The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2634-243X.htm

Received 21 March 2022
Revised 4 May 2022
Accepted 5 May 2022

The distribution of *nafs* in modern Standard Arabic and Classical Arabic: a corpus-based study

Amani Mejri
University of Debrecen, Doctoral School of Linguistics, Institute of English and American Studies, Debrecen, Hungary

Abstract

Purpose – This corpus-based study provides a descriptive account of the distribution of the polysemous noun *nafs* in two Arabic varieties, Modern Standard Arabic (MSA) and Classical Arabic (CA). The research objective is to survey the use of *nafs* as a reflexive marker in local binding domains and as a self-intensifier in NP-adjoined positions.

Design/methodology/approach – The consulted corpora are Timespamped JSI Web corpus for MSA and Quran corpus for CA. While attending to corpora size differences, MSA and CA exhibit a pattern of difference and similarity in *nafs* diffusion.

Findings – In the modern variety, *nafs* is pervasively used as reflexive marker in canonical binding domains, along with a less frequent, yet notable, intensifier user, and these uses are partially and cautiously attributed to the specific genre in which they occur. In CA, *nafs* is mainly recurrent as a polysemous noun, along with extensive use as a reflexive marker in local binding settings. As an intensifier, *nafs* is totally non-existent in the CA corpus, in the same way as it is in absentia in VP-constituent extraction in MSA.

Originality/value – Examining whether *nafs*, as a reflexive marker, deviates from canonical binding in Arabic the way English reflexive pronouns do. Building a general account of this distribution is relevant in understanding the explicit (syntactic) and implicit (discourse-based) dimensions of reflexive marker and self-intensifier processing and interpretation in Arabic as a first and second language.

Keywords Local binding, Reflexive marker, Intensifier, Classical Arabic, Modern standard Arabic

Paper type Research paper

1. Introduction

In the standard binding theory, non-reflexive and reflexive pronouns are interpreted based on the syntactic environments in which they occur. Reflexives are assigned their interpretive value by being bound by an antecedent in their clause. Pronouns, on the other hand, do not require a clause-mate antecedent to be interpreted.

In $Mary_i$ *hurt herself*$_{i/*j}$, the reflexive pronoun *herself* is referentially dependent on and syntactically bound by the antecedent *Mary*. The binding configuration here can be explained by the structural relation of c-command (Chomsky, 1981). It pertains to the syntactic structure of the sentence where the antecedent, being higher in position, c-commands the anaphors, being lower in position. Its definition states this hierarchical relation between the binder and the bindee as follows:

(1) $\forall$ *x c-commands y* $\Leftrightarrow$ *x does not contain y and the first branching node dominating x also dominates y*

(2) $\forall$ *x binds y* $\Leftrightarrow$ *x c-commands and is coindexed with y.*

© Amani Mejri. Published in *Saudi Journal of Language Studies*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

The author would like to thank the anonymous reviewers for their critical comments on an earlier version of this manuscript.

Saudi Journal of Language Studies
Vol. 2 No. 2, 2022
pp. 84-106
Emerald Publishing Limited
e-ISSN: 2634-2448
p-ISSN: 2634-243X
DOI 10.1108/SJLS-03-2022-0038

C-command requires that the NP *Mary* is dominated by the inflectional phrase (IP), which is the branching node that dominates the reflexive *herself* as well. Both NPs are dominated by IP in the positions of [Spec, IP] and VP internal argument, where the subject is a c(onstituent)-commanding and satisfying the antecedence of the bound element (Huang, 1987). Having the first requirement of c-command satisfied, both NPs are also coindexed in terms of Φ-features, hence the NP, *Mary*, binds the reflexive, *herself*, at the clause level. Coindexation here means that both NPs share the same referential value: the reflexive and its binder share the same reference of person, gender and number, which satisfies the classical binding condition of indexing (Chomsky, 1981).

Along with the structural relation of c-command, it is important to note that locality is another integral element defining the standard binding of anaphors (Felser and Cunnings, 2012). These binding conditions are in fact established through the binding Principles A and B.

(1) *Principle A: An anaphor is bound in a local domain*

(2) *Principle B: A pronoun is free in a local domain* (Chomsky, 1981).

According to these principles, anaphors need to be clause-bound, meaning that they are in the clause that contains their binder, given their referential deficit. The locality constraint of anaphors is a binding condition. It is measured by the syntactic distance between the antecedent and the antecedee. The notion of governing category determines whether these NPs are close enough to satisfy this binding condition (Charnavel, 2020; Haegaman, 1994).

This theoretical background of local anaphoric binding has long been the basis of various studies that examined the distribution of reflexive and non-reflexive pronouns in English (Huang, 1987; Pollar and Sag, 1992). The same canonical picture of binding has also been cross-linguistically investigated in languages such as French (Charnavel, 2020) and Chinese (Huang, 2000), among other languages. However, studies in Arabic that have addressed the distribution of reflexive and non-reflexive pronouns are still in their developmental stage (Alenazy, 2021; Hammami *et al.*, 2009), which makes Arabic under-researched languages in this regard. This negligence has also led the Arabic language to be under-researched in first and second language acquisition, particularly because language acquisition studies have been nurtured by linguistic theory whether in English (Chien and Wexler, 1990; among a wide range of studies) or for example Italian (McKee, 1992), and whether they are online or offline (Patterson *et al.*, 2014). However scarce as they might be, some studies have addressed the acquisition of binding in Arabic; they focused on particular varieties of the Arabic language such as Qatari (Mustafawi and Mahfoudhi, 2002) or Saudi Arabic (Bolotin, 2002).

Against this background, this corpus-based study investigates and provides a descriptive account of the distribution of reflexive pronouns in Modern Standard Arabic (MSA) and Classical Arabic (CA). The aim is to provide an account of this distribution which will potentially help in expanding the comparison among the different Arabic varieties. Such an endeavor also aims at setting some theoretical bases which can be used to inform language acquisition studies and pedagogical policies that pertain to acquiring, learning and teaching the Arabic language and its different varieties. Another aim of this paper is to examine whether some language change, either semantic or syntactic, occurred at the level of pronoun distribution in the transition from CA to MSA. While these research aims are pursued within the scope of the Arabic language, some constructions will be compared to English in order to explain the deviance from or convergence with the canonical. When a comparison is made to an English construction, or we do not intend to make a sort of comparison with English, given that it might not be translatable. Rather the goal of this research is to present a descriptive account of reflexive pronoun distribution in MSA and CA.

2. Theoretical background

The standard binding theory introduced so far determines the antecedence relation between an NP that serves as a binder and an NP that serves as a bindee in argument positions.

This binding relation is defined in a minimal governing category with an output of coindexation assignment that is relevant for NP interpretation (Lasnik, 1989). The minimal governing category stands for the functional unit that contains the verb (being the governor) and its arguments (being the external argument [the subject] and the internal argument [the object]). Both arguments share the same referential value technically expressed by coindexation. This binding requirement defines coargument binding, where the binder and the bindee are both in argument positions vis-à-vis the governor. In this way, potential binders and bindees need to comply with the requirements of c-command and locality.

Given the syntactic structure of strict binding or emphatic use of reflexives, the following structures represent the constructions to be examined coupled with examples from English that might not be exactly found in Arabic, but the research interest is to survey the general characteristics of the local and non-local, if any, environments accommodating reflexives in Arabic:

(1) *Local binding positions*

a. Omar_i hurt himself_i

b. Hoda believes Omar_j's description of himself_j

c. Himself_j, Bill_j saw in the mirror

(2) *Reflexives as intensifiers (NP adjoined position)*

a. The president_i himself_i opened the meeting

The canonical English reflexive occurrences are indicative of the endophoric nature of reflexives, which is expected to be part of the Arabic language in similar reflexive events that lead to strict binding (a, b and c). Examples (a) and (b) represent the typical domains of binding presented in the binding theory. While in (a) the IP is the local domain where the reflexive *himself* is bound by its c-commanding antecedent, in (b) the complex NP is the binding domain that hosts the antecedent *Omar*, being the binder and the governor, and the bound reflexive *himself*. In this particular example, the binding domain shifts from the tensed clause to the non-tensed clause, while still satisfying the binding requirements of locality, c-command, coindexed interpretation of the relevant NPs. Sentence (c) in its turn is another representation of the typical binding across a tense clause, with a particular position held by the reflexive after being fronted. In fact, the bound NP here is base-generated as an internal argument of the verb *saw*, but it undergoes focus-motivated topicalization that places it in a position preceding the binder.

Another occurrence of reflexives is also a subject of interest in this study: reflexives as intensifiers (2a). They are used for emphatic purposes and in different syntactic positions and are called self-intensifiers (Gast and Siemund, 2006). The reflexive pronoun in similar examples is attached to the NP it modifies, and it assumes the intensifying function that contrasts the NP, *the president*, with the alternative referents involved in the specific event mentioned in the sentence. NP-juxtaposed reflexive pronouns assume, in this way, a pragmatic function that pertains to the discourse in which the sentence occurs, such as sentence (2a) (*the president*, and not his assistant, *opened the meeting*). Although intensifying reflexives can also occur in non-juxtaposed positions and can have different roles (for example agentive), this study focuses on the NP-juxtaposed intensifying reflexive that bears, along with pragmatic input, a syntactic connection to its binder: both share the same identity index (person, gender and number). This use is expected to be part of the Arabic language, which will be presented in the results section.

As the language examined in this research is Arabic in two varieties_ MSA and CA (Quranic) _ it is necessary to provide a brief overview of the development of these varieties

followed by an introduction to the pronominal system in the language. To start with, Arabic is a Semitic language that witnessed various historical evolutions from the pre-Islamic to the post-Islamic era (Owens, 2013) In the Islamic period, the Quran is the holy book for Muslim scriptures that was collected in the 7th century, and CA was used until the 9th century (Kaik *et al.*, 2018). Through the period of Islamization, Arabic was being spread to different geographical areas. However, from that period to the current modern one, the Arabic language has undergone development from CA to MSA. In each region where Arabic spread the MSA form has interfered with the lexis, phraseology and syntax of the regional a dialect. MSA remains the official language of the Arab countries, while they preserve their own varieties of Arabic (Holes, 2004). From another perspective, Arabic in these two varieties, the MSA and the CA, is divided into the Old and New Arabic, while attending to the rise and uniqueness of the diglossic situation in the Arab world (Owens, 2013).

With this development in the history of the language, the Arabic pronominal system is the same in the two varieties that are examined here. It includes three types of explicit pronouns. They are detached (disjoint), attached (joint as affixes) and reflexive pronouns (Igaad and Tarrad, 2019; Hammami *et al.*, 2009). Detached pronouns occur independently as separate morphemes in nominative and accusative case positions. Moreover, the Arabic pronominal system distinguishes between the dual and plural features for second and third-person pronouns. Unlike English, it equally distinguishes between masculine and feminine features for plural second person and plural third person pronouns (Mejri, 2020; Mustafawi and Mahfoudhi, 2002). The nominative Arabic pronouns are *'ana*, *nahnu*, *'anta*, *'anti*, *'antum*, *'antunna*, *huwa*, *hiya*, *huma*, *hum* and *hunna*.

In terms of agreement, first-person pronouns (*'ana/I* and *nahnu/we*) do not have gender features, unlike second person pronouns (*'anta* and *'anti*/you singular masculine and feminine, respectively; *'antum* and *'antunna*/you plural masculine and feminine, respectively; and the dual, *'antuma*/you with no gender distinction), and third-person pronouns (*huwa* and *hiya*/he and she singular masculine and feminine, respectively; *hum* and *hunna*/they plural masculine and feminine, respectively; and the dual *huma*, with no gender distinction), which do have this distinction (Igaad and Tarrad, 2019).

The joint personal pronouns occur in nominative, dative and accusative case positions. These nominative suffixes are *'alef* (one letter 'a), *waw* (one letter w) and *nūn* (one letter n). The dative and accusative joint pronouns are *-ī*, *-nā*, *-ka/-ki*, *-kumā*, *-kum*, *-hu*, *-hā*, *-humā*, *-hum* and *-hunna*. In fact, the nominative attached personal pronouns are considered clitics and attached to the verb as affixes in SVO sentence formats (Hammami *et al.*, 2009).

– 'Al-awlad kharaj-u
The children go.PAST-3PM
'The children went out' *SVO*

Dative and accusative joint pronouns are attached to verbs, nouns and prepositions, such as:

– Tarz-hu
See-PRSNT-2SM him
'You see him'

– Kitabu-ha
Book.3SF
'Her book'

– La-ha
For.3SF
'For her'

As far as reflexive pronouns are concerned, in MSA they are complex morphemes composed of the grammatically invariant morpheme *nafs* to which a dative joint pronoun is attached. Prior to explaining the structure of reflexive pronouns in Arabic, it is necessary to start with defining the polysemous word *nafs* and to accentuate the uses that are relevant to this study. In the online Arabic dictionary *Al-Maany*, the noun *nafs* has several meanings among which are:

(1) Soul in English, with the plural form "anfus"

(2) Blood

(3) The same thing/person/entity

(4) Being envied by someone

(5) Being remarkably ethical

(6) An individual

From this multitude of meanings of the noun *nafs*, the meaning of self or soul in English is the primary sense which is directly relevant to the research. The third meaning, which stands for the intensifier use in English, is the another definition that is also directly relevant to this study. The rest of the definitions are excluded as they do not contribute to the understanding of how reflexive pronouns are distributed and interpreted in the Arabic language.

Given that in Arabic there are no reflexive pronouns in the entire pronominal system (Igaad and Tarrad, 2019), the noun *nafs/soul*, when attached to dative joint pronouns, has a reflexive meaning and use, hence is used as a reflexive marker (presented in Appendix 1). To illustrate this use, the following example represents a simple reflexive event where both the external argument of the verb (subject) and the internal argument (object) are coreferential:

1. Ahmad jara-ha nafsa-hu
Ahmad hurt.3SM himself
'Ahmad hurt himself'

Nafsa-hu in this sentence is the reflexive marker that indicates that the Agent *Ahmad* and the *Patient* (himself) are semantically equivalent, yet in different argument positions. NP1, *Ahmad*, is the binder of NP2, *nafsa-hu*. As Arabic is a language that allows different word order, the categorization of NP1 and NP2 might be subject to the word order in which they occur, which is SVO, but it can potentially be OVS among other possible orders (AlBzour, 2015).

Reflexive markers in Arabic are also used with state verbs (such as *hasiba/think* or *believe*; *khala/believe*; *zaaima/suppose*; *wajada/find*; etc.). Some examples to illustrate this use include:

2. Ramy wajada nafsa-hu fi wafhi'in saab
Ramy find.PAST.3SM himself in (a) situation difficult
'Ramy found himself in a difficult situation'

3. Al-mumathila khalat nafsa-ha jamila
The-actress think.PAST.3SF herself beautiful
'The actress thought herself beautiful'

In these examples, the reflexive marker is always bound by a local antecedent in the same domain. Although reflexive binding is linear, with the reflexive following the binder, it is possible to have different positions based on word order.

The reflexive marker in Arabic can also be used in intensification contexts, which means that a particular NP is being emphasized against a set of entities in the same setting, as defined by Gast and Seimund (2006). In the same way English reflexive pronouns are used in reflexive events and intensification contexts, the reflexive marker in Arabic, *nafs* + *joint pronoun*, assumes both functions.

4. Al-rai'is nafsa-hu iftata-ha al-ijtima'a
The-president himself open.PAST.3SM the-meeting
'The president himself opened the meeting'

The same English example (2a) in section 2 is translated into Arabic. The joint point between both examples is that the reflexive pronoun is used as a self-intensifier in an NP-juxtaposed position that distinguishes the agent from a universe of entities in the same context. This pragmatic usage is valid in Arabic, and the same syntactic conduct of the English self-intensifier is also detected with the Arabic reflexive marker: joined to the NP is accentuates, hence locally bound by it.

3. Literature review

Examining pronominal distribution based on Chomskyan binding theory in general and reflexive pronoun distribution in particular has been an established endeavor in linguistic studies (Baker, 1995; Charnavel and Sportiche, 2016; Huang, 2000). However, MSA and CA are under-researched in this regard. In particular, the distribution of the noun *nafs* and its relevance in informing how reflexive pronoun binding is achieved in these two Arabic varieties has received limited attention (Al-Raba'a, 2021; Kremers, 1997; Plunkett, 1996). This review presents the studies that have addressed some aspect of the binding theory in the Arabic language.

In a recent study, Alenazy (2021) examined the distribution of possessive pronouns in MSA and linked it to the distribution of reflexive pronouns. His claim is based on the assumption that possessive pronouns in Arabic are bound in the same way as pronouns, and their binding affects word order in the sentence. Alenazy states that the position of the possessive pronoun can either allow for or block its movement. To elaborate, if the subject is coreferential with the same anaphoric element in the object position, then optional movement is allowed and the sentence can have the VOS order as shown in example (5), along with the dominant word order VSO.

5. Dahana bayt-a-hu at-tajir-u
paint.PAST.3SM house.acc.his the-merchangt.nom
'The merchant painted his house' (Alenazy, 2021, p. 1019)

Within the same line of thought, Alenazy postulates that, in the opposite case, where the anaphoric element is part of the subject and corefers to the object, movement becomes a proviso: the object should be moved to a position preceding the subject in a VOS word order akin to (6):

6. Dahana al-bayt-a sahibu-hu
paint.PAST.3SM house.acc. owner.nom-its
'The house's owner painted it' (Alenazy, 2021, p. 1020)

To account for this movement and its implication word order, Alenazy indicates that the structural notion of c-command is a central element in determining the position of possessive pronouns in Arabic, which also regulates the distribution of reflexive markers.

In a similar context that treats the distribution of possessive and reflexive pronouns based on structural considerations in binding theory, Plunkett (1996) shows that the binding domain is central in defining the position of the bound element, whether it is a possessive or a reflexive pronoun. In Arabic, the VP is introduced as the binding domain where these pronominal elements are bound. Her argument adopts the view that binding in Arabic is achieved in the VP domain, unlike English which might have a wider binding domain than the VP. To prove this point, Plunkett uses possessive and reflexive pronouns to highlight the strict locality condition that marks the Arabic language:

7. yuhibu-hu
love.PRSNT.acc-him
'(He) loves him'

8. Zaydun yhi-bu nafsa-hu
Zaydun love.PRSNT.3SM himself

According to Plunkett, pronoun binding in (7) and (8) is achieved within the VP domain, and the broader binding domains are not part of the Arabic language. The data used to corroborate this claim is based on pronoun binding, including joint (object, possessive) and disjoint pronouns (reflexive markers like *nafsa-hu*).

Turning to the specific distribution of reflexive pronouns in Arabic, little is known about the canonical and non-canonical diffusion of *nafs* (as a reflexive marker and intensifier) in both CA and MSA. Alternatively, some studies address the environments in which *nafs*, as a reflexive marker, occurs in MSA and one of its regional varieties, namely Jordanian Arabic. AlBzour (2015) presents the binding stipulation of *nafs* in MSA and Jordanian Arabic by defining its domain and by revisiting previous claims that advocate a complete similarity between English and Arabic pronominal distribution in terms of Principle A. AlBzour states that in MSA, *nafs* serves as a reflexive marker with the strictest binding account (the antecedent and the bindee being syntactically very close) and self-intensifier, which stands for non-verbal emphasis in Arabic (*tawkeed ma'anawi*). Within this descriptive account, he also supplies a preliminary view of the behavior of *nafs*. He claims that moving it to a front position in a sentence, which is part of MSA and not Jordanian Arabic, depends mainly on pragmatic reasons rather than syntactic reasons only. Notwithstanding the original research purpose of the study, the data used to examine this distribution and to extract linguistic assumptions about Arabic is not documented in a corpus and does not present the patterns in which *nafs* occurs in the examined varieties.

Recognizing that an Arabic language corpus is central to revise the distribution of *nafs*, Kremers (1997) addresses this issue in a corpus that combines untagged journalism and literary Arabic texts of more than 3 million words. Kremers does not focus on quantifying the use of *nafs* in the corpus, based on the fact that the aim is not to compare 2 or more distributional patterns. Rather the main focus is to delineate the environments in which this polysemous noun occurs. The environments which accommodate *nafs* show that it can be used as a reflexive marker, particularly when it is used with a finite verb in an object position. This particular use indicates the binding requirement of its distribution, which is in line with the claims made about the *nafs* being a locally bound reflexive marker. In this study, Kremer (1997) postulates that throughout the linguistic transition from pre-classical to CA, this

reflexive marker moved from an optionally to an obligatorily bound position, yet he does not provide any accounts of *nafs* in CA.

In consideration of the fact that little in known on the diffusion of *nafs* in the classical as compared to the modern variety, this study seeks to answer the following research questions:

RQ1. What are the distributional patterns of reflexive markers *nafs* in MSA and CA?

RQ2. What are the distributional similarities and differences in MSA and CA?

4. Methodology

This corpus-based study uses Timestamped JSI Web Corpus 2014–2021 for MSA and includes 5,395,395,993 tokens, 4,849,103,095 words and 142,215,846 sorted sentences from journalism sources. The Timestamped JSI Web Corpus is a perpetually growing corpus based on the consistent updating and integration of web articles into the corpus (Trampus *et al.*, 2012). The second corpus is the Quran for CA. It includes 128,843 words and 6236 sorted sentences. It was built by Alqassem (2013) based on the Quranic Arabic Corpus (Dukes, 2009). The two corpora used for this study differ in their size. This is why they will be first examined separately and then compared.

To investigate the distribution of reflexive pronouns in the two corpora, the word *nafs* was searched in each corpus. In the first phase of the search, the word-sketch tool was used to summarize the lexico-syntactic behavior of the word (Kilgarriff *et al.*, 2014). This tool is used to examine the collocational behavior of the reflexive marker *nafs* (Appendixes 2 and 3). The occurrence and position of the nouns and verbs collocating with *nafs* indicate the lexical and syntactic uses of the word. Whenever *nafs* is followed by a definite noun, the lexical use is identified. Examples include:

9. Fi nafs al-siyaaq
   In same the-context
   'In the same context'

10. Min nafs al-masdar
    From same the-source
    'From the same source'

Extractions (9) and (10) represent the third definition in section 2, which is "the same thing, entity, or person". In the same noun-right position of the noun *nafs*, the word sketch tool displays the emphatic/intensifier use that is also recovered from the context, as in (11) and (12).

11. Al-hukuma nafsa-ha
    The-government itself
    'The government itself'

12. Al-laa'eb nafsa-hu
    The-player himself
    'The player himself'

Based on this corpus feature, two major uses of the noun *nafs* are identified, and will be further examined and discussed on in the Results section, based on the frequency values.

Using the same tool, *nafs* appears in verb-right positions with two possibilities suggested based on the inflectional form of *nafs* and the subject of the verb. If no pronoun is attached to *nafs,* it is used lexically to mean *soul,* specifically definitions 1 and 2 in section 2. For example:

13. Qatala nafsan
Kill.PAST.3SM (a) soul
'He killed a person'

If a joint personal pronoun is attached to *nafs* and is coindexed with the subject of the verb the reflexive/syntactic use is identified. Examples from the word sketch tool include:

14. Fajjara nafsa-hu
Explode.PAST.3SM himself
'He exploded himself'

15. Mathala nafsa-hu
Represent.PAST.3SM himself
'He represented himself'

(14) and (15) are instances of the noun *nafs* being a reflexive marker where a joint pronoun is attached to it, serving in this way the required reflexive use of the term. Verbs occurring with *nafs* as a reflexive pronoun will be further classified according to their semantic fields in both corpora.

Given that the two corpora used in this study differ in size, the option of random sampling is used in order to thoroughly examine the distribution and use of *nafs*. The random sample is comprised of 500 sentences retrieved from the MSA Corpus, being the larger text entity in this study. Upon being informed of the grammatical and collocational diffusion of the noun *nafs* using the work sketch tool, we manually annotated the sample to extract the frequent uses of the noun *nafs* based on the definitions charted in this study. Table 1 demonstrates the annotation scheme adopted in this research and classifies the use of *nafs* as follows:

In this scheme, the use of *nafs* is quantified and extracted based on its occurrence in compliance with every indicated domain. The findings obtained from this sample will be presented in the results section.

Constructions	Examples	Usage	Annotation
Local binding: tensed domain	Omar hurt himself	Reflexive marker	Local binding
Local binding: non-tensed domain	Hoda believes Omar$_j$'s description of himself$_j$	Reflexive marker	Local binding
Local binding: VP-constituent extraction	Himself$_j$, Bill$_j$ saw in the mirror	Reflexive marker	Local binding: topicalization
Intensifier	The president$_i$ himself$_i$ opened the meeting	Intensifier	Intensifier
Polysemous noun	*Qatala nafsan* (he killed a person/someone)	Content use	Content use

Table 1.
The annotation scheme

5. Results

5.1 The distribution of nafs in Modern Standard Arabic

As already indicated, using the word sketch tool has revealed a set of syntactic environments in which *nafs* is used to convey reflexivity (as a reflexive marker) or intensification (as a self-intensifier) along with its content use. Table 1 sketches the reflexive use of *nafs* with psychological verbs, akin to *wajada*/*find*, *i'itabara*/*consider* and *a'arafa*/*know*, and other action-oriented verbs, such as *shanaqa*/*hang*, *fajjara*/*explode* and *mathala*/*represent*. The normalized frequency values presented in Table 2 shows the most to the least frequent uses.

Sorting on frequency, psychological verbs seem to be more frequent with *nafs* a reflexive marker than the set of action-based verbs. This finding resonates with Igaad and Tarrad's (2019) proposition that the reflexive use of *nafs* is mostly detected with psychological verbs in MSA.

Nafs is also detected in noun-right positions, which trigger not the reflexive use but rather the intensifier use. Table 3 summarizes some of the noun-right occurrences of *nafs* with the normalized frequency values. In this position, *nafs* has a very limited frequency pattern when compared with verbs, hence the reflexive use.

Table 3 shows that *nafs* is used as an intensifier, which is different from the common use or meaning of *the same* (considered to be an error in Arabic). The emphatic or intensifier meaning here stands for the definition of intensification: the NP to which the reflexive marker is juxtaposed is being contrasted with a set of entities within the same context. To elaborate, *al-hukuma nafsa-ha* (= *the government itself*) does not stand for *the same government* (*nafs al-humuka*, where *nafs* is in a noun-left position and means *the same*) but rather emphasizes and contrasts the NP, *al-hukuma* (*government*), within its respective context. The intended meaning that is mainly sought and reported in this study, and in Table 3 in particular, is, for example, *the government itself* rather than *the same government*.

To proceed with the examination of a context-based distribution of *nafs*, the random sample is used to extract the frequency of these uses and to further study the environments accommodating them in MSA. Using the concordance tool in this sample, a detailed account is supplied on the grammatical behavior that is initially detected within the word sketch tool. Based on this data, the distribution of *nafs* is reported in strict binding contexts, being coargument binding and intensification examples, being NP-juxtaposed positions.

Psych verbs	Frequency	Action-oriented verbs	Frequency
Find (one self)	14.61	Impose	3.96
Consider	3.96	Introduce	3.08
Know	3.87	Explode	2.64
		Hang	1.97
		See	1.93
		Describe	1.38
		Represent	1.11

Table 2. Verb-right occurrences of *nafs*

Noun-right	Frequency
Government itself	4.38
Player himself	3.67
System itself	3.75
Person himself	3.16

Table 3. Noun-right occurrences of *nafs*

Prior to presenting the contexts of each usage, it is important to demonstrate the frequency of each occurrence of *nafs* in the random sample. Figure 1 sketches the regularity of each use.

Figure 1 indicates that the lexical use (which includes the definitions excluded from this study) is the most frequent in the Timestamped Arabic Web corpus. The other uses that concern this study seem to be of equal distribution, meaning that *nafs* as reflexive marker is as frequent as *nafs* being an intensifier. The regularity of the last two uses stems also from the nature of the studied journalism web corpus and implies that these two uses are recurrent in MSA. The following sections will be devoted to commenting on each use separately.

5.2 Local binding: nafs as a reflexive marker

Given the distribution of *nafs* in verb-right and noun-right positions, two different uses are detected and charted. In the verb-right positions, *nafs* is used as a reflexive marker and abides by the syntactic rules of binding, which are outlined in the first examined construction in this study (a). In similar environments, *nafs* is expected to be subject to strict binding, where the reflexive marker and its antecedent are locally coreferential, as shown in the subsequent extractions.

16. Yajidu nafsa-hu fi wadh'in sayi'
Find.PST.3SM himself in situation bad
'He found himself in a bad situation'

17. Faradhat nafsa-ha
Impose.PST.3SF herself
'She imposed herself'

The joint feature between (16) and (17) is that the antecedent is not overtly spelled out in Arabic, as is always the requirement in languages similar to English. Arabic, as a pro-drop language, allows similar constructions, where the antecedent, which is supposed to bind the reflexive marker, is recovered from the verb morphology. Although, these instances represent one of the strict binding occurrences in MSA, they also represent one feature of binding in pro-drop languages.

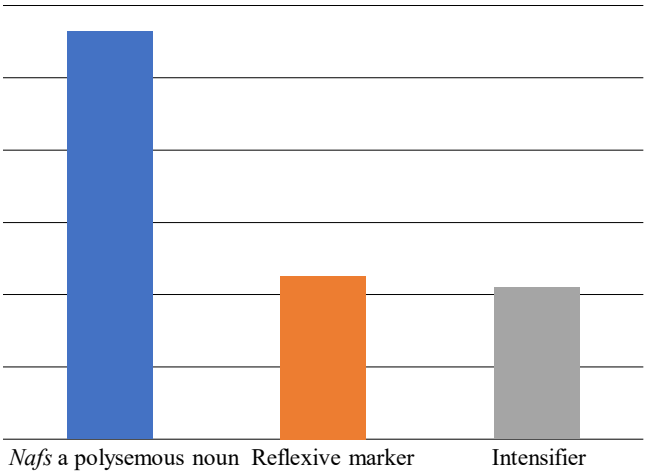


Figure 1.
The distribution of *nafs* in the random sample

In other examples, the antecedent is overtly spelled out, and the binding of the reflexive marker, *nafs*, satisfies the canonical binding picture. Examples to illustrate this occurrence from the random sample include:

18. Wa la'alla hedhihi 'al-lahjat$_i$ faradhat nafsa-ha$_i$
And maybe this dialects impose.PST.3P herself
'And maybe these dialects imposed themselves'

19. Inna manahijahum 'al-dīnīyya$_i$ tafridhu nafsa-ha$_i$ 'alayhim
That orientation-3PM the-religious impose.PRSNT.3P herself on-them
'Their religious orientations impose themselves on them'

As recurrent binding constructions, (18) and (9) prove that MSA is one of the languages that fit within the model of strict binding evoked by Chomsky (1981). Regardless of the status of the antecedent, these examples represent the very typical tensed binding, where reflexive binding is achieved. Following Plunkett (1996), The VP is the tensed domain where binding in Arabic occurs and the two coreferential NPs are in argument positions, as in *impose*, *their religious orientations* and *themselves* in *their religious orientations impose themselves on them*.

Beyond the tensed binding domain, complex NPs can also be binding domains. Investigating this domain as well has led to the finding the reflexive marker *nafs* is not recurrent in NP domains. The following example represents the sole construction found in the studied sample:

20. Laysa hunākā dhulmun 'ashaddu min dhulmi 'al-insen li-nafsi-hi
No there-is oppression tougher from oppression the-human-being to-himself
'There is no tougher oppression than that of the human being to himself'

In (20), the complex NP, *dhulmi 'al-insen li-nafsi-hi*, which occurs as a post-preposition (min/from), is where both the antecedent *'al-insen* and the reflexive marker *nafsi-hi* are coreferential. *'al-insen* is the accessible subject and governor of the reflexive *nafsi-hi*, and the canonical binding requirements are satisfied in this domain. However, this example is not repeatedly used in MSA.

To recapitulate, the examined examples of local binding exemplified in (1a), (1b) and (1c) do not all eventuate as part of MSA. While binding in the tensed domain is part of the modern variety, the non-tensed NP domain is marked by an extremely limited occurrence and VP-constituent extraction is totally absent in the examined random sample. As demonstrated, the latter construction does not deviate from canonical binding, but rather represents a movement of the bindee in the front position for emphasis. Based on the random sample, it is safe to claim that focus-based movement of the reflexive marker *nafs* is not frequently used in MSA, given its total absence in the examined body of text. Notably, this sort of movement does not seem to be part of the examined text in this study, which AlBzour's (2015) claim that reflexive topicalization is also part of MSA. In fact, it seems that this specific use is part of well-determined genres of the Arabic text.

5.3 Reflexives as intensifiers

In noun-right positions, *nafs* acts as a self-intensifier in an adnominal position. Akin to English, these self-intensifiers are attached to the NP they modify in a sort of local binding where the modified NP and the self-intensifier share the same referential value. The following examples show some of these occurrences in MSA:

21. Ta'arradha Blather nafsa-hu li-'al-iqaf min janib lajnat
Incur.PST.3SM Blather himself to-arrest by (the) commission
'al-qyam al-tābi'a lil-FIFA'
Value of FIFA
'Blather himself was stopped by the FIFA's Values Commission'

22. Wa sa'yi-hi bi-dirāyatin tāmma 'ila taksim Turkey nafsa-hā min khilel
And seek.NOUN.3SM by-awareness full to divide Turkey herself through
'al-mumārasāt 'al-majnuna 'al-lati irtakaba-hā
the-practices the-crazy which commit.PST.3SM
'And he sought with full awareness to divide Turkey itself through the crazy practices he committed'

In these examples, *nafs* assumes the intensifying function, where the emphasized NPs assume the role of the binder. At this level, it should be stressed that these self-intensifiers do not imply the meaning of *the same* in Arabic (which is a common error in the language). Rather these sentences show that *nafs* here is used to contrast the emphasized NP with a set of entities within its respective context. By way of illustration, in *Blather nafsa-hu* in (21) it is obvious that the emphasis is laid on *Blather* and not any other individual from the same context. (Obviously, the self-intensifier here cannot be confused with the erroneous meaning of *the same Blather*, given that it is juxtaposed to a proper noun.) The exact pattern is also observed in (22) with the NP *Turkey nafsa-ha*, which equally testifies to be an example where the NP *Turkey* is being underscored rather than being intended as *the same Turkey*. These sentences not only typify *nafs* as a self-intensifier, they equally point to the same locality feature that binding has in MSA. From the examined constructions and their occurrence, it appears that Arabic is one of the languages that is marked by the strict locality of the reflexive marker and the self-intensifier. Both occurrences are connected to their antecedents in environments that stipulate syntactic closeness between the binder and the bindee. This feature in particular and the other constructions will be further examined in CA as well.

5.4 The distribution of nafs *in Classical Arabic*

Repeating the same process of data investigation in the Quran corpus, the word sketch tool is used to examine the collocations in which the word *nafs* occurs. Upon scrutiny, *nafs* occurs with verbs in a verb-right pattern (*lose*, *wrong*, *intrigue*, *convert*, *shelter/protect*, *provide*, *to be*), and with nouns in a noun-right pattern. However, it should be noted that the latter occurrence is limited in this corpus, when compared with MSA findings. Table 3 outlines some occurrences with verbs that signal the reflexive use of *nafs* in binding environments. The other occurrence that needs to be numerically reported is that with nouns in noun-right position, which is presented with the normalized frequency values in Table 4.

Table 4.
Nafs in classical Arabic

Nafs with verbs	Frequency	Nafs with nouns	Frequency
Lose	1.02	Good (for)	0.38
Wrong	2.05	Earth/land (a noun can intervene)	0.38
Intrigue	0.64	God (a noun can intervene)	0.51
Convert	0.51		
Shelter/protect	0.38		
Provide	051		
To be	1.02		

Although the noun-right pattern of reflexives is detected in CA, its incidence is less frequent than in MSA given the size of the corpus. The absence of reflexives in the emphatic use, compared to MSA, implies that the classical text is replete with the religious spirit of instruction that necessitated the syntactic/reflexive use of *nafs* (combined with the other use of the term indicated in the definitions of *nafs* outlined in section 2). This difference can also be detected in a distribution pattern that can be found in CA, which is verb-left. This specific example complies with one of the prototypical occurrences of reflexive pronouns in general after being fronted in some sentences (*himself, he saw in the mirror*). An example of this distribution is found mostly with the verb *to wrong*, which will be further commented on in what follows.

In order to represent the distribution of *nafs* as a reflexive marker and as a (polysemous) noun in the Quran corpus, Figure 2 shows that the reflexive use is remarkably higher (72%) than the other uses. (The claims are cautiously made given the size difference between corpora in this study).

Having established the higher frequency of the reflexive use of *nafs* in the CA corpus, it is necessary to turn to scrutinizing the environments in which it occurs. However, it should be noted that the examined classical corpus does not include examples where *nafs* functions as a self-intensifier, which is the major difference between the corpora in this study. While *nafs* is used syntactically as a reflexive pronoun and a polysemous noun, its emphatic use as a self-intensifier is not part of this classical text.

5.5 Local binding: nafs as a reflexive marker

Notwithstanding its apparent similarity with MSA, *nafs* as a reflexive marker in CA is actually different in some respects. The difference is apparent in the fact that the examined local binding environments are all part of CA, unlike the modern variety. Examples from the classical corpus are outlined in what follows.

The first typical occurrence of *nafs* as a bound reflexive marker is in coargument binding position, as shown in the succeeding illustrations.

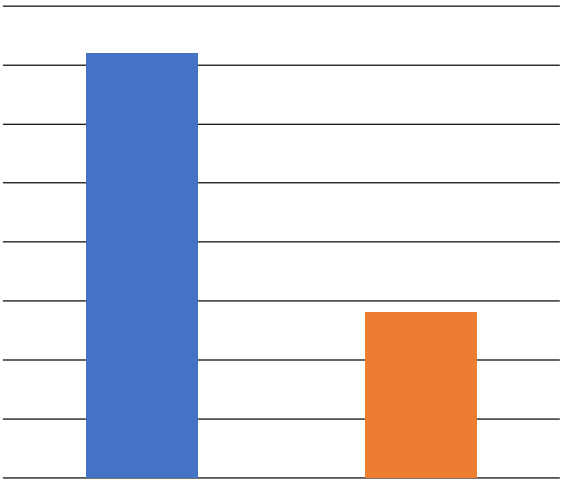


Figure 2.
The distribution of *nafs* in the Quran Corpus

23. Kadhab-ū 'ala anfusi-hum
Lie.PST.3P on themselves
'They lied to themselves (Al-an'am 24)'

24. Qawmun dhalamū anfusa-hum
People wrong.PST.3P themselves
'People wronged themselves' (Ali 'Imran, 117)

(23) and (24) represent the tensed binding domain of reflexive pronouns. *Nafs* in these sentences is bound by a covert (23) an overtly expressed (24) antecedent, in a canonical binding domain. A quintessential example is *qawmun* and *anfusa-hum* with the verb *dhalamu* (*to wrong onself*), which is shown as an example of a verb that might be followed by the use *nafs* as a reflexive marker (among possibly other uses).

In another incidence, *nafs* as a reflexive marker, is detected in a complex NP, such as (25)

25. Inna numli lah-um khayran min anfusi-him
That.1P dictate.PRSNT.1P them good from themselves
'We respite good for themselves' (Ali 'Imran, 178)

The reflexive marker *anfusi-him* is part of the complex NP *lah-um khayran min anfusi-him*, which is bound by the joint pronoun attached to the preposition *for* (*lah-um*). The syntactic closeness of the binder (*la-h-um*) and the bindee (*anfusi-him*) is the indication that this is an instance of strict reflexive binding, identified as the second typical non-tensed domain in the theory. (26) is another instance of NP binding domain:

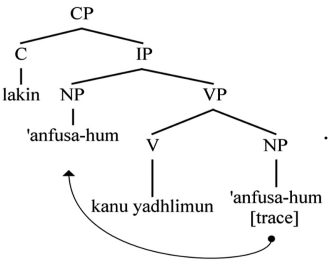
26. 'Idh bu'itha fihum rasūlun min anfusi-hum
As sent.3SM in.3P messenger from themselves
'As He raised up a messenger from among themselves' (Ali 'Imran, 164)

The prepositional phrase, *fihum rasūlun min "anfusi-hum*, accommodates the binder as an inflection attached to the preposition *fi/in* to be *fi-hum*, and the bindee as a part of the NP complement '*anfusi-hum*. Although (26) is the second occurrence presented in this piece of research, it should be highlighted that this occurrence is not very frequent when compared to the tensed domains of binding, which resonates with the MSA findings.

The last local binding instance that is investigated in this study is VP-constituent extraction, where the reflexive pronoun is moved from its original position to precede its binder. Although this occurrence is not detected in MSA, it is identified in CA, as revealed in the following sentence.

27. Lākin anfusa-hum kānū yadhlimun
But themselves be.3P wrong.PST.3P
'But they used to wrong themselves' (Al-Naḫl 33)

Example (27) is the realization of *nafs* as a reflexive marker that has undergone focus-motivated movement. In its turn, illustration (29) demonstrates the topicalization process that lands the reflexive marker *'anfusa-hum* in a position prior to its binder, and equally shows the base-generated position of the reflexive marker. VP-constituent extraction is another example of the typical tensed binding domain that can be detected in CA, unlike its counterpart MSA.



As fronting the reflexive marker is used to achieve some emphatic purposes, it should be noted that this fronting pattern occurs mainly with the verb *wrong* in the classical corpus of this particular study, which implies the centrality of emphasis needed in the CA corpus that is motivated by the essence of religious instruction.

The final examined occurrence of *nafs* as a self-intensifier is not part of the Quranic corpus. Upon scrutiny, there is no incidence of *nafs* as an intensifier in CA. Given that the only investigated classical text is the Quran, it is not possible to generalize this finding. Rather it is necessary to search other classical texts to confirm or nullify this claim. As far as the Quranic corpus is concerned, *nafs* occurs mainly in verb-right positions as a reflexive marker, along with its polysemous uses. In this regard, Table 5 summarizes all the occurrences of *nafs* in MSA and CA that are detected in the studied corpora.

6. Discussion

This study focuses on examining the distribution of the polysemous noun *nafs* in two Arabic varieties, MSA and CA. Specifically, the research investigates the binding occurrences of *nafs* in two Arabic corpora. These occurrences include local reflexive binding within the tensed and the NP domain (coargument binding and VP-constituent extraction), and the intensifier/emphatic use of self-forms. This division is purely syntactic for reflexive binding and syntactic and pragmatic for the use of the self-intensifier. Only the relevant definitions (1 and 3 in section 2) of *nafs* are considered in this study.

The study findings reveal that *nafs* in verb-right position is used as a reflexive pronoun and abides by local binding, and when detected in noun-right position, it is used as a self-intensifier. At this level, the distributional patterns detected in both Arabic varieties need to be compared to each other to detect any similarities and differences between them. A preliminary thought might be that the studied varieties are very similar in the diffusion of *nafs*, given that the structures that are examined are proved to be part of both of them. Yet the differences noted between the classical and modern varieties cannot be overlooked. While some constructions are present in one variety, they are not part of the other. From a structural point of view, MSA seems to have a very strict distribution of *nafs*. The only noted occurrences it has are within strict binding domains. Notwithstanding this pattern of occurrence, the examined Modern Arabic text limits the distribution of *nafs* as a reflexive

Constructions	Features	MSA	CA
Local binding: tensed domain	IP/VP hosts the binder, the governor and the bindee	√	√
Local binding: non-tensed domain	Complex NPs host the binder, the governor and the bindee	√	√
Local binding: VP-extraction	The reflexive undergoes topicalization movement	×	√
Intensifier	A self-form juxtaposed to the NP it modifies	√	×

Table 5.
Summary of the results

marker that precedes its binder, hence the absence of the VP-constituent extraction. Another remarkable finding, forwarded cautiously in this study, is that *nafs* as a reflexive marker is only pervasive in the tensed domain, while it is very limited in the non-tensed domain. Therefore, *nafs*, used as a structural reflexive marker, is limited in its distribution in MSA to one major domain, the tensed domain, which Plunkett (1996) defines as the VP domain in particular. Likewise, CA evidence supports this distributional pattern of *nafs*. Its occurrence in the tensed coargument binding is higher in frequency than the other occurrences, which resonates with MSA implying that in the process of the Arabic linguistic transition, this usage of *nafs* has been preserved (Kremers, 1997). Studies that match this line of thought, where *nafs* is structurally distributed in well-defined binding contexts, show that this occurrence is also part of Standard Arabic and its varieties (Alenazy, 2021; Alkafri, 2019; Mustafawi and Mahfoudhi, 2002).

Turning to the difference that lies between the two examined varieties, it is shown that there are two constructions that are part of one Arabic variety only. The first construction is VP-constituent extraction which is not part of the studied MSA corpus but is moderately frequent in the Quranic corpus. While the Quranic text is replete with religious instruction and emphasis that requires the topicalization of the reflexive marker *nafs*, the examined corpus of MSA consists of journalism texts meant to be impartial and that do not necessitate the same emphasis using *nafs*. This finding opposed the general assumption that AlBzour (2015) contends: *nafs* focus-movement is part of MSA. A very careful inference here is that this movement can be found in particular varieties of Arabic such as CA and can be found in well-defined genres in Arabic, which still need to be thoroughly charted. The other instance that is part of MSA is *nafs* as a self-intensifier. As another syntactic and pragmatic occurrence of *nafs*, this occurrence is frequent in MSA. Such a fact can be justified by the nature of the corpus that needs this sort of emphasis given the nature of news and journalism (that can also be further examined in literary texts such as prose and poetry). The fact that this occurrence is not detected in the Quranic corpus can be accounted for given the nature of the Quran. As an authoritative text (Abu-Raiya, 2012), it does not set the ground for some pragmatic and context-bound interpretations that need the kind of emphasis served by self-intensification (contextual contrast of the emphasized entity). Rather the implicit emphasis or intensifier use is instantiated in expressions *kul* and *jamee*'

In the same way the VP-constituent extraction is not part of MSA due to the nature of the examined corpus, *nafs* as a self-intensifier is not part of the Quran given the nature of this text. This claim is not definitive and it needs to be further investigated in a different MSA and CA corpora, prior to positing any instances of syntactic and/or semantic loss or emergence in the transition from the classical to the modern variety of Arabic.

Based on the findings collected from this study, it is safe to postulate that *nafs* as a polysemous noun serves a multitude of syntactic and pragmatic functions. Given that many languages are examined in terms of their anaphoric distribution and proved to abide by or deviate from the canonical picture of binding (Huang, 2000), the Arabic language seems to have one of the strictest accounts of binding. This is evident in the distribution of *nafs* as a local reflexive marker within a well-defined syntactic domain, and in the diffusion of *nafs* as a self-intensifier found only in NP-juxtaposition. In both environments, *nafs* is very closely placed to its binder, whether it is a syntax-based position (reflexive) or syntax-as well as discourse-bound position (self-intensification).

7. Conclusion

This study is a corpus-based analysis of the distribution of *nafs* in MSA and CA. Given the nature of this endeavor, the analysis conducted in this study is exploratory in nature and

outlines the similarities and differences between each variety in terms of the diffusion and use of *nafs*. Results reveal that it is used as a reflexive marker and self-intensifier with almost the same frequency rate as in MSA. However, the use of *nafs* as a polysemous noun is far more frequent than the investigated uses. On the contrary, the distribution of *nafs* in CA is different from that in MSA. The dominant use is the reflexive one, whereas the intensifier use is completely absent and the content use is considerably less frequent when compared to modern Arabic.

The comparison undertaken in this study accentuates the differences and the similarities between both varieties by showing which constructions are found in MSA and CA. Although the nature of the examined corpora is divergent, this juxtaposition reveals how each Arabic variety abides by and the canonical binding picture. While both CA and MSA seem to follow strict formal binding, Modern Arabic exhibits the strictest account of *nafs* distribution in reflexive and NP-juxtaposed positions, while CA exhibits a more flexible diffusion of *nafs* in positions that exceed the binder. This claim is solely based on the texts consulted in this study, and generalizing it requires scrutinizing other types of Arabic texts.

The findings of this analysis capitalize on the fact that *nafs* in VP-constituent extraction is absent from MSA while *nafs* as self-intensification is absent from CA. The structural difference in the distribution hinges upon two major aspects: while a specific structure is no longer part of MSA although it was part of CA, another structure emerges. The implication is that given this analysis, it is necessary to further explore the diffusion of *nafs* and to examine other Arabic texts (from different genres) in order to see whether it is subject to linguistic changes in the transition from CA to MSA.

8. Pedagogical implications and future research

This study sets out to describe and analyze the distribution of *nafs* in MSA and in Quranic Arabic based on corpora, where the comparison and discussion examine the syntactic patterns of this distribution. It appears that are differences in the occurrence of *nafs* between the varieties and they should be taken into consideration in pedagogical practices of teaching Arabic. The fact that there are syntactic and discourse-based differences in the distribution of *nafs* in the classical and modern versions of Arabic points to the linguistic transition and its consequences making certain occurrences more frequent than the others.

With growing interest in language acquisition and particularly pronoun acquisition cross-linguistically, the findings of this study can be relevant in defining the specific distribution of *nafs* as a reflexive pronoun and its acquisition in Arabic as first and a second language. Various studies addressed this matter (Mustafawi and Mahfoudhi, 2002), yet their linguistic background is based on a set of data that is compared to English rather than a sort of data that documents the entire distribution of the term in one of the Arabic varieties. For this reason, this study offers a preliminary distributional account of *nafs* in Arabic which needs to be further developed in different varieties of the language and based on other sorts of texts.

Based on the findings of this research, it is tempting to examine why some structures are found in one variety and not in the other. To offer a comprehensive account, the research agenda that announces itself includes studying the absence of the intensifying function of *nafs* in CA and the implications, and examining whether *nafs*, as a reflexive marker, deviates from canonical binding in Arabic (including its different regional varieties) the way English reflexive pronouns do. Building a general account of this distribution is relevant in understanding the explicit (syntactic) and implicit (discourse-based) dimensions of reflexive marker and self-intensifier processing and interpretation in Arabic as a first and second language.

References

Abu-Raiya, H. (2012), "Towards a systematic Qura'nic theory of personality", *Mental Health, Religion and Culture*, Vol. 15 No. 3, pp. 217-233, doi: 10.1080/13674676.2011.640622.

Al-Raba'a, B.I.M. (2021), "Anaphoric binding in Modern Standard Arabic: a phase-based analysis", *Canadian Journal of Linguistics*, Vol. 66 No. 3, pp. 374-411, doi: 10.1017/cnj.2021.22.

AlBzour, N.N. and AlBzour, B.A. (2015), "Arabic uniglossia: diglossia revisited", *Studies in Literature and Language*, Vol. 10 No. 3, pp. 7-12, doi: 10.3968/n.

Alenazy, M.A. (2021), "Binding relations and their implications for word order in Arabic", *Theory and Practice in Language Studies*, Vol. 11 No. 9, pp. 1018-1024, available at: http://irep.iium.edu.my/91905/1/TPLS%20Surmounting%20Stereotypes.pdf.

Alkafri, A.A.L. (2019), "Similarities and differences between child and adult L2 learners in the acquisition of English reflexives", *Open Journal of Modern Linguistics*, Vol. 09 No. 06, doi: 10.4236/ojml.2019.96038.

Alqasem, Z. (2013), *Unifying Quranic Analyses into a Single Database*, project dissertation, University of Leeds.

Baker, C.L. (1995), "Contrast, discourse prominence, and intensification, with special reference to locally free reflexives in British English", *Language*, pp. 63-101, doi: 10.2307/415963.

Bolotin, N. (2002), "Acquisition of binding in L1 Arabic", in Dilworth, B. and Benmamoun, E. (Eds), *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics*, Vols XIII-XIV, John Benjamins, Standford and Berkeley, California, pp. 211-218.

Charnavel, I. (2020), "Logophoricity and locality: a view from French anaphors", *Linguistic Inquiry*, Vol. 51 No. 4, doi: 10.1162/ling_a_00349.

Charnavel, I. and Sportiche, D. (2016), "Anaphor binding: what French inanimate anaphors show", *Linguistic Inquiry*, Vol. 47 No. 1, pp. 35-87, available at: https://www.jstor.org/stable/43695776?seq=1.

Chien, Y.C. and Wexler, K. (1990), "Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics", *Language Acquisition*, Vol. 1 No. 3, doi: 10.1207/s15327817la0103_2.

Chomsky, N. (1981), *Lectures on Government and Binding*, Dortrecht, Foris.

Dukes, K. (2009), *The Quranic Arabic Corpus*, Leaman, Oliver.

Felser, C. and Cunnings, I. (2012), "Processing reflexives in a second language: the timing of structural and discourse-level constraints", *Applied Psycholinguistics*, Vol. 33 No. 3, pp. 571-603.

Gast, V. and Siemund, P. (2006), "Rethinking the relationship between SELF-intensifiers and reflexives", *Linguistics*, Vol. 44 No. 2, pp. 343-381, doi: 10.1515/LING.2006.013.

Haegeman, L. (1994), *Introduction to Government and Binding Theory*, Wiley-Blackwell.

Hammami, S., Belguith, L. and Ben Hamadou, A. (2009), "Arabic anaphora resolution: corpora Annotation with coreferential links", *International Arab Journal of Information Technology*, Vol. 6 No. 5, pp. 480-488, available at: https://ccis2k.org/iajit/PDF/vol.6,no.5/7.pdf.

Holes, C. (2004), *Modern Arabic: Structures, Functions, and Varieties*, Georgetown University Press, Georgetown University.

Huang, C.T.J. (1987), "Remarks on empty categories in Chinese", *Linguistic Inquiry*, Vol. 18 No. 2, pp. 321-337, available at: http://www.jstor.org/stable/4178540.

Huang, Y. (2000), *Anaphora: A Cross-Linguistic Study*, Oxford University Press, New York.

Igaab, K.Z. and Tarrad, I.R. (2019), "Pronouns in English and Arabic: a contrastive study", *English Language and Literary Stududies*, Vol. 9 No. 1, pp. 53-69, doi: 10.5539/ells.v9n1p53.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J. and Suchomel, V. (2014), "The Sketch Engine: ten years on", *Lexicography*, Vol. 1 No. 1, pp. 7-36, doi: 10.1007/s40607-014-0009-9.

Kremers, J. (1997), *How Arabs Speak to Each Other about Themselves: A Study of NAFS and Bad in Modern Standard Arabic*, Katholieke Universiteit, Nijmegen.

Kwaik, K., Abu, S., Motaz, C.S. and Dobnika, S. (2018), "A lexical distance study of Arabic dialects", *Procedia Computer Science*, Vol. 142, pp. 2-13, doi: 10.1016/j.procs.2018.10.456.

Lasnik, H. (1989), *Essays on Anaphora*, Springer, Dordrecht.

McKee, C. (1992), "A comparison of pronouns and anaphors in Italian and English acquisition", *Language Acquisition*, Vol. 2 No. 1, pp. 21-54, available at: https://www.jstor.org/stable/20011365?seq=1.

Mejri, A. (2020), "The interpretation of English reflexive and non-reflexive pronouns by Tunisian students", *Argumentum*, Vol. 16, pp. 241-259.

Mustafawi, E. and Mahfoudhi, A. (2002), "The development of binding principles: new findings", *Cahiers Linguistiques d'Ottawa*, Vol. 30, pp. 91-111.

Owens, J. (2013), *The Oxford Handbook of Arabic Linguistics*, Oxford University Press, Oxford.

Patterson, C., Trompelt, H. and Felser, C. (2014), "The online application of binding condition B in native and non-native pronoun resolution", *Frontiers in Psychology*, Vol. 5, pp. 1-16, doi: 10.3389/fpsyg.2014.00147.

Plunkett, B. (1996), "Binding in Arabic and the lexical clause", in Eid, M. (Ed.), *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics*, John Benjamins, Massachusdets, pp. 85-98, doi: 10.1075/cilt.134.08plu.

Pollard, C. and Sag, I. (1992), "Anaphors in English and the scope of the binding theory", *Linguistic Inquiry*, Vol. 23, pp. 261-305, available at: https://www.jstor.org/stable/4178768.

Trampus, M. and Novak, B. (2012), "The internals of an aggregated web news feed", *Proceedings of 15th Multiconference on Information Society*, available at: https://ailab.ijs.si/dunja/SiKDD2012/Papers/Trampus_Newsfeed.pdf.

Corresponding author
Amani Mejri can be contacted at: amany.mejri@arts.unideb.hu

Appendix 1
***Nafs* as a reflexive marker in Arabic and its English equivalents**

Person	Number	Gender	P-pronouns	R-Pronouns
1	Singular	Neutral	‘Ana (I)	Nafs-ī (myself)
1	Plural	Neutral	Nahnu (We)	‘Anfusa-nā (ourselves)
2	Singular	Masculine	‘Anta (You)	Nafsa-ka (yourself)
2	Singular	Feminine	‘Anti (You)	Nafsa-ki (yourself)
2	Dual	Neutral	‘Antuma (You)	Nafsa-kuma (yourselves)
2	Plural	Feminine	‘Antunna (You)	‘Anfusa-kunna (yourselves)
2	Plural	Masculine	‘Antum (You)	‘Anfusa-kum (yourselves)
3	Singular	Neutral	Huwa (He)	Nafsa-hu (himself)
3	Singular	Feminine	Hiya (She)	Nafsa-hā (herself)
3	Dual	Neutral	Huma (They)	Nafsa-humā (themselves)
3	Plural	Feminine	Hunna (They)	Anfusa-hunna (themselves)
3	Plural	Singular	Hum (They)	Anfusa-hum (themselves)

Appendix 2
***Nafs* in MSA**

1473481	lebanese-forces...	, و فخامة الرئيس لا يستقيل , و مجلس النواب لا يحل	نفس	ه , و اذا فعل لا تفيد الانتخابات ل ان ها لا تغير شيئا " .
1473482	lebanese-forces...	، انطلاقا من هذا الواقع , تلفت المعلومات الديبلوماسية	نفس	ها , الى ان كل ما مرت ب ه المبادرة الفرنسية , قد بات
1473483	lebanese-forces...	الحديث عن ها . <s></s> و من هنا , تلاحظ المعلومات	نفس	ها , ان الملف اللبناني عموما , و ليس فقط ملف تاليف
1473484	shorouknews.com	يط بين ها " الراوى " , الذي يكاد يعادل محفوظ الواقعى	نفس	ه , و " الحارة " و " النكية " ب غموض ها و احلام ها و ,
1473485	almalnews.com	سليم 152 وحدة فى هذا المشروع . <s></s> فى الوقت	نفس	ه بلغ اجمالى النفقات الاستثمارية ل الانشاءات خلال الـ
1473486	almasryalyoum.c...	خيرة علماء الازهر , ف عنى ب تعليم ابن ه , و تعهد ه ب	نفس	ه , و جعل ه يتردد على حلقات العلم ب الازهر , فتردد ع
1473487	aljazeera.net	اسي اليساري المعروف يانيك جادو في هذا السياق رغم	نفس	ها التفاؤلي حيث قال " ربما لن ننجح , لكن دعو نا نحاو

, و ذلك ل ضمان توفير مساحة كافية و تطبيق اجراءات التباعد الاجتماعي . <s></s> فيما حصلت الكورية الجنوبية يو يونغ يون , على اوسكار افضل ممثلة مساعدة عن دور ها في فيلم " ميناري ". <s></s> سول يفوز ب جائزة اوسكار افضل فيلم رسوم متحركة هذا العام <s></s> و يجد غاردنر **نفس** ه بعد سقطة , في طابور سماوي لا نهاية ل ه هو ب مثابة غرفة انتظار ما بعد الموت , قبل ان يهوي ب الخطا الى " عالم ما قبل الولادة " , حيث من المفترض ان تكتسب كل " روح " بشرية شخصية ها , ب صفاة ها و

Appendix 3
Nafs in CA

, و ذلك ل ضمان توفير مساحة كافية و تطبيق اجراءات التباعد الاجتماعي . </s><s> فيما حصلت الكورية الجنوبية يو يونغ يون , على اوسكار افضل ممثلة مساعدة عن دور ها في فيلم " ميناري ". </s><s> سول يفوز ب جائزة اوسكار افضل فيلم رسوم متحركة هذا العام </s><s> و يجد غارندر نفس ه بعد سقطة , في طابور سماوي لا نهاية ل ه هو ب مثابة غرفة انتظار ما بعد الموت , قبل ان يهوي ب الخطا الى " عالم ما قبل الولادة " , حيث من المفترض ان تكتسب كل " روح " بشرية شخصية ها , ب صفاة ها و