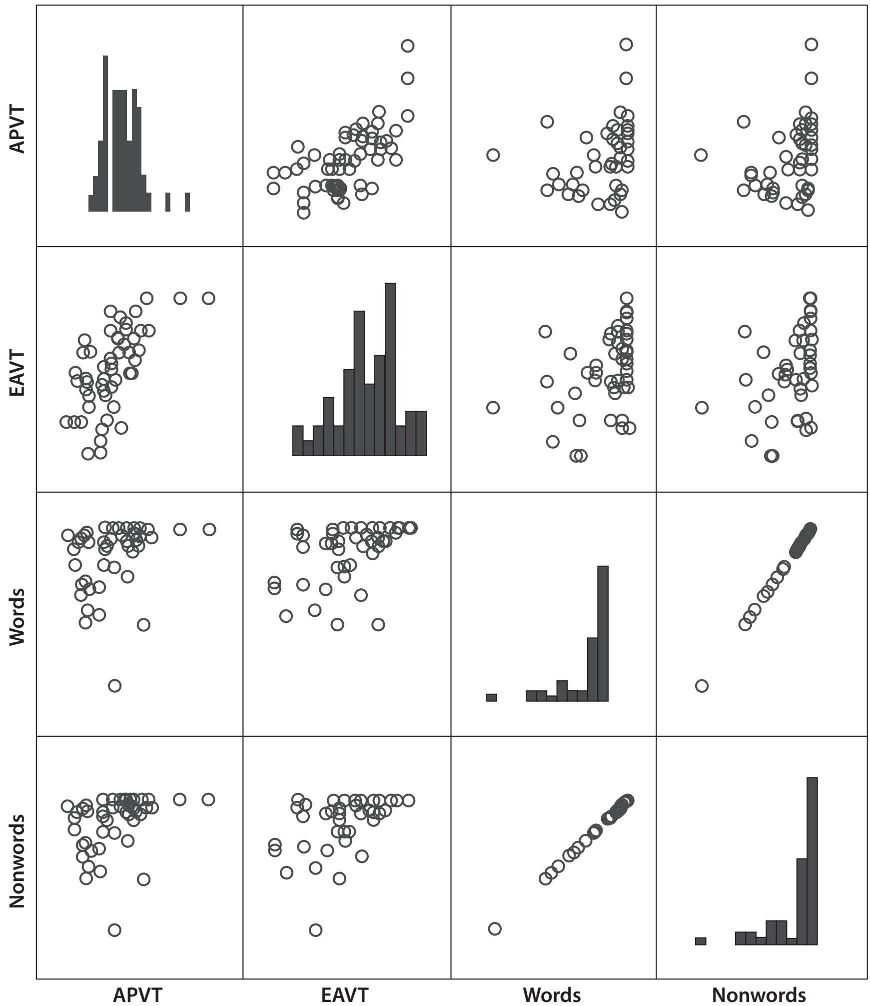


Figure 4. A scatterplot of the correlation between Arabic Picture Vocabulary Test (APVT), the Arabic Expressive Vocabulary Test (AEVT), word and nonword repetition tests for all children (TD and CL) using PPC scoring method

3.4 Regression analysis on the PPC scoring method for TD and CL children

A multiple linear regression was calculated to predict the receptive vocabulary skills measured by the children's scores on the APVT (DV, dependant variable) based on children's performance on word repetition and nonword repetition (IV, independent variables), to see if the children's performance on word/nonword repetition skills can predict their receptive vocabulary size. A significant regression equation was found (F (1, 42) = 5.168, p = .028), with an R^2 of .110. The children's performance on APVT is equal to (−20.412 + .553) (nonword repetition). The children's performance on nonword repetition can predict 11% of the variance of the APVT, however word repetition was excluded, as its predictive value was not significant. Another set of regression was conducted with the AEVT as the DV and word and nonword repetition as the IV. A significant regression equation was found (F (1, 42) = 19.236, p < .001), with an R^2 of 0.314. The AEVT predicted weight is equal to (−39.301 + .733) (nonword repetition). The children's performance in nonword repetition can predict 31% of the variance of the AEVT. Word repetition was excluded as its predictive value was not significant.

An identical multiple linear regression was calculated for the clinical group. A non-significant regression equation was found (F (1, 13) = .226, p = .0643), with an R^2 of .017. The children's performance on APVT is equal to (14.86 + .085) (nonword repetition). The children's performance on nonword repetition can predict 1.7% of the variance of the APVT. Another set of regressions was conducted with the AEVT as the DV this time and the word and nonword repetition as IV. A non-significant regression equation was found (F (1, 13) = .627, p < .443), with an R^2 of 0.046. The AEVT predicted weight is equal to (8.153 + .108) (nonword repetition), therefore children's performance on the nonword repetition can predict 4.6% of the variance of the AEVT.

4. Discussion

This section will summarize the findings in light of nonword repetition theories, taking into consideration the implication of the performance on word and nonword repetition, nonword repetition in relation to receptive and expressive vocabulary.

4.1 Summary of findings

We examined the phonological skills and verbal working memory skills of children using the Word Nonword Repetition tests (WNRep). The test's success in identifying children at risk of language impairment was evaluated. The real word repetition task in this experiment was designed to control for articulation skills and to investigate its utility as a clinical marker for language impairment along with NWR. The results showed that the TD group (n = 44, mean of age = 38 moths) performed significantly better on word and nonword repetition when compared with the CL group (n = 15, mean of age = 43 months). Both word and NWR are potentially useful to differentiate typically developing children from the clinical ones. Furthermore, the TD and CL groups did significantly better on word than on nonword repetition, which suggests that both groups were sensitive to lexical familiarity. The interaction between the group type (TD and CL) and the word type (word vs. nonword) was not significant indicating that both groups were similarly sensitive to the different word types.

The results additionally showed that there was a significant effect of word length in TD and CL groups in both word and nonword repetition; the longer the word or the nonword was the more difficult it was to repeat for both groups of children. The interaction between word type and word length was significant. Accuracy of nonword repetition decreased as the number of syllables increased and the gap between the word vs nonword scores was significant at one-syllable word/nonword, two-syllable word/nonword and three-syllable word nonword.

We found almost identical results for the whole word correct (WWC) and the percentage of phoneme correct (PPC) methods, however correlations and regression reached higher levels of significance and with good predictive value for receptive and expressive vocabulary performance in the WWC scoring methods. In general, there was a medium to strong correlation between word and nonword repetition tests and the APVT and AEVT in TD children and across different scoring methods. As for the CL group, there was a strong correlation between APVT and AEVT, and between word and nonwords, but there was no significant correlation between either of the two repetition tasks and the two vocabulary tests using either of the two different scoring methods. The difference in correlation and regression results across the two different scoring methods in this study was minimal and did not alter the outcomes. However, it is important to take into consideration that the selection of the scoring method may influence results, though it did not in this current study.

Overall, the performance of the TD and CL groups in the first study showed a similar profile to results seen in other languages (Botting & Conti-Ramsden, 2001; Conti-Ramsden & Hesketh, 2003; Dollaghan & Campbell, 1998; Weismer et al., 2000; Gathercole & Baddeley, 1990b; Gray, 2003) as the TD group scored better on word repetition than they did on nonword repetition. The predictive value of both word and nonword repetition for both vocabulary tasks was high in TD group but not in the CL, possibly due to the great variability in the performance of the children in the CL group.

4.2 Theoretical interpretations and implications

The nonword repetition task in the WNRep test was designed to measure effects of phonological short term memory (Gathercole & Baddeley, 1990a) by manipulating word length in syllables, while also looking at effects of wordlikeness (word vs nonwords). This design was used to examine if the phonological loop as presented by Baddeley (2003) is responsible for providing temporary storage of unfamiliar phonological forms as well as sorting and processing sound combinations. Therefore, any impairment in this part of working memory causes a deficit in the phonological representation, affecting the process of learning new words.

The results of this study were partially consistent with the PSTM theory, as the longer the word or nonword was the more difficult it was for children with language impairment to repeat. In addition, children who performed poorly on nonword repetition had lower vocabulary sizes when compared with other children who did well on nonword repetition. However, the fact that children with language impairment performed significantly worse than TD children even on one and two-syllable nonwords does not seem to be compatible with PSTM, as most studies in English showed that differences in performance on NWR usually start at three syllables and above. This poor performance on two and even on one-syllable nonwords indicates that there might be other more important considerations or processes involved in NWR in Arabic, such as effects of roots and patterns, which are closely related to wordlikeness effects. Shaalan (2010) found similar results in that children with SLI performed significantly less accurately on NWR on two syllable nonwords. This latter interpretation is more compatible with phonological processing accounts (Snowling et al., 1991; Chiat, 2001).

The results of real word repetition in our study raise questions regarding the processes underlying the repetition skills based on the working memory model. For example, whether the phonological loop in the working memory model is mainly responsible for providing temporary storage of unfamiliar phonological forms only, or whether deficits in the phonological loop only cause a deficit in learning new words or there are possibly other deficits in different components of the model. The working memory model fails to answer these questions, especially when we consider real word repetition. The model also fails to explain how the semantic, syntactic and lexical components interact with NWR, or how short-term memory is linked to long term memory. It is not possible to counter or explain the underlying processes involved in nonword repetition in one model like the working memory model, because the nonword repetition test was found to be a flexible measure that was designed to measure different language aspects (e.g., phonology, morphology) by manipulating the stimuli.

4.3 Word and nonword repetition as clinical markers

The nonword repetition task in the first study succeeded in distinguishing children with language impairment from typically developing ones; the WNRep was sensitive to children's repetition skills, which in turn is related to their language skills, as shown in many studies (e.g., Botting & Conti-Ramsden, 2001; Conti-Ramsden & Hesketh, 2003; Dollaghan & Campbell, 1998; Weismer et al., 2000; Gathercole & Baddeley, 1990b; Gray, 2003). Therefore, WNRep might be a potential clinical marker for language impairment in Gulf Arabic. This finding, however, cannot be generalized, as the CL group in the current study may not be representative of all children at risk of language impairment in Arabic. Most of the times, and due to the lack of standardised assessment tools in Arabic, only children with moderate to severe language impairment are referred to assessment or discovered by their parents to have language deficits. Therefore, it is recommended that the utility of the WNRep test in identifying children at risk of language impairment should warrant further investigation with a larger sample of children. It is recommended to investigate the diagnostic accuracy of the WNRep in a population study that includes a full range of ages. The preliminary results of WNRep might be consistent with the previous studies that investigated NWR as a clinical marker for children with language impairment, however it is still not clear what the factors that make some nonword tests succeed in discriminating children with language impairment better than others are. Some factors might affect the utility of nonword repetition as a clinical marker, such as test characteristics and design (e.g., wordlikeness, nonword length and nonword complexity, numbers of participants, age, and type of language impairment). For example, Graf Estes et al. (2007) reported that the CNRep test (Gathercole & Baddeley, 1996) was more successful in differentiating children with SLI from TD children when it was compared to another NWR test, namely the NRT (Dollaghan & Campbell, 1998).

On the other hand, the real word repetition task in the WNRep showed similar results to those found in NWR, although the items in this task were familiar and real words. Just like nonword repetition, it seems that real word repetition could be a potential clinical marker in children with language impairment in Gulf Arabic, as these children scored significantly lower than the typically developing children. This may support the findings reported by Dispaldro et al. (2013) who argued that real word repetition can potentially be a useful diagnostic assessment tool to distinguish Italian TD children from children with SLI, despite some fundamental differences in the two tasks. Dispaldro et al. (2013) argued that the main skills required for word repetition differ from those required for nonword repetition. Real word repetition relies on existing phonological and semantic knowledge accumulated from learning previous vocabulary, while nonword repetition relies mostly on phonological short-term memory (PSTM). However, both tasks share some skills, such as speech perception, oral-motor skills, use of PSTM and lexical and linguistic knowledge. These shared skills probably allowed real word repetition to succeed as a diagnostic tool that distinguished the CL group from the TD group in Gulf Arabic speaking children.

The results of real word repetition could also be attributed to typological differences in Arabic. Arabic is a root and pattern language, and to form new words children may use their knowledge about root derivation, where one root is manipulated extensively to produce various items that are semantically related to the root. In Arabic new words are strongly related to children's lexical knowledge and this may explain why real word repetition could be effectively used as a clinical marker that could predict vocabulary knowledge.

4.4 Relations with receptive and expressive vocabulary

The significant correlation between the Arabic Picture Vocabulary Test (APVT) and nonword repetition reported in typically developing children found here is consistent with several studies (Briscoe et al., 2001; Coady & Evans, 2008; Gathercole et al., 1991, 1992). However, in contrast to our results, some studies found no correlation between expressive vocabulary and nonword repetition (e.g. Briscoe et al., 2001; Conti-Ramsden, Botting, & Faragher, 2001; Stokes et al., 2006).

The CL group showed weak correlations between performance on word and nonword repetition and their scores on the APVT and AEVT. This was attributed to the presence of three children from the CL group who performed in line with average scores of the TD children on receptive and expressive vocabulary. When the correlation analysis was conducted again excluding the three children, results showed significant correlations between word and nonword repetition and the APVT and the AEVT.

The small clinical (CL) group sample could also be a reason why the correlation between the repetition tasks and the vocabulary tasks was not significant. Having a larger sample may help overcome the individual differences among participants, especially that the performance of each child with language impairment may vary across different tests. Though the criteria of selecting the CL group in this study was strict and similar to the criteria used in many studies (e.g., Chiat & Roy, 2007; Shaalan, 2010), the duration of speech therapy received was not considered in the criteria.

4.5 Clinical implications

This study has some clinical implications that could help in assessing and identifying children with language impairment, especially in the domain of word and nonword repetition and vocabulary skills. This study also evaluates the utility and efficiency of using different scoring methods for NWR and the relationship between the two NWR tasks and expressive and receptive vocabulary tests in Gulf Arabic. Examination of the development of phonological skills as represented by NWR across different ages in Gulf Arabic shows that they are in line with those reported in other languages.

The WNRep is a short and easy assessment tool that can be used with young children with less effort and time, especially when it is compared to other speech and language assessment tests. Furthermore, the WNRep could be useful as a screening tool that can be used in schools and kindergartens by teachers. The test showed good predictive values for receptive and expressive vocabulary for the TD group and it is expected to have similar predictive values for the CL if it is conducted with a larger group. Testing larger groups will aid the assessment of test utility and more measures are needed to examine the reliability and validity of these tests.

One should be cautious when interpreting these results and more investigation is required to replicate these findings and examine the utility of this test. One reason for this is that the children included in the clinical group (CL group) are not representative of all children at risk of language impairment. These children tend to have severe difficulties and their language deficits are more observable by parents and clinicians. We know that children with language impairment have varying levels of difficulties and some may even perform within normal range on some tests of nonword repetition. Some studies that examined larger numbers of children (some of whom met the criteria of SLI) found that some of these children performed within normal range on nonword repetition and some of the TD children were found to do poorly on nonword repetition (see Weismer et al., 2000; Gathercole, 2006 for a review and discussion of this). Therefore, while NWR is a good tool to identify children at risk of SLI or language impairment, we cannot conclude that it is sufficient to rule language impairment in our out. However, it is one of the tools that should be combined with other assessment tools.

Two scoring methods were used in this study, namely the whole word correct method (WWC) and the percentage of phonemes correct (PPC). The aim of using these two different scoring methods was to investigate if this would make any difference in analyzing the output of the word and NWR tests. Results showed there was no significant contribution of the scoring methods employed to the groups' results on the word and NWR test. Therefore, it is recommended in clinical settings to use the WWC scoring method as it is easier to apply and less time consuming. On the other hand, using the PPC scoring method could be useful for research purposes as it is more informative about details of children's response to nonword repetition.

5. Conclusion

This study examined the viability of a word and nonword repetition test (WNRep) as a diagnostic tool to distinguish Gulf Arabic speaking children with language impairment (CL) from typically developing (TD) peers. Results of the WNRep showed that this test could be a potentially useful diagnostic tool. Furthermore, both groups' performance on WNRep significantly correlated with the Arabic receptive and expressive vocabulary tests. Comparing the results of the two scoring methods used in the present study showed no significant difference between the methods.

Our results provide some initial information about the performance of TD Gulf Arabic speaking children and those with language impairment on word and nonword repetition and how their performance on these tasks correlates with receptive and expressive vocabulary tests. These results show significant effects of wordlikeness as children had better NWR scores on words vs. nonwords. However, wordlikeness effects in Arabic are influenced by non-liner morphology and therefore it is important to carefully examine the effects of both roots and patterns in NWR and how each one of them influence NWR.

References

Alqattan, S. (2015). Early phonological acquisition by Kuwaiti Arabic children. Ph.D. dissertation, University of Newcastle, UK.

Amayreh, M. M. (2003). Completion of the consonant inventory of Arabic. *Journal of Speech, Language, and Hearing Research*, 46, 517–529. https://doi.org/10.1044/1092-4388(2003/042)

Amayreh, M. M., & Dyson, A. T. (2000). Phonetic inventories of young Arabic-speaking children. *Clinical Linguistics & Phonetics*, 14, 193–215. https://doi.org/10.1080/026992000298823

Archibald, L., & Gathercole, S. (2006). Nonword repetition: a comparison of tests. *Journal of Speech, Language, and Hearing Research*, 49, 970–983. https://doi.org/10.1044/1092-4388(2006/070)

Ayyad, H. (2011). Phonological development of typically developing Kuwaiti-Arabic speaking pre-schoolers. Ph.D. dissertation, University of British Columbia, Canada.

Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, 36, 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 8, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Bishop, D. V. M. (1997). *Uncommon understanding: Comprehension in specific language impairment*. Hove: Psychology Press.

Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders*, 49, 381–415. https://doi.org/10.1111/1460-6984.12101

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry*, 37, 391–403. https://doi.org/10.1111/j.1469-7610.1996.tb01420.x

Bishop, D. V. M., Snowling, M. J., Thomspon, P. A., Greenhalgh, T., & The CATALISE Consortium. (2016). CATALISE: a multinational and multidisciplinary Delphi consensus study: identifying language impairment in children. *PLOS One*, 11, 7. https://doi.org/10.1371/journal.pone.0158753

Botting, N., & Conti-Ramsden, G. (2001). Nonword repetition and language development in children with specific language impairment (SLI). *International Journal of Language and Communication Disorders*, 36, 421–432. https://doi.org/10.1080/13682820110074971

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology*, 63, 44–78. https://doi.org/10.1006/jecp.1996.0042

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics*, 22, 441–469. https://doi.org/10.1017/S0142716401003083

Briscoe, J. M., Bishop, D. V. M., & Norbury, C. (2001). Phonological processing, language, and literacy: a comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry*, 42, 329–340. https://doi.org/10.1111/1469-7610.00726

Bukshaisha, F. A. (1985). An experimental phonetic study of some aspects of Qatari Arabic. Ph.D. dissertation, University of Edinburgh, UK.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes*, 16(2–3), 113–142. https://doi.org/10.1080/01690960042000012

Chiat, S., & Roy, P. (2007). The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language and Hearing Research*, 50, 429–443. https://doi.org/10.1044/1092-4388(2007/030)

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, 43, 1–40. https://doi.org/10.1080/13682820601116485

Conti-Ramsden, G., & Hesketh, A. (2003). Risk markers for SLI: a study of young language-learning children. *International Journal of Language & Communication Disorders*, 38, 251–263. https://doi.org/10.1080/136828203100092339

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment. *Journal of Child Psychology and Psychiatry*, 6, 741–748. https://doi.org/10.1111/1469-7610.00770

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185. https://doi.org/10.1017/S0140525X01003922

DeVellis, R. F. (2012). *Scale development: Theory and applications*. Los Angeles: Sage Publications.

Delattre, P. (1971). Consonant gemination in four languages: An acoustic, perceptual, and radiographic study: I. *IRAL: International Review of Applied Linguistics in Language Teaching*, 9, 31–52. https://doi.org/10.1515/iral.1971.9.2.97

Dispaldro, M., Benelli, B., Marcolini, S., & Stella, G. (2009). Real-word repetition as a predictor of grammatical competence in Italian children with typical language development. *International Journal of Language & Communication Disorders*, 44, 941–961. https://doi.org/10.1080/13682820802491794

Dispaldro, M., Deevy, P., Altoé, G., Benelli, B., & Leonard, L. B. (2011). A cross-linguistic study of real-word and non-word repetition as predictors of grammatical competence in children with typical language development. *International Journal of Language & Communication Disorders*, 46, 564–578. https://doi.org/10.1111/j.1460-6984.2011.00008.x

Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-word and nonword repetition in Italian-speaking children with specific language impairment: A study of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, 56, 323–336. https://doi.org/10.1044/1092-4388(2012/11-0304)

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics*, 16, 211–222. https://doi.org/10.1017/S0142716400007098

Dollaghan, C., & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41, 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (Third Ed.). Circle Pines, MN: American Guidance Service.

Ebbels, S. (2014). Introducing the SLI debate. *International Journal of Language and Communication Disorders*, 49, 337–380. https://doi.org/10.1111/1460-6984.12119

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of speech, language, and hearing research*, 47, 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Frisch, A., & Zawaydeh, B. (2001). The psychological reality of OCP-Place in Arabic. *Language*, 77, 91–106. https://doi.org/10.1353/lan.2001.0014

Frisch, S., Pierrehumbert, B., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory*, 22, 179–228. https://doi.org/10.1023/B:NALA.0000005557.78535.3c

Gallagher, A. (2014). What's in a name? Some thoughts on Reilly *et al.* (2014). Commentary on Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J. M. and Wake, M., 2014, Specific language impairment: a convenient label for whom? *International Journal of Language and Communication Disorders*, 49, 416–451. https://doi.org/10.1111/1460-6984.12102

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition*, 23, 83–94. https://doi.org/10.3758/BF03210559

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27, 513–543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28, 200–213. https://doi.org/10.1016/0749-596X(89)90044-2

Gathercole, S. E., & Baddeley, A. D. (1990a). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29, 336–360. https://doi.org/10.1016/0749-596X(90)90004-J

Gathercole, S. E., & Baddeley, A. D. (1990b). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, 81, 439–454. https://doi.org/10.1111/j.2044-8295.1990.tb02371.x

Gathercole, S. E., & Baddeley, A. D. (1996). *The children's test of nonword repetition*. The Psychological Corporation.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, 12, 349–367. https://doi.org/10.1017/S0142716400009267

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28, 887–898. https://doi.org/10.1037/0012-1649.28.5.887

Gathercole, S. E., & Adams, A. M. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language*, 33, 672–688. https://doi.org/10.1006/jmla.1994.1032

Gathercole, S. E., Willis, C., Baddeley, A. D. & Emslie, H. (1994). The children's test of nonword repetition: a test of phonological working memory. *Memory*, 2, 103–127. https://doi.org/10.1080/09658219408258940

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999a). Phonotactic Influences on Short-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25. 84–95. https://doi.org/10.1037/0278-7393.25.1.84

Gathercole, S. E., Service, E., Hitch, G., Adams, A.-M., & Martin, A. (1999b). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, 13, 65–77. https://doi.org/10.1002/(SICI)1099-0720(199902)13:1<65::AID-ACP548>3.0.CO;2-O

Graf Estes, K., Evans, J., & Else-Quest, N. (2007). Differences in nonword repetitions performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50, 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Gray, S. (2003). Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, 36, 129–151. https://doi.org/10.1016/S0021-9924(03)00003-0

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, 59, 267–333. https://doi.org/10.1006/brln.1997.1819

Hoff, E., Core, C., & Bridges, K. (2008). Nonword repetition assesses phonological memory and is related to vocabulary development in 20-to 24-month-olds. *Journal of Child Language*, 35, 903–916. https://doi.org/10.1017/S0305000908008751

Horohov, J. E., & Oetting, J. B. (2004). Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics*, 25(1), 43–65. https://doi.org/10.1017/S0142716404001031

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

Kohnert, K., Windsor, J., & Ebert, K. D. (2009). Primary or "specific" language impairment and children learning a second language. *Brain and language*, 109, 101–111. https://doi.org/10.1016/j.bandl.2008.01.009

Leonard, L. (1998). *Children with language impairment*. Cambridge, MA: MIT Press.

Martin, N. A., & Brownell, R. (2000). *Expressive one-word picture vocabulary test-4*. Novato, CA: Academic Therapy Publications.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, 91, 3–19. https://doi.org/10.1037/0022-0663.91.1.3

Mustafa, I., AlZayaat, A., Hamed, A. Q., & AlNajaar, M. (2004). *AL Mu'jam AL Waseet*. Cairo: Maktabat al-Shorouq al-Dawliyah.

Ragheb, M., & Davis, S. (2010). The acquisition of word-final clusters in Cairense Arabic: An OT analysis. Paper presented in *Arabic Linguistics Symposium*, Austin, TX.

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F., Morgan, A., Goldfeld, S., Nicholson, J., & Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language and Communication Disorders*, 49, 416–451. https://doi.org/10.1111/1460-6984.12102

Rice, M. L. (2014). Advocating for SLI. Commentary on Bishop, D. V. M., Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders*, 49, 381–415. https://doi.org/10.1111/1460-6984.12101

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2-to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language and Hearing Research*, 47, 223–234. https://doi.org/10.1044/1092-4388(2004/019)

Santos, F. H. D., Bueno, O. F. A., & Gathercole, S. E. (2006). Errors in nonword repetition: bridging short-and long-term memory. *Brazilian Journal of Medical and Biological Research*, 39, 371–385. https://doi.org/10.1590/S0100-879X2006000300008

Shaalan, S. (2010). Investigating grammatical complexity in Gulf Arabic speaking children with Specific Language Impairment (SLI). Ph.D. Dissertation, University College London, UK.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords and phonological processes: Some comments on Gathercole, Willis, Emslie and Baddeley. *Applied Psycholinguistics*, 12, 369–373. https://doi.org/10.1017/S0142716400009279

Stokes, F., & Klee, K. (2009). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *Journal of Speech, Language and Hearing Research*, 52 (4), 872–882. https://doi.org/10.1044/1092-4388(2009/08-0030)

Stokes, S. F., Wong, A. M., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research*, 49, 219–236. https://doi.org/10.1044/1092-4388(2006/019)

Taylor, C. L. (2014). The SLI construct is a critical link to the past and a bridge to the future. Commentary on Bishop, D. V. M., Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders*, 49, 381–415. https://doi.org/10.1111/1460-6984.12101

Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 43, 865–878. https://doi.org/10.1044/jslhr.4304.865

An exploratory longitudinal study of vocabulary development in bilingually exposed children with autism spectrum disorder (ASD) in the United Arab Emirates

Saleh Shaalan, Kerry Egan, Daniel Gould and Pam Olsen
Mohammed Bin Rashid Center for Special Education –
Operated by The New England Center for Children©

This is a longitudinal study of the vocabulary development of 29 children with autism spectrum disorder (ASD) in the UAE, who are aged 44 to 77 months and enrolled in a bilingual intensive intervention program. These children were assessed at entry, 6 months, 12 months, 18 months, and 30 months using receptive and expressive vocabulary tests in Arabic and English. MANOVA results and follow-up tests showed significant increases in vocabulary skills across times and in both languages. Pearson correlation coefficients showed strong positive correlations between the four vocabulary tests. Results confirmed findings of no detrimental effects of bilingual intervention reported in other languages. Moreover, results showed that speaking abilities at the start of the program did not correlate with their expressive vocabulary at the end of the study. We discuss limitations of this study and some considerations when conducting similar research in the future.

Keywords: bilingualism, ASD (autism spectrum disorders), vocabulary

1. Introduction

The United Arab Emirates (UAE) is a very cosmopolitan country with a total population of 9.6 million inhabitants (World Bank, 2019) with a majority of expatriates coming from various parts of the world. Figures from Abu Dhabi and Dubai put the percentage of non-Emiratis at 81% and 92% respectively (Statistics Centre Abu Dhabi, 2017; Dubai Statistics Center, 2018). The rest are UAE citizens, who mostly speak Emirati Arabic. The high percentage of expatriate population is also found

https://doi.org/10.1075/sal.10.08sha

in other Gulf Cooperation Council countries (GCC), which witnessed tremendous changes in their economies due the discovery of oil in the last century. This led to rapid socio-economic changes, as reflected in the significant increases in income accompanied by rapid developments and a continuing influx of expatriates from all over the world.

Most government schools in the UAE teach Arabic and English from the kindergarten and English is widely used in businesses, universities, and in the community, especially in big urban centers like Dubai and Abu Dhabi. Although Arabic is the official language of the Government of the UAE, English is considered by many as the lingua franca of the UAE (Hopkyns, 2017). Moreover, many parents choose to send their children to international schools, most of which use English as the main medium of instruction (Al-Issa & Dahan, 2011).

These unique cultural and linguistic characteristics of the UAE context raise many challenges to service providers in various sectors, especially in health and education (El-Amouri & O'Neill, 2011; Raddawi & Meslem, 2015). One area where this cultural and linguistic diversity can pose some challenges is the identification, assessment, and intervention practices for individuals with language impairments in general and those with autism spectrum disorder (ASD) in particular.

1.1 Autism spectrum disorder

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that is characterized by significant deficits in social interaction with the presence of repetitive, restricted, and stereotypic behaviors (American Psychiatric Association, 2013). It exists across races and ethnicities. Figures from the US put the prevalence rate of ASD at 1 in 59 children (Baio, Wiggins, Christensen, et al., 2018). One study of ASD in the United Arab Emirates (UAE) estimated that 1 out of every 146 children in Dubai is diagnosed with ASD (Al-Abbady, Hessian, & Alaam, 2017). However, more recent and comprehensive epidemiological studies of the prevalence of ASD in the UAE in general are needed.

Deficits in social communication and interaction are a major component in the diagnostic criteria of ASD (APA, 2013). Children with ASD present with a large amount of variability in terms of their language development. Most of these children speak later than their typically developing peers and a minority of them have adequate speech and language skills (Kim, Paul, Tager-Flusberg, & Lord, (2014). It is estimated that 20–50% of individuals with autism do not develop functional speech, (Kim et al., 2014) although these numbers are affected by the characteristics of the children being studied (e.g., whether they were enrolled in an early intervention program) and how functional speech is defined. While children with autism with functional speech seem not to differ from typically developing children in articulation skills and syntactic development (Kim et al., 2014), their language skills are characterized by deficits in pronouns (pronouns reversals), word use, and the presence of some odd intonational patterns (such as monotonous speech) and vocal quality. Their language is also characterized by the presence of significant receptive language deficits (relative to their expressive language skills) (See Kim et al., 2014 for an overview of language skills in individuals with ASD).

1.2 Autism spectrum disorder and bilingualism

Bilingualism is defined as the regular use of two or more languages in everyday life (Grosjean, 1992). Bilingualism is a multidimensional phenomenon that is influenced by a multitude of factors, such as timing and amount of exposure to the two languages, support provided in each language, and social status of each language (e.g., majority language vs home language) among others. Studies of language development in typically developing bilingual children did not find that these children lag behind their typically developing peers in terms of various aspects of their language development milestones, when they are provided with adequate exposure in the two languages (Bialystok, 2001; De Houwer, 2009; and Genesee, 2006). Although typically developing (TD) bilinguals may have less vocabulary compared to their TD monolingual peers in each language in the first three years (Pearson, Fernandez, & Oller, 1993), they are not different when their conceptual vocabulary (the total number of unique vocabulary items) from their two languages are combined.

Until the last decade, there was a paucity of studies that examined the language skills of individuals with ASD who come from multilingual backgrounds and such studies remain limited in their numbers. Kay-Raining Bird, Genesee, and Verhoeven (2016) reviewed studies that examined language development in three populations: those with developmental language disorder (previously known as specific language impairment, SLI), Down Syndrome, and autism spectrum disorders (ASD) and found that bilingual individuals with developmental disorders were overall not different from their monolingual peers with developmental disorders. Kay-Raining Bird, Lamond, and Holden (2012) used parent questionnaires to examine language abilities of bilingual individuals with ASD between 2 and 22 years old with varying abilities and they found that monolingual and bilingual children with ASD had similar abilities in the areas of expressive and receptive language skills, and literacy skills. Hambly and Fombonne (2012) used the Vineland Adaptive Behavior Scales-II (VABS-II; Sparrow, Cicchetti, & Balla, 2005) to examine the language abilities of simultaneous and sequential bilingual children with ASD and found that they had similar language abilities to monolingual children with ASD. Similar results were reported by Ohashi, Mirenda, Marinova-Todd, et al.

(2012); Petersen, Marinova-Todd, and Mirenda (2012); and Valicenti-McDermott et al. (2013). In a review of language development in bilingual children with ASD, Lund, Kohlmeier, and Durán (2017) found only seven empirical studies that met their inclusion criteria, and which compared monolingual and bilingual children with ASD. They concluded that there was no difference between these two groups of children with the possibility that bilingual children with ASD having a slight advantage in vocabulary development.

There are very few studies that compared vocabulary skills of bilingual children with ASD to those of monolingual children with ASD (Hambly & Fombonne, 2012, 2014; Petersen et al., 2012). Petersen et al. (2012) studied 14 English-Chinese bilinguals preschoolers and 14 English monolingual preschoolers with ASD and found no significant difference in their vocabulary size in both production and comprehension. Hambly and Fombonne (2012) compared the vocabulary skills of monolingual children with ASD with simultaneous and sequential bilinguals with ASD using the McCarther Communicative Development Inventory (MCDI, Fenson et al., 1993) and found no significant difference between the three groups of children. When examining the various factors influencing bilingual expressive vocabulary size in multilingual children, Hambly and Fombonne (2014) found that recent direct language exposure estimates accounted for 69% of the variance in the second language vocabulary, while indirect language (passive) exposure did not guarantee expressive bilingual vocabulary. Therefore, they argued that current language exposure might relate more strongly to bilingual vocabulary development than to the onset of bilingualism (simultaneous vs. sequential).

In Gulf Emirati Arabic and in other varieties of Gulf Arabic, we do not have information about language development in general and vocabulary development in particular in monolingual or bilingual children with ASD. Even when it comes to monolingual TD children, we do not have enough information about their vocabulary development.

Despite the lack of evidence for negative effects of bilingualism on various aspects of the language abilities of children with autism and despite the recommendations by professional organizations that support multilingualism in individuals with ASD and recommend not changing the home language if it differs from the school/community language (ASHA, 2014; Fredman, 2011), this has not fully transformed into the majority of clinical practices. Studies that looked into ASD in individuals from multilingual backgrounds reported that parents were commonly advised to use only one language with their child with autism (Drysdale, van der Meer, & Kagohara (2015); Kay-Raining Bird et al. (2012); Kremer-Sadlik (2005). This practice has been encountered when dealing with children with ASD in the UAE.

In this chapter we report the results of an exploratory longitudinal study that follows the vocabulary development of 29 bilingually (Arabic/English) exposed

Emirati children with ASD in the UAE. These children are enrolled in an intensive program for children with ASD and related disorders. This Abu-Dhabi Government funded-program is one of the rare programs where instruction is provided bilingually (English and Arabic) to children with ASD by qualified professionals who speak Arabic and English at native or near-native levels of proficiency.

In this longitudinal study we examined the patterns of bilingual vocabulary development in these Emirati Arabic speaking children with autism who were enrolled in a bilingual intensive behavioural program in the UAE. We examined if the findings reported in most studies in other languages were replicated, i.e., whether bilingual instruction would have no detrimental effect on their vocabulary skills.

It is expected that these children come to the program with varying speaking abilities, as some will be Non-speaking (non-verbal), others will be Minimally Speaking while at the end of continuum some of these children will be Speaking (verbal). So, we attempted to assess whether the speaking ability of the students would make a difference when it comes to bilingual instruction for children with ASD in the UAE. We examined if their speaking ability (measured based on the number of words produced on the English or Arabic expressive vocabulary tests) at the time of joining the program would correlate with their expressive vocabulary test scores 30 months from the start of the intervention program.

Finally, this exploratory study pinpointed to areas of further research in the provision of bilingual intervention for children with ASD in general, and those in the UAE and the GCC countries in particular due to the unique social, demographic, and linguistic characteristics of these countries.

2. Methods

This is a descriptive longitudinal study where we follow 29 UAE children with autism for 30 months from the time they join a bilingual behavioural intervention program to examine their vocabulary skills in Arabic and English. Due to the unique nature of the program, it was not possible to find a control group since most autism programs in the UAE offer their services in Arabic or English only.

2.1 Participants

Twenty-nine Emirati students with ASD participated in the study (21 male and 8 female[1]). Students' mean age was 63 months (5 years and 3 months) with a range from 44 to 77 months, see Table 1 for a summary of participants' characteristics.

1. IRB approvals were obtained for this study.

Table 1. Summary of the characteristics of participants (n = 29) and their baseline performance (when they entered the program)

	Age (months)	Vocabulary Tests Raw Scores at Entry				Speaking Abilities at Entry		
		ER	EE	AR	AE	Non-speaking	Minimally speaking	Speaking
M	63.6	13.2	9.4	9.8	6.8	12	4	13
SD	10.0	22.5	18.5	16.4	11.3			
Range	44–77	0–87	0–77	0–68	0–48			

Note. *ER = English receptive vocabulary test (raw scores from the ROWPVT-4).*
EE=English expressive vocabulary test (raw scores from the EOWPVT-4)
AR = Arabic receptive vocabulary scores (raw scores from the APVT, Shaalan, 2010)
AE=Arabic expressive vocabulary test (raw scores from the Arabic Expressive Vocabulary Test (AEVT, Khater, 2018).

Twenty-eight of these students came from Emirati speaking households and one came from a household where they speak another variety of Arabic (Levantine Arabic), however this student's family reported they spoke with him in English only. The students were successive admissions who belonged to three different cohorts that joined the program in three consecutive periods from Spring 2015 to Fall 2016. Students joined the study as long as they met the following criteria:

1. They were diagnosed with autism by a certified physician in the UAE and their diagnosis was confirmed by a clinical psychologist.
2. These children with autism did not have other significant concomitant disorders.
3. They were aged between 3 and 6;6 (6 years and 6 months) years old.

Thirty-seven students met these criteria when they joined the program during this period. Initially two were excluded based on the presence of significant accompanying medical conditions (cerebral palsy and brain atrophy) and of the 35 remaining students six students left the study because they were discharged from the program at various stages of the project. Their data was not included in the final results. Therefore, the results reported here are of 29 students only.

2.1.1 *Speaking ability criteria and participants' bilingual exposure*

All participants' communication skills were thoroughly assessed when enrolled at the program and their speaking abilities were classified into three categories. Non-speaking (non-verbal), Minimally Speaking (minimally verbal), and Speaking (verbal). The criteria adopted were modified from Lord, Risi, and Pickles (2004) and were as follows: Children who used less than five spontaneously produced, different words per day (in either English or Arabic) were classified as Non-speaking. Children who used more than five words but less than 20 words on a daily basis were considered Minimally Speaking, and those with more than 20 words per day were considered Speaking. Words produced in echolalia and in vocal stereotypes were not counted. The distribution of these students by their speaking abilities is shown in Table 1.

A questionnaire that was adapted from COST Action (2011), which in turn was based on ALDeQ (Paradis, Emmerzael, & Sorenson Duncan, 2010) and ALEQ (Paradis, 2011), was distributed to parents. Only 16 parents filled in this questionnaire which examined the patterns of language use and communicative partners of these students in both languages. Out of these 16 parents, two (12.5%) reported that they changed the language they used with their children from Arabic to English following advice they received from health professionals. Four parents (25%) reported that they changed the home language from Arabic only to English and Arabic after their children were diagnosed with ASD and three parents (18.7%) mentioned they already used Arabic and English at home before their children were diagnosed. Seven parents (43.7%) reported that they used Arabic only before and after diagnosis.

In multilingual homes, parents tended to speak Arabic with their children, while their children interacted in English with domestic helpers and sometimes with their siblings who mostly go to English medium schools. Even in Arabic Speaking only homes, most parents reported that their children interacted with their domestic helpers at home in English. We did not collect information about the language proficiency of these domestic helpers, however most of them speak nonnative English or nonnative Arabic, or a variety known as Pidgin Gulf Arabic (Smart, 1990), characterized by lack of many grammatical and lexical complexities of Arabic or English.

2.2 Procedures

To measure the effects of bilingual instruction on the vocabulary skills of Emirati children with autism, each child was tested using four vocabulary tests in English and Arabic. Testing was conducted upon entry into the program (on average two weeks after joining the program) and before developing their individual educational plans. Children were given this time to make sure they were familiar with testing settings in order to maximize their collaboration. The following tests were used to assess students' vocabulary in English and Arabic.

2.2.1 *Measures*

The Receptive One Word Picture Vocabulary Test -4th Ed. (ROWPVT-4, Martin & Brownell, 2010). This test was designed for individuals between 2 years old and adults in the United States and the norms were based on US population. The test consisted of 190 items.

The Expressive One Word Picture Vocabulary Test-4th Ed. (EOWPVT-4) (Martin & Brownell, 2010). This test was also designed and normed for English speaking individuals in the US. It was designed for individuals between two years old and adults. This test was co-normed with the ROWPVT-4 and consisted of 190 items.

The Arabic Picture Vocabulary Test (APVT) (Shaalan, 2010, 2017). This is a receptive vocabulary test that was developed for Gulf Arabic and a few words were changed to match Emirati Gulf Arabic. The test was designed to be used with children aged between 2 and 10 years old and consisted of 132 items.

The Arabic Expressive Vocabulary Test (AEVT; Khater, 2018). This is an expressive vocabulary test and the administration is similar to the EOWPVT (Martin & Brownell, 2010). It was designed for Gulf Arabic speaking children aged between 2 and 10 years old and consisted of 64 items.

The receptive vocabulary tests consisted of presenting the student with four pictures on one page and the student was required to point to the picture that matched the one produced by the speech-language pathologist. When administering expressive vocabulary tests, each student was presented with one picture only and s/he was asked to name it (e.g., a student was presented with a picture of 'ball' and was asked 'what is this?"). Administration was similar to the standard procedures of the original tests with the exception that examiners were allowed to repeat a test item when the student was not paying attention.

For the English receptive and expressive vocabulary tests no norms were reported as these tests were normed on a different population, therefore we opted to report raw scores only. Raw scores indicated the number of items the student correctly identified or labelled and therefore it was possible to compare students' performance across different times based on the number of pictures the students identified or labelled. Due to the lack of standardized tests in Emirati Arabic, we also reported raw scores instead of standard scores in the Arabic receptive and expressive vocabulary tests.

Each student was tested at five different points in time following entering the program: at entry, 6 months, 12 months, 18 months, and 36 months. All the tests were conducted by qualified speech-language pathologists.

Families were not asked to change the way they interact with their children. However, when they asked about the position of the program on bilingualism, we encouraged these families to use their home language (which was mostly Emirati Arabic) with their children and we alleviated their concerns about the alleged negative impact of bilingualism on children with autism.

2.2.2 *About the intervention program*

The intervention program was guided by evidence-based practices in the field of applied behavior analysis (ABA), such as the use of discrete trial teaching, incidental and natural environment teaching, and task analysis (Cooper, Heron, &

Heward, 2007). The program consisted of 30 hours of instruction per week where students received therapy based on individualized educational plans (IEPs) that were designed by the clinical teams and approved by the students' parents. Each student had an IEP that comprised skills in the following areas: discrimination, communication skills, activities of daily living, self-help skills, academic skills, social skills, and community skills. Specific training on vocabulary instruction was provided based on student's needs and functioning and therefore not all of them received direct instruction on vocabulary learning. All communication and vocabulary objectives were in Emirati Gulf Arabic (for Arabic objectives) and American English (for English objectives). The only exception was when working on reading or academic objectives in Arabic where Modern Standard Arabic (MSA) was used.

The therapy consisted of 1:1 and group instruction and was carried out in classrooms, mobility rooms, cafeteria, gym and the pool, and in community settings. Students were based in different classes at the program. Each classroom comprised 6 students and 6 teachers and a supervisor. Approximately 40% of the teachers were Arabic speakers and 60% were English speaking teachers and the ratio differed among classrooms. Most of the direct therapy was conducted by classroom teachers who received extensive training in how to teach children with autism. The other members of the clinical team consisted of board-certified behavior analysts, speech-language pathologists, occupational therapists, and special education teachers. The role of speech-language pathologists was as consultants to the teachers and they did not provide direct therapy using pull-out model of service delivery. Teachers were not given specific instruction and generally they were not aware of student's participation in the study. It is estimated that students were exposed to Arabic 30% of the time at school and 70% to English. The decision to provide bilingual intervention was based on best practices in bilingual services for children with developmental disorders, where home language should be supported and changed (Fredman, 2011) and bilingual intervention is preferred over monolingual intervention (Ebert, Kohnert, Pham, Disher, & Payesteh, 2014; Kohnert, Yim, Nett, Kan, & Duran, 2005; Thordardottir, 2010).

3. Results

3.1 Descriptive statistics

Table 2 shows the performance of all participants on the four vocabulary tests at the following times: entry, six months, 12 months, 18 months, and 30 months. No standard scores were reported throughout this study as none of the four tests had standardized scores for this population (bilingually exposed children with ASD in the UAE).

Figures 1–4 illustrate the performance of all participants on each test. Since the four tests differ in their design and the age groups they target, their raw scores are not comparable, therefore we preferred to display the graph for each test separately.

The data from the descriptive statistics and Graphs 1–4 show a steady increase in the vocabulary skills of the group of participants with clearly rising slopes, especially in the English receptive and expressive vocabulary skills and the Arabic receptive vocabulary skills. On the Arabic expressive vocabulary test the slope is less sharp when compared to the other three tests. Table 2 shows the high variability in students' performances on the vocabulary tests as shown in their high standard deviations across all tests and very wide range of scores seen in all tests.

Table 2. Descriptive statistics for all participants (*n* = 29)

		ER	EE	AR	AE
Time 1 (entry)					
	M	13.2	9.4	9.8	6.8
	SD	22.5	18.5	16.4	11.3
	Range	0–87	0–77	0–68	0–48
Time 2 (6 months)					
	M	19.9	15.6	12.6	8.3
	SD	24.2	22.5	14.6	12.5
	Range	0–77	0–87	0–56	0–51
Time 3 (12 months)					
	M	26.5	20.9	16.9	10.3
	SD	28.3	27.4	19.3	14.2
	Range	0–107	0–101	0–89	0–53
Time 4 (18 months)					
	M	28.5	24.6	19.3	11.8
	SD	26.5	30.1	21.0	14.7
	Range	0–90	0–110	0–87	0–55
Time 5 (30 months)					
	M	33.6	29.4	20.9	14.8
	SD	31.9	32.9	20.3	16.2
	Range	0–109	0–117	0–67	0–60

Note. ER = English receptive vocabulary test (raw scores from the ROWPVT-4). EE = English expressive vocabulary test (raw scores from the EOWPVT-4). AR = Arabic receptive vocabulary scores (raw scores from the APVT, Shaalan, 2010). AE = Arabic expressive vocabulary test (raw scores from the Arabic Expressive Vocabulary Test (AEVT; Khater, 2018).

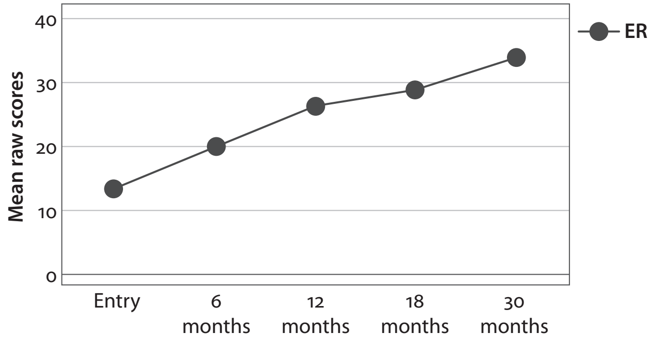


Figure 1. Performance of all participants (n = 29) on the English receptive vocabulary test (ROWPVT)

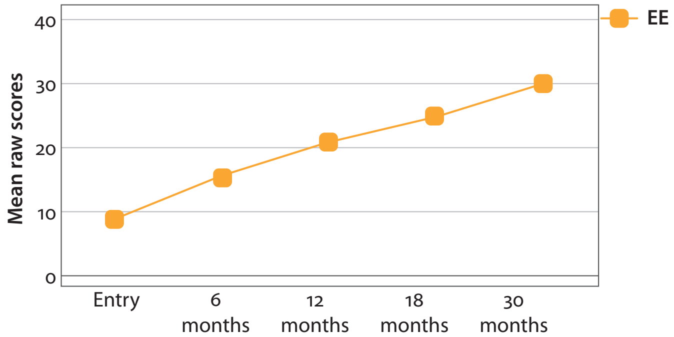


Figure 2. Performance of all participants (n = 29) on the English expressive vocabulary test (ROWPVT)

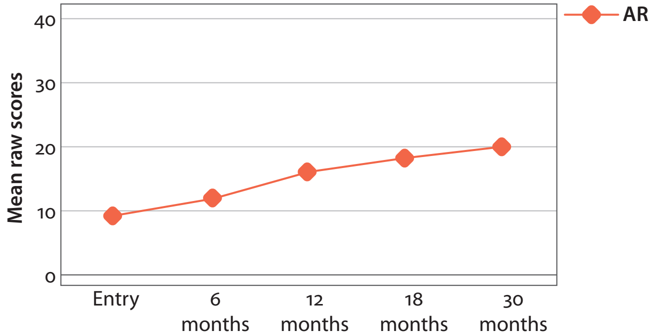


Figure 3. Performance of all children (n = 29) on the Arabic receptive vocabulary test (APVT)

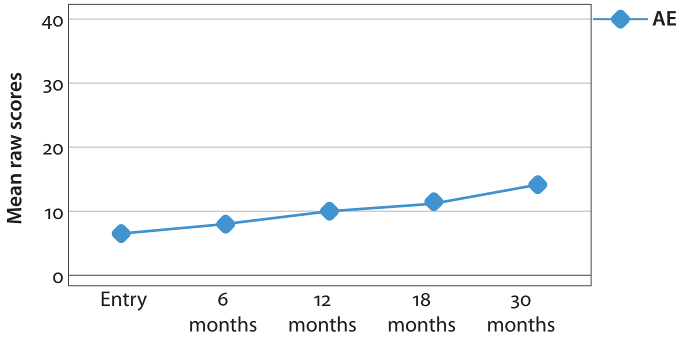


Figure 4. Performance of all children (n = 29) on the Arabic Expressive vocabulary test (AEVT)

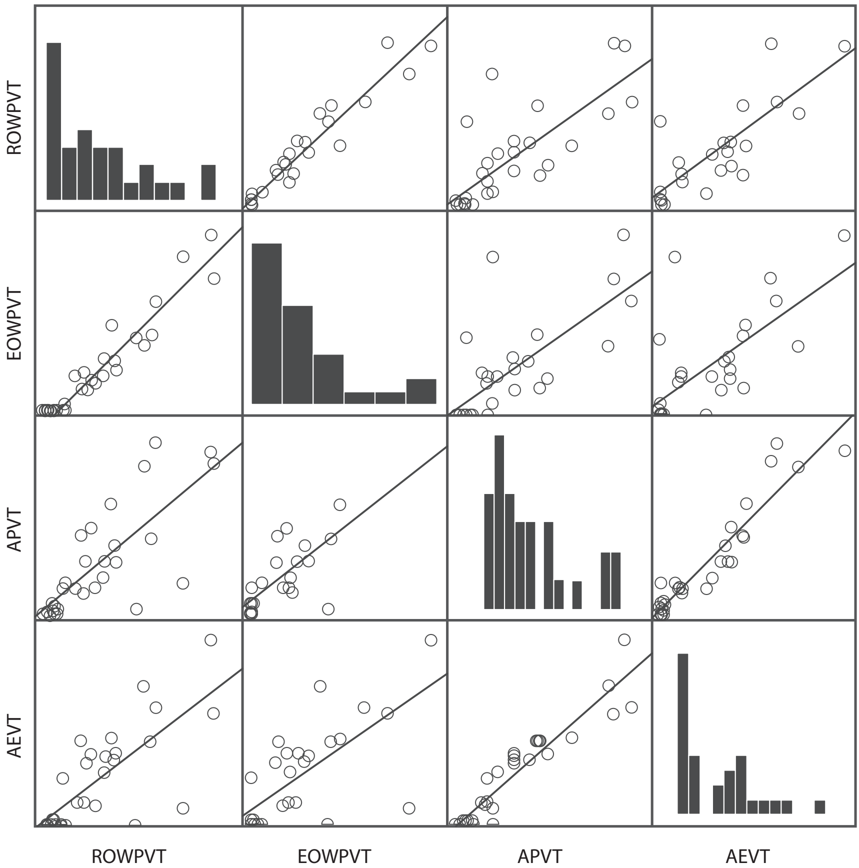


Figure 5. A scatterplot matrix showing the strong positive correlations between the four vocabulary tests in English and Arabic for all participants (n = 29) at 30 months (Time Point 5). ROWPVT = English Receptive vocabulary tests, EOWPVT = English Expressive vocabulary test, APVT = Arabic Receptive vocabulary test, AEVT = Arabic Expressive (AE) vocabulary test

3.2 Correlations between the four vocabulary tests

Table A1 in Appendix A shows the correlations between all the vocabulary tests at the different time points: entry, six months, 12 months, 18 months, and 30 months for all participants ($n = 29$).

The correlations were all significant and positive as increases in one vocabulary test were seen when another vocabulary test increased too and this was seen even among tests from a different language. However, stronger correlations were seen between tests from the same language. For example, the correlation between English Receptive (ER) score at 30 months (time 5) and English Expressive (EE) at the same time was .96, which is considered very high. Similarly, the correlation between Arabic Receptive (AR) and Arabic Expressive (AE) at the same time point was .94. These strong positive correlations between the tests are illustrated in the scatterplot matrix in Figure 5:

3.3 Multivariate analysis of variance (MANOVA)

A one-way repeated measure multivariate analysis of variance (MANOVA) was conducted to determine whether there was a statistically significant difference in vocabulary scores (measured in receptive and expressive vocabulary tests in both languages) over the five time points: entry, 6 months, 12 months, 18 months, and 30 months. Prior to conducting the repeated measure MANOVA, a series of Pearson correlations was conducted between all the dependent variables as MANOVA is based on the assumption that these dependent variables (the four vocabulary test scores) correlated with each other. Table A1 in Appendix A shows a strong correlation between all four tests and across different time points. The one-way repeated measures MANOVA showed there was a statistically significant difference in vocabulary scores $F(16,333.6) = 6.77$, $p < .001$, $\eta^2 = .19$) in this group of participants who are enrolled in a bilingual behavioral intervention program.

A series of one-way ANOVAs were conducted on each of the four vocabulary tests as follow-up tests to the MANOVAs. They were all statistically significant as shown in the following Table 3. Therefore, we can conclude that there was a significant effect of time across the four tests.

Table 3. Results of one-way ANOVAs the four vocabulary tests in Arabic and English

Vocabulary test	ANOVA results	Significance	Effect size
ER	$F(2.7, 77) = 22.28$	$P < .001$	.44
EE	$F(1.6, 44.9) = 22.72$	$P < .001$	44
AR	$F(2.8, 78) = 12.54$	$P < .001$	.31
AE	$F(2.1, 59.5) = 17.19$	$p = .001$	.38

Post-hoc analyses using pairwise comparisons were conducted to compare the means of tests across the five periods (as shown in Appendix B). Results showed that within each test there were significant differences between the means across different time points for most of the times, indicating a strong effect of time on participants' performance. In most cases, the vocabulary scores at the next time point were significantly better than the previous time point, indicating continuous improvement in the scores. On the English receptive (ER) vocabulary tests (ROWPVT), there were significant differences between all means at all different time points (entry, 6 months, 12 months, 18 months, and 30 months). See Table B1 in Appendix B. On the English expressive vocabulary test (EOWPVT), there were significant differences between the mean scores except between means scores at times 3 (12 months) and 4 (18 months), see Table B2 in Appendix B.

On the Arabic receptive (ER) vocabulary test (APVT), there were significant differences between all means except between time 1 (entry) and 2 (6 months). See Table B3 in Appendix B. There were significant differences between the mean scores on the Arabic expressive (AE) vocabulary test (AEVT) except between time 3 (12 months) and 4 (18 months), see Table B1 in Appendix B. These differences in mean raw scores across all tests and time points are summarized as follows:

ER: time point 5>time point 4>time point 3>time point 2> time point 1
EE: 5>4=3>2>1
AR: 5>4>3>2=1
AE=5>4=3>2>1

Therefore, this shows that children significantly increased their vocabulary scores on the four tests throughout the different stages of this study with no evidence of decrease in their scores on the four tests.

As seen in Table 2 there is a considerable variability in participants' performance on the four vocabulary tests. Therefore, it is crucial to examine some individual participants' results as ASD is characterised by very prevalent heterogeneity. One way to address this variability is by examining these participants' performance based on their speaking abilities and the other way is by looking at some participants' individual performances.

3.4 Variability in children with autism: Results based on speaking abilities

Table 4 shows the distribution of the participants based on their speaking abilities upon entry and 30 months later at the end of the study.

Table 4. Distribution of speaking abilities of the 29 participants at entry and 30 months later

Speaking ability	Number at entry	Number 30 months after entry
Non-speaking	12	9
Minimally speaking	4	1
Speaking	13	19

Table 4 shows that out of the 12 Non-speaking (NS) students who started the program, two of them joined the Speaking Group and one joined the Minimally Speaking (MS) group at the end of the study. Nine remained as NS at the end of the study. All the four students who started as MS joined the Speaking (S) group by the end of the study. The Speaking group started with 13 students and it increased to 19 after two students joined from the NS group and all the four students in the MS group joined this group.

Table C1 and Table C2 in Appendix C show the Pearson correlation coefficients between the tests in time point 1 (entry) and time point 5 (30 months) for the Non-speaking group (n-12) and the Speaking Group (n = 13) respectively. Table C1 shows that for the Non-speaking group there was no correlation between any of the two expressive vocabulary tests (the EOWPVT in English and the AEVT in Arabic) at time 1 and any of the four vocabulary tests at time 5 (in the shaded rows). This is contrary to the Speaking Group in Table C2 where both of these expressive vocabulary tests correlated significantly with the vocabulary tests at time point 5 (in the shaded rows). The Minimally Speaking group consisted of four students only and therefore no correlation coefficients were reported.

In the following section we will study the trajectories and characteristics of these three groups.

3.4.1 *The non-speaking group*

Participants who had less than 5 words used functionally on a daily basis on either language were considered Non-speaking. There were 12 NS students who met these criteria. Most of these students did not produce any words as reported by parents and as seen in their initial assessments when they joined the program. Table 5 shows the performance of this group across the different stages of the study. Three of the children from the NS group were no longer considered so at the end of the study (as mentioned in Table 4), with two of them moving to the Speaking group and one joining the Minimally Speaking group (who produced less than 20 different words/day).

Table 5. Descriptive Statistics for participants identified as Non-speaking at entry (n = 12)

		ER	EE	AR	AE
Time 1 (entry)					
	M	1.3	0.0	0.2	5.8
	SD	2.5	00	0.6	7.7
	Range	0–5	0	0–2	0–0
Time 2 (6 months)					
	M	3.8	1.7	0.5	0.0
	SD	7.5	3.9	1.0	0.0
	Range	0–21	0–11	0–3	0–0
Time 3 (12 months)					
	M	6.8	1.6	3.1	0.0
	SD	12.1	3.7	3.8	0.0
	Range	0–37	0–10	0–12	0–0
Time 4 (18 months)					
	M	9.3	2.8	3.4	0.3
	SD	14.4	6.8	3.3	0.6
	Range	0–50	0–21	0–11	0–2
Time 5 (30 months)					
	M	8.7	5.3	6.3	2.6
	SD	12.7	10.6	6.5	6.7
	Range	0–45	0–30	0–22	0–23

The rest of the students (9) remained with very limited expressive vocabulary in both languages. One of them named one picture only in English and the rest of the nine students did not name any items and all remained Non-speaking at the conclusion of the study at 30 months. Their receptive vocabulary scores were not different throughout the study and remained flat and almost unchanged. See Table D1 in Appendix D for the individual performance of all these students on the four vocabulary tests.

Due to the large variability seen within this group of students, we will review the performance of two students who both started as Non-speaking, but whose paths diverged considerably at six and 12 months, as shown in Figure 6 and Figure 7. Participant 15 was aged 49 months (4 years, 1 month) at entry, while participant 16 was 47 (3 years, 11 months) at entry. Participant 15 had a raw score of 4 on the English receptive vocabulary test (ER) and 0 on the English expressive vocabulary test (EE), the Arabic receptive (AR) tests, and the Arabic expressive (AE) tests at

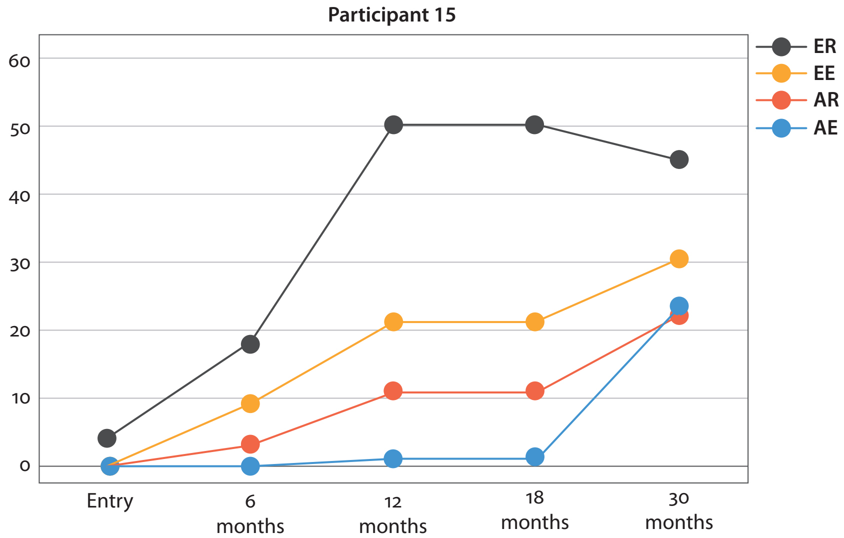


Figure 6. Participant 15's performance on the four vocabulary tests from entry to 30 months

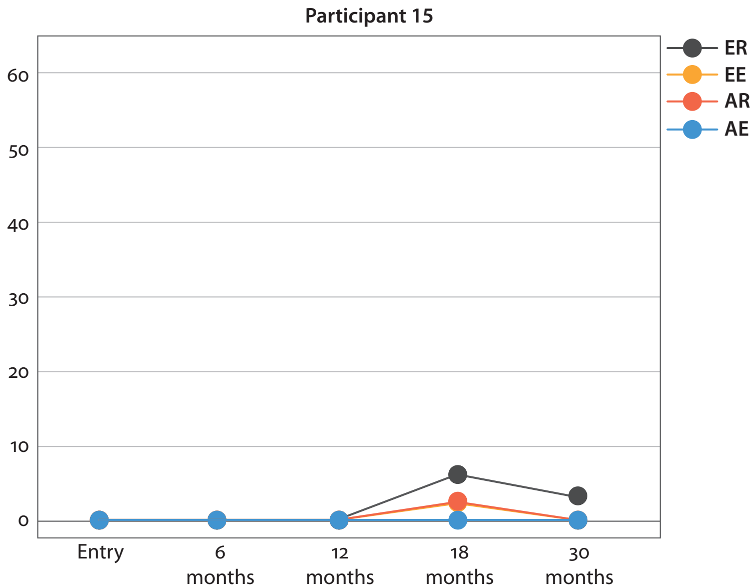


Figure 7. Participant 16's performance on the four vocabulary tests

entry. Six months later he showed good improvements on the English tests and 12 months later his scored continued to improve with an increase in his Arabic receptive vocabulary test scores. At 30 months we saw for the first time a large improvement in his Arabic expressive vocabulary test scores when he had a raw score of 23. At 30 months from entry this student was communicating effectively in both languages and for a variety of functions. He met the criteria for bilingualism (the regular use of two languages on a daily basis), although English remains his dominant language. On the other hand, Participant 16 remained Non-speaking with more or less no change in his tests scores across the four vocabulary tests as depicted in Figure 7, although he has shown improvement in other areas of language and communication skills (e.g., following directions).

3.4.2 *Minimally speaking children*

At the beginning of the study four participants were considered Minimally Speaking based on the criteria used for this study (they had between 5 to 20 different words per day at the beginning of the study). Table 6 and Figure 8 show the performance of the Minimally Speaking participants on the four vocabulary tests across the five different time points. See also Table D2 in Appendix D to examine the individual performance of all students in the MS group on the four vocabulary tests and across the different time points.

At the end of the study, all these Minimally Speaking children became Speaking with two of them having higher raw scores on the Arabic expressive test (AEVT) compared to its English counterpart (EOWPVT). One of the two students obtained higher raw scores on the English expressive vocabulary, while the other showed

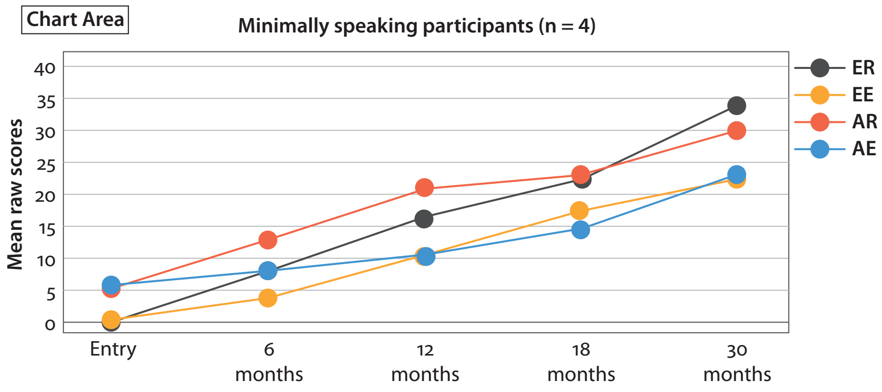


Figure 8. The performance of the four Minimally speaking students on the four vocabulary tests across different time points

Table 6. Performance of the Minimally speaking group on the four vocabulary tests at entry, 6 months, 12 months, 18 months, and 30 months

		ER	EE	AR	AE
Time 1 (entry)					
	M	0.3	0.3	5.0	5.8
	SD	0.5	0.5	5.6	7.7
	Range	0–1	0–1	0–12	0–17
Time 2 (6 months)					
	M	8.0	3.8	12.5	8.0
	SD	8.7	4.3	4.0	8.6
	Range	0–16	0–8	9–18	0–20
Time 3 (12 months)					
	M	16.3	10.0	20.3	10.5
	SD	15.0	11.5	8.3	10.5
	Range	0–31	0–20	11–31	0–25
Time 4 (18 months)					
	M	22.0	16.8	22.5	14.5
	SD	16.1	13.8	8.7	12.4
	Range	1–37	2–30	15–30	1–30
Time 5 (30 months)					
	M	33.5	22.3	29.3	22.3
	SD	21.6	18.4	22.0	16.7
	Range	11–63	0–45	12–58	6–45

equal proficiency in both languages in the areas of receptive and expressive vocabulary. The latter participant's score on the Arabic Expressive (ER) test was better than 12 of the 13 students in the Speaking group (see Participant's 18 scores in Table D2 in Appendix D). All these children use both languages on a daily basis, with varying levels of proficiency in these two languages.

Two of these children started with a preference for Arabic language and after 30 months their profiles were very different as shown in Figures 8 and 9. For example, Participant 18 started with a strong preference for Arabic with very poor performance on the English tests. However, by the end of the study she became a proficient bilingual and her vocabulary scores had increased considerably. The other one (Participant 6) still had Arabic as his dominant language, with very limited expressive vocabulary in English. His receptive language improved, albeit slowly when compared to Participant 18.

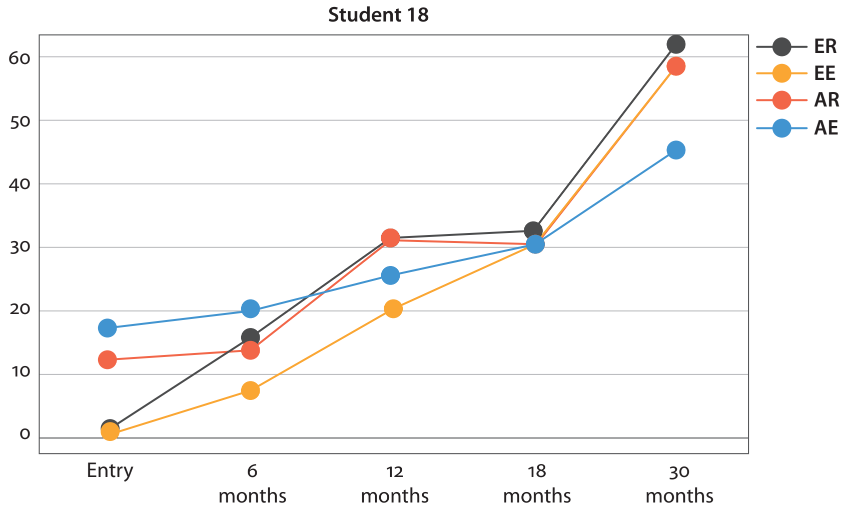


Figure 9. Participant 18's performance on the four vocabulary tests from entry to 30 months

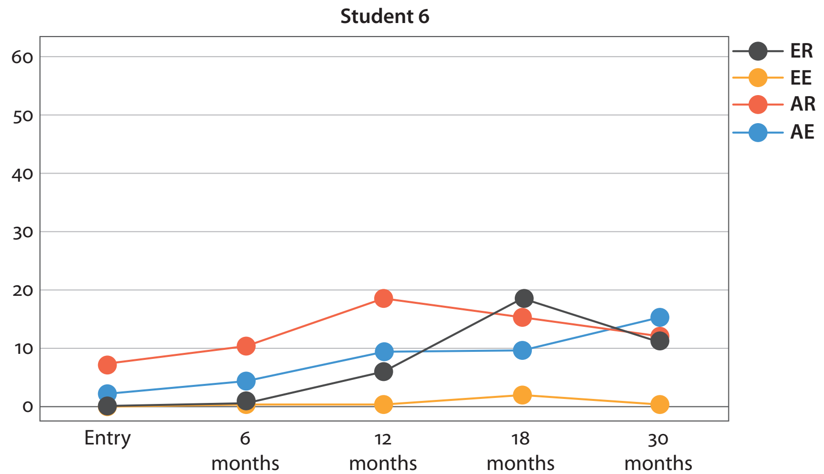


Figure 10. Participant 6's performance on the four vocabulary tests

3.4.3 *Speaking children*

Children were considered Speaking if they used 20 different words or more per day. At the beginning of the study 13 students were considered Speaking.

Two of these children started with good bilingual proficiency (defined as having comparable vocabulary raw scores in both languages), but the rest were proficient in one language only (3 Arabic, 8 English). Currently, 10 of these 13 children can communicate for a variety of functions in both languages. However, three of them remain communicating mostly in English at home and school and they do not prefer to speak Arabic due to limited proficiency. Language preference and language proficiency information was based on each student's SLPs and schoolteachers' reports as we were not able to obtain more detailed information from parents. These reports were consistent with students' scores on the Arabic and English vocabulary tests. See Table D3 in Appendix D for their individual performance on the four vocabulary tests across different time points.

Table 7. Performance of the participants identified as Speaking at entry (n = 13)

		ER	EE	AR	AE
Time 1 (entry)					
	M	28.8	20.5	20.2	13.4
	SD	26.6	23.6	20.1	13.8
	Range	0–87	0–77	0–68	0–48
Time 2 (6 months)					
	M	38.3	32.0	23.7	16.0
	SD	24.9	25.1	14.6	14.7
	Range	4–77	1–87	0–56	0–51
Time 3 (12 months)					
	M	47.8	42.1	28.6	19.7
	SD	27.5	28.5	22.3	15.6
	Range	13–107	3–101	3–89	0–53
Time 4 (18 months)					
	M	48.3	47.2	33.1	21.5
	SD	24.0	31.2	23.2	15.1
	Range	15–90	14–110	3–87	0–55
Time 5 (30 months)					
	M	56.6	53.8	31.8	23.7
	SD	30.1	33.0	20.9	15.7
	Range	23–109	16–117	4–67	0–60

4. Discussion and analysis

This exploratory longitudinal study followed the vocabulary development of 29 bilingually (English/Arabic) exposed children with ASD (average age: 5 years and 3 months), who are enrolled in a bilingual intensive behavioral intervention program in the UAE. These children were tested at 5 time points: entry, 6 months, 12 months, 18 months, and 30 months.

Despite the large variability and heterogeneity known in ASD and seen in this sample of students, the results show that most of these children made positive vocabulary gains in both languages in this intensive behavioral intervention program where both Arabic (the home language for most of these children and a language of instruction at the program) and English (the lingua franca of this community and the dominant language of instruction at the program) were supported. As seen in Table 8 the number of bilingual children doubled at the end of the study indicating that when bilingual instruction is provided it can help children with autism acquire and use the languages of their communities. The number of monolingual children (English/Arabic) has decreased from 11 to 7 and it is expected that most of the remaining monolingual students will improve their communication skills in the other language as long as they continue to receive instruction and support in these languages. Nine out of the 29 students (31%) remained Non-speaking after 30 months of intensive bilingual intervention. Even though these children did not show any improvement in their spoken and receptive vocabulary, there is strong evidence that they improved their overall communication skills especially in following directions (Shaalan, Gould, Egan, & Olsen, 2016) and other aspects of comprehension skills (e.g., responding to greetings, following daily routines, etc.).

These results support the findings of most studies of bilingualism in children with ASD which did not report any evidence of negative impact of bilingual exposure on their language skills (see Drysdale et al., 2015; Kay-Raining Bird et al. 2012; Hambly & Fombonne, 2012, 2014). Therefore, the recommendation to provide bilingual intervention to bilingually exposed children with autism seems to be justified based on these findings. Moreover, it is possible that bilingual intervention is even preferred to monolingual intervention as we saw that some students who started as monolinguals became bilingual in a relatively short period of time, especially when both languages were supported by the family and the community and children received direction intervention in both of these languages.

Table 8. The overall distribution of bilingual, monolingual English or Arabic, and Non-speaking children before and after the study

Bilingual status	**At entry**	**30 months**
Bilingual (E/A)	6	13
Monolingual Arabic	7	3
Monolingual English	4	4
Non-speaking	12	9

The study of the performance of the three groups of students and the correlations among their performance across tests shows that speaking ability at the beginning of the program was not a reliable indicator of their speaking ability 30 months later. These participants were aged between 44 months (3;8) and 77 months (6;5) at the beginning of the study. Three of the 12 students who started this study in the Non-speaking group moved out of this group at the end of the study and one of them had a better performance than some of the students who started in the Speaking Group. Similar results were seen in the students who started in the Minimally Speaking group as three out of the four students showed good bilingual abilities by the conclusion of the study. Moreover, the correlation between the expressive vocabulary test scores at the beginning of the study and at the end of the study was not significant for the Non-speaking group, while it was significant for the Speaking group. Therefore, the recommendation to provide bilingual intervention to bilingually exposed children with autism seem to be justified regardless of the speaking abilities of children at the beginning of the intervention.

Since this is an exploratory study, we aimed to examine some of the various variables interacting with and affecting the development of vocabulary skills in bilingual children with ASD in the UAE. One important variable that needs to be examined more carefully is the influence of recent high quality and intensive input on the language skills in general and vocabulary skills in particular in children with ASD. This possible strong effect of *recent* high-quality input may explain some of the findings of the study, such as the relative small increase in Arabic expressive vocabulary and increase in the number of Arabic monolingual children who turned bilingual by the end of the study.

Overall, we saw that children's scores on the Arabic Expressive (AE) vocabulary test (the AEVT) seemed to increase slowly compared to the other tests. This could be due to the different designs used in the tests as the AEVT has considerably fewer items (60 items only) compared to 190 items in the two English tests (ROWPVT and EOWPVT) and 132 items in the Arabic Receptive (AR) test (the APVT). It is also possible that the larger improvement seen in English compared to Arabic tests is due to the influence of recent high-quality input provided by the intervention program, where 70% of the time students were exposed to English versus 30% of the time in Arabic. The effects of recent input have been reported in another study by Gonzalez-Barrero and Nadig (2018) who found that current amount of language exposure accounted for 62% of the variance in vocabulary skills not only for TD children but also for bilingual children with ASD. Moreover, Bedore, Peña, Griffin, and Hixon (2016) found a stronger contribution for current input when compared to age of exposure, especially for older children (Grade three vs Grade 1 and 2). Similar findings were also reported by Hambly and Fombonne (2014) and Reetzke, Zou, and Sheng (2015). This shows the malleable nature of bilingual language development where language skills can be influenced by many variables whose contribution can change throughout different stages of language development. This effect of timing of exposure (recent versus old) should be studied along with the effects of age of exposure, which takes into account the time of the student's first exposure to the languages in question.

Another revelation of this exploratory study is that conventional descriptors of multilingualism (e.g., 'home language', 'heritage language', 'dominant language', 'prestigious language') may not apply accurately in the case of the UAE and possibly other GCC countries due to the unique linguistic, demographic, and cultural characteristics of these societies. Children in the UAE are exposed to at least 4 or 5 varieties of English and Arabic from very early childhood in most of the times: Emirati Gulf Arabic, Modern Standard Arabic, Pidgin Gulf Arabic, Global English, and nonnative English. While Emirati Gulf Arabic is the main language used at home by the UAE native population and is widely used among Emiratis and in Government institutions, it is still not the language used by the majority of the inhabitants in the UAE. Instead, most inhabitants of big cities in the UAE are exposed to Global English (Hopkyns, 2017) and nonnative English when interacting within their community settings (public or private schools, shopping centres and other public spaces) and each community has its own language that is used among its community members (e.g., East Asian communities, which together constitute the biggest nonnative community, speak Hindi, Urdu, Malayalam and other Indian languages). Therefore, the traditional terms used to describe multilingualism in other cultures (e.g., simultaneous bilingualism vs. sequential bilingualism; L1 vs. L2; home language vs. school language; minority/heritage language vs. majority language; Kay- Raining Bird et al., 2016) may not be adequate to describe the multilingual reality in the UAE. For example, although a child who is born in a UAE family and speaks Arabic only might be exposed to another language (nonnative varieties of English or Arabic, such as Pidgin Gulf Arabic) from birth, he or she may not be considered a simultaneous bilingual even though s/he may meet some of the criteria of simultaneous bilingualism.

5. Limitations and future directions

One of the major limitations of this longitudinal study is the lack of control groups. Ideally, such a study would require two control groups: Arabic only and English only control groups in order to examine the effects of bilingualism on vocabulary development, preferably in a program with similar level of intensity. However, this current intervention program is very unique in this region in terms of the availability of qualified Arabic speaking and English speaking professionals in the areas of special education, behavior analysis, speech-language pathology, occupational therapy, who can provide bilingual intervention at such high standards.

Another important limitation of this study was the lack of quantitative information on language exposure and vocabulary development at home. Parents were given questionnaires to quantify participants' exposure to English and Arabic in order to determine language dominance. It was not possible to examine the relationship between exposure and outcomes as the multilingualism situation was very complex in most households. Additionally, the questionnaire used was not completed by 13 out of 29 parents; and even among parents who filled in the questionnaire, many provided inaccurate information or contradictory answers (e.g., many said the child was exposed to Arabic only at home, then they said the domestic helpers were speaking to him in English). Therefore, other methods for collecting data about exposure (historical and current) are needed. These could include interviews or using technology to collect such information.

Thirdly, due to the unique linguistic and cultural situations in the UAE, the common terms used in studies of multilingualism in other countries or cultures may need to be modified to be suitable to the UAE context.

Fourthly, the current study covers only vocabulary skills in both languages and does not include systematic examination of the effects of exposure on the syntax, morphology, and phonology of both languages. This may warrant further examination in future studies.

Finally, it is recommended to examine other aspects of bilingual intervention such as the bidirectional effects and cross-linguistic transfer between the two languages of intervention (Haman et al., 2017; Yang, Cooc, & Sheng, 2017) and the effects of other nonnative input that these children receive in nonnative English or nonnative Arabic, including Pidgin Gulf Arabic as there are some preliminary data suggesting nonnative input of low language quality could have some negative effects on language development in children with ASD (Hambly & Fombonne, 2012).

References

Al-Abbady, K., Hessian, H. M., & Alaam, M. W. (2017). Prevalence, trend, determinants and prediction of autism spectrum disorders among Dubai population, diagnostic approach and management contexts. *Pediatric Health Research*, 2, 2. https://doi.org/10.21767/2574-2817.100016

Al-Issa, A., & Dahan, L. (2011). Global English and endangered Arabic in the United Arab Emirates. In A. Al-Issa & L. Dahan (Eds) *Global English and Arabic: Issues of language* (pp. 1–22). Bern Switzerland: Peter Lang Publishers. https://doi.org/10.3726/978-3-0353-0120-5

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders. 5th ed*. Arlington, VA: American Psychiatric Association.

American Speech Language Hearing Association (ASHA). (2014). *Bilingual service delivery*. Retrieved from http://www.asha.org/Practice-Portal/Professional-Issues/Bilingual-Service-Delivery/

Baio, J., Wiggins, L., Christensen, D. L., et al. (2018). Prevalence of autism spectrum disorder among children aged 8 Years – Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveill Summ*, 67, 1–23. https://doi.org/10.15585/mmwr.ss6706a1

Bedore, L. M., Peña, E. D., Griffin, Z. M., & Hixon, J. G. (2016). Effects of age of English exposure, current input/output, and grade on bilingual language performance. *Journal of Child Language*, 43, 687–706. https://doi.org/10.1017/S0305000915000811

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge, New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511605963

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis*. (2nd ed.). Upper Saddle River, N.J: Pearson.

COST Action IS0804 (2011). *Questionnaire for parents of biligual children (PABIQ)*. http://www.bi-sli.org/

De Houwer, A. (2009). *An Introduction to bilingual development*. Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781847691705

Drysdale, H., van der Meer, L., & Kagohara, D. (2015). Children with autism spectrum disorders from bilingual families: a systematic review. *Review of Journal of Autism and Developmental Disorders*, 2, 22–38. https://doi.org/10.1007/s40489-014-0032-7

Dubai Statistics Center. (2018). Number of population estimated by nationality. Retrieved from https://www.dsc.gov.ae/en-us/Themes/Pages/Population-and-Vital-Statistics.aspx?Theme=42&year=2018#DSC_Tab1

Ebert, K., Kohnert, K., Pham, G., Disher, J., & Payesteh, B. (2014). Three treatments for bilingual children with primary language impairment: Examining cross-linguistic and cross-domain effects. *Journal of Speech, Language, and Hearing Research*, 57, 172–186. https://doi.org/10.1044/1092-4388(2013/12-0388)

El-Amouri, S., & O'Neill, S. (2011). Supporting cross-cultural communication and culturally competent care in the linguistically and culturally diverse hospital settings of UAE. *Contemporary Nurse*, 39, 240–55. https://doi.org/10.5172/conu.2011.39.2.240

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *MacArthur communicative development inventories: User's guide and technical manual*. Baltimore: Paul H. Brookes Publishing Co.

Fredman, M. (2011). Recommendations for working with bilingual children-Prepared by the Multilingual Affairs Committee of IALP. *Folia Phoniatrica et Logopaedica*, 58, 458–464.

Genesee, F. (2006). Bilingual first language acquisition in perspective. In P. D. McCardle & E. Hoff (Eds.). *Childhood bilingualism: Research on infancy through school age* (pp. 45–67). Clevedon, England: Multilingual Matters. https://doi.org/10.21832/9781853598715-005

Gonzalez-Barrero, A. M., & Nadig, A. (2018). Bilingual children with autism spectrum disorders: the impact of amount of language exposure on vocabulary and morphological skills at school age. *Autism Research*, 11, 1667–78. https://doi.org/10.1002/aur.2023

Grosjean, François. (1992). Another view of bilingualism. In R. J. Harris (Ed.) *Advances in psychology, cognitive processing in bilinguals* (pp. 51–62). Oxford: North-Holland. https://doi.org/10.1016/S0166-4115(08)61487-9

Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., Mieszkowska, K., Łuniewska, M., Kołak, J., Miękisz, A., Kacprzak, A., Banasik, N., Foryś-Nogala, M. (2017). How does L1 and L2 exposure impact L1 performance in bilingual children? Evidence from Polish-English migrants to the United Kingdom. *Frontiers Psychology*, 4, 1444. https://doi.org/10.3389/fpsyg.2017.01444

Hambly, C., & Fombonne, E. (2012). The impact of bilingual environments on language development in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42, 1342–1352. https://doi.org/10.1007/s10803-011-1365-z

Hambly, C., & Fombonne, E. (2014). Factors influencing bilingual expressive vocabulary size in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 8, 1079–1089. https://doi.org/10.1016/j.rasd.2014.05.013

Hopkyns, S. (2017). A conflict of desires: Global English and its effects on cultural identity in the United Arab Emirates. Ph.D. dissertation. University of Leicester, UK.

Kay-Raining Bird, E., Lamond, E., & Holden, J. (2012). Survey of bilingualism in autism spectrum disorders. *International Journal of Language & Communication Disorders*, 47, 52–64. https://doi.org/10.1111/j.1460-6984.2011.00071.x

Kay-Raining Bird, E., Genesee, F., & Verhoeven, L. (2016). Bilingualism in children with developmental disorders: A narrative review. *International Journal of Language & Communication Disorders*, 63, 1–14. https://doi.org/10.1016/j.jcomdis.2016.07.003

Khater, M. (2018). Early phonological skills as a predictor for receptive and expressive vocabulary size in Gulf Arabic speaking children. Ph.D. dissertation. City, University of London, UK.

Kim, S. H., Paul, R., Tager-Flusberg, H., & Lord, C. (2014). Language and communication in autism. In F. R. Volkmar, R. Paul, S. Rogers, and K. A. Pelphrey (Eds), *Handbook of autism and pervasive developmental disorders*. (4th Ed.). New Jersey: Wiley Publishers. https://doi.org/10.1002/9781118911389.hautc10

Kohnert, K., Yim, D., Nett, K., Kan, P. F., & Duran, L. (2005). Intervention with linguistically diverse preschool children: A focus on developing home language(s). *Language, Speech, And Hearing Services in Schools*, 36, 251–263. https://doi.org/10.1044/0161-1461(2005/025)

Kremer-Sadlik, T. (2005). To be or not to be bilingual: autistic children from Multilingual Families. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Proceedings of the 4th international symposium on Bilingualism* (pp. 1225–1234). Somerville, MA: Cascadilla.

Lord, C., Risi, S., & Pickles, A. (2004). Trajectory of language development in autistic spectrum disorders. In M. L. Rice & S. F. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 7–29). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Lund, E. M., Kohlmeier, T. L., & Durán, L. K. (2017). Comparative language development in bilingual and monolingual children with autism spectrum disorder: A systematic review. *Journal of Early Intervention*, 39, 106–124. https://doi.org/10.1177/1053815117690871

Martin, N., & Brownell, R. (2010). *The receptive one word picture vocabulary test, 4th ed*. Novato, California: ATP Assessment.

Ohashi, J. K., Mirenda, P., Marinova-Todd, S., Hambly, C., Fombonne, E., Szatmari, P., Bryson, S., Roberts, W., Smith, I., Vaillancourt, T., Volden, J., Waddell, C., Zwaigenbaum, L., Georgiades, S., Duku, E., & Thompson, A. (2012). Comparing early language development in monolingual- and bilingual- exposed young children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 6(2), 890–897. https://doi.org/10.1016/j.rasd.2011.12.002

Paradis, J. (2011). Individual differences in child English second language acquisition: comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1, 213–237. https://doi.org/10.1075/lab.1.3.01par

Paradis, J., Emmerzael, K., & Sorenson Duncan, T. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, 43, 474–497. https://doi.org/10.1016/j.jcomdis.2010.01.002

Pearson, B. Z., Fernandez, S. C., & Oller, R. E. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43, 93–120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x

Petersen, J. M., Marinova-Todd, S. H., & Mirenda, P. (2012). Brief report: An exploratory study of lexical skills in bilingual children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 42, 1499–1503. https://doi.org/10.1007/s10803-011-1366-y

Raddawi, R., & Meslem, D. (2015). Loss of Arabic in the UAE: Is bilingual education the solution? *International Journal of Bilingual & Multilingual Teachers of English*, 3, 85–94. https://doi.org/10.12785/ijbmte/030203

Reetzke, R., Zou, X., & Sheng, L. (2015). Communicative development in bilingually exposed Chinese children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, 58, 813–825. https://doi.org/10.1044/2015_JSLHR-L-13-0258

Shaalan, S. (2010). Investigating grammatical complexity in Gulf-Arabic speaking children with specific language impairment. Ph.D. dissertation. University College London, London, UK.

Shaalan, S. (2017). Reliability and validity of four Arabic language tests: a comparison of performance of Qatari school-aged children with and without language impairment. *Arab Journal of Applied Linguistics*, 2, 20–48.

Shaalan, S., Gould, D. D., & Egan, K., & Olsen, P. (2016). Bilingual intervention using ABA methods in the United Arab Emirates: Preliminary findings. Paper presented at the *Bi-Annual Convention of the European Association for Behaviour Analysis*. Enna, Sicily, Italy.

Smart, J. (1990). Pidginization in gulf Arabic: A first report. *Anthropological Linguistics*, 32, 83–119.

Sparrow, S., Cicchetti, D., & Balla, D. (2005). *Vineland adaptive behaviour scales*. (2nd ed.). Minneapolis, MN: Pearson Assessment.

Statistics Centre Abu Dhabi. (2017). *Statistical year book population 2017*. Retrieved from https://www.scad.gov.abudhabi/Release%20Documents/Statistical%20Yearbook%20-%20Population%20-%20EN.pdf

Thordardottir, E. (2010). Towards evidence-based practice in language intervention for bilingual children. *Journal of Communication Disorders*, 43, 523–537. https://doi.org/10.1016/j.jcomdis.2010.06.001

Valicenti-McDermott, M., Tarshis, N., Schouls, M., Galdston, M., Hottinger, K., & Seijo, R. (2013). Language differences between monolingual English and bilingual English-Spanish young children with autism spectrum disorders. *Journal of Child Neurology*, 28, 945–948. https://doi.org/10.1177/0883073812453204

World Bank. (2019). United Arab Emirates Data. https://data.worldbank.org/country/united-arab-emirates?view=chart. Accessed 15 September 2019.

Yang, M., Cooc, N., & Sheng, L. (2017). An investigation of cross-linguistic transfer between Chinese and English: a meta-analysis. *Asian-Pacific Journal of Second and Foreign Language Education*, 2, 15. https://doi.org/10.1186/s40862-017-0036-9C

Appendix A. Correlations table

Table A1. Correlations between the four vocabulary tests across the 5 time points for all participants (n = 29). ER = English receptive test (ROWPVT), EE = English expressive test (EOWPVT), AR = Arabic receptive test (APVT), and Arabic expressive test (AEVT). 1 = entry, 2 = 6 months, 3 = 12 months, 4 = 18 months, 5 = months

Correlations																				
	ER1	**ER2**	**ER3**	**ER4**	**ER5**	**EE1**	**EE2**	**EE3**	**EE4**	**EE5**	**AR1**	**AR2**	**AR3**	**AR4**	**AR5**	**AE1**	**AE2**	**AE3**	**AE4**	**AE5**
ER1	1	.892**	.871**	.834**	.841**	.850**	.937**	.918**	.870**	.868**	.757**	.731**	.774**	.760**	.563**	.716**	.688**	.662**	.578**	.544**
ER2	.892**	1	.951**	.949**	.913**	.849**	.936**	.959**	.966**	.943**	.683**	.768**	.799**	.792**	.622**	.699**	.681**	.654**	.610**	.585**
ER3	.871**	.951**	1	.951**	.898**	.844**	.954**	.962**	.959**	.943**	.790**	.800**	.869**	.865**	.688**	.760**	.747**	.733**	.714**	.675**
ER4	.834**	.949**	.951**	1	.917**	.778**	.889**	.928**	.944**	.938**	.708**	.754**	.819**	.811**	.664**	.690**	.687**	.681**	.630**	.653**
ER5	.841**	.913**	.898**	.917**	1	.812**	.872**	.909**	.934**	.964**	.716**	.828**	.842**	.858**	.776**	.760**	.767**	.759**	.736**	.751**
EE1	.850**	.849**	.844**	.778**	.812**	1	.921**	.883**	.895**	.848**	.846**	.827**	.861**	.880**	.645**	.838**	.818**	.773**	.735**	.658**
EE2	.937**	.936**	.954**	.889**	.872**	.921**	1	.968**	.948**	.905**	.808**	.791**	.842**	.857**	.631**	.787**	.760**	.719**	.696**	.621**
EE3	.918**	.959**	.962**	.928**	.909**	.883**	.968**	1	.982**	.958**	.749**	.771**	.834**	.837**	.618**	.739**	.723**	.691**	.674**	.597**
EE4	.870**	.966**	.959**	.944**	.934**	.895**	.948**	.982**	1	.970**	.749**	.813**	.872**	.868**	.678**	.771**	.762**	.727**	.716**	.664**
EE5	.868**	.943**	.943**	.938**	.964**	.848**	.905**	.958**	.970**	1	.750**	.816**	.854**	.849**	.741**	.759**	.765**	.754**	.728**	.703**
AR1	.757**	.683**	.790**	.708**	.716**	.846**	.808**	.749**	.749**	.750**	1	.879**	.891**	.914**	.796**	.897**	.901**	.930**	.852**	.808**
AR2	.731**	.768**	.800**	.754**	.828**	.827**	.791**	.771**	.813**	.816**	.879**	1	.911**	.939**	.857**	.880**	.899**	.921**	.900**	.861**
AR3	.774**	.799**	.869**	.819**	.842**	.861**	.842**	.834**	.872**	.854**	.891**	.911**	1	.957**	.785**	.904**	.906**	.889**	.859**	.831**
AR4	.760**	.792**	.865**	.811**	.858**	.880**	.857**	.837**	.868**	.849**	.914**	.939**	.957**	1	.824**	.889**	.893**	.896**	.874**	.835**
AR5	.563**	.622**	.688**	.664**	.776**	.645**	.631**	.618**	.678**	.741**	.796**	.857**	.785**	.824**	1	.848**	.860**	.904**	.885**	.946**
AE1	.716**	.699**	.760**	.690**	.760**	.838**	.787**	.739**	.771**	.759**	.897**	.880**	.904**	.889**	.848**	1	.988**	.949**	.931**	.879**
AE2	.688**	.681**	.747**	.687**	.767**	.818**	.760**	.723**	.762**	.765**	.901**	.899**	.906**	.893**	.860**	.988**	1	.974**	.956**	.894**
AE3	.662**	.654**	.733**	.681**	.759**	.773**	.719**	.691**	.727**	.754**	.930**	.921**	.889**	.896**	.904**	.949**	.974**	1	.954**	.919**
AE4	.578**	.610**	.714**	.630**	.736**	.735**	.696**	.674**	.716**	.728**	.852**	.900**	.859**	.874**	.885**	.931**	.956**	.954**	1	.918**
AE5	.544**	.585**	.675**	.653**	.751**	.658**	.621**	.597**	.664**	.703**	.808**	.861**	.831**	.835**	.946**	.879**	.894**	.919**	.918**	1

** Correlation is significant at the 0.01 level (2-tailed).

Appendix B. Pairwise comparisons

Table B1. Pairwise comparisons: Scores of the English receptive vocabulary test (ROWPVT) at different time points for all participants (*n* = 29)

Measure			
(I) factor 1	**(J) factor 1**	**Mean difference (I-J)**	**Sig.**[b]
1	2	−6.552*	.003
	3	−13.207*	.000
	4	−15.207*	.000
	5	−20.276*	.000
2	1	6.552*	.003
	3	−6.655*	.001
	4	−8.655*	.000
	5	−13.724*	.000
3	1	13.207*	.000
	2	6.655*	.001
	4	−2.000	.228
	5	−7.069*	.011
4	1	15.207*	.000
	2	8.655*	.000
	3	2.000	.228
	5	−5.069*	.045
5	1	20.276*	.000
	2	13.724*	.000
	3	7.069*	.011
	4	5.069*	.045

* The mean difference is significant at the 0.05
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).
Note. 1 = Time point 1 (entry), 2 = Time point 2 (6 months), 3 = Time point 3 (12 months), 4 = time point 4 (18 months), 5 = time point 5: 30 months

Table B2. Pairwise comparisons: Scores of the English expressive vocabulary test (EOWPVT) at different time points for all participants (*n* = 29)

(I) factor1	**(J) factor1**	**Mean difference (I-J)**	**Sig.**[b]
1	2	−6.345*	.001
	3	−11.690*	.000
	4	−15.414*	.000
	5	−20.172*	.000
2	1	6.345*	.001
	3	−5.345*	.001
	4	−9.069*	.000
	5	−13.828*	.000

(I) factor1	(J) factor1	Mean difference (I-J)	Sig.[b]
3	1	11.690*	.000
	2	5.345*	.001
	4	−3.724*	.002
	5	−8.483*	.000
4	1	15.414*	.000
	2	9.069*	.000
	3	3.724*	.002
	5	−4.759*	.004
5	1	20.172*	.000
	2	13.828*	.000
	3	8.483*	.000
	4	4.759*	.004

Based on estimated marginal means
* The mean difference is significant at the
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).
Note. 1 = Time point 1 (entry), 2 = Time point 2 (6 months), 3 = Time point 3 (12 months), 4 = time point 4 (18 months), 5 = time point 5: 30 months.

Table B3. Pairwise comparisons: Scores of the Arabic receptive vocabulary test (APVT) at different time points for all participants ($n = 29$)

(I) factor1	(J) factor1	Mean difference (I-J)	Std. error	Sig.[b]
1	2	−2.759	1.458	.069
	3	−7.103*	1.639	.000
	4	−9.552*	1.667	.000
	5	−11.103*	2.286	.000
2	1	2.759	1.458	.069
	3	−4.345*	1.584	.010
	4	−6.793*	1.643	.000
	5	−8.345*	2.014	.000
3	1	7.103*	1.639	.000
	2	4.345*	1.584	.010
	4	−2.448*	1.136	.040
	5	−4.000	2.420	.109
4	1	9.552*	1.667	.000
	2	6.793*	1.643	.000
	3	2.448*	1.136	.040
	5	−1.552	2.283	.502
5	1	11.103*	2.286	.000
	2	8.345*	2.014	.000
	3	4.000	2.420	.109
	4	1.552	2.283	.502

Based on estimated marginal means
* The mean difference is significant at the
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).
Note: 1 = Time point 1 (entry), 2 = Time point 2 (6 months), 3 = Time point 3 (12 months), 4 = time point 4 (18 months), 5 = time point 5: 30 months.

Table B4. Pairwise comparisons: The Arabic expressive vocabulary test (AEVT) at different time points for all participants (*n* = 29)

(I) factor1	(J) factor1	Mean difference (I-J)	Sig.[b]
1	2	−1.483*	.001
	3	−3.483*	.001
	4	−4.966*	.000
	5	−7.966*	.000
2	1	1.483*	.001
	3	−2.000*	.004
	4	−3.483*	.000
	5	−6.483*	.000
3	1	3.483*	.001
	2	2.000*	.004
	4	−1.483	.081
	5	−4.483*	.001
4	1	4.966*	.000
	2	3.483*	.000
	3	1.483	.081
	5	−3.000*	.018
5	1	7.966*	.000
	2	6.483*	.000
	3	4.483*	.001
	4	3.000*	.018

Based on estimated marginal means
* The mean difference is significant at the
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).
Note. 1 = Time point 1 (entry), 2 = Time point 2 (6 months), 3 = Time point 3 (12 months), 4 = time point 4 (18 months), 5 = time point 5: 30 months.

Appendix C. Correlations based on speaking abilities

Table C1. Pearson correlations coefficients between the four vocabulary tests at entry (time point 1) and at the end of the study (time point 5) in the Non-speaking Group (*n* = 12): 30 months from entry (Time Point 5) in children classified as Non-speaking at entry. ROWPVT = English receptive test, EOWPVT = English expressive test, APVT = Arabic receptive test, AEVT = Arabic expressive test. 1 = entry, 5 = 30 months

	ROWPVT	EOWPVT	APVT	AEVT	ROWPVT	EOWPVT	APVT	AEVT
ROWPVT	1	−.181	−.135	.[a]	.861**	.977**	.768**	.861**
EOWPVT	−.181	1	.924**	.[a]	−.166	−.209	−.247	−.162
APVT	−.135	.924**	1	.[a]	−.116	−.155	−.157	−.121
AEVT-1	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]
ROWPVT	.861**	−.166	−.116	.[a]	1	.923**	.900**	.963**
EOWPVT	.977**	−.209	−.155	.[a]	.923**	1	.876**	.893**
APVT	.768**	−.247	−.157	.[a]	.900**	.876**	1	.830**
AEVT	.861**	−.162	−.121	.[a]	.963**	.893**	.830**	1

** Correlation is significant at the 0.01 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

Table C2. Correlations between the four vocabulary tests at entry (Time Point 1) and at the end of the study in the Speaking-Group (n = 13): 30 months from entry (Time Point 5) in children classified as Non-speaking at entry. ROWPVT = English receptive test, EOWPVT = English expressive test, APVT = Arabic receptive test, AEVT = Arabic expressive test (AEVT1 = entry, 2 = 6 months, 3 = 12 months, 4 = 18 months, 5 = month

	ROWPVT	EOWPVT	APVT	AEVT	ROWPVT	EOWPVT	APVT	AEVT
ROWPVT	1	.776**	.632*	.616*	.871**	.852**	.480	.448
EOWPVT	.776**	1	.785**	.811**	.856**	.854**	.655*	.704**
APVT-	.632*	.785**	1	.863**	.594*	.630*	.840**	.883**
AEVT	.616*	.811**	.863**	1	.648*	.636*	.835**	.925**
ROWPVT	.871**	.856**	.594*	.648*	1	.939**	.575*	.485
EOWPVT	.852**	.854**	.630*	.636*	.939**	1	.563*	.471
APVT	.480	.655*	.840**	.835**	.575*	.563*	1	.911**
AEVT	.448	.704**	.883**	.925**	.485	.471	.911**	1

** Correlation is significant at the 0.01 level (2-tailed).

Appendix D.

Table D1. The performance of the non speaking students (n = 12) on the four vocabulary tests across different time

ID points	Sex	Age (mo)	Raw Scores at Entry				6 Months				12 Months				18 Months				30 Months			
			ER	EE	AR	AE	ER	EE	AR	AE	ER	EE	AR	AE	ER	EE	AR	AE	ER	EE	AR	AE
004	M	65	0	0	0	0	0	0	1	0	4	0	2	0	0	0	2	0	3	1	6	1
010	M	59	0	0	0	0	0	0	0	0	0	0	3	0	7	0	2	0	7	0	4	0
011	M	56	4	0	0	0	21	11	0	0	37	10	12	0	23	13	4	2	19	25	12	7
013	M	44	0	0	0	0	0	0	0	0	1	0	0	0	12	0	0	0	12	7	14	0
015	M	47	4	0	0	0	18	9	3	0	27	9	8	0	50	21	11	1	45	30	22	23
016	M	49	0	0	0	0	0	0	0	0	0	0	0	0	2	0	6	0	0	0	3	0
023	F	53	0	0	0	0	0	0	0	0	3	0	0	0	4	0	3	0	0	0	2	0
024	M	66	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	6	0	0	0
026	M	48	2	0	0	0	2	0	0	0	4	0	4	0	2	0	4	0	3	0	1	0
028	M	61	5	0	2	0	5	0	2	0	6	0	4	0	6	0	7	0	4	0	3	0
032	M	77	0	0	0	0	0	0	0	0	0	0	0	0	5	0	2	0	2	0	2	0
036	M	53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	6	0

Note. ER = English receptive vocabulary test (raw scores from the ROWPVT-4).
EE = English expressive vocabulary test (raw scores from the EOWPVT-4)
AR = Arabic receptive vocabulary scores (raw scores from the APVT; Shaalan, 2010)
AE = Arabic expressive vocabulary test (raw scores from the Arabic Expressive Vocabulary Test (AEVT), Khater, 2018).

Table D2. The individual performance of the minimally speaking students (n = 4) on the four vocabulary tests across different time points

ID	Sex	Age (mo)	Raw Scores at Entry				6 Months				12 Months				18 Months				30 Months			
			ER	EE	AR	AE	ER	EE	AR	AE	ER	EE	AR	AE	ER	EE	AR	AE	ER	EE	AR	AE
006	M	77	0	0	7	2	1	0	10	4	6	0	18	9	18	2	15	9	11	0	12	15
018	F	74	1	1	12	17	15	7	13	20	31	20	31	25	32	30	30	30	63	45	58	45
020	F	73	0	0	1	0	16	8	9	0	27	20	21	0	37	27	30	1	31	21	12	6
037	M	77	0	0	0	4	0	0	18	8	1	0	11	8	1	8	15	18	29	23	35	23

Note. ER = English receptive vocabulary test (raw scores from the ROWPVT-4).
EE = English expressive vocabulary test (raw scores from the EOWPVT-4)
AR = Arabic receptive vocabulary scores (raw scores from the APVT, Shaalan, 2010)
AE = Arabic expressive vocabulary test (raw scores from the Arabic Expressive Vocabulary Test (AEVT, Khater, 2018).

Table D3. The performance of speaking students (*n* = 13) on the four vocabulary tests across different time points

ID	Sex	Age (mo)	Raw Scores at Entry				6 Months				12 Months				18 Months				30 Months			
			ER	EE	AR	AE	ER	EE	AR	AE	ER	EE	AR	AE	ER	EE	AR	AE	ER	EE	AR	AE
001	M	72	0	0	14	2	4	1	23	7	**15**	**3**	**23**	**16**	**15**	**14**	**23**	**17**	26	16	22	20
005	M	65	48	34	55	20	48	54	36	23	76	58	39	37	66	58	60	31	71	74	67	38
012	M	50	54	0	0	0	34	34	0	0	40	48	4	0	44	36	3	0	58	50	4	0
017	M	64	0	0	0	0	19	21	12	0	43	32	10	0	28	29	18	18	24	27	10	7
019	F	58	0	0	12	14	7	5	15	13	13	10	16	16	19	15	20	20	23	17	32	27
021	F	75	14	14	15	17	41	21	21	21	45	38	21	25	54	51	18	27	42	58	44	28
022	F	72	17	19	4	0	29	18	14	0	22	22	3	4	27	24	13	4	37	25	16	17
027	M	66	87	77	68	48	77	87	56	51	107	101	89	53	90	110	87	55	107	117	64	60
029	M	67	3	6	21	10	6	3	18	19	21	15	21	25	29	14	26	25	38	37	22	22
031	M	74	30	41	26	25	50	56	26	29	60	54	38	26	54	67	48	34	68	52	31	27
033	M	66	45	36	11	3	75	49	18	5	75	82	37	6	85	94	36	5	89	103	14	5
034	F	71	46	38	22	24	64	43	44	27	59	57	41	32	62	70	54	33	109	89	60	36
035	F	72	31	1	14	11	44	24	25	13	46	27	30	16	55	31	24	11	44	35	28	21

Note. ER = English receptive vocabulary test (raw scores from the ROWPVT-4).
EE=English expressive vocabulary test (raw scores from the EOWPVT-4)
AR = Arabic receptive vocabulary scores (raw scores from the APVT, Shaalan, 2010)
AE = Arabic expressive vocabulary test (raw scores from the Arabic Expressive Vocabulary Test (AEVT, Khater, 2018).

Index

This volume is the first systematic attempt to survey current progress in the relatively new field of Experimental Arabic Linguistics. While experimental work on Arabic linguistics has appeared sporadically in several venues in the past, the chapters in this book provide a more coherent picture of the exciting directions which the field is pursuing. They provide insights into the complex nature of the Arabic language and how native speakers process it, using cutting-edge experimental methodologies in the fields of phonetics, psycholinguistics, and typical and atypical language development. This volume is of particular interest to scholars, researchers, and students at both the undergraduate and graduate level, in the fields of linguistics and language studies and can be a point of reference for scholars and researchers in the fields of theoretical and experimental Arabic linguistics.

ISBN 978 90 272 0884 2

9 789027 208842

John Benjamins Publishing Company