Patterns and Functions of Total Reduplication in Classical Arabic and Modern Standard Arabic: A Corpus-Based Study

Mohammed Modhaffer (Sana'a)/Sivaramakrishna Challavenkata (Mysore)

Abstract

In this paper, we investigate the form, salient patterns and core functions of word-level total reduplication in Classical Arabic (CA) and Modern Standard Arabic (MSA). Using a multi-genre corpus of 469 million words, we extract total reduplication (TR) candidates into an SQL database, manually filter them, and perform concordance search to identify the patterns and functions. Data analysis reveals **nine** patterns and **eleven** functions of TR and compares their relative frequency in each variety. The functions of TR are mapped into **two** broad categories: morphological and semantic/pragmatic. Results show an interesting variation in terms of top functions being favored by the two varieties. While TR is favored by CA to express **serial ordering**, MSA is noticed to favor it to express **intensification**. The empirical findings of this study provide a reliable quantification of the status of TR in CA and MSA which is rather difficult to obtain by theoretical means: on the one hand, TR in Arabic is not as productive as in other languages such as Indonesian. On the other hand, it is more common in CA than in MSA because the latter usually resorts to using loose phrases to express the same concepts expressed by TR in CA.

1 Introduction

The Reduplication is the "doubling of some part of a morphological constituent (root, stem, word) for some morphological purpose" says Inkelas (2014: 114). Yet, Inkelas' definition restricts the object of reduplication to "morphological constituents" and the motivation for reduplication to "morphological purposes". According to Classical Arabic grammatical theory, however, reduplication can operate on morphological as well as syntactic objects while it can be used to serve morphological and semantic purposes.

This study aims to answer the following questions:

1. What is the status of total reduplication in Classical Arabic vs. Modern Standard Arabic?
2. What is the form of total reduplication in Arabic?
3. What are the morphological functions of total reduplication in Classical Arabic and Modern Standard Arabic?
4. What are the pragmatic functions of total reduplication in Classical Arabic and Modern Standard Arabic?

In addition, this study aims to examine the real use of total reduplication in Arabic as attested in the corpus, and list the patterns according to the frequency ranking.

Throughout this paper, the term TR will refer to total reduplication in word level, CA to Classical Arabic, MSA to Modern Standard Arabic. When using the word "Arabic" alone, it refers to Standard Arabic (both CA and MSA).

2 Theoretical Background

One of the common problems reduplication raises for Natural Language Processing (NLP) applications is the difficulty of identification and extraction. Yet, it is quite safe to argue that this problem is language-dependent. That is, a given language may possess a complex reduplication system while another one may show a straightforward system of reduplication (i. e. Arabic in case of TR). For instance, it has been reported that recent development in Indonesian (Western Malayo-Polynesian, Sundic) orthography uses a "superscript 2 to indicate reduplication: *orang*^{2} for *orang-orang* ('persons') vs. *orang* ('person')" (Bussmann 2006, s. v. *diacritic*). A superscript is most likely to be stripped (amongst other numbers and punctuation marks) before the text corpus is manipulated, and it is not likely to remain a superscript in a text file where most of NLP data is stored.

Wang (2005) asserts that most research into reduplication is form-oriented, not functional or pragmatic. In general, reduplication has been extensively investigated within the morpho-phonological frame. Following Broselow and McCarthy's (1983) "A Theory of Internal Reduplication", many researchers, in the 1980s and 1990s, focused on forms and patterns of (internal) reduplication in many languages of the world, including Arabic. Yet, little attention has been paid to the semantic functions of reduplication. Even in the present-day publications, Arabic remains almost unnoticed. For example, in his survey of the meanings and functions of TR across several languages, Kallergi (2015) did not cite even one example of Arabic reduplication.

Suçin (2010) conducted a contrastive study of Turkish and Arabic reduplication. A serious weakness with his approach, however, is that he followed the tradition of CA grammarians in his characterization of what constitutes reduplication. He cited examples of reduplication consisting of three or more words. For instance

(1) ðahabat ʔibiluhu *ʃaðaran maðaran baðar* (Suçin 2010: 210]
'His camels scattered in all directions.'

The three words *ʃaðaran maðaran baðar* were regarded as reduplication by CA grammarians. In modern lexicography, however, these words are an instance of set combinations and collocations which acquired an idiomatic meaning. Similarly, *ħasanun ħasanun ħasanun* (gloss: "What a beauty!" [ibid, p. 210]) cannot be regarded as reduplication because these words are actually instances of repetition. That is because they do not add anything new to the original meaning of the word *ħasanun*. In Yemeni colloquial Arabic, native speakers use the word *tˁajjib* to mean 'OK' and sometimes it is repeated several times and such repetition does not add any semantic value to the original meaning of the single word *tˁajjib*. Finally, Suçin's

characterization of ʔitbaːʕ "subordination" as a form of reduplication is not appropriate. These limitations show the need to study Arabic TR in the framework of modern linguistics.

With respect to corpus-based investigations into reduplication, few studies adopted corpus-based approach. In most of these studies, the used corpora were often of small size. For example, Harley and Layva (2009) investigated the form and meaning of verbal reduplication in Hiaki (Uto-Aztecan, United States of America) using 343 verb forms and sentences. Further, Wang (2005) used a corpus of 1687 items to analyze English discourse in relation to reduplication and repetition. Chakraborty and Bandyopadhyay (2010) used a corpus of 14,810 tokens (3675 distinct word forms) in designing a rule-based system for identification of Bengali (Indo-Aryan, Bangladesh and India) reduplication and its semantic analysis.

To the best of our knowledge, no previous study has attempted a corpus-based approach to Arabic TR.

2.1 Types of reduplication

Reduplication is divided by linguists into two main types: **partial reduplication** and **total (or full) reduplication**. Partial reduplication is reduplicating a part of a constituent, either prefixing, infixing or suffixing. For example, Macdonald and Darjowidjojo (1967) state that in Modern Indonesian "the first consonant of the root, followed by the vowel /ë/ is prefixed to the root:

(2) *laki* 'male' → *lëlaki* 'male, man, husband'

Total reduplication copies the entire constituent without change as in Arabic

(3) *ʔalf* 'a thousand' → *ʔalf ʔalf* 'a million'

Rubino (2005: 15) divides reduplicative constructions into three types: simple, complex and automatic. In his terms, "a simple [reduplicative] construction is one in which the reduplicant matches the base from which it is copied without phoneme changes or addition". In this sense, TR in Arabic is "simple reduplicative constructions". A complex reduplicative construction is that in which "sound change or phoneme order reversal" occurs. An instance of a complex reduplication is the Bahuvrihi compounds in Mangarayi (Australian):

(4) *gurjag* ('lily'') → *gurjurjagɟi* (''having lots of lilies'')
baŋgal ('egg') → *banŋgaŋgalj*i ('having lots of eggs')

The third type of reduplication in Rubin's taxonomy is automatic reduplication which is "obligatory in combination with another affix, and which does not add meaning by itself to the overall construction; the affix and the reduplicated matter together are mono-morphemic, e. g. the Ilocano aginCV- prefix to express pretense: *singpet* ('behave') → *aginsi-sinpet* 'to pretend to behave'" (ibid: 18).

2.2 Brief description of Arabic

Standard Arabic is composed of Classical Arabic and Modern Standard Arabic. CA is the variety of the Holy Qur'an. It served as the medium of communication, literature, trade and commerce during the golden era of Islamic Empire (7th Century – 13th Century circa). MSA is a

revival copy of CA and it came into existence in the 19th Century. In terms of spelling and morphology, MSA does resemble CA to a large extent, but both differ in terms of structure, where MSA is said to use a simpler structure. For instance, the following structure longer appears in MSA texts:

(5) ʔaʕtˁaj-ta-ni:-ha:
give[PAST-2+SG+MASC+NOM]-me[ACC]-it[ACC]
'You gave it to me.'

Instead, in MSA the above structure is expressed in a way similar to the following:

(6) ʔanta ʔaʕtˁaj-ta-ha: l-i:
You gave-NOM-it to-me-GEN
"You gave it to me."

In the present-day educational system of Arab world, MSA is learnt at elementary and upper-elementary school onwards, while CA is learnt at higher levels of education such as graduate and post-graduate programs in Arabic language and literature. MSA is learnt explicitly through textbooks and it is rather difficult. However, CA is learnt explicitly through classical books and manuscripts which date back to the 7th Century (i. e. Seebawayh's era) and it is very hard to learn even for native speakers of Arabic. A substantial part of the vocabulary of CA, which had been employed by the Abbasid author Al-Jaħiz̟ (died 869), is no longer employed by any contemporary Arab author. A native speaker of Arabic pursuing a post-graduate program in Arabic literature would hardly understand the books of Al-Jaħiz̟ without recourse to **Lisaanul Arabi** – the standard dictionary of Arabic.

Whatever be the case, there exists some overlap between CA and MSA in terms of lexicon and structure. This may be attributed to the fact that the Holy Qur'an is still read and learnt by every (Muslim) native speaker of Arabic. That is, the Holy Qur'an and the huge body of religious and literary texts which are written in CA have served as an archive for CA. For more information on both CA and MSA, see Versteegh (2014), Watson (2002), Bateson (1967) and Al-Huri (2015).

Nowadays, every newborn child in the Arab world is raised speaking the regional variety of Colloquial Arabic of their home country (with lots of sub-varieties in the same country/region). Colloquial Arabic is composed of 22 regional varieties and further sub-varieties in each regional colloquial variety. For instance, Yemeni colloquial Arabic is further subdivided into San'ani dialect, Taizi dialect, Tehami dialect, Hadramout dialect and the dialects of eastern parts of the country.

There exists a big difference between MSA and its 22 regional colloquial varieties in terms of structure and lexicon. Further, the 22 regional colloquial varieties of MSA are quite different from each other and sometimes it is not easy for two speakers of different regions, for instance from Morocco and Yemen, to understand each other if they each speak in their regional colloquial varieties. As a consequence, MSA is used as a lingua franca for communication among the MSA speakers who come from different regions. In each regional colloquial variety, there exists an intersection with MSA in terms of lexicon; that is, there is a vast quantity of

lexical doublets, and in some cases lexical triplets, too. The following data set illustrates the lexical doublets found in Yemeni Colloquial Arabic and MSA:

(7)	**Modern Standard Arabic**	**(Yemeni) Colloquial Arabic**	
	qitˁtˁ-un	dimmi:, biss	'cat'
	cat-INDEF	cat	
	birk-at-un	ma:ʒil	'pond'
	pond-FEM-INDEF	pond	
	ðiʔb-un	θaʕajl	'fox'
	fox-INDEF	fox	

2.3 Reduplication in Arabic

Classical Arabic grammarians classified reduplication under the grammatical category of *tawkeed* 'emphasis' which is of two types: lexical and semantic. *Al-tawkeedul lafzi* 'lexical emphasis' is the reduplication of the same constituent in which the reduplicant and the reduplicated are of the exact orthographic shape, .e. g. *ruwajd-an ruwajd-an* 'slowly slowly – i. e. very slowly'. *Al-tawkeedul maʕnawi* 'semantic emphasis' is another type of emphasis in which the meaning of the reduplicant is similar to the meaning of the reduplicated constituent while the two are not exactly of the same orthographic shape. In the following example, *nafsuhu* 'himself' is an instance of semantic emphasis.

(8)	*ħaḍara*	*zajd-un*	*nafs-uh-u*
	came	Zaid-NOM	self-his-NOM
	'Zaid came **himself**'		

Al-tawkeedul lafzii (lexical emphasis) devised by CA grammarians is known in modern linguistic theory as total reduplication in which two constituents of the same orthographic shape appear in juxtaposition to each other. In CA grammatical theory, this type of reduplication is licensed to operate on all parts of speech categories (which are nouns, verbs and particles) as recognized by the early CA grammarians such as Ibn Aqeel (died 1367, 1980). In addition, TR is licensed to operate on full sentences.

In his *Alfiat* (literally 1000 verses of Classical Arabic Grammar), Ibn Malik (died circa 1273, 2012) prescribed the rules of *tawkeed* (i. e. emphasis) in lines 530 – 533. We give below the line which prescribes the rules of TR:

530 – وما من التوكيد لفظي يجي مكررا كقولك ادرجي ادرجي

530 – wa ma: minal tawkeedi lafzijjan jaʒi: mukarraran kaqawluki ʔudriʒi: ʔudruʒi:

(Translation: 'And regarding the lexical emphasis, it [allafz, i. e. the word] comes repeated as your [3+SG+FEM] saying *ʔudruʒi: ʔudruʒi:* "insert, insert"'; translation is ours). It has to be noted that a sentence in Arabic may be constituted from one word. That is due to the drop features. The example given by Ibn Malik is a clear instance of it.

With the advent of modern linguistic theory, the three parts of speech categories of Arabic have been expanded into eight categories, viz. noun, pronoun, verb, adjective, adverb, preposition,

particle and interjection[1]. Theoretically speaking, TR in Arabic (both CA and MSA) is licensed to operate on every category *except prepositions*, viz. nouns, pronouns, verbs, adjectives, adverbs, particles, interjections, phrases, clauses and full sentences. Examples are given below:

Nouns:

(9) ʔakal-tu *t-tufaːħ-at-a* *t-tufaːħ-at-a*
I ate-NOM the-apple-FEM-ACC the-apple-FEM-ACC
'I ate *just* the apple.'

Pronouns:

(10) *ʔanta* *ʔanta* *ʔallaːhu-u*
You you Allah-NOM
'You are Allah' (Supplication)

Verbs:

(11) *naʒaħ-a* *naʒaħ-a* *l-muʒtahid-u*
succeeded-ACC succeeded-ACC the-diligient-NOM
'The diligent passed'

Adjectives:

(12) ʔa*l-ħadiːθ-u* ***tˤ-tˤawiːl-u*** ***tˤ-tˤawiːl-u***
the-sayin-NOM the-long-NOM the-long-NOM
'The **very long** Hadeeth'

Adverbs:

(13) maʃajt-u *ruwajd-an* *ruwajd-an*
I waked-NOM slowl-ACC slowly-ACC
'I walked very slowly.'

Particles:

(14) *laː laː* ʔaxuːn-u l-ʕahd-i
not not I cheat-NOM the-oath-GEN
'I never ever break oath.'

Interjections:

(15) *hajhaːta* *hajhaːta*
impossible impossible
'quite impossible'

Phrases:

(16) *fi-l-lajl-i* *fi-l-lajl-i* tasːħ-u l-ʕabar-aːt-i
in-the-evening-GEN in-the-evening flow.3.SG.FEM-NOM the-tears-GEN
'In the evening, the tears flow.'

---

[1] The classical two types of sentences *ʒumlatun ʔismiyyatun* 'nominal sentence' and *ʒumaltun fiʕliyyatun* 'verbal sentence' have been redefined to yield four types of sentences: affirmative, negative, interrogative and interjective sentences. What was known in CA grammar as the *ʃibhu ʒumaltin* 'semi-sentence' is now known as a phrase or a clause. Further, the *ʒumlatun ʔismiyyatun* 'nominal sentence' is now known as *verbless sentence* or *equative sentence*

Verbless sentence:

(17) *hija l-fa:ʔiz-at-u* *hija l-fa:ʔiz-at-u*
she the-winner-FEM-NOM she the-winner-FEM-NOM
'She is the winner'

Sentences with verbs:

(18) *ʒa:ʔa* *l-matˁar-u* *ʒa:ʔa l-matˁar-u*
came.3.SG.MASC the-rain-NOM
'The rain came, i.e. it is going to rain'

Examples (16), (17) and (18) show that Arabic licenses TR in para-word level. However, we will confine our investigation to the word-level bigram TR.

3 Methodology

This section gives an account of the methodology adopted throughout this investigation. Section 3.1 describes the corpora: counts and genres. Section 3.2 lists the extraction algorithm. Section 3.3 provides details of manually filtering the candidates and discarding the false positives. Section 3.4 explains the concordance search of every TR. Finally, details of the IPA transcription and POS tagging of the true TR candidates are furnished in Section 3.5.

3.1 The Corpora

In terms of text corpora, a total of 469,093,365 words were investigated, of which 290,184,738 tokens are CA texts and the remaining 172,747,927 are MSA texts. The following table shows the genres and their counts in the corpora.

Genre	Tokens	Variety
Literature	28325819	CA
Jurisprudence	26216999	CA
Prophet's Sayings	81648668	CA
History	27499943	CA
Lexicons	16114913	CA
Prophet's Biography	22798962	CA
Holy Qur'an's Explanation	87579434	CA
Encyclopedic texts	11436078	MSA
Law	12685940	MSA
Defense	18977700	MSA
Medical	12136714	MSA
Newswire	117511495	MSA
Total	**469,093,365**	

Table1: Corpora genres and counts

CA texts were extracted from 5000 e-books belonging to the Shamela Library which can be obtained for free.[2] The Shamela Library was classified by human classifiers. The e-books were

[2] http://sourceforge.net/projects/albahhet/files/ShamelaEpub/shamela_epub.tar.gz/download [06.12.2018].

converted into UTF-8 text files. Then the texts were cleaned from punctuation marks, vocalization marks (diacritics) and symbols.

MSA newswire texts were retrieved from the corpus collected by Dr. Ahmed Abdelali. It contains 113 million tokens and it can be obtained for free[3]. The remaining four million tokens of newswire as well as the remaining genres were crawled from the World Wilde Web. MSA texts are, by default, not vocalized, and the punctuation marks as well as symbols were simply stripped at the time of crawling.

3.2 Extraction

For the TR in Arabic, the whole constituent is reduplicated irrespective of case, person, number, gender or definiteness. As such, extracting the candidates is intuitive and straightforward. Figure (1) shows the extraction algorithm which is being implemented in Python 3.4.

```
# Algorithm to Extract Arabic Total (Bigram) Reduplication
# Input: Text files (cleaned and stripped)
# Output: Total reduplication candidates, ranked from the highest to the lowest.
# Method:
Step 1: Declare a container CONT for reduplication candidates
Step 2: Open text files in UTF-8 formatting and strip new line characters
Step 3: Tokenize the text to produce an array of words ARR
Step 4: Assign index to 0
Step 5: Loop over ARR.
 Compare ARR[index] with ARR[index + 1].
 If they are the same, then append them to the CONT.
 Increment index by 1
Step 6: Calculate frequencies of candidates in CONT and rank them from the highest to the lowest.
Step 7: Save the candidates and their frequency into SQL database.
```

Figure 1: Extraction Algorithm

The results of the extraction process are shown in Table 2 below:

	Total Extracted		**>= 5**		**After filtering**		**After concordance**	
	CA	MSA	CA	MSA	CA	MSA	CA	MSA
Bigrams	25192	8130	3615	1476	2897	1062	**374**	**132**

Table 2: Counts of the extracted TR candidates across CA and MSA

The final number of reduplication entries in both CA and MSA bigrams evaluated to 506 whereas the remaining candidates did not qualify as reduplication. It has to be noted that stripping the punctuation marks from the corpora resulted in more than 3000 false positive candidates in CA texts and 1300 more in MSA texts. Furthermore, all the candidates whose frequency is less than five have been discarded as insignificant. These two factors straightforwardly justify why the final TR candidates dropped dramatically from 3861 to 374 in CA, and from 1603 to 132 in MSA.

[3] http://aracorpus.e3rab.com/argistestsrv.nmsu.edu/AraCorpus.tar.gz [06.12.2018].

3.3 Manual filtering

The final number of reduplication entries in both CA and MSA bigrams and trigrams evaluated to 506 whereas the remaining candidates did not qualify as TR instances. Principal reasons include:

3.3.1 Object of a transitive verb whose subject is a compound or a genitive construct [X, Y] where Y = object

(19) *jadxul-u* *ʔahl-u* ***l-ʒann-at-i*** ***l-ʒann-at-i***
enter-NOM owners-NOM the-heaven-FEM-GEN the-heaven
'Believers will enter heaven'

In the above example, *lʒannati lʒannati* was deemed as possible TR candidate. However, the first *lʒannati* is part of a genitive construct ʔahlu *lʒannati* and the second *lʒannati* is a direct object of the verb *jadxulu*.

3.3.2 Parenthetic sentences or phrases

(20) daxal-u: **ʕalaj-hi** (**ʕalaj-hi** s-sala:m-i)
they entered-NOM on-him (on-him the-peace-GEN)
'They visited him – peace be upon him'

Due to stripping the punctuation marks from "ʕalajhi – ʕalajhi", it was extracted as potential TR candidate. It has to be noted that the definite article /ʔal/ assimilated to /s/ before the phoneme /s/, i.e. /ʔal-sala:m-i/ → /s-sala:m-i/.

3.3.3 Metalanguage

In the context of explaining a linguistic or a grammatical construction, the same construction may follow the metalinguistic construction immediately, e.g.

(21) bi-ħaðf-i **ʕala: ʕala:** ħadd-i qawl-i ʃ-ʃa:riħ-i
by-omitting-GEN on on just-GEN saying-GEN the-annotator-GEN
'by omitting [the preposition] "on", as the annotator said'

3.3.4 Explanation

This comes in the form of X : X is …., e.g. *"halaka: halaka means died"*. Due to stripping all the punctuation marks, it would look as *"halaka halaka means died"*. These constructions are abundant in lexicons corpus.

3.3.5 End of a sentence and beginning of another

In this case, the end of a previous sentence is the same as the beginning of an immediately adjacent one. For example:

(22) taħarra:-hu **kaða:lika. kaða:lika** qa:l-a r-ra:ʁib-u
he watched-it like that. Like that said-ACC the-raʁib.PN-NOM
'He guessed it like that. ʔarraʁibu (an author) said like that.'

3.3.6 Syntactic reduplication

Due to prefixes, suffixes, infixes and the drop features in Arabic, a phrase or a sentence may be composed of one word, e.g.

(22) raʔaj-tumu:-h raʔaj-tumu:-h
saw-you.PL.MAS-him
'Did you *really* see him?'

Syntactic reduplication in Arabic falls outside the scope of this paper.

3.4 Concordance and category selection

To check every single entry of TR in context, two representative samples were drawn from the CA and MSA texts for the sake of concordance. Each sample was composed of 50 million words. All TR candidates were searched in the concordance corpus and the purposes of reduplications were determined based on the results returned by the search. Most of the false TR candidates were dropped in this stage as the concordance records showed the context in which the entries occurred. The procedure of selecting and defining the TR categories is shown in the list below:

1. TR candidates were stored in an SQL table.
2. A Python script imported the TR candidates and a loop was executed to perform the search for all the TR candidates in the representative corpora.
3. The output was saved in a text file.
4. For each TR candidate, the concordance records were manually examined.
5. For each concordance record, the purpose of the TR candidate in question was determined from the context. Mostly, only one purpose was identified despite there existed hundreds of records for the TR candidate in question. For some TR candidates, several purposes may exist. However, we selected only the most frequent one.
6. The table of TR candidates was manually updated after examining each TR candidate.
7. Finally, the functions of TR were grouped and ordered using a query in the SQL management studio. The graphs were plotted from the statistical summaries generated by the query commands.

Table 2 shows an example of the SQL table of TR entries:

EntryID	HeadWord	Transcription	Singleton	POS	FreeCount	Gloss	Purpose	Type
1	ألف ألف	ʔalf ʔalf	ʔalf 'thousand'	N	3365	million	Word Formation	Morphologi-cal

Table 2: Example of records in TR SQL database

3.5 Transcription and POS Tagging

The final TR candidates were manually transcribed using the International Phonetic Alphabet (IPA) symbols. Part of Speech Tags were assigned manually, too.

4 Data Analysis

This section presents the data analysis. Section 4.1 lists the functions of TR in CA. Section 4.2 lists the functions of TR in MSA. Section 4.3 compares the status of TR in CA and MSA. Finally, Section 4.4 plots the frequency distribution of the nine grammatical categories in which TR operates.

4.1 Functions of TR in Classical Arabic

For CA, 374 entries were analyzed, 11 functions were attested in the data, and they are listed in Table 3 below. For each entry in Table 4, we provide illustrative examples in dataset (23) with the same serial number as in Table 4.

S. No.	Functions	Count of Unique Entries	Percentage %	Free Count of Bigrams	Type
1.	Serial Ordering	109	31.41	7907	Semantic/ Pragmatic
2.	Word Formation	100	28.82	7554	Morphological
3.	Distributivity	66	19.02	3852	Semantic/ Pragmatic
4.	Emphasis	28	8.07	3522	Semantic/ Pragmatic
5.	Intensification	18	5.19	2928	Semantic/ Pragmatic
6.	Supplication	6	1.73	2835	Semantic/ Pragmatic
7.	Warning	5	1.44	138	Semantic/ Pragmatic
8.	Oath	5	1.44	362	Semantic/ Pragmatic
9.	Graduality	5	1.44	1439	Semantic/ Pragmatic
10	Subordination	3	0.86	613	Semantic/ Pragmatic
11	Fraction of Fraction	2	0.58	45	Semantic/ Pragmatic
	Total	**374**	**100%**	**31195**	

Table 3: Functions of TR in Classical Arabic

As seen in the above table, **serial ordering** is the top purpose of TR in CA, scoring 31.41%. Interestingly enough, **word formation** takes in the second position with 28.82%. **Distributivity** occupies the third place scoring 19.02%. **Emphasis**, which is the literal translation of the Arabic *tawkeed* ('emphasis' – the classical notion of reduplication), finds only a fifth place at 8.07%. This confirms that TR in Arabic, as in many other languages, serves many functions other than emphasis. Below are examples of the functions of TR in CA.

Dataset (23): Examples of TR functions in CA			
Single Entry and Gloss	**Reduplication Entry**	**POS Tag**	**TR Gloss**
1. Serial Ordering			
θala:θ-an three-ACC	θala:θan θala:θan	Num Num	three after three
2. Word Formation			
bajna 'between'	bajna bajn	Adv Adv	in between
3. Distributivity			

qitˁʕ-at-an piece-FEM-ACC 'a piece'	qitˁʕatan qitˁʕatan	N N	one piece for each one
4. Emphasis			
bara:ki 'sit down'	bara:ki bara:ki	V V	sit down + EMPHASIS
5. Intensification			
ʔabad-an eternity-ACC	ʔabadan ʔabadan	Adv Adv	never ever
6. Supplication			
sallim 'save'	sallim sallim	V V	Oh God. Save us!
7. Warning			
ħadaʔ 'a tribe in Yemen'	ħadaʔ ħadaʔ	N N	be careful and cautious
8. Oath			
ʔad-damm-u the-blood-NOM 'the blood'	ʔaddammu ddammu	N N	a pledge to protect someone
9. Graduality			
ʃajʔ-an thing-ACC 'a thing'	ʃajʔan ʃajʔan	N N	slowly, gradually
10. Subordination			
rasu:l-u messenger-NOM 'a messenger'	rasu:lu rasu:lu	N N	messenger of the messenger
11. Fraction of Fraction			
luqm-at-u bite-FEM-NOM	luqmatu luqmatu	N N	a portion of a spoon of food

4.2 Functions of TR in Modern Standard Arabic

For MSA data, 132 entries were analyzed, five functions were revealed and they are listed in table 4 below. For each entry in Table 4, we provide an illustrative example in dataset (24) with the same serial number as in Table 4.

S. No.	Purpose	Count of Unique Entries	Percentage%	Free Count of Tokens	Type
1.	Intensification	70	53.03	5747	Semantic/ Pragmatic
2.	Word Formation	29	21.97	4058	Morphological
3.	Serial Ordering	28	21.21	849	Semantic/ Pragmatic
4.	Graduality	4	3.03	60	Semantic/ Pragmatic
5.	Subordination	1	0.76	36	Semantic/ Pragmatic
	Total	**132**	**100%**	**10750**	

Table 4: Functions of TR in Modern Standard Arabic

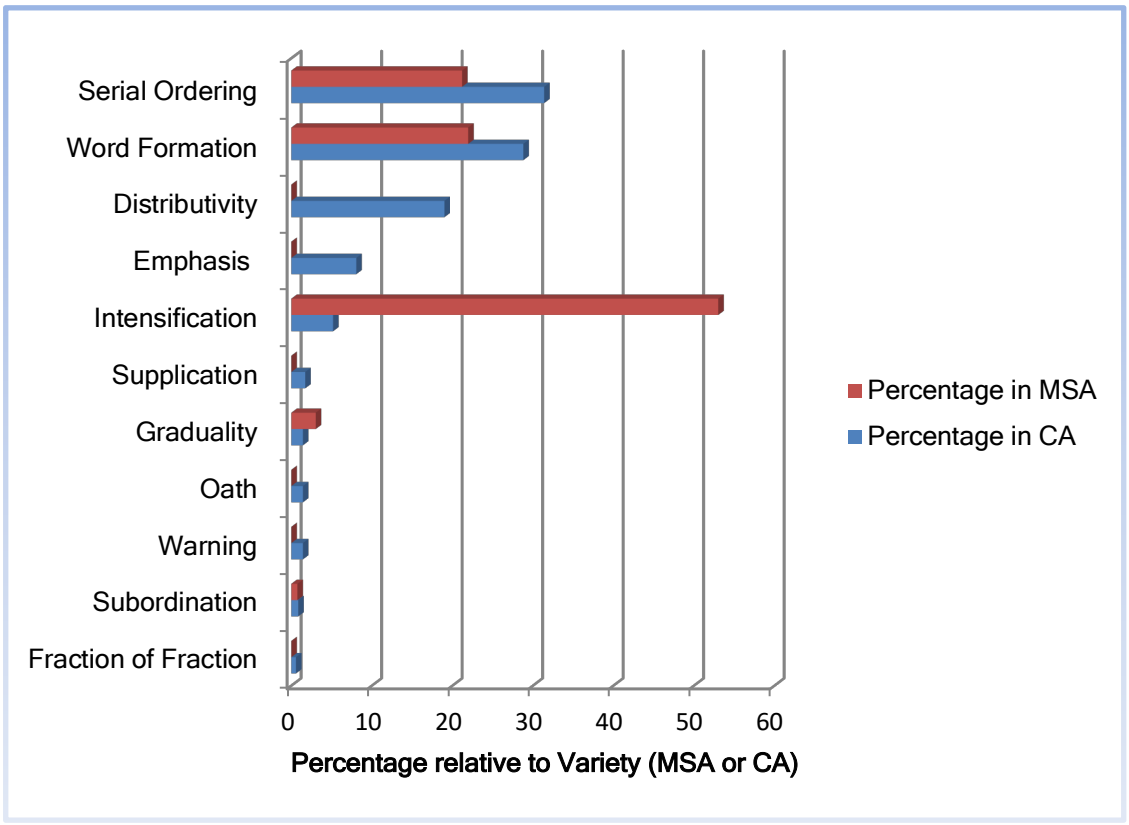
As seen in Table 4 above, **intensification** achieved 53.03% which is more than half of the overall percentage. MSA speakers tend to use TR to indicate that the feature of a given constituent has extraordinarily exceeded the normal quantity or quality. **Word formation** takes the second place with 21.79%, **serial ordering** the third with 21.21% which is a close percentage to that of word formation. **Graduality** and **subordination** display lower percentages indicating that they are rarely expressed by TR. Examples of TR in MSA are listed in dataset 24 below.

Dataset (24): Examples of TR in MSA

Singleton and Gloss	Reduplication Entry	POS Tag	TR Gloss
1. Intensification			
ʒidd-an serious-ACC 'very'	ʒiddan ʒiddan	Adv Adv	very very
2. Word Formation			
ʒaw 'air'	ʒaw ʒaw	N N	air to air
3. Serial Ordering			
fard-an individual-ACC	fardan fardan	N N	one by one
4. Graduality			
qali:l-an little-ACC 'little'	qali:lan qali:lan	Adj Adj	gradually
5. Subordination			
walijj 'master'	walijj walijj	N N	deputy of deputy

4.3 Status of TR in CA and MSA

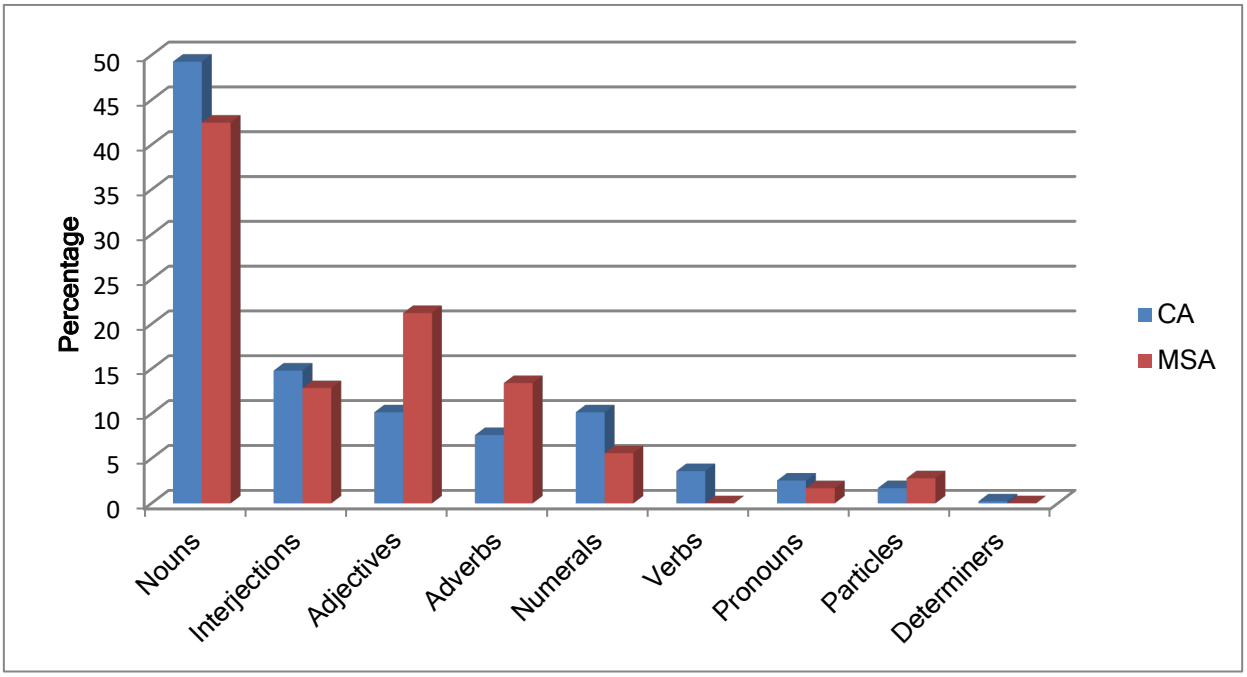
As seen in Graph 1 below, CA outscores MSA in employing TR to express **serial ordering**, **word formation** and **distributivity**. On the other hand, MSA outscores CA in employing TR to express **intensification** and **graduality**. Distributivity, emphasis, supplication, warning, oath, and fraction of fraction are no longer expressed by TR in MSA.



Graph 1: Status of TR in CA and MSA

4.4 Grammatical categories (POS tags)

Nine grammatical categories (parts of speech tags) were attested in both CA and MSA data and their distribution is displayed in Graph 2 below (left-hand is CA and right-hand is MSA). It is shown that **Nouns** are most favored by TR in both CA and MSA. **Interjections** take second place in MSA but fourth in CA. **Adjectives** occupy third place in CA but second in MSA. **Adverbs** achieve fourth place in CA but third in MSA. **Numerals** maintain fifth position in both CA and MSA. **Verbs** appear in sixth position in CA but completely disappear from MSA data. **Pronouns** maintain seventh position in both CA and MSA. **Particles** appear in eighth position in CA but have been promoted into the sixth position by MSA. Finally, **determiners** are seen in the last position in CA, but they completely disappear from MSA data.



Graph 2: PoS Distribution across CA and MSA

5 Results and discussion

This study has arrived at the following results and conclusions:

First: a difference between CA and MSA is noticed in terms of the top purpose of TR. While CA favors TR to express **serial ordering** (31.41%), MSA has been observed to favor TR to express **intensification** (53.03%).

Second: using TR as a means of **word formation** is very common in CA (100 out of 374 entries, 28.82%), whereas it is slightly less common in MSA (29 out of 132 entries, 21.97%).

Third: despite being common in CA (66 out of 374 entries, 19.02.26%), expressing **distributivity** by means of TR has completely disappeared in MSA data. Other functions of TR which are no longer attested in MSA texts are **emphasis**, **supplication**, **warning**, **oath** and **fraction of fraction**. This is suggestive of the fact that MSA might have developed other means to convey the same concepts expressed by TR in CA. The following dataset demonstrates how native speakers of MSA express some of the concepts as opposed to CA.

Dataset (25): Differences between CA and MSA in expressing some TR functions

TR in Classical Arabic	Equivalent non-TR expression in MSA	Function
1. kati:b-at-an kati:b-at-an battalion-SG+FEM-ACC 'battalion opposing the other'	kati:b-at-an muqa:bil-a kati:b-at-an in front of-ACC 'battalion in front of battalion'	Distributivity
2. ʔal-harab-u lharabu the-escape-NOM 'escape.IMPERATIVE'	ʔuhrubu: escape.IMPERATIVE.2.PL.MASC	Warning

3. sˁabr-an sˁabr-an patience.SG.MASC-ACC 'be patient'	jaʒib-u ʔan na-sˁbir must-NOM that 1+PL-patient 'we should be patient'	Emphasis

Fourth: total reduplication operates in all Arabic part of speech categories except **prepositions**. In CA data, the ranking of TR categories is as follows: nouns > interjections > adjectives > numerals > adverbs > verbs > pronouns > particles > determiners. Nouns achieve 49.26%, while the remaining categories display 50.74%. In MSA data, TR of verbs and determiners has not been attested in the data although it is permitted by MSA grammar. The ranking of TR categories in MSA data is as follows: nouns > adjectives > adverbs > interjections > numerals > particles > pronouns. Again, nouns form the top category in MSA data as they score 42.46%, while the remaining categories score 57.54%. That interjections get demoted by MSA can be explained as a shift in preference of MSA TR from closed class categories to open class categories.

With respect to the research questions raised at the beginning of this paper, the first question was about the status of TR in CA and MSA. As it has been shown, TR is not as quite productive in Arabic as in other languages such as Indonesian. In terms of quantification, data analysis showed that TR is employed by CA more than MSA. Variation has been attested in terms of the functions TR serves in both CA and MSA.

The second question was about the form of TR in Arabic. As shown in this paper, TR takes the form of *[X, X]*, and *[X]* may stand for any grammatical category **except prepositions**. The fact that prepositions are the only grammatical category on which TR cannot operate is really intriguing. However, examining the semantics of prepositions may help to offer an explanation. That is, Arabic prepositions do not carry full semantic thought as other grammatical categories do. For instance, TR can operate on verbs because they convey complete semantic unit of thought due to the drop feature. In Example (11) above, *naʒaħ-a* 'passed' carries a pro drop feature which means 'he' and the interpretation is 'he, the diligent, passed'. Another reason why TR cannot operate on Arabic prepositions may be attributed to their syntactic nature. That is, an Arabic preposition must take an argument and it cannot take itself as an argument because it cannot assign genitive case to itself. The following example offers further illustration:

(26)	a. ʔajna	waɗaʕ-t-u	l-kita:b-a
	where	you.put.NOM	the-book-ACC
	'Where did you put the book?'		
	b. *ʕala:	ʕala: t-ta:wil-at-i	
	on	on the-table-FEM-GEN	
	c.	t-ta:wil-at-i ʕala: t-ta:wil-at-i	
		'On the table. On the table.'	

It is clear that the sentence in 26.b is not accepted in Arabic because the preposition *ʕala:* 'on' cannot assign genitive case to itself and the correct sentence is 26.c where the preposition *ʕala:* 'on' is seen taking *t-ta:wil-at-i* 'the table' as an argument to which the genitive case was assigned.

The third question was about the morphological functions of TR in CA and MSA. As shown by data analysis, **word formation** including **onomatopoeia** serves morphological functions in both CA and MSA.

The fourth and last question was about the semantic or pragmatic functions of TR in CA and MSA. As attested by data analysis, **serial ordering**, **distributivity**, **emphasis**, **intensification**, **supplication**, **warning**, **oath** and **graduality** serve semantic or pragmatic functions in CA. In MSA, semantic and pragmatic functions of TR are served by **intensification**, **serial ordering**, **graduality** and **subordination**.

References

Al-Huri, Ibrahim (2015): "Arabic Language: Historic and Sociolinguistic Characteristics". *English Language and Literature Review:* 1/4: 28–36.

Bateson, Mary (1967): *Arabic Language Handbook*. Vol (3). Georgetown University Press.

Broselow, Ellen/ McCarthy, John (1983): "A Theory of Internal Reduplication". *The Linguistic Review* 3/1: 25–88.

Bussmann, Hadumod (2006): *Routledge Dictionary of Language and Linguistics*. London: Routledge.

Chakraborty, Tanmoy/Bandyopadhyay, Sivaji (2010): "Identification of reduplication in Bengali corpus and their semantic analysis: A rule-based approach". In: *Proceedings of 23rd International Conference on Computational Linguistics*: Beijing, Coling 2010 Organizing Committee: 73–76.

Harley, Heidi /Leyva, Maria (2009): "Form and Meaning in Hiaki (Yaqui) Verbal Reduplication". *International Journal of American Linguistics* 75/2: 233–272.

Ibn Aqeel (died 1367, 1980): ʃarħ ʔibn ʕaqi:l (Annotation of ʔibn ʕaqi:l, translation is mine). Vol. 3. Cairo: Daaru Turaath.

Ibn Malik (died 1273, 2012): ʔawđaħ ʔamasaalik ʔila ʔalfiat Ibn Malik. *Dar Alkutub Al'ilmiah. (Almaktaba Ashaamila*. https://archive.org/details/awda7-almassalik [02-03-2019].

Inkelas, Sharon (2014): *The Interplay of Morphology and Phonology*. Oxford: Oxford University Press.

Kallergi, Haritini (2015): *Reduplication at the Word Level: The Greek Facts in Typological Perspective*. Berlin/Boston: de Gruyter.

Macdonald, Roderick/Darjowidjojo, Soenjono (1967): *A Student Reference Grammar of Modern Formal Indonesian*. Washington: Georgetown University Press.

Rubino, Carl (2005): "Reduplication: Form, Function and Distribution". In: Hurch, Bernhard (ed.): *Studies on Reduplication*: Berlin/New York: de Gruyter: 11–29.

Suçin, Mehmet (2010): "Turkish and Arabic Reduplications in Contrast". *Australian Journal of Linguistics:* 30/2, 209–226.

Versteegh, Kees (2014): *The Arabic Language*. Edinburgh: Edinburgh University Press.

Wang, Shih-ping (2005): "Corpus-based Approaches and Discourse Analysis in Relation to Reduplication and Repetition". *Journal of Pragmatics* 37/4: 505– 540.

Watson, Janet (2002): *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.