Productivity and Predictiveness of Grammatical Gender Assignment on Modern Standard Arabic's Nouns

Abstract

Like many other Semitic languages, Arabic language is known as having two genders to separate the masculine from the feminine, and distance itself from accepting the mistake in using them or confusing them, in order to preserve the integrity of the communication process and its continuation. Moreover, the linguistic research has indicated that there are difficulties in assigning grammatical gender in the Arabic language, as it is the case in many languages, which is related to distinguishing between masculine and feminine, the other is related to the multiplicities in the structural forms of gender whether natural or grammatical. Thus, the topic of masculinization and feminization is one of the most obscure topics that refuses to lead to an analytical principle. In addition, this paper is an attempt trying to present a proposed mechanism based on the use of an automatic learning algorithm according to a semantic network to build relationships between nouns, through which the gender is specified in the nouns in Modern Standard Arabic. The results of the application of the proposed mechanism confirm the possibility of predicting the gender of the nouns in a somewhat large extent, as the gender of the nouns cannot be completely predicted. This confirms that the difficulty continues, and the researches and investigations need to remain open, for more visions and new grammatical methods, to solve this issue.

مستخلص

تحرص اللغة العربية على الفصل بين المذكر والمؤنث،وتنأى بنفسها عن قَبول الخطأ فيهما أوالخلط بينهما،حفاظا على سلامة عملية التواصل واستمرارها، وقد دلّت الأبحاث اللغوية، على وجود صعوبات في تعيين الجنس النحوي في العربية؛ تتعلق بتعيين المؤنث من المذكر من الجنس نفسه، وتعدّد الأوجه النحويّة التركيبية، سواء أكان جنسًا حقيقيًا طبيعيًا أم جنسًا مجازيًا نحويًا، فموضوع التذكير والتأنيث من أكثر الموضوعات غموضًا وتمنّعًا على الانقياد لمبدأ تحليليّ، ويُعنى هذاالبحث بمحاولة تقديم آلية مقترحة تستند على استخدام خوارزمية تعلّم آليّة وفقًا لشبكة دلاليّة لبناء العلاقات بين الأسماء، يتمّمن خلالها تعيين الجنس (النوع) في الأسماء باللغة العربية المعاصرة، وتؤكد نتائج تطبيق الآلية المقترحة إمكانية التنبؤ بجنس (نوع) الأسماء بشكل كبير إلى حدّ ما؛ إذ لا يمكن التنبؤ بجنس (نوع) الأسماء بشكل كلي تام؛ وهو ما يؤكد أن الصعوبة مستمرة، وبحاجة أن يظل البحث والاستقصاء مفتوحًا لمزيد من الرؤى، وجديد من التقعيد.

1. Introduction

This paper discusses how gender is assigned to nouns in Modern Standard Arabic (MSA). Arabic has either derived nouns, which contain verbal nouns and verbal derivatives or non-derived nouns (substantives), which contain proper nouns and borrowed nouns. Arabic nouns inflect for gender, number, case or definiteness. Some other different nouns do exist as categorized as derived nouns which carry names related to their derivational processes.

2. Background

Gender has received researchers' attention for a long period of time. The difference between masculine and feminine is recognized as for human and animals naturally due to the existence of both masculine and feminine among them. Then, this extends to has its impact on human's languages, particularly on nouns, which come up as masculine and feminine as in Arabic (Abduattawab, 1970). The term "gender" is generated or borrowed from Latin "Kgenus" which just refers to gender as being male and female. Gender is not only used to feature nouns but also used to categorized pronouns and adjectives. Arabic, among other Semitic languages, has to contain two gender: masculine and feminine (Amaayrah, 1993). Masculine and feminine are used to assign gender to nouns, adjectives, gerunds, locatives and pronouns where they all carry the feminine mark indicators on them while referring to meanings or things in the universe. Thus, verbs and prepositions (including all functional morphemes, derivational morphemes and inflectional morphemes) are not considered in the study of gender system in languages, particularly Arabic. Verbs refer to actions linked directly to the SUBJECT (not only the doer) where they sometimes carry gender mark indicators. These gender mark indicators have nothing to do with the gender system. Rather, they are used as part of the agreement system (Subject/verb agreement). On the

other hand, other morphemes such functional morphemes or bound morphemes do not contain any gender system and some do not even carry meanings by themselves.

Arabic (MSA) is based on two categories (masculine and feminine) where there are no neutral pronouns. MSA relies on having masculine as the default gender. Nouns are divided into three kinds with relation to number: singular (masculine and feminine), dual (masculine and feminine) and plural (masculine and feminine). A noun is assigned to either masculine (as in *raʒul:* "man" or *kitāb*: "book") or feminine (as in *ʾimrʾah:* "woman" or *dār*: "house"). *raʒul:* "man" is known in MSA as (*muḏakar ḥaqiqi:* natural masculine) because it contains natural gender assignment and it has its feminine counterpart. However, *kitāb:* book is a assigned to masculine grammatically and it does not have any feminine counterpart. On the other side, *ʔimraʔh:* woman is also know in MSA as (*muaʾnt ḥaqiqi:* natural feminine) and so it comes under natural gender assignment where it has it masculine counterpart. *dār*: "house", however, is assigned to feminine grammatically and has no masculine counterpart.

Feminine gender shows three main feminine mark indicators: **-t** (**-h** in English script) as in خديجة: *ḥadīʒah*, the short /a/ (*ʾalif ʾalmaqṣūrah*) as in سلمى: *Salmā* and the long /ā/ (*ʾalif almadd*) as in حسناء: *ḥasnā*. For those feminine nouns, which have no feminine mark indicators, are known as *muaʾnt maʿanawi* (unreal feminine); that is, they are treated as *muaʾnt ḥaqiqi:* natural feminine. There is no rule to assign gender to those nouns. This type is called *maʒāzī* unreal (similar to grammatical gender) where nouns are assigned to gender depending on hearing and/or language books and dictionaries.

Therefore, gender on nouns is seen to be divided into two main categories with regard to gender assignment system. The first category consists of natural gender

4

assignment including both masculine and feminine and grammatical gender assignment including both masculine and feminine. The second category is based on the form in which nouns might have morphological feminine mark indicators or might not. Nouns appear in five different kinds:1- natural feminine (e.g. *Umm*: "mother"); 2- grammatical feminine with no gender mark indicator (e.g. *kabid*: "liver"); 3- grammatical feminine with gender mark indicator (e.g. *maʿdah*: stomach); 4- natural masculine (e.g. *ʾabb*: father); 5- grammatical masculine (e.g. *raʾs*: head).

In the second century BC, Protagoras is considered the first linguist who payed attention to gender assignment apart from natural gender assignment system. Then, Thrax proposed that those feminine nouns have different suffixes from other nouns which refer to masculine. He divided gender in Greek according to nouns' suffixes rather than on their natural type in which they showed to contain masculine, neutral and feminine (Mario pie, 1970).

Arabic linguists have differentiated between gender with relation to animacy where animate contains both masculine and feminine based on natural gender assignment, in general. An inanimate show both masculine and feminine based on grammatical gender assignment.

Assara'an (1997) and Anees (1987) proposed that gender in Semitic languages have both masculine and feminine as the only two types in gender system including all nouns. Other Latin languages such as French, Spanish, Italian and portugees show to have types of gender: natural gender and grammatical gender where they all include both masculine and feminine. Some other Indo-European languages like German, Greek and Russian divide gender into three different kinds which are masculine, neutral and feminine. Some Eastern African languages extend gender types to consist of

sixteen types where each type has specific (distinctive) morphological features. For example, "big and strong" has a layer where as "small and weak" has another different layer. Some other languages such as Turkish and Persian do not show the difference between the different types of gender on the form (nouns). Rather, there must be gender indicator or tangible evidence.

Joseph (1950) claims that the distribution and the difference between genders among languages is generated from ancient views, which are based on religious motivations or other unknown ones. Then, it has been used as it is without understanding any reasoning behind it. Old Arabic grammarian school and linguists (such as *Alfaraa, Assajistani, Almabrad, Abi Bakr Alanbaari, Abn Khaalwiyah, Abn Jini, Abn Faris, Abi Albarak Alanbaari, Abn Alhaajib Araazi*) allocated some works to the study of gender; among many other modern Arabic linguists like *Mahmmad Hussain, Ibrahim Assamrrai, Ramadan AbduAttawab, Ibrahim Barraakat, AbdulAlmuna'am Annajaar, Mahmood Ukaashah* and *Ismail Amaayirah*.

On the other hand, other western linguists such as Grosjean Dommergues, Cornu, Guillelmon, and Bessson (1994) and Bates, Devescovi, Hernandez, and Pizzamiglio (1994). Greenberg (1978) tried to explain the origin of gender, too. Aitchison (1987), McMahon (1994) and Aikhenvald (2000), among many others, had put great efforts to find out reasoning and explanation for the gender system in languages. They tried to explain the processing of gender in comprehension and explain how to use content to turn it into form. Yet, the historical view is not enough to account for and explain the presence of gender system.

Other linguists, following other two approaches: linguistic nativism and linguistic functionalism, connect gender to its communicative functions. So, linguists

like Köpcke and Zubin (1984); Mills (1986); Bates and MacWhinney (1989); Corbett (1991) proposed that grammatical gender is important and necessary to link the referents to the context. Also, some other linguists such as (Wijnen & Deutsch, 1987; Desrochers, 1986) claim that grammatical gender can play an important role in forming correct structures consisting of words, phrases, clauses and sentences where they all need to be assigned appropriately in the right projection(s) in any tree diagram. Word recognition process carries as much importance as other processes. So, some psycholinguists claim that gender has its impact on the word recognition process (Colé & Segui, 1994; Grosjean et al., 1994). Thus, it is obvious that they all agree that gender is considered core in languages when related to the different field of linguistics studies. In addition, previous studies claim that there is no logical reasoning to differ between masculine and feminine, and so, grammatical gender is less logical and more arbitrary (Zubin, 1992; Köpcke & Zubin, 1984).

Corbett (1991) proposed that languages follow a systematic assignment to gender based on semantic factors or on semantic, morphological and phonological factors. He claimed that some languages assign gender to nouns systematically while some other languages show that gender assignment arbitrary. However, those languages, which were under study, show a high percentage of prediction but through information required independently in the lexicon (Corbett, 1991, p. 68).

Therefore, such huge efforts, which have been achieved to study thoroughly gender assignment, assure the difficulty of setting rules for the gender system and show the sensitivity of the topic in the research area. This difficulty and sensitivity is still exist and research needs to go further and deeper to solve some of the issues in gender assignment system, which this research is trying to do with respect to Arabic.

It is essential to state that gender assignment is one of the vague topics and irresolvable as being with no clear analysis and regulations to apply to all languages or even to any language by itself. Talafihah and Alagtash (2014) state that grammatical gender assignment in Arabic is still obstacle without being able to recognize masculine and/or feminine with the different possible syntactic structures in the language.

Having problems in grammatical gender assignment may lead to a misunderstanding or a misleading discourse analysis. Fandrees (1950, 127) states that misusing and/or mixing gender is one of the most shocking areas to the hearer, which ultimately and consequently causes understanding the discourse. Thus, most Arabic linguists see that knowing and differentiating between masculine and feminine is a way to understand syntax of Arabic. Abu Bakr Alanbaari (1970: 87) suggests that wrong grammatical gender assignment is equal to wrong case marking in syntax.

Moreover, the difficulty can be seen when using grammatical gender, as mentioned before, while writing and/or converse. There is no tool to utilize figuring out what is masculine and what is feminine. Hence, the importance of this research exists to assign grammatical gender on nouns which carry no feminine mark indicators and attempt to solve some of the issues and overcome some of the difficulties in MSA.

3. Previous studies

Gender studies, as stated before, has been studied in many languages from different language families. Dixon (1968:105) suggests that studying grammatical gender took other diminutions by investigating and understanding other languages other than the Indo-European languages. Corbett (1991) considers gender assignment system crucial when it is linked to the structure of the lexicon. He suggests that gender assignment is a rule-based system after covering many languages, concluding by

introducing a rule- based model for Russian nouns. Aikhenvald (2000) emphasizes on the semantic features while studying grammatical gender. Rice (2006), on the other hand, has done the same type of the studies where he connected gender assignment to the constraints-based model. He suggests gender assignment hierarchy in German.

After that, two different theories: traditional theory and the cognitive theory, appeared. The traditional theory tends to the symbolic rules whereas the cognitive theory focuses on the analogy between the linguistics items. Bybee (1985; 1988; 1995a; 1995b; 1998) suggests the network model where he claims that there is a connection between the linguistics item and the brain. Langacker (1987) suggests the symbolic rules, too.

It is important to state that Olstad's study (2011) is considered unique of its type as he produces a supervised machine learning algorithm to test the predictiveness of Contini-Morava' analysis (1994). Then, his focus was on the Swahili noun class.

Some languages show to contain a big number of genders. For example Fula show around twenty genders (Corbett, 2006:753). However, Semitic languages show only two genders (masculine and feminine), as stated before. Lipinski (2001) suggests that Semitic languages may contain common gender. Afroasiatic languages, in general, stress on the importance of semantic features to determine gender. Heine (1982) suggests the big relation between the size and the grammatical assignment of feminine gender. Castellino (1975:354), among many other linguists, produces number of criteria such as masculine/feminine, animate/inanimate, human/non-human, individual/collective, concrete/abstract, singular/plural and minor/major.

Arabic, with around 200 million speakers, is a Semitic language. It shows two main varieties Classical Arabic and Modern Standard Arabic (MSA) where they share many similarities features (Badawi, Carter, & Gully, 2004; Harrama, 1983; Ryding,

2005). Arabic has shown to have many studies regarding gender for a long period. However, studies which are using technology and software programs are rare, except Alzahrani (2019) who studied grammatical gender assignment on Saudi dialects. Hence, the importance of the current study becomes urgent.

4. Objective of the study

This paper suggests an electronic way to study gender through some software programs and some semantics features of nouns connecting them together to help in the production of nouns and the prediction of gender in MSA.

5. Hypothesis of the study

This paper states clearly that gender assignment system in MSA is not purely arbitrary. Thus, there should be some systemic rules, which might help, by logical connecting between nouns with their semantic features, predicting grammatical gender to those nouns with no feminine mark indicators.

6. Methodology

This study is based on both descriptive and analytical methods. As a start, I have used two Arabic dictionaries (ʾlmuʿʒam ʾlwaṣīṭ (2011) and lisān ʾlarab (2007)) to collect a number of Arabic nouns. Then, those nouns were entered into an Excel sheet o sort them out according to some semantics features like gender, size, position, shape, etc. More than 65,000 nouns were collected. After that, the big task is to remove all the nouns which carry any feminine mark indicators or show natural gender assignment. Another important task is to gather nouns into groups creating semantic networks and forming some meaning relations as can be seen in the following figures:
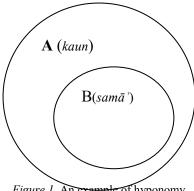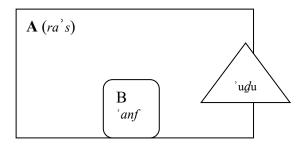
Figure 1. An example of hyponomy.

Figure 2. An example of semantic network.

Figure 1 shows two different nouns but they still share a relation. So, *kaun*:"space" should contain the other noun *sam*ā:" sky"; that is, when we think of *kaun*:"space", definitely the image that comes to our minds all nouns related to it such *sam*ā:"sky", *qamar*: "moon", etc. We can claim here that *kaun*:" space" is the domain noun for many nouns, which come under its branching. On the other side, Figure 2 shows an example of semantic network between a noun *raʾs*: "head" with other nouns such as *ʾanf:* "nose" and *ʾuḏun*: "ear". Discussion on this matter is provided in the coming sections.

Arabic nouns classes are lexically specified where the noun class assignment using semantic features is predictable via some basic rules. Hence, gender assignment can be predictable, to some degree, when trying to study the different noun classes

11

linking them with the semantic feature value sets. This study does not include nouns, which are assigned to gender morphologically.

To test the predictiveness of gender assignment on MSA, I have used a supervised machine-learning algorithm (weka). I have used the excel sheet with all the list of nouns, which have been sorted out, to suit the software package Weka (Witten & Frank, 2005). Then, the program does the analysis and produces the results according to the given details and/or information. The following tables are provided to show the work done on the excel sheet before importing it to the software program Weka.

Table 1. *One Example From the Semantic Feature Database*

Noun	Human	Animal	Plant	Inanim.	Shape	Size	Number	Derived	Loan
ʒabal	n	n	n	tool	long	big	sg	n	n

Table 2. *One Example From the Semantic Feature Database*

Noun	Human	Animal	Plant	Inanim.	Shape	Size	Number	Inside	Derived	Loan
ḥaʒar	n	n	n	tool	&	small	sg	n	n	n

Table 3. *One Example From the Semantic Feature Database*

Noun	Human	Animal	Plant	Inanim.	Shape	Size	Number	Inside	Vertical	Derived	Loan
sāq	n	n	n	tool	long	small	sg	n	yes	n	n

7. Results and discussion

As stated earlier, this study suggests a new method to study nouns in MSA. The aim is mainly to test the Arabic nouns' productivity and the predictiveness of grammatical gender assignment. All nouns which show natural gender assignment (e.g. *ʾabb*: "father") are removed trying to reach to very close and/or exact figures in order to come up with a good conclusion. Also, all nouns, which carry any feminine mark indicators like *-t* (-h in English script) as in خَديجَة: *ḫadīʒah*, the short /a/ (*ʾalif ʾalmaqṣurah*) as in سلمى: *Salma* and the long /a:/ (*ʾalif ʾalmadd*) as in حسناء: *ḥasnā*, are removed. All details and the semantic feature value sets which were used are sensitive

to the software. So, I have also removed any noun which cannot have enough information to be entered in the excel sheet, that is, some abstract nouns do not have a complete image in the human's mind as it is unseen and/or may be also unknown. Examples of this type are many but I would rather give the noun *hawā*: "air". This nouns, from the first glance, resemble those feminine nouns which end with the feminine mark indicator /ā/; however, it is masculine and we have less information including the semantic features to add. I would consider this noun as an exception due to the lack of information related such noun. Therefore, these nouns were also removed, (at least at this stage). As it is the case in (Alzahrani, 2019), MSA shows to have a very high percentage of predictiveness, as will be seen in the discussion below.

Alzahrani (2019) has produced some semantic features to classify nouns and then group the collected nouns to predict their genders. These main features are ± [animacy], ± [big] and ± [center]. Results are motivating although they show that grammatical gender assignment is not fully predictable. However, it is something logical and I was expecting so.

MSA data was huge; as being more than 65,000 nouns were tested. This study shows that 62% of MSA can be predicable. To achieve this, a noun should go through different levels in order to assure its gender. At the beginning, it is important to sate again that ± [animacy] feature is one of the easiest to deal with because of the nature of the research. The study, as mentioned earlier, focuses on those nouns which must carry -[animacy] as a feature; otherwise, they can be easily predictable and would add to the results showing high percentage of predictiveness as shown in the following figure:
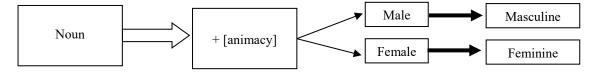
Figure 3. Gender assignment on animate nouns.

Therefore, I have tested all the collected nouns against ± [animacy], ± [big] and ± [center] semantic features. Nouns, which show uncertainty or do not agree with the hypothesis, are test using other features, as will be discussed.
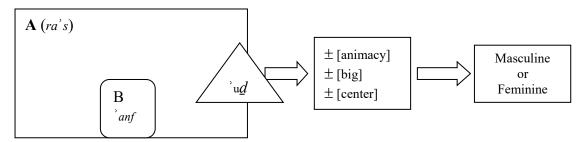

Figure 4. An illustration of the tested semantic features.

Figure 4 shows one of the examples on some nouns, which share some relations and some sematic features. To test them, we need to group them. So, *raʾs*:"head", for example, is + [big] with relation to *ʾanf*: "nose". If we consider the noun *raʾs*:"head" is the domain in this small group (which is given as an example; otherwise, it can be a bigger group containing many related nouns). Thus, it can be generalized that most nouns which are + [big] are masculine; otherwise, they are feminine. That is to say, most nouns - [big] are feminine. Consider the following figure:
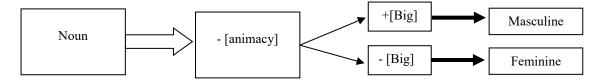

Figure 5. Gender assignment according to size.
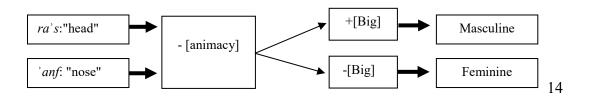

14

Figure 6. An example of gender assignment according to size.

Semantic networks are essential, in this work, to be able to decide which noun is + [big] and which one is -[big], as a group. So, *raʾs*:"head" is + [big] in this semantic network. Then, it is masculine. However, the noun *ʾanf*: "nose" is -[big] in this network but it is still masculine. It is so not because of the -[big] but by adding some more features.

We have introduced ± [animacy], ± [big] semantic features but we still face problem while assigning gender to a noun such as noun *ʾanf*. Hence, another level of semantic features is used to come up with an accurate prediction and limit the chances of wrong predictions. The next feature I used, in this analysis, is ±[Center]. This feature is used for those nouns which have clear and enough information with relation to other group of nouns of the same semantic network.

The noun *ʾanf* is given the feature as -[big] but is still assigned to masculine. However, this noun is +[Center] with relation to the semantic network. So, we suggest that a noun, which is -[big] and +[Center], is assigned to masculine. Other nouns, which are -[big] and -[Center] like *ʾuḏun*: "ear", are assigned to feminine as their gender. However, this does not mean that some nouns show the same features and are assigned to masculine. Yet, they will be tested under more features to find the best analysis why they seem violating the rules.

To sum up, if a noun shows +[big], it should be assigned to masculine regardless it is ± [center], as shown in Figure 8 below. When it is hard to predict correctly its gender using ± [big], we run the test of ± [center] where +[center] is assigned to masculine and -[center] is assigned to feminine, as it is illustrated in Figure 9 below.

Then, if we still have problem with the noun's gender, we need to use another level of semantic features, as will be shown below. Consider the following figures:

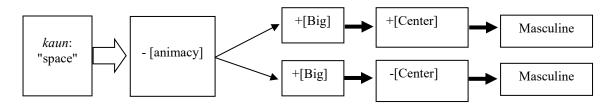*Figure 7.* Gender assignment according to size.

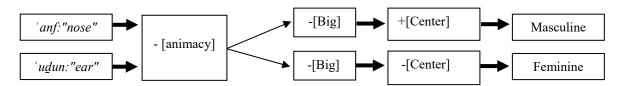*Figure 8.* An example of gender assignment according to size ignoring position/location.

Figure 9. An example of gender assignment according to position/location.

Figure 10 shows +[Big] is the main focus if we compare it with ±[Center]. The overall is masculine. On the other hand, -[Center] is the focus to determine feminine gender, particularly for thos nouns which show +[Big]. Thus, Figure 10 shows that we sometimes need two or more levels (if we consider of animacy) in order to determine nouns' gender in MSA. It is important to reiterate that nouns must be grouped into small groups having relations to assign the different semantic feature value sets to them. Otherwise, the process fails. Table 4 shows a summary gender assignment in MSA considering ±[Big] and ±[Center] features.
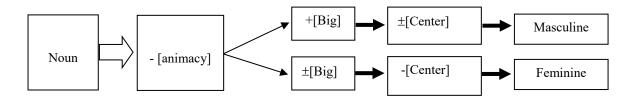
Figure 10. Gender assignment according to position/location to determine feminine.

Table 4. *Gender assignment system in MSA*

	Animacy	Big	Center	Overall
Noun	-	+	+	M
	-	+	-	M
	-	-	+	M
	-	-	-	F

Having covered the two main semantic features (±[Big] and ±[Center]) does not solve the problem with other nouns in MSA which violate the hypothesis showing wrong predictions and may result in wrong statistics. Some nouns in MSA do not match with the previous semantic features due to the lack of the information and/or the limited information, we have. Yet, one more feature ±[inside] is added to the semantic feature value sets. It is created because some MSA nouns show a kind of relation which may or may not be seen but known. For example, the noun *kawn*: 'space' has to contain *samāʾ*: 'sky'. So, both nouns show +[Big] to the human' mind. However, we may claim that *kaun*: 'space' is +[Big] but *samāʾ*: 'sky' is -[Big] with comparison to the domain noun *kaun*: 'space'. Nevertheless, I suggest that *samāʾ*: 'sky' is +[Big] to match what human's mind tells us about the noun separately although I have stated earlier that this

17

study aims to group the nouns into groups in order to assign semantic features to them. Consider the following figure:
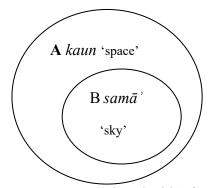

Figure 11. An example ±[inside] feature.

Figure 11 shows that *samāʾ*: 'sky' cannot be to identified as +[Center] nor –[Center] due to the lack of information, as mentioned earlier. Meanwhile, ±[Big] feature is confusing according to the human's mind. Most probably it is given +[Big] to satisfy those who claim it is a big entity. Then, it should be assigned to feminine as its gender. However, I suggest that this noun should go to the next level of labeling using ±[inside] feature. Here, I claim that any noun has +[Inside] should be assigned to feminine as it is gender. Now, putting all the previous features ±[Big] and ±[Center] aside. *samāʾ*: 'sky' is one of the nouns that comes under the noun *kaun*: 'space'. It is one of the wider image and/or nouns which are components of *kaun*: 'space'. So, *samāʾ*: 'sky', which is ±[Big], ±[Center] and +[Inside], is feminine, as the following figure illustrates:
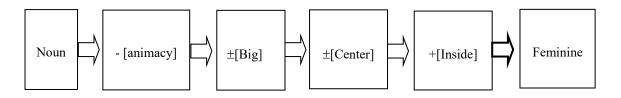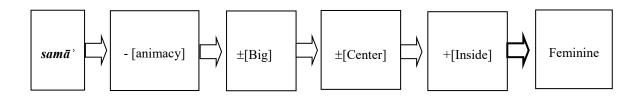

Figure 12. Illustration of +[Inside] feature.

Figure 13. An example of +[inside] feature.

Having said that, one can predict that any noun with -[Inside] is assigned to masculine. For example, *kaukab*: "planet" is –[Inside] when comparing it with *sama:ʔ*: 'sky'. So, it is assigned to masculine, as Figures 14 and 15 show below.
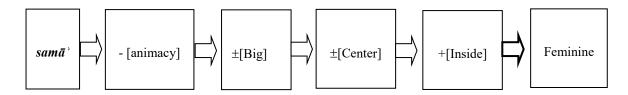

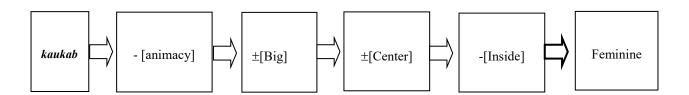*Figure 14.* Illustration of -[Inside] feature

Figure 15. An example of -[Inside] feature

Nouns have gone through two main levels of sorting out and labeling with the three features ±[Big], ±[Center] and ±[inside]. Table 5 below summaries all the discussion about gender assignment system in MSA.

Table 5.*Gender assignment system in MSA with ±[Inside] feature*

	Animacy	Big	Center	Inside	Overall
	-	+	+	+	M
	-	+	-/?	+	F
	-	+	+	-	M
Noun	-	+	-	-	M
	-	-	+	+	M
	-	-	-	+	F
	-	-	+	-	M
	-	-	-	-	F

However, there are some other nouns which are still violating this process of analysis. That is to say, we may find some nouns with +[Inside] but they are assigned to masculine, as their gender. Also, some other nouns show to have -[Inside] feature but they are assigned to feminine, as their gender. Therefore, I suggest another level of semantic feature using ±[Vertical] and ±[Long]. Consider the following figures:

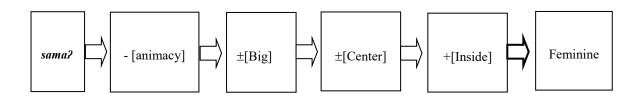*Figure 16.* Illustration of -[Inside] feature

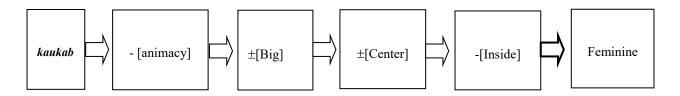| kaukab | ⇒ | - [animacy] | ⇒ | ±[Big] | ⇒ | ±[Center] | ⇒ | -[Inside] | ⇒ | Feminine |

Figure 17. An example of -[Inside] feature

After introducing the main semantic feature value sets, some nouns are found to have some issues while assigning grammatical gender. It is very common and expecting that there remain some nouns whose gender is not assigned using the previous levels and way of analysis. Corbett (1991; 13) calls this "semantic residue".

Thus, I have introduced ±[Long] and ±[Vertical] feature as a solution. Some languages from different family members tend to use shapes of the entities to help classifying nouns, verbs, etc. For example, nouns are divided into four classes using the same way in Jingulu (Pensalfini 2003: 159-68).

Therefore, those nouns, which needed more analysis, were put under these features. Any noun show to have +[Long] and +[Vertical] is assigned to masculine taking into consideration that this noun is not assigned to the correct gender after applying the previous semantic features. An example of such noun is the word ʾuṣbuʿ: "finger". It is both +[Long] and +[Vertical]. So, it can be assigned to masculine as its gender. Some other nouns may show +[Long] and -[Vertical] but they are still assigned to masculine. On the other hand, other nouns which show to have -[Long] and -[Vertical] or -[Long] and +[Vertical] are assigned to feminine as their genders as shown in Table 5 below. Consider the following figures:
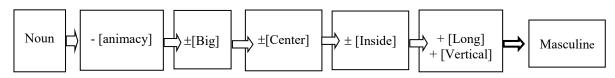
| Noun | ⇒ | - [animacy] | ⇒ | ±[Big] | ⇒ | ±[Center] | ⇒ | ± [Inside] | ⇒ | + [Long] + [Vertical] | ⇒ | Masculine |

Figure 18. Illustration of masculine linked to +[Long] and +[Vertical]

Noun ⇒ - [animacy] ⇒ ±[Big] ⇒ ±[Center] ⇒ ± [Inside] ⇒ + [Long] - [Vertical] ⇒ Masculine

Figure 19. Illustration of masculine linked to +[Long] and +[Vertical]

ba:b "door" ⇒ - [animacy] ⇒ ±[Big] ⇒ ±[Center] ⇒ ± [Inside] ⇒ + [Long] + [Vertical] ⇒ Masculine

Figure 20. Examples of masculine linked to +[Long] and +[Vertical]

qārib "boat" ⇒ - [animacy] ⇒ ±[Big] ⇒ ±[Center] ⇒ ± [Inside] ⇒ + [Long] + [Vertical] ⇒ Masculine

Figure 21. Examples of masculine linked to +[Long] and +[Vertical]

Noun ⇒ - [animacy] ⇒ ±[Big] ⇒ ±[Center] ⇒ ± [Inside] ⇒ - [Long] - [Vertical] ⇒ Feminine

Figure 22. Illustration of feminine linked to +[Long] and +[Vertical]

Noun ⇒ - [animacy] ⇒ ±[Big] ⇒ ±[Center] ⇒ ± [Inside] ⇒ -[Long] +[Vertical] ⇒ Feminine

Figure 23. Illustration of feminine linked to +[Long] and +[Vertical]

ḥawḍ "sink" ⇒ - [animacy] ⇒ ±[Big] ⇒ ±[Center] ⇒ ± [Inside] ⇒ - [Long] - [Vertical] ⇒ Feminine

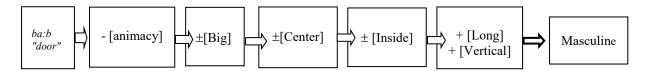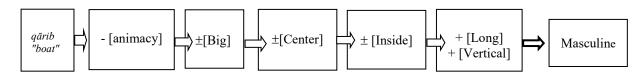*Figure 24.* Examples of feminine linked to +[Long] and +[Vertical]

Figure 25. Examples of feminine linked to +[Long] and +[Vertical]

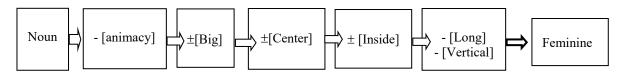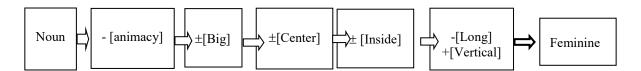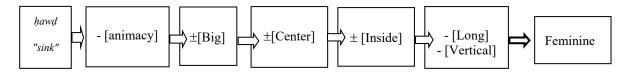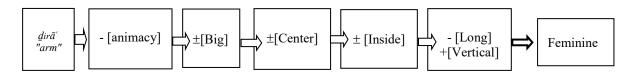It is always important to remember that this does not apply to all nouns in MSA and there might be some exceptions, which need more investigations and explanation. However, this is what this paper has discovered so far. Table 7 shows Gender assignment system in MSA using all the given semantic features, which are given in this study.

Table 6. *Gender assignment system in MSA with +[Long] and +[Vertical] features*

	Animacy	Big	Center	Inside	vertical	long	Overall
Noun	-	+	+	+	+	+	M
	-	+	+	+	-	+	M
	-	+	-	+	-	-	F
	-	+	-	-	+	-	F

Table 7. *Gender assignment system in MSA with ALL features*

Noun	Animacy	Big	Center	Inside	vertical	long	Overall
	-	-	+	+	-	-	F
	-	+	+	+	-	-	F
	-	+	-	+	-	-	F
	-	+	-	-	-	-	F
	-	-	-	+	-	-	F
	-	-	+	-	-	-	F
	-	+	-	-	+	-	F
	-	?	?	?	+	-	F
	-	-	-	-	-	-	F
	-	+	+	+	+	+	M
	-	+	+	-	-	-	M
	-	+	+	-	+	+	M
	-	+	-	-	+	+	M
	-	-	-	+	+	+	M
	-	-	-	-	+	+	M
	-	+	+	+	-	+	M
	-	?	?	?	-	+	M

8. Conclusion and suggestions

Like other languages in the world, MSA show to contain natural gender assignment to nouns such *ʾab*: "father" and *Umm*: "mother, Which is known and recognized through sex. However, not all nouns carry +[animate] feature and can be assigned to gender easily through meaning nor through form. There are some nouns which are -[animate] and there are assigned to gender in MSA with even knowing a reason. Therefore, it is hard to claim that gender assignment, in MSA, is purely and completely arbitrary. Also, it is not possible to claim that grammatical gender assignment is a 100% predictable.

This study attempts to produce a new vision in the way one can analyze different nouns in Arabic. To achieve this study, I have set some phases to take a noun through to determine its gender using some semantic features: ± [animacy], ± [big], ± [center], ±[Inside], ±[Long] and ±[Vertical]. A supervised machine-learning algorithm (weka) is used to analyze the data. The study came up with some interesting findings. One of them is that most masculine noun should show +[Big] feature but it is seen through the research that some nouns violate this rule. Then, ± [center] is introduced to solve the issue. It is seen that most +[Center] nouns are assigned to masculine, with some exceptions which needed ±[Inside] feature to be used. A noun which carry +[Inside] is mostly feminine. Then, I have introduced a table to analyze nouns having these features together. Most nouns with +[Big], -[Inside] and +[Center], +[Big], +[Inside] and +[Center] or +[Big], +[Inside] and -[Center] are assigned to masculine. Some other nouns which are +[Big], -[Inside] and -[Center], -[Big], +[Inside] and -[Center] or -[Big], +[Inside] and +[Center] are assigned to feminine. Nevertheless, some nouns appear to behave differently. Thus, two more features were added to other semantic

features. ±[Long] and ±[Vertical] are used to analyze those nouns which are still hard to predict their genders. It is seen that those nouns should be +[Long] and –Vertical] to be assigned to masculine; otherwise, they are feminine.

Finally, it is good to state that some MSA may show some violation which is obviously expected because this study shows that 62% of MSA can be predicable. There should be further investigations to involve more nouns and might add some more semantic features in order to find a reason for the difficulty of the predictiveness. Also, it is important to say that there are some nouns show that they accept both masculine and feminine as their gender like the noun *sāq*: "leg". Therefore, it can be taken as a reason why this study still showing only 62% as a percentage of the prediction.

References

Aikhenvald, A.Y. (2000). *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Aitchison, J. (1987). *The language lifegame*: *Prediction, explanation and linguistic change*. In W. Koopman, F. van der Leek, O. Fischer, and R. Eaton (Eds.), *Explanation and language change* (pp. 11-32). Amsterdam: John Benjamins.

Alanbaari, A. (1970). *albalaaghah fe alfarq bein almuthakar wa almua'nath. Review: Ramadan* AbduAttuawaab. Egypt, Daar alkutub Publisher.

Alzahrani, S. (2019). Grammatical Gender Assignment on Arabic Nouns: Saudi Dialects. , *US-China Foreign Language Vol. 17* (6), 251-270.

Amaayrah, I. (1993). *Dhaahirat alta'aneith, bein alughah alarabaiah wa allughaat assaamiyah"Dirasah lughawiyah ta'asiliyah*, T:2, Jordan, Daar Hanein publisher.

Anees, I. (1987). Min asraar alughah, T.6, Cairo, Egyptien Anjiloo Library.

Assaraan, M. (1997). Elm alughah muqadimah lilqaari. Ed. 2, Cairo, Daar alfikr alarabi.

Badawi, E., Carter, M., & Gully, A. (2004). *Modern written Arabic: A comprehensive grammar.* London, UK: Routledge.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney and E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3-73). Cambridge, UK: Cambridge University Press.

Bates, E., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1994). *Gender priming in Italian* (Tech. Rep. No. 9402). San Diego, CA: University of California, Center for Research in Language.

Bybee, J. L. (1985). *Morphology*. Amsterdam & Philadelphia: Benjamins.

Bybee, J. L. (1988). Morphology as lexical organization. In M. Hammond and M. Noonan (Eds.), *Theoretical morphology* (pp. 119-141). New York: Academic Press.

Bybee, J. L. (1995a). Diachronic and typological properties of morphology and their implications for representation. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 225-246). Hillsdale (New Jersey): Erlbaum.

Bybee, J. L. (1995b). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425-455.

Bybee, J. L. (1998). The emergent lexicon. *Chicago Linguistic Society, 34*, 421-435.

Brockelmann, C. (1977). Fiqh alughaat assamyiah. Traaaaanslated by *AbduAttawaab, Ramadan*. Riyadh, AlRiyadh University Press.

Castellino, G. (1975). Gender in Cushitic. In J. T. Bynon (Ed.), *Proceedings of a Colloquium Held by the historical section of the Linguistics Association* (pp. 333-359). Great Britain: University of London.

Colé, P., & Segui, J. (1994). Grammatical incongruency and vocabulary types. *Memory & Cognition, 22*(4), 387-394.

Contini-Morava, E. (1994). Noun classification in Swahili. *Research Reports, Second Series*. Publications of the Institute for Advanced Technology in the Humanities, University of Virginia.

Corbett, G. G. (1991). *Gender*. Cambridge, UK: Cambridge University Press.

Corbett, G.G. (2006). Gender, grammatical. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (Vol. 4, pp.749-756). Amsterdam: Elsevier.

Desrochers, A. (1986). Genre grammatical et classification nominale. *Canadian Journal of Psychology, 40*(3), 224-250.

Dixon, R.M.W. (1968). Noun classes. *Lingua, 21*, 104-125.

Greenberg, J. H. (1978). How does a language acquire gender markers? In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Eds.), *Universals of human language: Word structure* (pp. 47-82). Stanford: Stanford University Press.

Grosjean, F., Dommergues, J. Y., Cornu, E., Guillelmon, D., & Besson, C. (1994). The gender-marking effect in spoken word recognition. *Perception & Psychophysics, 56*(5), 590-598.

Harrama, A. M. (1983). Some aspects of negation in modern standard Arabic (Unpublished master's thesis, University of Arizona, Tucson).

Heine, B. (1982). African noun class systems. In H. Seiler and C. Lehmann (Eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen* (Vol. 1, pp. 189-216). Tübingen: Gunter Narr Verlag.

Joseph, V. (1950). Alughah. *Translated by: Abdulhameed Adawaakhli wa Mahammad Aqsaas.* Cairo, Egyptien Anjiloo Library.

Köpcke, K. M., & Zubin, D. A. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen: En Beitrag zur natürlichen Klassifikation. *Linguistische Berichte, 93*, 26-50.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites Stanford* (Vol. 1). California: Stanford University Press.

Lipiński, E. (2001). *Semitic languages outline of a comparative grammar (OLA 80)* (2nd ed.). Leuven: Peeters Publishers.

Ibn Mandhour, M. (2007). *Lisa:n alarab, Almaktabah altawqifiyah.* Cairo. Egypt.

Maʤmaʕ Aluʁat Alarabiyah. (2011). *AL-Muaʤam.AL-Wasˤitˤ.* (5th ed.). Cairo. Egypt.

Mario, P. (1970). "Lughaat albashar usoolha wa tabiatha wa taturha", Cairo, American University.

McMahon, A. M. S. (1994). *Understanding language change.* Cambridge: Cambridge University Press.

Mills, A. E. (1986). *The acquisition of gender: A study of English and German*. Heidelberg/Berlin: Springer-Verlag.

Olstad. J. (2011). *On words and sounds: A selection of papers from the 40th PLM.* Cambridge: Cambridge Scholars Publishing.

Rice, C. (2006). Optimizing gender. *Lingua, 116*, 1394-1417.

Pensalfini, R. (2003). *A Grammar of Jingulu, an Aboriginal Language of the Northern Territory.* The Australian National University, Canberra ACT: Pacific Linguistics.

Ryding, K. C. (2005). *A reference grammar of modern standard Arabic*. Cambridge, UK: Cambridge University Press.

Talafihah, A. ; Alagtash, A. (2014). Mabaadi alarabiyah fe Ttaneeth Alasmaa wa Ttathkeeraha: Nadhrah fe muajahat suabat attaleim. Majalat aladaab wa aluloom alensaaniyah.

Wijnen, F., & Deutsch, W. (1987). The recognition of grammatical gender in Dutch. Paper presented at the *Tagung experimentell arbeitender Psychologen*, April 13, Aachen, Germany.

Witten, I., & Eibe F. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Fransisco. Morgan Kaufman.

Zubin, D. A. (1992). Gender and noun classification. In W. Bright (Ed.), *International encyclopedia of linguistics* (Vol. 2, pp. 41-43). Oxford: Oxford University Press.