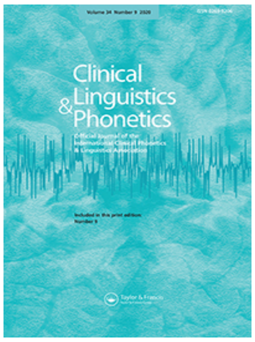

Clinical Linguistics & Phonetics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/iclp20

A linguistically-driven response categorisation protocol for Arabic nouns and verbs: clinical and research applications

Tariq Khwaileh, Eiman Mustafawi & Yusuf Albustanji

CLINICAL LINGUISTICS & PHONETICS
2020, VOL. 34, NO. 9, 861–877
https://doi.org/10.1080/02699206.2019.1708464

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

A linguistically-driven response categorisation protocol for Arabic nouns and verbs: clinical and research applications

Tariq Khwaileh [a], Eiman Mustafawi [a], and Yusuf Albustanji[b]

[a]Department of English Literature and Linguistics, College of Arts and Sciences, Qatar University, Doha, Qatar; [b]Speech-Language Pathology Unit, Johns Hopkins Aramco Healthcare, Dhahran, Saudi Arabia

ABSTRACT

Research into language breakdown in Arabic has been growing in the last two decades. The field, however, remains challenged by the lack of assessment materials, normative databases, and standards for categorising responses from participants with language impairment. The aim of this paper is to introduce a linguistically driven protocol for categorising responses from Arabic-speaking patients and healthy participants in language production tasks. The protocol is informed by Arabic morpho-syntactic/morpho-phonological features such as inflection for gender, person, and number; and on a larger scale agreement in noun phrases and verb phrases. The emerging error categories are applicable to different Arabic varieties and sub-varieties. The data supporting the resulting error categorisation protocol stem from responses from patients with aphasia performing various production tasks.

ARTICLE HISTORY
Received 8 July 2018
Revised 27 December 2018
Accepted 19 December 2019

KEYWORDS
Arabic; response; category; protocol; clinical linguistics; aphasia

Background

In the domains of speech and language pathology, experimental linguistics and psycho-neurolinguistics, the use of linguistically informed clinical and research tools are essential for drawing accurate conclusions about the nature of language breakdown and identifying functional loci of different patterns of performance. Such tools may include language assessment and profiling batteries, therapeutic materials, and response categorisation protocols. A response categorisation protocol is a classification scheme used to categorise responses/errors from healthy and impaired speakers, and can be used alongside clinical assessment tools to support qualitative analysis of patients' responses, enabling researchers and clinicians to identify and classify types and forms of errors produced.

Studies on categorising behavior have discussed the functions of categories, the different methods through which they can be applied, and the categorisation strategies to consider while developing a protocol (Chorney, McMurtry, Chambers, & Bakeman, 2015; Saldaña, 2009). Chorney et al. (2015) highlighted four important components to include in the process of developing a categorisation scheme and to include in the accompanying manual; a list of the categories (i.e. labels), their operational definitions, and the sampling strategy of the researcher's choice.

First, labels are essential to develop and to assign to different responses. In developing the list of labels, Bakeman and Gottman (1997) note that there are differences in behaviors that can be reported in forms of labels and which may vary in terms of concreteness and

CONTACT Tariq Khwaileh tariq.khwaileh@qu.edu.qa Department of English Literature and Linguistics, College of Arts and Sciences, Qatar University, PO Box 2713, Doha, Qatar
Supplemental data for this article can be accessed on the publisher's website.

granularity. In terms of concreteness, categories may represent behaviors that are physically based (i.e. facial expressions, body language, etc.), or behaviors that are socially based, which would require human judgement as they are constructed concepts. Granularity refers to the scope within which these behaviors should be categorised. Bell & Bell (1989) suggested that behaviors can either be macro- or micro-categorised. The former is defined as a categorisation strategy that captures responses at a general level, whereas the latter captures a more specific level (i.e. utterance by utterance). The second component is essential to include in the development and categorisation of labels is operational definitions. Operational definitions describe the observable features of responses through specific examples of actions within a given context (Chorney et al., 2015). The third component important to consider while developing such a protocol is the sampling (or recording) strategy the researcher aims to employ. Chorney et al. (2015) suggest that the research question of the study would determine this, as several factors may or may not be important to the study in question.

To the best of the authors' knowledge, response categorisation protocols are not readily available for the Arabic language. Previous studies on Arabic typical and atypical language processing (e.g. Albustanji, 2009; Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi, Prunet, & Béland, 2002; Khwaileh, 2011; Khwaileh, Body, & Herbert, 2015, 2017; Khwaileh, Body, & Herbert., 2014; Mimouni, Kehayia, & Jarema, 1998, 1997) made use of error categories that are idiosyncratic in nature. Furthermore, some error categories (e.g. semantic, phonological and morphological errors) used in studies on Arabic are adopted from research on English (e.g. Khwaileh, 2011), a language typologically distant from Arabic. The development of an Arabic response categorisation protocol is crucial to research, clinical practice and experimental linguistics in the Arab region. A reliable protocol would have to take into account the linguistic features of the Arabic language, and be informed by the language's peculiar morpho-phonology and syntax, according to the most up to date theoretical frameworks.

Arabic morpho-phonology

Within the study of Semitic morphology, there have been two dominant views describing for Semitic morphology: the root-template/pattern and stem-word views (Benmamoun, 1999, 2003; Heath, 2003; McCarthy, 1981; McCarthy & Prince, 1990). It is important to take into account the conflicting views regarding the morphological make-up of Arabic words. These views are important to consider and understand prior to developing a response categorisation protocol for Arabic. It is also crucial to identify the appropriate theory (with established evidence) that would drive the development of the response categorisation protocol described in this paper.

The predominant view on Arabic words has been the root-pattern view. Arabic morphology has been understood under the assumption that words are composed of consonantal roots, C(onsonant)V(owel)-skeletons and vocalic patterns that carry the inflectional and derivational information of the Arabic word. McCarthy (1981, 1983 1986), McCarthy & Prince (1990) provided an insight into this phenomenon in Semitic languages, e.g. Arabic and Hebrew, and this type of morphology is referred to as nonconcatenative morphology as opposed to concatenative morphology which assumes that words are built through concatenating morphemes serially/sequentially, which is the case in the Indo-European languages. Non-contcatenative morphology, on the other hand, refers to morphology of building words without stringing morphemes sequentially. In the Arabic non-concatenative morphology words are built upon the foundations of three fundamentals: a consonantal root, a CV skeleton, and a vocalic pattern (Boudelaa & Marslen-Wilson, 2015, 2013, 2011, 2004, 2001; Frost, Forster, & Deutsch, 1997; Frost, 1998; McCarthy & Prince, 1990; Mimouni et al., 1998). The consonantal root conveys the semantic content, the CV-skeleton determines the morphological and phonological structure of the word, and the vocalic pattern provides grammatical information such as tense, number, and gender. For example, the word [kitaab] 'book' has the CV skeleton [CVCVVC], consonantal root /ktb/ and the vocalic pattern /-i-aa-/ (e.g. the root ktb used in /kitab/ 'book', /kataba/ 'he wrote', /kaatib/ 'writer').

The second view on Arabic morphology was proposed by Benmamoun (1999, 2003), who suggests that Arabic words are compiled in a word to word fashion. He maintains that there is no need to develop a different model for the Arabic language. Benmamoun (1999, 2003) supports a word-stem view; in his view, Arabic words are no different than words from other Indo-European languages, they are composed of lexemes and morphemes and they are formed through adding and subtracting processes. However, for Arabic, it requires a different process since the derivational implementation for Arabic is different to that of English. Instead of affixing a suffix or a prefix, in Arabic the prosodic part of the word is affixed. Heath (2003) is in support of this view. He distinguishes between formal elements needed for morphological derivation from those needed for a parsing of a surface level, he stated that "for nouns, there is no reasonable way to separate consonants and vowels into different levels, … since the vowels cannot be motivated grammatically" (Heath, 2003, p. 115). Heath's model is based on the argument that Arabic has many underived stems whose phonological substance is composed of consonants and vowels, and that there are Arabic words that share consonantal sequences/roots yet they are not semantically related. Both Benmamoun and Heath argue against the grammatical status of the vocalic pattern and state that vocalic patterns cannot be carrying any grammar of tense and aspect in verbs or numbers in nominal patterns. They present the case of number in Arabic nouns by claiming that singular nouns in Arabic are not marked by any morphemes, and it cannot be argued that there is a default singular pattern since singular nouns' vocalic patterns vary in an unpredictable way. This leads to questioning the assumption that vocalic patterns bear grammatical features of a given word (see Appendix A for examples).

Both views described above have been based on formal analysis of Arabic data. In response to these views, experimental linguists and neuropsychologists carried out a number of experimental studies to investigate the nature of Arabic morphology. In a series of studies using psycholinguistic/experimental paradigms, Boudelaa and Marslen-Wilson (2001) looked into the nature of morphological units in the Arabic mental lexicon through a cross modal-masked priming experiment with healthy speakers. The aim of their study was to investigate whether consonantal root, argued to be the basis of semantic fields in Arabic, is a facilitator of the other related roots. The authors conclude that the consonantal root is important not only because it conveys semantic meaning, but also because it provides the language processor with the canonical timing morpheme with three consonant slots to fulfil the structural canonical request for word production by the language processor. This is consistent with similar studies on Hebrew (Frost et al., 1997) but inconsistent with research on non-Semitic languages.

In addition, Boudelaa and Marslen-Wilson (2004) reported a bigger priming effect between word pairs sharing skeletal morpheme as opposed to those sharing a vocalic pattern morpheme. They also mention a similarity in the size of priming effects in the skeleton and the word pattern conditions. The authors maintain the cognitive effectiveness of the CV skeleton. They also maintain that the CV-Structure of English and other languages is a phonological structure but in Arabic it is a morphemic structure, supporting that languages with non-concatenative morphology require a different theory to account for their word formation processes.

More recently, Boudelaa and Marslen-Wilson (2011) conducted a series of masked priming experiments investigating morphemic decomposition in Arabic to check whether the productivity of the morphemes (consonantal root or vocalic pattern) plays a role in lexical processing and decomposition of Arabic words. The first set of experiments revealed that nominal vocalic patterns indeed do prime when they are in a context of productive root. The findings from the second set of experiments confirmed the view that the presence and amount of vocalic pattern priming is entirely determined by the productivity of the roots with which the vocalic pattern co-occurs. Overall, Boudelaa and Marlsen-Wilson (2011) found that priming was determined primarily by the productivity of the root and not the vocalic pattern, and that the consonantal root is a determinant and main driver of the decomposition of Arabic words. In a recent neuroimaging study, Bozic Boudelaa and Marslen Wilson (2010) showed a left fronto-temporal brain network (typically involved in dealing with linguistic complexity) to be activated by all morphologically complex words in Arabic, suggesting special neural correlates for processing words with non-concatenative morphology.

The results reported in the above studies call into question the stem-based approaches to Arabic morphology. That is; the psycholinguistic data from healthy Arabic speakers are better interpreted using the root-pattern framework. Studies investigating Arabic-speaking participants with brain damage have also provided further evidence that the root-pattern view is a feasible framework on which clinical and research tools can be based (e.g. Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi et al., 2002; Khwaileh, 2011; Khwaileh et al., 2015, 2017; Mimouni et al., 1998, 1997). The processing evidence in support of the root-pattern framework has been prevailing from clinical and neuropsychological data.

Mimouni et al. (1997) investigated the ways in which agrammatism is manifested in Algerian Arabic through the error analysis of connected speech samples taken from three participants with Broca's aphasia. The authors report a pattern of errors that can be explained by the root-pattern framework, viewing Arabic words as consisting of three morphemes: the consonantal root, the vocalic pattern and a CV skeleton. This was reflected in the participants' patterns of substitution errors of morphologically complex forms, which is a feature of Semitic languages. Mimouni et al. (1998) investigated lexical representation, morphological relatedness, and modes of access in Algerian Arabic through an auditory morphological priming paradigm. The authors ran word recognition tasks of singular and plural nouns in 24 healthy participants and 2 Algerian-speaking agrammatic patients. They reported a differential processing of sound (regular) and broken (irregular) plurals evident in the reaction times, indicating whole word access for broken (irregular) plurals and decomposition into word and suffix for suffixed sound (regular) plurals. Mimouni et al. (1998) report that organization of the Algerian Arabic lexicon is compatible with the root-pattern framework. Furthermore, the differential processing (based on reaction time and error analysis differences) found can be best explained by McCarthy's model of the lexicon of Arabic.

Prunet, Béland, and Idrissi (2000) conducted an in-depth error analysis of paraphasias from ZT, a bilingual speaker (Arabic-French) with deep dyslexia following a cerebrovascular accident. They report a dissociation between the consonantal roots and vocalic patterns in ZT. His reading displays the characteristics of the deep dyslexia in both Arabic and French. His production consists of semantic, visual, and morphological errors, and shows a concreteness effect in reading aloud and impaired reading of non-words. ZT's errors of exchange changed the order of root consonants but never within the vowels of the vocalic patterns, suggesting that these two are separate cognitive entities. These findings are also in support of a root-pattern model rather than a stem-based view of the Arabic lexicon. ZT's case was studied through different experimental paradigms, and there is a general consensus that the errors he produced could be accounted for by a root-pattern framework only (Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi et al., 2002).

Further evidence in support of the root-pattern framework came from a study on processing regular and irregular plural forms in Levantine Arabic (Khwaileh et al., 2015). Khwaileh et al. (2015) carried out a picture-naming paradigm consisting of 90 pictures representing nouns in their singular, dual and plural forms. The participants were three Levantine Arabic speakers with aphasia following cerebrovascular accidents. The picture-naming task yielded error patterns of vocalic pattern substitutions only in the broken plurals, as omissions would result in phonologically illegal non-words. This error type can only be explained within the root-pattern framework. Bound suffix omissions occurred within the sound plurals but were significantly less frequent in all three participants.

To the best of the author's knowledge, none of the published experimental (psycholinguistic) and clinical (neuropsychological) studies on Arabic have supported the stem-based view on Arabic morphology. Evidence from natural data (healthy and impaired participants) supports a root-pattern framework on Arabic morphology, this approach is currently the major driving theory describing for Arabic morphology and therefore is the basis on which a response categorisation protocol should be based. The fact that this framework has been supported extensively by language processing data warrants devising clinical and assessment tools driven by the root-pattern theoretical framework rather than the stem-based approach, which has been supported by self-introspection data, and lacked processing data support, to date.

Syntax of arabic

There are two types of Arabic sentences; nominal sentences which begin with a noun or pronoun, and verbal sentences, which begin with a verb. Arabic is known for its sound and broken plurals (Idrissi, 1997; McCarthy & Prince, 1990). Sound plurals are derived through suffixation of masculine /-ʊ:n/ to the singular stem, as in singular /l-mʊʕəllɪm/ '(male) teacher' plural /l-mʊʕəllɪm ʊ:n], and the feminine {-aat} to the singular stem, as in singular / l-mʊʕəllɪm -ɑ/ '(female) teacher' plural /l-mʊʕəllɪm – a:t/. Comparatively, 'broken plurals' involve a non-linear morphological process, the consonantal roots are mapped onto specific plural templates, e.g. sing. /kalb/ ‘dog’ pl. /kıla:b/, sing. /manzil/ ‘house’ pl. /mana:zil/, sing. / qabr/ ‘grave’ pl. /qubu:r/).

According to Alkuhlani et al. (2011), Arabic nominals (i.e. nouns, adjectives, and verbs) inflect for definiteness (definite/indefinite), gender (masculine/feminine), and number (singular/dual/plural). This inflection is carried out through suffixation and vocalic pattern changes [i.e. the root /xtm/: /xa:tam/- singular /xawa:tım/-plural]. Arabic morpho-syntactic agreement is built around agreement of gender and number features for nouns and their adjectives; and verbs and their subjects.

Arabic adjectives agree with the nouns that they modify in definiteness, gender and number except for plural irrational (non-human) nouns, which always take feminine singular adjectives (Alkuhlani, Habash, 2011). This is demonstrated in /al-kara:si al-kabi:r-a(t)/ (the big (feminine singular adjective) chairs (masculine plural non-human noun)). Arabic verbs have the same kind of relationship with their subjects as nouns have with their adjectives. The only exception is that in an order in which the verb precedes and it is in past tense form, the subject agreement would only agree in gender. That is, with such word order, the verb keeps its singular form regardless of the number of the subject (e.g. / ʔakal-a al-ʔawla:d/ (the boys(definite masculine plural) ate (paste tense verb)).

Arabic verb phrases are far more complex than their noun phrase counterparts; verb phrases carry more agreement patterns. Arabic verbal sentences contain a verb and a subject (which may be a noun, pronoun, adverbial or relative clause). The verb agrees with the subject in number and gender when the subject precedes the verb. When a verb precedes a plural subject, the verb remains singular but agrees in gender and is therefore asymmetrical between singular and plural subjects. Agreement patterns that are grammatical in Subject-Verb order are not so in the Verb-subject order, as goes for overt plural subjects. Arabic allows subjects to be dropped (i.e., covert subjects like: /katab-tʊ- l-dars/ wrote-1st person .singular definite -lesson-accusative/ ‘I wrote the lesson’). In such a case, the verb completely agrees with the dropped subject. Thus, it is clear that word order interacts with number agreement. As per Holes (2004), Arabic verbal phrases also carry aspect and factuality, negation, mood and modality, time and tense.

The current study investigates only the categories of verbs and nouns. This is due to the fact that Arabic nouns and verbs together are a rich source for the investigation of the breakdown of inflectional processes, and are therefore key to developing an Arabic error categorisation protocol. Further, nouns and verbs are backed with extensive theoretical framework and evidence; providing the current study with a theory-driven background. The reason as to why Arabic adjectives are not emphasized on is that there exists little to no agreement as to whether they should be considered a distinct word class. Further, there seems to be no real notion of an adjective as a stand-alone category in Arabic, this is illustrated by how any Arabic adjective can be used as a noun or proper noun. Adjectives act only as modifiers, and at sentence-level; they agree with the subject.

Aim and scope of this paper

Most of the research on Arabic processing (Albustanji, 2009; Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi et al., 2002; Khwaileh, 2011; Khwaileh et al., 2014, 2015; Khwaileh, Mustafawi, Howard, & Herbert, 2016; Khwaileh et al., 2017; Khwaileh, Mustafawi, Herbert, & Howard, 2018; Khwaileh, Mustafawi, Howard, & Herbert, 2019;

Mimouni et al., 1998, 1997) lack use of standardized material and response categorisation protocols for error analysis. The aim of this paper is to develop a linguistically informed and data-driven response categorisation protocol for language breakdown of the Arabic language. The protocol is based on the root-pattern view on Arabic morphology due to the considerable processing evidence supporting this view, as discussed above.

A response categorisation protocol for all grammatical categories (nouns, verbs, adjectives, adverbs) at different modalities (production and comprehension) would be enviable, but unviable in this study. In the current study, focus is shed on the production of nouns, and verbs at the single word level on the one hand and agreement on a phrase level on the other. Developing a protocol for comprehension is irrelevant in the current context, since responses from comprehension tasks cannot be categorised for morpho-syntax.

Methodology

Participants

Fifty-one healthy native Gulf Arabic speakers were recruited in order to pilot the protocol developed during the First cycle. The purpose of including them is to yield alternative correct and other acceptable responses. Their mean age was 30.9 years old (min: 22; max: 61 years old). Females formed the majority of our sample (Female = 70%, Male = 30%). All participants were educated in Modern Standard Arabic during school (min: 12; max: 22 years of education). All participants reported normal development of speech and language, and normal hearing and vision. Participants were recruited through email invitations initially. Then, interested participants were asked to sign a consent form and fill out a questionnaire about their background. A confidentiality form was also signed by the researcher to confirm that all data are kept confidential and anonymous. This study was approved by Qatar University Ethics Review Panel.

Eight participants with aphasia who spoke the same variety were also recruited from two hospitals to yield possible patterns of breakdown. Participants had been attendants of speech and language clinics within the hospitals and were contacted through their speech and language therapists. Prior to their participation, they received an information sheet about the study, and were invited to sign a consent form. Their caregivers were involved in the recruitment process if patients were not able to communicate effectively. This experiment was ethically approved by the hospital's Review Panel. Table 1 shows Patients' information.

As reported by their speech and language therapists, all eight participants were adult native speakers of Gulf Arabic. Six participants were literate and two were illiterate. They

Table 1. Background information of participants with aphasia as reported by their speech and language therapists.

Participant code name	Sex	Age	Dialect	Education (in years)	L1	L2	Hemiplegia	Etiology
GA1	F	60	Gulf Arabic	12	Arabic	NA	Right side	CVA
GA2	F	62	Gulf Arabic	Illiterate	Arabic	NA	Right side	CVA
GA3	M	58	Gulf Arabic	16	Arabic	NA	Right side	CVA
GA4	F	29	Gulf Arabic	12	Arabic	English	Right side	CVA
GA5	M	84	Gulf Arabic	18	Arabic	English	Right side	CVA
GA8	M	51	Gulf Arabic	16	Arabic	English	Right side	CVA
GA12	F	76	Gulf Arabic	Illiterate	Arabic	NA	Right side	CVA
GA13	M	53	Gulf Arabic	16	Arabic	English	Right side	CVA

were reported to have normal development of speech and language prior to their injury. Six patients were diagnosed with non-fluent aphasia, and two with fluent aphasia. All participants were more than 6-months post onset at the time of the study. Their hearing and vision were normal or adjusted to normal.

Materials

Materials were a range of production tasks including two picture-naming tasks (nouns and actions), picture description, reading, repetition, and a recitation of Al-Fatiha verse from the Holy Quran. All these tests were taken from the Gulf Arabic Aphasia Test, Khwaileh et al. (2016). The picture-naming paradigm consisted of 24 object pictures and 20 action pictures. Pictures were configured to be 885 pixels (width) by 600 pixels (height) for presentation on a laptop screen with a screen resolution of 1024 by 768 pixels. Each of the patients group was presented with a reading task which included 32 simple words, 32 morphologically complex words, 20 function words, and 15 non-words. The presentation order was randomized and each word appeared individually on a laptop screen in written MSA (Modern Standard Arabic) form, which is the standard form of Arabic, unlike the picture naming test and picture description tasks in which the participant would produce responses using Gulf Arabic. Participants were asked to read each word out loud.

A repetition task was presented through auditory input to the patients group. The list of words that were presented included 20 simple words, 20 morphologically complex words, and 15 non-words. Participants were told they will hear each word only once, and are expected to repeat it. The non-words task was included in the repetition and reading tests to capture possible responses from patients; researchers and clinicians tend to use non-words to test the functionality of sub-lexical routes in models of lexical processing.

Patients were asked to recite the Al-Fatiha verse solely on their own. This verse from the Holy Quran is read from memory by Muslims a minimum of 17 times a day at each prayer of the five prayers Muslims observe on a daily basis. It was used to yield errors in the production of highly frequent and lexicalized utterances in Arabic. They were encouraged to recite it from memory and without asking for help or offering a cue by the tester.

The patients group were presented with a black and white drawing of a shopping mall with visitors and workers carrying out various actions which is a scene in harmony with the local culture. Participants were asked to describe the picture using as many details as possible. Scores were recorded according to how many types of word categories, inflections, and errors were produced (i.e. content words, nouns, verbs, phrases, sentences, tense markers, fillers, neologisms, etc.).

General procedure

The study consisted of two phases: phase one included the first cycle of developing the error categories informed by previous literature. The first cycle of error categories selection was based on studies reporting error analysis from Arabic patients with aphasia (Albustanji, 2009; Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi et al., 2002; Khwaileh, 2011; Khwaileh et al., 2015, 2017; Mimouni et al., 1998, 1997). The above-mentioned studies used a qualitative approach and reported their errors in details. Table 2 shows the main categories that were included during the first cycle.

Table 2. Category types adapted during the First cycle.

Category types	Description
A. Single word level coding	
Single Noun and Verb Lexical Error Patterns	
1.1 Correct response	The target response is produced correctly
1.2 Visual error	The response is visually related to the target
1.3 Semantic error	The response is semantically related to the target
1.4 Phonological error	The response is phonologically related to the target
1.5 Unrelated error	The response is completely unrelated to the target
1.6 Other error	The response type does not fit any of the previous category types
1.7 Tip of Tongue	Indication of knowing the object/action but cannot produce it
1.8 Unknown	Indication of not knowing the object/action and cannot produce it
1.9 No response	Failure to respond to the presented stimulus
Morpho-syntactic Errors in Single Nouns & Morpho-syntactic Errors in Single Verbs	
2.1 Correct response	The target response is produced correctly in its morpho-syntactic form
2.2 Incorrect gender	The response is produced with a different gender form to the target
2.3 Incorrect number	The response is produced with a different number form to the target
2.4 Incorrect person	The response is produced with a different person form to the target
2.5 Morpho-phonological errors in the consonantal root	There is a morpho-phonological error in the consonantal root.
2.6 Morpho-phonological error in Nouns	The target noun is produced with a morpho-phonological error
2.7 Morpho-phonological error in Verbs	The target verb is produced with a morpho-phonological error
2.8 Other error	The response type does not fit any of the previous category types
2.9 No response	Failure to respond to the presented stimulus
B. Phrase level coding	
Noun-Adjective agreement errors	
1.1 Correct response	The target phrase is produced correctly.
1.2 Morphosyntactic errors	There is a number/gender/person disagreement on the head or the modifier
1.3 Other response	The response type does not fit any of the previous category types
1.4 No response	Failure to respond to the presented stimulus
Verb agreement errors	
2.1 Correct response	The target phrase is produced correctly.
2.2 Morphosyntactic errors	There is a gender/person/tense/number disagreement on the subject or verb
2.3 Other error	The response type does not fit any of the previous category types
2.4 No response	Failure to respond to the presented stimulus

During the first cycle, the main categories that were included for single-word level tasks included single noun and verb lexical error patterns, morpho-syntactic errors in single nouns, and morpho-syntactic errors in single verbs. On the other hand, the main categories that were developed for the phrase-level task included noun-adjective agreement and subject-verb agreement. Each category and sub-category type were assigned a number code which corresponded to a specific behavior.

The second phase included the second cycle of developing the protocol in which healthy participants and patients with aphasia were administered the picture naming tasks, reading aloud task, repetition task, the Al Fatiha recitation task, and picture description. The purpose of the second cycle was to test error categories resulting from previous cycle against data from healthy speakers and patients with aphasia. It further aimed to revise the error categories, and test the reliability of these categories through asking different raters to rate the responses using the categories.

Responses were collected and coded using the coding protocol that was developed during the first cycle. The healthy group was only tasked with the picture-naming paradigm in order to yield types of response categories found to be produced by a typical speaker of the variety to find out the types of alternative correct and acceptable responses. The patients group did all the production tasks in order to yield possible patterns of error on the production level. Literate participants did all tests, while illiterate participants did all but the reading task.

To ensure that participants understood the instructions of each test, we used prompts in Arabic, translation: "Let us go through a number of practice items; What is this picture of? What does this word say? Can you repeat this word?" If participants did not understand the requirement of the test, further explanation, and practice items were given, prior to starting the actual test. All instructions and administration of the tasks were in Arabic. Responses from speakers with aphasia were filtered for pauses, hesitations, gesture responses, no responses, and restarts. The coding system accounted for such responses through introducing categories relevant to such responses.

Inter rater reliability

All tasks were recorded, and three raters were asked to categorise the responses using the developed protocol, independently. One rater is a speech and language therapist, the other was a linguist who is involved in the project, and the third was a linguist who is not involved in the development of the protocol. All three raters were given the same instructions and were asked to assign each response a category using the protocol in hand. In cases where the rater could not assign a category to the response, they were asked to take a note of the response and suggest a new category for discussion with the research team at the end of the rating process.

Results

The tasks described above yielded a wealth of data for the validation and standardization of the error categories presented in the current response categorisation protocol. We focused on the use and functionality of the response-categorising protocol under investigation in this study to check the reliability and validity of the developed tool. The sensitivity of the protocol in detecting different responses produced by both healthy and impaired participants.

Overall, the results (Table 3) show almost perfect performance for each of the naming tasks within healthy speakers, yielding correct, alternative correct and acceptable responses. All the error categories represented in the final coding protocol were informed by the data collected from participants. Their overall performance was considerably below that of healthy participants. Their performance was better for both object and action naming, and poorer on the reading task indicating that the reading task was more difficult

Table 3. Overall results for healthy and aphasic groups.

Task (total items)	Aphasic group (n = 8) *Accuracy mean (%)*	Healthy group (n = 51) *Accuracy mean (%)*
Picture naming (44)	**23.9 (54.3%)**	**42.5 (97%)**
Objects (24)	12.1 (50%)	23.1 (96%)
Actions (20)	11.8 (59%)	19.4 (97%)
Reading (99)	**74.4 (75%)**	
Simple words (32)	26.7 (83.4%)	
Morphologically complex words (32)	22.8 (71.2%)	
Function words (20)	15.7 (78.5%)	
Non-words (15)	9.2 (61.3%)	
Repetition (55)	**47.5 (86.3%)**	
Simple words (20)	18.4 (92%)	
Morphologically complex words (20)	18.4 (92%)	
Non-words (15)	10.7 (71.3%)	
Recitation of Al Fatiha verse (15)	**13.9 (92.7%)**	

for patients; with the simple word reading being easiest and the non-word reading being the most difficult. Furthermore, there was a significantly high number of morpho-phonological errors in the morphologically complex reading task, in comparison to the other tasks. Furthermore, the results showed that the repetition of simple and morphologically complex words were easier to repeat than non-words. Results also show that patients overall are below norm at reciting Al-Fatiha verse from the holy Quran. The performance of the patients with aphasia on these tasks was very informative for the response categorisation protocol. The obtained categories in the protocol were revisited in line with the feedback of the raters, based on the participants' responses.

Furthermore, results from the picture-description task show that non-fluent patients with aphasia showed better performance in producing nouns than verbs, and content words than function words, which is a sign of agrammatism. Eighty-two per cent of the errors produced by patients in the picture description task were accounted for by the existing response categorisation protocol developed in the first cycle described above. However, 18% of the responses warranted revision of the protocol; hence changes to the categories were implemented in cycle two below.

New categories (after the second cycle)

Following the two groups' performances on each task, we added the following new subcategories to encompass response types that were not categorised in previous studies on Arabic (Albustanji, 2009; Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi et al., 2002; Khwaileh, 2011; Khwaileh et al., 2015, 2017; Mimouni et al., 1998, 1997). These categories were also piloted along with the rest of the categories in Appendix A, in order to establish validity and reliability of the coding protocol. The following subcategories were added to the protocol following feedback from the three raters, and based on new encounters with patients, examples are provided in Appendix A:

- **1.3.3.1** Phonological error in producing sounds related to the target word: when the participant produced a sound that is associated with the target presented to them.

- **1.4.4** Phonological error using the Conduite d'approche (effort to approximate the target response): when the participant produced part of the word multiple times in an attempt to encode phonemes of the target word or phrase.
- **1.6.1** Self-correction: when the participant initially mis-produced the target item then corrected themselves.
- **1.6.2** Gesture action: when the participant expressed a word using a gesture instead of verbally.
- **2.3.6** Number inflection error: when the participant produced the target word in its incorrect number form as a result of a change to the vocalic pattern of the target.
- **2.5.2** Morpho-phonological error in substitution of the vocalic pattern: when the participant committed a substitution in the vocalic pattern of the target word.
- **2.5.3** Morpho-phonological error with a change in the root: when the participant produced the target word with a change in one of the phonemes of the consonantal root.
- **2.5.4** Morpho-phonological error with an omission in the root: when the participant committed an omission in one of the phonemes of the consonantal root.
- **2.6.2** Definiteness error in nouns: when the participant produced the target noun with a definite article when not required to do so.
- **2.6.3** Definiteness error in nouns: when the participant produced the target noun without a definite article when required to do so.
- **2.7.1** Morpho-phonological error with perfective form in verbs: when the participant produced the target verb in perfective tense instead of imperfective tense.
- **2.7.2** Morpho-phonological error with perfective form in verbs: when the participant produced the target verb in imperfective tense instead of perfective tense.
- **2.7.3** Morpho-phonological error in active tense for verbs: when the participant produced the target verb phrase in active tense instead of passive tense.
- **2.7.4** Morpho-phonological error in active tense for verbs: when the participant produced the target verb phrase in passive tense instead of active tense.
- **2.7.5** Morpho-phonological error in progressive forms for verbs: when the participant produced the target verb in a non-progressive form instead of in the progressive form.
- **2.7.6** Morpho-phonological error in gemination for verbs: when the participant produced the target verb in a geminate form.

Inter rater reliability

The second cycle of the development process involved piloting and refining the coding protocol with the responses that were produced by the two participant groups that were tested. Three raters using the developed response categorisation protocol scored their performance independently. Two linguists with clinical assessment training were involved along with the speech and language therapist. One was part of the current project and the other was not. The linguist external to the project went over the categories assigned by the other two raters and checked for accuracy. This strategy is supported by Chorney et al. (2015) who suggest that lack of previous exposure to the data would increase the likelihood of objective coding towards the responses.

Table 4. Reliability scores (correlations between the interraters per participants).

Task	Healthy group sample									
A. Picture-naming	**C2**	**C5**	**C14**	**C23**	**C38**	**C43**	**C47**	**C48**	**C49**	**C51**
Object naming	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Action naming	1.000	1.000	.892	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Task	*Aphasic group*									
A. Picture-naming	**GA1**	**GA2**	**GA3**	**GA4**	**GA5**	**GA8**	**GA12**	**GA13**		
Object naming	.993	.975	.971	.967	1.000	.999	.771	.916		
Action naming	.835	.874	.999	.830	1.000	1.000	1.000	.999		
B. Reading	**GA1**	**GA2**	**GA3**	**GA4**	**GA5**	**GA8**	**GA12**	**GA13**		
Simple word reading	.999	NA	.783	1.000	.953	.882	NA	1.000		
Morphologically complex word reading	.969	NA	.782	.556	.823	.942	NA	.925		
Function word reading	.948	NA	1.000	.815	1.000	1.000	NA	.999		
Non-word reading	.999	NA	1.000	.698	.999	.935	NA	1.000		
C. Repetition	**GA1**	**GA2**	**GA3**	**GA4**	**GA5**	**GA8**	**GA12**	**GA13**		
Simple word repetition	.622	1.000	.975	1.000	1.000	.552	.909	1.000		
Morphologically complex word repetition	.767	1.000	.983	1.000	1.000	.975	.975	1.000		
Non-word repetition	.999	1.000	.859	.798	.644	.933	.845	.649		
D. Al Fatiha recitation	**GA1**	**GA2**	**GA3**	**GA4**	**GA5**	**GA8**	**GA12**	**GA13**		
	1.000	1.000	1.000	1.000	1.000	1.000	.785	1.000		

Overall reliability scores are presented in Table 4. Results showed a strong interrater agreement for the healthy participants in the picture-naming task. The coding only differed in one instance during the action-naming task, due to the auditory unintelligibility of the response given by participant C14. The results had a high inter-rater agreement in almost all subtests, with spearman's correlations above 0.782 (all statistically significant). In addition, Cronbach's alpha values were higher than .600, except for one patient (SA8) in the single word repetition task and another patient (SA4) in the morphologically complex word reading task. These scores indicate medium inter-rater agreement (according to Cohen, 1990). They were computed for all production tasks except Picture description, as it was an open-ended task with no target response.

Validity

Validity refers to whether the response-categorisation protocol is measuring what it is supposed to measure. The construct validity of the current protocol is supported by the root-pattern theory which has extensive processing evidence supporting it (Berent, Vaknin, & Marcus, 2007; Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi et al., 2002; Khwaileh, 2011; Khwaileh et al., 2015, 2017; Mimouni et al., 1998, 1997), unlike the stem-word theory for Arabic morpho-syntax. This gives the current protocol a solid theoretical grounding and basis that has been previously used in numerous experiments.

Face validity of the protocol is comprehensive in its current version. The protocol aims to score any single-word level for nouns and verbs, as well as a noun and verb phrases from typical and atypical participants. In devising the response categories, we factored in all potential error patterns depending on the morpho-syntax of the bare word categories, and agreements within any Arabic noun or verb phrase. We tested 51 healthy participants and compared performances between a sample of 10 healthy participants and eight patients participants on production tasks in the second cycle. During both cycles of our development, we revised and updated the categories and subcategories of the protocol to include categories that encompass all the types of responses produced by those participants.

Discussion

The aim of this paper is to introduce a linguistically informed and data-driven response-categorisation protocol for Arabic to help researchers and clinicians categorise responses from healthy and brain damaged participants in language production tasks. The categories developed are driven by extensive processing evidence on Arabic non-linear morphology. Psycholinguistic evidence has supported the non-concatenative nature of Arabic morphology (e.g. Boudelaa & Marslen-Wilson, 2000, 2001, 2004, 2010, 2011, McCarthy, 1981), as has neuropsychological evidence (e.g. Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi et al., 2002; Khwaileh, 2011; Khwaileh et al., 2015, 2017; Mimouni et al., 1998, 1997). The existence of such clinical data based on the root-pattern framework drove the current coding protocol.

To the best of our knowledge, this is the first protocol of its type for the Arabic language. The breakdown of language has been investigated extensively in English and other Indo-European languages, using clinical and research tools developed for those languages. The morpho-syntax and phonology of those languages are different from that of Arabic. In the past 20 years, morpho-syntactic breakdown in Arabic has been under investigation (Albustanji, 2009; Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi et al., 2002; Khwaileh, 2011; Khwaileh et al., 2015, 2017; Mimouni et al., 1998, 1997) placing pressure on researchers to develop tools for the Arabic language and its different varieties.

Furthermore, the developed protocol has fulfilled the requirements of a behavior coding protocol in line with Chorney et al. (2015). The currently developed protocol is valid and has medium to high reliability as recommended by previous studies (Chorney et al., 2015; Saldaña, 2009). The results from the reliability and validity scores indicate that the categories in the current coding protocol are were appropriate, as shown in most high interrater agreement scores where lower scores only occurred as a result of auditory unintelligibility of the recordings rather than a conflict in using the categories that were developed.

The categorisation protocol categorises both word- and phrase-level responses based on the non-concatenative nature of Arabic morphology, fulfilling the condition of a focused coding. This is also in line with the focused coding condition recommended by Saldaña (2009) and Chorney et al. (2015), who maintain that a common/shared concept needs to be present between all categories. It also fulfills the elaborative coding condition, which takes categories that have already been developed in a previous study in order to modify or improve them, since the categories were derived from responses reported in previous literature on Arabic (e.g. Béland & Mimouni, 2001; Idrissi & Kehayia, 2004; Idrissi et al., 2002; Khwaileh, 2011). This method is important for the aim of building on a previous study and elaborating on major theoretical findings.

Overall, the developed protocol takes into account both typical and atypical production of nouns and verbs, and categorises errors into both language-based (e.g. phonological errors, semantic errors, morpho-syntactic errors), and non-language-based errors (e.g. visual errors). The codes are informed by Arabic linguistic features such as inflection for gender, person, and number; and on a larger scale – agreement in noun phrases and verb phrases.

The protocol described in this study is modeled on very basic principles of the Arabic noun and verb phrases morpho-syntax, giving it the flexibility to be adopted for research from different Arabic varieties and sub-varieties. This in turn gives the tool high scalability and applicability throughout Arabic-speaking regions. Such a theoretically informed and evidence-driven coding protocol assures its applicability in both clinical and research fields. Nevertheless, it only acts as a baseline for the different types of responses that can be produced by an Arabic healthy or impaired speaker, and to assess their types of deficits. It is noteworthy to maintain that the reliability of the Arabic coding protocol needs to be achieved for all varieties of Arabic.

The daily clinical assessment and intervention interaction with individuals with aphasia reassure the well-known coat that "we can only be sure to improve what we can actually measure". This protocol provides a primary fingerprint/scoring-reference guideline for clinicians toward standardizing the process of assessing types of errors in qualitative and quantitative terms. Although the protocol is not holistic since it does not include all the Arabic linguistic features, yet, it is multidimensional measure tool as it covers the most common linguistic categories that are usually found to be in error among speakers with aphasia. This tool is an important step toward a more linguistically specific, well utilized-focused language evaluation tool, as it is practically useful for primary intended users who are the speech language pathologists/therapist in clinical settings. The tool facilitates a more linguistic profile analysis of expected errors. Furthermore, it guides clinicians toward a more standardized approach for measuring outcomes of their intervention strategies. Finally, it increases their inter and intra-rater reliability in scoring their analysis.

Acknowledgments

This project was supported by the Qatar National Research Fund; Grant NPRP7-1506-3-390. The authors would like to thank Ms. Shereen El Buy and Ms. Noor Numan for their help in completing this paper.

Disclosure statement

The authors report no conflict of interests.

Funding

This research was funded by a Qatar National Research Fund (National Priorities Research Program), Grant Number NPRP7-1506-3-390.

ORCID

Tariq Khwaileh http://orcid.org/0000-0002-2590-6920
Eiman Mustafawi http://orcid.org/0000-0001-5271-9461

References

Albustanji, Y. M. (2009). *Agrammatism in Jordanian Arabic Speakers* (Unpublished Ph.D thesis). Ohio State University.

Alkuhlani, S., & Habash, N. (2011). *A corpus for modeling morpho-syntactic agreement in Arabic: Gender, number and rationality*. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: short papers (pp. 357–362), Portland, Oregon.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge, MA: Cambridge university press.

Béland, R., & Mimouni, Z. (2001). Deep dyslexia in the two languages of an Arabic/French bilingual patient. *Cognition*, *82*(2), 77–126. doi:10.1016/S0010-0277(01)00148-2

Bell, D. C, & Bell, L. G. (1989). Micro and macro measurement of family systems concepts. *Journal of Family Psychology*, *3*(2), 137–157.

Benmamoun, E. (1999). Arabic morphology: The central role of the imperfective. *Lingua*, *108*, 175–201. doi:10.1016/S0024-3841(98)00045-X

Benmamoun, E. (2003). The role of the imperfective template in Arabic morphology. In J.Shimron (Ed.), *Language processing and language acquisition in a root-based morphology* (pp. 99–114). Amsterdam: John Benjamins.

Berent, I., Vaknin, V., & Marcus, G. F. (2007). Roots, stems, and the universality of lexical representations: Evidence from Hebrew. *Cognition*, *104*(2), 254–286.

Boudelaa, S., & Marslen-Wilson, W. D. (2000). *Non-concatenative morphemes in Language Processing: Evidence from Modern Standard Arabic Proceedings of Spoken Word Access Processes* (pp. 23–26), Nijmegen, The Netherlands: Max Planck Institute.

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

Boudelaa, S, & Marslen-Wilson, W. D. (2001). Morphological units in the arabic mental lexicon, cognition. *81*, 65–92.

Boudelaa, S, & Marslen-Wilson, W. D. (2004). Abstract morphemes and lexical representation: The CV-skeleton in arabic. *Cognition*, *92*, 271–303.

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

Boudelaa, S., & Marslen-Wilson, W. D. (2010). Aralex: A lexical database for Modern Standard Arabic. *Behavior Research Methods*, *42*(2), 481–487. doi:10.3758/BRM.42.2.481

Boudelaa, S, & Marslen-Wilson, W. D. (2013). Morphological structure in the arabic mental lexicon: Parallels between standard and dialectal arabic. *Language and Cognitive Processes*, *28*(10), 1453–1473.

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

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

Chorney, J. M., McMurtry, C. M., Chambers, C. T., & Bakeman, R. (2015). Developing and modifying behavioral coding schemes in pediatric psychology: A practical guide. *Journal of Pediatric Psychology*, *40*(1), 154–164. doi:10.1093/jpepsy/jsu099

Cohen, J. (1990). Things I have learned so far. *American Psychologist*, *45*, 1304–1312. doi:10.1037/0003-066X.45.12.1304

Idrissi, A.(1997). "Plural formation in Arabic." In M. Eid and R. Ratcliffe (Eds.) *Perspectives on Arabic LinguisticsX*, (pp. 123–145). Amsterdam: John Benjamins.

Frost, R. (1998). Towards a strong phonological theory of visual word recognition: True issues and false trails. Psychological Bulletin,*123*, 71–99.

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

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

Holes, C. (2004). *Modern Arabic: Structures, forms, and varieties* (p. 419). Washington, DC: Georgetown University Press.

Idrissi, A., & Kehayia, E. (2004). Morphological units in the Arabic mental lexicon: Evidence from an individual with Deep Dyslexia. *Brain and Language*, *90*, 183–197. doi:10.1016/S0093-934X(03)00431-0

Idrissi, A., Prunet, J. F., & Béland, R. (2002). Metathesis errors in patients speech: Implications for the Status of Abstract Roots in Arabic. *Brain and Language*, *83*(1), 90–92.

Khwaileh, T., Body, R., & Herbert, R. (2017). Lexical retrieval after Arabic aphasia: Syntactic access and predictors of spoken naming. *Journal of Neurolinguistics*, *42*, 140–155. doi:10.1016/j.jneuroling.2017.01.001

Khwaileh, T., Mustafawi, E., Herbert, R., & Howard, D. (2018). Gulf Arabic nouns and verbs: A standardized set of 319 object pictures and 141 action pictures, with predictors of naming latencies. *Behavior Research Methods*, *50*(6), 2408–2425. doi:10.3758/s13428-018-1019-6

Khwaileh, T, Mustafawi, E, Howard, D, & Herbert, R. (2019). Imageability, familiarity, and age of acquisition ratings for abstract nouns, verbs, and adjectives in Arabic. *The Mental Lexicon*, *13*(3), 354–387.

Khwaileh, T. A. (2011). *Lexical retrieval and morpho-syntactic processing in healthy speakers and speakers with aphasia: Evidence from Arabic* (Ph.D. Dissertation). Sheffield, UK: The University of Sheffield.

Khwaileh, T. A., Body, R., & Herbert., R. (2014). A normative database and determinants of lexical retrieval for 186 Arabic nouns: Effects of psycholinguistic and morpho-syntactic variables on naming latency. *Journal of Psycholinguistic Research*, *43*(6), 749–769. doi:10.1007/s10936-013-9277-z

Khwaileh, T. A., Body, R., & Herbert, R. (2015). Morpho-syntactic processing of Arabic plurals after aphasia: Dissecting lexical meaning from morpho-syntax within word boundaries. *Cognitive Neuropsychology*, *32*(6), 340–367. doi:10.1080/02643294.2015.1074893

Khwaileh, T. A., Mustafawi, E., Howard, D., & Herbert, R. (2016). An aphasia battery for Qatari/Gulf Arabic. *Front. Psychol. Conference Abstract: 54th Annual Academy of Aphasia Meeting*. doi:10.3389/conf.fpsyg.2016.68.00026

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

McCarthy, J. J. (1983). A prosodic account of Arabic broken plurals. In I. Dihoff (Ed.), *Current Trends in AfricanLinguistics I*, (pp. 289–320) Foris, Dordrecht.

McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, *17*(2), 207–263.

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

Mimouni, Z., Kehayia, E., & Jarema, G. (1997). Agrammatic aphasia in Arabic. *Aphasiology*, *11*, 125–144. doi:10.1080/02687039708248460

Mimouni, Z., Kehayia, E., & Jarema, G. (1998). The mental representation of singular and plural nouns in Algerian Arabic as revealed through auditory priming in agrammatic patients patients. *Brain and Language*, *61*(1), 63–87. doi:10.1006/brln.1997.1879

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

Saldaña, J. (2009). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage Publications Ltd.