

Grammatical Gender Assignment on Arabic Nouns: Saudi Dialects

Salih Alzahrani Taif University, Makkah, Saudi Arabia

Gender assignment has been studied in many languages and the assignment system is considered important particularly when it is related to determining the structure of the lexicon. Semitic languages show two genders, namely masculine and feminine. Nevertheless, it has been claimed that some nouns in Semitic languages can have common gender. Some linguists suggest number of criteria based on the classifications for number and gender in Semitic languages: masculine/feminine, animate/inanimate, human/non-human, individual/collective, concrete/abstract, singular/plural, and major/minor. Arabic, like other Semitic languages, determines gender of its nouns either through meaning or through form. So, grammatical gender is a way of analyzing nouns classes with relation to semantic features, like masculine, feminine, animacy, etc. It is a property of individual nouns regardless of their existing referents (if there are). There have been efforts by many researchers, trying to explain the processing of gender in comprehension. This paper discusses how gender is assigned to nouns in Saudi dialects where a supervised machine learning algorithm is used to test the predictiveness of gender in Arabic. The data show interestingly that the presence of some semantic and morphological features helps to predict the gender of nouns. It is not fully predictable but it is the start point for further investigations.

Keywords: Arabic, Saudi dialects, semantics, nouns, grammatical gender assignment

Introduction

This paper discusses how gender is assigned to nouns in Saudi dialects. Saudi dialects have two main kinds of nouns. The first kind is called derived nouns, which include verbal nouns and verbal derivatives. There are also other types of nouns included as derived nouns. They are named according to their meanings after the derivational processes. The second one is non-derived nouns (substantives), which include proper nouns and borrowed nouns. Saudi dialects' nouns inflect for gender, number, or definiteness but not for case.

This study starts providing background about Arabic morphology; nominal morphology followed by discussion on the different types of nouns and noun inflection focusing on gender.

Salih Jamaan Alzahrani, Ph.D., assistant professor, Foreign Language Department, Taif University, Taif University, Makkah, Saudi Arabia.

Background

Morphology is concerned on forming and shaping the rules and processes through which meaningful unit of language is set in word structure. Morphology, in general, is focused on two main parts. Derivational morphology, which studies how words are built, is at one side. On the other side comes inflectional morphology which deals with the interaction with syntax providing more details about categories, like gender, number, and case.

From a derivational perspective, Arabic morphology is generally based on roots. This is true of all word classes: verbs, nouns, and adjectives. A root may contain three consonants or four consonants. They must appear in a specific fixed order and carry the lexical meaning of the word. These roots are known as "trilateral" (involving three consonants) and "quadrilateral" roots (involving four consonants) as shon in the following examples:

(1) a. $/t b \chi / \rightarrow t a b \chi$ "cooking"

b. /hnds/ \rightarrow muhandis "an engineer"

Roots are always discontinuous because vowels are interspersed to form a meaningful stem; however, the sequence of consonants in each root must be invariable. For example, in the root /t $b\chi$ /, /t f/ always comes first, followed by /b/ and then / χ /. Nevertheless, they might appear in other various forms differently.

Section One

Nominal Morphology

Like modern standard Arabic, nouns in Saudi dialects can be divided into two main types. The first comprises derived nouns, including verbal nouns and verbal derivatives. Derived nouns show another type of nouns, which are categorized and labeled according to their meanings after various derivational processes. It includes nouns, such as unit nouns, instant nouns, and diminutive nouns. Substantives (the non-derived nouns) form the second main type of nouns. It contains both proper nouns and borrowed nouns.

Derived nouns. Derived nouns may appear in different types and patterns, depending on the type of root—trilateral or quadrilateral—as it is the base form. They express the notion of the verb and denote an action or a state related to the underlying verb. Consider the following examples:

(2) a. ribħ "benefit"

b. tardamah "translation"

On the other hand, verbal derivatives are derived from verbal roots. They express both the verb and its subject or object. They are participles which function as nouns (Ryding, 2005). They can be either active (3a) or passive participles (3b) and used as nouns, adjectives and adverbs, or even verb substitutes. They can occur in both genders according to their functions and depending on the animacy of their referents (human beings or abstract). Consider the following examples:

(3) a. ka:tib "a writer"

b. maxlu:g "creature"

There are seven different derived nouns named according to their meanings after the derivation processes. They are unit nouns, instant nouns, nouns of place, occupational nouns, instrumental nouns, diminutive nouns, and common nouns. Consider the following examples: (4) a. *na*χ*l*-ah: a palm tree (Unit noun)

b. *be:* **Z**-*ah*: a sale (Instant noun)

c. ma-ktab: office (Nouns of place)

d. rassa:m/-ah: artista (Occupational noun)

e. *mi-fta:*ħ: key (Instrumental noun)

f. xwe:tim: a small ring (Diminutive noun)

g. *Da Dar-ah*: a tree (Common noun)

Non-derived nouns. Non-derived nouns (i.e., substantives) include proper nouns and borrowed nouns as shown in the following examples:

(5) a. *Salim*

b. Sarah

- c. Atheer
- b. bank

As can be seen in the above examples, proper nouns are always definite. They may have feminine names with the feminine marker /-ah/ (5b), but it can appear without it, too (5c). On the other hand, borrowed nouns are "solid-stem nouns" because they cannot be analyzed into roots and/or patterns (Ryding, 2005).

Inflectional Morphology

Inflectional morphology, as mentioned above, presents the various ways of how words express certain grammatical categories in terms of gender, number, case, tense, and definiteness. Tense is out of the scope of this article as it focuses on verbs.

Nominal morphology. Nouns, in Saudi dialects, inflect for gender, number, and definiteness, but not for case. The following sections explain these aspects briefly. However, the focus is on gender inflection. Thus, it is discussed at the end.

Number. Saudi dialects show nouns inflect for singular, dual, and plural, as illustrated below.

Mass nouns. Uncountable nouns usually have one form, as shown in (6a). However, some dialects sometimes show singular and plural forms of some mass nouns, as shown in (6b).

(6) a. *laħam*: meat

b. *laħam-ah*: a piece of meat

Singular. Singular nouns are those countable nouns that may have both dual and plural forms as shown in (7).

(7) madrasah: a school

Dual. Dual nouns are formed by suffixing the dual morphemes /-e:n/, which is used in the masculine, as in (8) or /-ti:n/, which is used for feminine nouns marked with the feminine suffix /-ah/ as in (9). Other feminine nouns which are not marked with the feminine suffix /-ah/ are marked dual with the masculine dual marker /-e:n/, as in (10a). This explains the ungrammatical form because it contains /-te:n/.

Before going any further, it is important to mention that Saudi dialects do not show case marking; therefore, they do not dual morpheme to distinguish case. Consider the following examples:

```
(8) galam-e:n: "two pens"
```

(9) madrasa-te:n: "two schools"

(10) a. *da:r-e:n*: "two houses"

b.^{*} da:r-te:n: "two houses"

Saudi speakers usually use singular and plural—that is, the plural form is used more generally to express dual and plural. Thus, dual is optionally expressed especially when the number is not important in the context. Consider the following examples:

(11) *Sind-ak Sju:n?* have-2SGM eye.PLF "Do you have eyes?"

(12) *Sind-ak Se:n-e:n* have-2SGM eye.DLF "Do you have two eyes?"

Plural. Plural nouns can be either sound plurals or broken plurals. Sound plural nouns do not show any internal changes when plural suffixes are attached. They show the masculine form with /-i:n/ suffix, which refer to male human beings as seen in (13a). They also show the feminine form in which the feminine singular marker /-ah/ is replaced with /-a:t/ as in (13b). Other feminine singular nouns are pluralized differently as they take the broken plural as in (13c)

(13) a. mudarris-i:n: "teachers"

b. mudarris-a:t: "teachers"

c. ni:r-a:n: "fires"

Some masculine nouns which refer to non-human entities are pluralized exceptionally with the feminine plural marker /-a:t/a shown in (14) below:

(14) mat ^fa:r-a:t: "airports"

On the other hand, broken plural nouns have internal changes when plural affixes are attached in which extra consonants may be added to form this kind of plural. Ablaut is an internal vowel change similar to the broken plural (Ryding, 2005, p. 144). Broken plural nouns show various patterns. Consider the following examples:

(15) a. Pa-Shur: "months"

b. Pa-ðru S: "arms"

(16) Pa-bwa:b/ bi:ba:n: "doors"

The above examples shows how two singular nouns with different patterns are pluralized using one pattern as 2a-fhur "months" and 2a- δru "arms" in (15a & 15b). Also, a singular noun can be pluralized using two different patterns as 2a-bwa:b/bi:ba:n "doors" in (16).

Definiteness. Definiteness is marked either by the definite article /*al*-/ as in (17), by having a noun as the first word in "genitive construct" (18) or by suffixing a possessive pronoun to a noun (19).

(17) *al- be:t*

DEF-house.SGM "The house"

(19) *Se:n-ah* al-gawiyah eye.SGM-3SGM.POSS DEF-sharp.SGF "His eye is sharp."

Gender. Nouns in Saudi dialects are distributed as masculine or feminine. However, some nouns can be treated as either masculine or feminine, as will be illustrated below. Gender in Arabic is semantically and morphologically assigned (Cowell, 1964). Semantically, gender falls back on animacy. That is, animate nouns, such as nouns referring to people and/or animals, usually have the grammatical gender corresponding to their natural gender as in (20), but for inanimate nouns, the grammatical gender is largely arbitrary (21).

(20) a. *?ab*: "father" (M)

b. Jamal: "camel" (M)

(21) a. sama: "sky" (F)

b. *3ama*: "camel" (F)

Morphologically, masculine nouns are the default (i.e., unmarked) whereas feminine nouns usually end with the singular feminine marker suffix /-*ah*/ (22). However, a few nouns have covert gender. So, their gender is recognized by agreement sequences (23). For example, you may use $\int a\chi \varsigma$. "person" to refer to male or you may sometimes use it to refer to female. However, syntax can determine its gender. That is, subject verb agreement.

(22) a. na:fið-ah: "window" (F)

b. *kalm-ah*: "a word" (F)

(23) *∫aχş*. "person" (M/F)

Gender is recognized all over the whole structure of a sentence. That is, it is not only nouns that show gender, but also, other constituents such as verbs, adjectives, demonstratives, and adverbs must agree in gender with the corresponding noun. In other words, a language, which has a gender system, must show to involve different agreements ultimately dependent on nouns of different classes (Corbett, 1991).

Masculine nouns. Some masculine nouns refer to male human (24a) beings and male animals (24b). Some other nouns are treated and used as masculine. Consider the following examples:

(24) a. *?a*χ: "brother" (M)

- b. Pasad: "lion" (M)
- c. *ba:b*: "a door" (M)

Feminine nouns. Some feminine nouns refer to female human beings (25a) and female animals (25b). Consider the following examples:

(25) a. *Auxt*: "sister" (F)

b. *na:gah*: "female camel" (F)

The feminine marker suffix /-*ah*/ is sometimes used to derive the feminine form from the masculine nouns. It applies to both female human beings (26a) and female animals (26b), as seen in the following examples:

(26) a. *3ad-ah*: "grandmother" (F)

b. *biss-ah*: "cat" (F)

Another kind of nouns is those nouns which are treated as feminine nouns whether they have corresponding masculine forms (26) or not (27). They may occur with the feminine marker suffix /-ah/(27a). However, they sometimes appear with no feminine marker (27b), as shown in the following examples:

(27) a. madrasah: "a school" (F)

b. *?ar d*: "land" (F)

Feminine nouns can also be determined by their referents and syntactic agreement (28), as mentioned earlier. They always do not carry any feminine marker suffixes. They include body parts (28a) and proper names of cities and/or countries (28b), as shown in the following examples:

(28) a. *sa*:*g*: "leg" (F)

b. Makkah (F)

c. *affams*: "the sun" (F)

The following examples show how some nouns are determined as feminine via syntactic agreement.

(29) af-fams ti-frig kul şabaħ

DEF-sun.SGF 3SGF.IPFV-rise every morning.SGM

(30) "The sun rises every morning."

in-kasar-at as-sa: S-ah.

PASS-break-3SGF DEF-watch-SGF

"The watch was broken."

In the above examples, it is seen that the verb *ti/rig*: "rises" in 0 carries the prefix */ti-/* as an agreement marker to indicate that the subject is feminine. Also, the verb *inkasarat*: "was broken" in 0 has the suffix */-at/* as a feminine agreement marker.

The last type of feminine nouns includes unit nouns and instant nouns which are derived from verbal nouns (31a) and collective nouns (31b) respectively. These nouns end with the feminine marker /-ah/. Consider the following examples:

(31) a. *naħl-ah*: "a bee" (F)

b. *diħk-ah*: "laugh" (F)

Finally, proper names for males and females are included as nouns. They are assigned to either masculine or feminine according to their referents whether the feminine marker /-ah/ exists or does not, as shown below:

(32) a. *Muhammad* (M)

b. *3ami:lah* (F)

с. *ћатгаћ* (М)

Furthermore, it is important to mention that some nouns are treated as both masculine and feminine. The syntactic agreement only shows their gender. Consider the following examples:

(33) *sikki:n*: "a knife" (F/M)

The above noun *sikki:n*: "a knife" does not have markers to show whether it is masculine or feminine. However, subject agreement is the only indicator, as the following example shows:

(34) as-sikki:n ji-gta $\frac{g}{ti}$ -gta $\frac{g}{t}$.

DEF-knife.SGM/F 3SGM/F.IPFV-cut

"The knife cuts."

The verb *ti-gta* $\frac{f}{ji-gta}$ $\frac{f}{s}$ "cuts" may have either $\frac{ji}{ji}$ as a masculine indicator or $\frac{ti}{ji}$ as a feminine indicator. Both forms are grammatically acceptable.

Section two

Introduction

Nouns and gender agreement system has been studied by many linguists. Yet, there is a space to present some studies on languages, which receive less attention or even not been studied. As a start, the author will adopt two definitions and/or explanations for noun class and grammatical gender.

Dixon (1968, p. 105) defined noun class as (reproduced from Contini-Morava, 1994):

• All nouns in the language are divided into a small and closed set of classes signaled by inflectional morphology.

• The class of a noun is obligatorily co-referenced on other elements in the sentence via grammatical agreement.

Corbett (1991, p. 1) suggested a definition for grammatical gender to distinguish it from natural gender.

To understand what linguists mean by "gender", a good starting point is Hockett's definition: "genders are classes of nouns reflected in the behavior of associated words" (1958:231). A language may have two or more such classes of gender. The classification frequently corresponds to a real-world distinction of sex, at least in part, but often too it does not ("gender" derives etymology from Latin genus, via Old French gendre, originally meant "kind" or "sort"). The word "gender" is not used for just a group of nouns but also for the whole category; thus we may say that a particular language has, say, three genders, masculine, feminine and neuter, and that the language has the category of gender.

Ibrahim (1973, p. 11) defined gender as "a syntactic category with sex as the corresponding notional category". Crystal (1987, p. 158) claimed that grammatical gender is a way of analyzing of nouns classes with relation to semantic features, like masculine, feminine, animacy, etc. Franceschina (2005) and Nichols (1992) provided similar definitions.

Therefore, it is important to say that grammatical gender carries a feature apart of agreement system or any other elements in a sentence, which may help determining gender (e.g., the syntactic behavior). It is a property of individual nouns regardless of their existing referents (if there are). There have been efforts by many researchers, trying to explain the processing of gender in comprehension, 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) and McMahon (1994), among many others, had tried to explain how languages use content to turn it into form; however, the historical view is not enough to account for and explain the existence of gender.

On the other side, linguistic nativism and linguistic functionalism are two other approaches to explain the existence of linguistics devices, such as gender in languages. The former focused on biological explanation of the human's innate language instinct, which includes language devices, like tense, case, and gender. They related this to the presence of such devices in unrelated languages. The latter claims that there is related to the human communication, but not a language-specific biological instinct.

Human cognition and emotion provide the basic meanings and communicative intentions that any natural language must encode, together with a universal set of processing constrains that sharply delimit the way that meanings and intensions can be mapped onto a real-time stream of gestures and/or sounds. (Bates & MacWhinney, 1989, p. 6)

258 GRAMMATICAL GENDER ASSIGNMENT ON ARABIC NOUNS: SAUDI DIALECTS

Therefore, unlike linguistic accidentalism, the other two approaches (linguistic nativism and linguistic functionalism) relate gender to its communicative functions. Some linguists including Köpcke and Zubin, (1984); Mills, (1986); Bates and MacWhinney, (1989); Corbett, (1991), among many others, claimed that grammatical gender helps to keep the track of the referents in any context. Some others claims that it helps in building up the structure of sentences and phrases by recognizing the words which can go together to form one projection in the tree diagram of a sentence (Wijnen & Deutsch, 1987; Desrochers, 1986). Other psycholinguists claim that it helps in the word recognition process (Colé & Segui, 1994; Grosjean et al., 1994).

It is clear from the previous studies that the most linguists believe that gender is arbitrary (Zubin, 1992). It is also staed in Bloomfield (1993):

The gender categories of most Indo-European languages...do not agree with anything in the practical world.... There seems to be no practical criterion by which the gender of a noun in German, French, or Latin could be determined. (Bloomfield, 1933, as cited in Köpcke & Zubin, 1984, p. 439)

Corbett (1991) claimed that gender assignment is systematic after covering many languages, trying to study the gender system in those languages.

Nouns may be assigned to genders according to semantic factors or according to a combination of semantic and formal (morphological and phonological) factors. While in some languages the rules are straightforward, in others they appear much less so. Nevertheless, in those languages which have been studied in depth, the gender of at least 85 percent of the nouns can be predicted from information required independ-ently in the lexicon. (Corbett, 1991, p. 68)

It is obvious that grammatical gender has received attention by many researchers for decades. However, there seems to be two main things, which are not yet to be contained in the previous studies. First, the arbitrariness of grammatical gender assignment is the predominant perception for most researchers, forgetting the properties and the semantic features of nouns, which lead to the production of nouns and the prediction of gender. Secondly, Arabic, as far as to the knowledge of the researcher, has not been under any type of study, related to this matter.

The following sections discuss a new method illustrating how gender is assigned on nouns in Saudi dialects. It is an attempt to pave the way to researchers to apply this approach on a wider range of vocabulary and utilize other semantic features to develop the idea.

Grammatical Gender Assignment

As stated before, in Section "Gender", nouns are classified as masculine or feminine in Arabic, including all the Arabic dialects all over the Arab world. However, gender assignment in Arabic has never been studied in terms of cognitive categorization so far. To the knowledge of the researcher, there has not been any study attempts to relate gender assignment to the semantic features and/or the cognitive categorization in Arabic. This paper, therefore, tries to set a kind of a general understanding of how gender is assigned in Arabic with relation to the human's brain processing based on some semantic features a noun may carry.

Gender assignment in other languages. Gender assignment has been studied in other languages and the assignment system is considered important particularly when it is related to determining the structure of the lexicon (Corbett, 1991). Corbett claims that gender assignment is a rule-based system. A number of Indo-European gender systems show regularities, which support the rule-based model. The studies suggest that

existing nouns show gender individually, and the presence of another mechanism for gender assignment on new nouns is necessary.

Rice (2006) linked gender assignment to the constraints-based framework of the Optimality Theory, providing evidence from German. He suggests a general gender assignment hierarchy in German (Rice, 2006, p. 1405).

Corbett suggests another model for gender assignment on Russian nouns. It depends on ordered rules. He claims that gender assignment is connected to the lexicon (Corbett, 1991, p. 14). There should be other information of an item (a noun) in the lexicon, which is needed to assign gender to that item. Corbett proposed a rule-based model for gender assignment. His focus was not part of the study investigating whether this model is part of the native speaker's mental system or not.

This leads us to another task, which needs more focus and research. Native speakers usually assign gender to nouns, which already exist in the language. However, the human brain is also able to assign gender to other nouns, which are new nouns (borrowed words) which come from other languages and may have different gender assignment system. Native speakers assign gender on new nouns depending on the same rules and regularities they use with nouns in their mother tongue languages. Therefore, they use the same existing gender assignment system.

In the traditional theory and the cognitive theory, they lack providing more information about the lexical items (nouns). The former depends on the symbolic rules while the latter is more linked to the analogy between the linguistics items. Some linguists, such as Bybee (1985; 1988; 1995a; 1995b; 1998) proposed a network model which may support the reason behind assigning gender to new nouns naturally. He suggests that there is a connection between different linguistic items in the brain. Thus, the morphological structure of nouns is crucial in the rule-based model. That is, the set of the symbolic rules take an important part of gender assignment. Langacker (1987, pp. 29-46) claimed that the symbolic rules can be established regardless of the existence of other stored information.

On the other hand, there has been a new trend to present gender assignment study in a different way in which various disciplines are needed to analyze data. One famous study shows that using a supervised machine learning algorithm is successful to test the predictiveness of Contini-Morava's analysis (1994). The focus was on Bantu noun class system. Olstad (2011) had applied a similar type of study to test Swahili noun class. Aikhenvald (2000) stresses the importance of semantics features in studying grammatical gender. Grammatical gender's studies extended to a wider understanding via studying other languages other than the Indo-European languages (Dixon, 1968, p. 105).

Languages, which show to contain noun classes depending on semantics features, such as animacy, sex, shape, size, etc, differ on the number of the classes (gender) they have. These noun classes are also known as gender classes. Some languages show to have two genders (masculine and feminine) like Arabic (See Section "Gender Assignment in Semitic Languages"). Some other languages, like German have three genders (masculine, feminine, and neuter). This number can extend to tens of genders, such as Bantu. Therefore, the semantics of these classes is based on some parameters: sex, humanness, rationality, and/or animacy. These features may appear as a language-specific or may come together in a combination of parameters. As this paper focuses on Arabic, the discussion of these similarities and differences is out of the scope.

Gender assignment in Semitic languages. As stated above, languages can contain a number of genders. Some languages show up to twenty different genders, such as Fula (Corbett, 2006, p. 753). Unlike, those languages, Semitic languages show two genders, namely masculine and feminine. Nevertheless, Lipiński (2001, pp. 235-242) claimed that some nouns in Semitic languages can have common gender.

"Semantics features" are crucial in Afroasiatic languages, too. For example, some languages show that feminine is associated with small size and diminutives (Heine, 1982, p. 190). Some linguists suggest number of criteria based on the classifications for number and gender in Semitic languages: masculine/feminine, animate/inanimate, human/non-human, individual/collective, concrete/abstract, singular/plural, and major/minor (Castellino, 1975, p. 354).

Gender assignment in Arabic. Arabic is a Semitic language with more than 200 million speakers (Ryding, 2005). Like all languages, Arabic has been through various changes, and has been shaped to include different varieties.¹ For example, Classical Arabic and Modern Standard Arabic are two varieties, which share similar phonological, morphological and syntactic features (Badawi, Carter, & Gully, 2004; Harrama, 1983; Ryding, 2005), however, they may differ to some degree in the lexical items.

No one can deny that Arabic is known for is rich morphology due to the complicated derivational and inflectional processes. Like other languages, Arabic can determine gender of its nouns either through meaning (See Section "Gender") or through form. However, the challenge is how to produce a model or way to help recognizing gender through a network containing of different information in the speaker's brain.

An example of how Arabic determines gender via meaning is shown in the following example:

(35) a. *?ax*: "brother" (M)

Some nouns refer to male human which is by default will be masculine. On the other hand, some other nouns refer to female human and will be assigned as feminine, as shown in the following examples:

(36) a. *Puxt*: "sister" (F)

On the other side, some Arabic nouns are assigned to masculine or feminine gender by recognizing the gender markers. There are three main feminine markers which appear at the end of nouns. One of these feminine markers is the suffix */-ah/* which is attached to nouns to derive the feminine form out of the masculine (which is the default). Consider the following examples:

(37) a. *3ad*: "grandfather" (M)

b. *3ad-ah*: "grandmother" (F)

Some other nouns have this feminine marker as part of the noun and it does not occur through any derivational process as shown in the following example:

(38) madrasah: "a school" (F)

The other two feminine markers are called *2alif*. It comes in to two different shapes as can be noticed in the following examples:

(39) a. s^saħr**a ?**. "desert" (F)

b. *ħalwa*: "dessert" (F)

Both above examples show *a*? and *a* at the end of the nouns and they are part of the form. They do not have

¹ See, for example, Ryding (2005) for more details about the history of Arabic.

and/or appear in another masculine form. This applies to Arabic including all the Arabic dialects. So, it is the case in Saudi dialects, too. Allen (2001, p. 35) suggested that all the feminine nouns must end with /-t/ attached to the root. Ibrahim (1973) suggested some rules to those nouns which are borrowed from other languages. Some linguists suggest some other rules which apply to number and gender assignment to loan words in Arabic, including Arabic dialects (Smeaton, 1973; Hafez, 1996; Al-Qinai, 2000; Al-Saidat, 2011; Bahumaid, 2015; Poplack, Sayahi, Mourad, & Dion, 2015; Hamdi, 2017).²

Gender assignment in Arabic dialects. Arabic dialects, including Saudi dialects, show similar gender assignment process. That is, they assign gender through meaning (semantics) as in (40a) or through form as in (40b). Consider the following examples:

(40) a. bint: "girl" (F)

b. *wald*: "boy" (M)

(41) da:mi Sah: "university" (F)

However, it is interesting to see some Saudi dialects assign different genders to the same exact noun. Those nouns normally cannot be of the nouns, which rely on animacy or other nouns which are identified through meaning to be assigned to masculine or feminine accurately. Consider the following examples:

Table 1

Examples of different gender assignment to the same noun

Noun	MSA	ZSA	HA
bat ^f n "abdomen"	М	М	F
mat ^f ar "rain"	М	М	<i>ma t ^farah</i> (F)

Examples in Table 1 show that Hijazi Arabic (HA) tends to assign feminine gender to nouns which are masculine in MSA and/or other Saudi dialects. When this applies to loan words, it is noticed that the variation in gender assignment in Saudi dialects expand to include a bigger number of nouns. Hamdi (2017) suggested that this variation in assigning gender to loan words is motivated by either semantic analogy or by the different sense. He also suggests that native speakers tend to assign gender to those loan words depending on the nearest equivalent form in the mother tongue language. Hence, this is one of the reasons behind conducting this project of gender assignment in Arabic. It is not yet clear how our brains process gender assignment as two different Saudi dialects (belongs to Arabic) differ in assigning gender to one noun. This leads us to the next section which discusses the new methods used to figure out the secret behind gender assignment in Arabic regardless of the different dialects.

Gender Assignment: A New Method

This section illustrates the new approach the researcher applied to study gender assignment, specifically, in Saudi Arabic dialects. As far as to the researcher's knowledge, the attempt is purely new and has not been seen in previous studies. The following sections explain in details how this study is conducted.

The data. There have been two stages to collect data. First, the author used two well-known Arabic dictionaries (*AL-Muadam. AL-Was ^sit ^s Arabic* [2011] and *lisa:an alarab* [2007]) to gather as many nouns as possible with no limitations. Then, XL is used to sort out the collected nouns in relation to some semantic features,

² The current study doesn't investigate any loan words. So, it is out of the scope of this paper.

such as size, gender, etc. the nouns which are collected were about 40,000 nouns. At the same time, 30 Saudi speakers (20 males and 10 females) were asked to fill in forms which contain daily life places and/or situations. They need to provide only nouns, for example, nouns they may have in a bedroom. They also have to provide other information, such as gender, size, shape, etc. There were around 10,000 nouns which are basically found in the data gathered from the dictionaries. The aim of this second data collection is to figure out how the Saudi speakers assign gender themselves to nouns regardless of the information provided in the dictionaries. The two different ways show that there are slight differences in gender assignment.

The second stage, the focus was to eliminate those nouns which have their gender assigned naturally such human. Then, the rest of the nouns have been sorted out according to their family. That is, nouns are grouped into semantic network to build up relations. Also, some other nouns are grouped according to their corresponding relations between the words' senses (if applicable). Consider the following figures:



Figure 2.An example of semantic network.

ba:b

The above figures illustrate the method the researcher followed in grouping the nouns before starting the analysis. Figure 1 shows that the noun $\hbar aml$ "pregnancy" should involve the noun $\hbar ami:n$ "fetus". So, it has been recognized to contain relation between the senses in a way or another. Figure 2, on the other hand, shows that the noun *be:t: "house"* should contain *ba:b: "door"* as part of its network. More details are provided in Section "Results and Discussion".

Hypothesis. The main focus, at least in this paper, is to test the productivity of nouns and the predictiveness of the gender assignment system in Arabic. It is assumed that noun classes are lexically specified and the noun class assignment is based on some basic principles. The distribution of nouns over the different classes depends on some semantic features through which it is predictable. This prediction varies according the different noun

classes, the feature value sets and the list of the nouns with given with given classes and semantic features, which help to formalize the rules for gender assignment, at least, on the tested data. It is important to exclude those nouns which are assigned to their genders morphologically by having any feminine gender marker.

Methodology. The data which were collected was huge and therefore the author implements a supervised machine learning algorithm (weka)³ to test the predictiveness of gender assignment on Arabic nouns (Saudi dialects). At the beginning, the author has used some features and values to classify nouns and either isolate them or extend them to contain a higher level in the relation and/or the network. After that, the data was imported to the software package Weka (Witten & Frank, 2005) in order to view if the hypotheses of predication of nouns is valid or not (more details will be discussed in the following sections). The following tables show the semantic features database:

Table 2

One Example From the Semantic Feature Database

Noun	Human	Animal	Plant	Inanim.	Shape	Size	Number	Derived	Loan
ba:b	n	n	n	tool	long	small	sg	n	n

Table 3

One Example From the Semantic Feature Database

Noun	Human	Animal	Plant	Inanim.	Shape	Size	Number	Inside	Derived	Loan
ba:b	n	n	n	tool	long	small	sg	n	n	n

Results and discussion. The hypothesis of this paper states clearly that gender assignment system in Arabic is not purely arbitrary. It can be generalizable in some nouns like those which carry any of the feminine gender markers attached to them. So, a noun which has $/t / and/or /h/^4$ attached to it should be assigned to feminine, such as *tuffaħ*: "an apple". As mentioned before, it should be noted that Arabic has three feminine gender markers⁵.

Therefore, the aim is to reach to a conclusion regarding the predictiveness of gender assignment system in Arabic. The target is (in this paper) on nouns which are abstract and do not have any markers attached to them, depending on some semantic features. Any kind of agreement system is not used to identify and/or assigning gender. Weka (machine learning algorithm) is used to test the predictiveness. Interestingly, the results show that some nouns, which carry some semantic features, can be predictable although it may have some exceptions. That is, it is not fully (100%) predictable.

The test shows that there are three main fundamental features help to predict gender of a noun in Arabic: size, position/location, and animacy.

The data, which were used and tested, show that more than 50% of the tested nouns agree with the hypothesis of the author. For the sake of the easiness of the discussion, the author introduces the following features, as tools, to test the pedictiveness of gender:

³ Weka is a collection of machine learning algorithms for data analysis. It has many functions, which can be used for many purposes, such as nouns' classifications.

⁴ It should be stated that $\frac{-h}{\text{ or } -t}$ sometimes appear not only as a suffix but also as part of the stem.

⁵ See Section "Gender Assignment in Arabic", for more details and examples.



Figure 3. An illustration of the tested semantic features.

The above features are used separately and then together in order to come up with a final model for gender assignment system in Arabic. Firstly, it is obvious that the discussion is all about those nouns which carry -[animacy]; otherwise, gender assignment is fully predictable. Consider the following figure:



Figure 4. Gender assignment on animate nouns.

The data show that each noun which is +[Big] is masculine; otherwise, it is feminine. So, we can say that most nouns which are small are feminine, as illustrated in Figure 5 below:



Figure 5. Gender assignment according to size.

The question is that how we identify size as being +[Big] or not. To solve this problem, it is crucial to say that we rely on semantic network to label nouns as +[Big]. So, one can say *be:t* "a house" is +[Big]. Then, we may assign masculine as its gender. However, is it the case for nouns, such as *ba:b* "a door"? The answer is that we rely on a deeper understanding of the network. That is, the noun *ba:b* "a door" should come under the noun *be:t* "a house" in our brain mapping of the semantic relations between these two nouns. Thus, the top one (*be:t* "a house") in the semantic network is masculine. Nevertheless, the other nouns may be assigned to masculine or feminine according to their relations to other nouns in the semantic network and/or to their positions, as will be explain below.

The last semantic feature we use is \pm [Center]. This feature is used to modify those nouns which appear to have masculine/feminine gender having +[Big] as a feature. That is, if the noun is +[Big] but it is at the same time is assigned to feminine gender. Alternatively, other nouns may appear carrying -[Big] and they are assigned to masculine gender. Thus, we rely on \pm [Center] feature to determine the gender of such nouns. Consider the following figures:



Figure 7. Gender assignment according to position/location.

Figure 6 shows that +[Big] controls gender assignment in Arabic nouns. So, +[Big] nouns are assigned to masculine regardless if they are ±[Center]. Figure 7, on the other hand, shows clearly that a noun which is -[Big] requires position/location (with relation to other set of nouns which it belongs to in the semantic network) (±[Center]) in order to be able to decide its gender. So, -[Big] plus +[Center] results in masculine. For example, the noun ra?s "a head" is consider to be -[Big] with relation to dism "body"; however, it appears to be +[Center] with relation to other nouns in the network (dism "body"). Then, we may easily assign it to masculine as its gender. On the other side, -[Big] plus -[Center] results in feminine. For example, the same noun ra?s "a head" is -[Big] with relation to dism "body"). Then, we may easily assign it to masculine as its gender. On the other side, -[Big] plus -[Center] results in feminine. For example, the same noun ra?s "a head" is -[Big] with relation to dism "body". The noun se:n "an eye" is -[Big] with relation to both ra?s "a head" and dism "body"; however, it occurs -[Center] with relation to ra?s "a head". Then, it is assigned to feminine as its gender. Consider the following table:

Table 4



Gender Assignment System in Arabic

Figure 8. Gender assignment according to position/location.

Figure 8 seems to be a generalization for the whole idea. However, it is important to state that Figure 8 is produced to cover another set of nouns which look like a violation for the previous rules. It seems that +[Big] is assigned to masculine regardless of \pm [Center] feature while \pm [Big] plus -[Center] is assigned to feminine. We rely on this last figure when we have a set of nouns which requires another feature \pm [inside]. Consider Figure 9:



Figure 9. An example ±[inside] feature.

The above figure explains the importance of \pm [inside] because some nouns might share \pm [Big] and/or \pm [Center]. Nevertheless, they are assigned to masculine or feminine genders. For example, *kawn* "space" is assigned to masculine because it is +[Big]. However, *sama 2*"sky" is assigned to feminine. The reason is that we don't rely on \pm [Big] and/or \pm [Center]. Rather, it is because *sama 2*"sky" is considered +[Inside] with relation to *kawn* "space". Consider the following figure:



Figure 10. Illustration of +[inside] feature.

On the other hand, a noun which appears to have -[Inside] as a feature should be assigned to masculine as its gender as it is the case in *kawn* "space", being the outside of *sama2*"sky". So, B (*sama2*"sky") is an inside part of the A (*kawn* "space") as in Figure 9. Consider the following figure:



Figure 11. Illustration of -[inside] feature.

It is important to mention that we use \pm [Inside] if we cannot determine the noun's gender by size and/or position/location. That is, if a noun seems to violate the rules of size and/or position, we rely on \pm [Inside] feature. Table 5 summarizes gender assignment system in Arabic considering the presence of all the semantic features which are used in analyzing the data.

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

 Table 5

 Gender Assignment System in Arabic With ±[Inside] Feature

Conclusion and Suggestions

Like MSA, Saudi dialects, assign gender through meaning (semantics) as in *bint*: "girl" (F) or through form as in *wald*: "boy" (M). Although some Saudi dialects assign different genders to some nouns, they cannot be of those nouns which are assigned to gender naturally or through meaning as in *mat* ^far "rain".

Although gender assignment, in Arabic, is not purely arbitrary, it cannot be claimed to be fully predictable. This study shows that nouns in Saudi dialects can be assigned to their genders through semantic features. The focus was on some semantic feature set: size, position/location and animacy. The results show that it is possible to rely on these values to decide genders of the tested nouns: \pm [animacy], \pm [big], \pm [center], or \pm [Inside]. Most +[Big] nouns are assigned to masculine if they appear -[Inside] and +[Center]. On the other hand, other nouns which are -[Big] and -[Center] are assigned to feminine as their gender. If it is hard to detect the size and/or the position/location within the noun's semantic network, it is possible to detect it and its gender through their hyponomies utilizing \pm [Inside]. Most +[Inside] nouns are assigned to feminine gender specially when they are -[Cneter] and -[Big]. This study excludes all those nouns which carry any feminine markers (/t/, /h/, /a/, or /2a/. As it is the case in most studies, there are some exceptions which need further studies and deeper investigations. In regard of gender assignment in Arabic, some more semantic features are needed to be tested to come up with some universal rules.

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.
- Allen, J. P. (2001). *Middle Egyptian: An introduction to the language and culture of hieroglyphs.* Cambridge: Cambridge University Press.

Al-Qinai, J. (2000). Morphophonemics of loanwords in Arabic. Studies in the Linguistics Sciences, 30, 2.

Al-Saidat, E. (2011). English loanwords in Jordanian Arabic: Gender and number assignment. Language Forum, 37(1), 59-72.

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

Bahumaid, S. (2015). Lexical borrowing: The case of English loanwords in Hadhrami Arabic. *International Journal of Language and Linguistics*, 2(6), 13-24.

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.
- 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.
- Cowell, M. W. (1964). A reference grammar of Syrian Arabic. Washington, DC: Georgetown University Press.
- 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.
- Crystal, D. (1987). The Cambridge encyclopedia of language. New York: Cambridge University Press.
- 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.
- Franceschina, F. (2005). Fossilized second language grammars: The acquisition of grammatical gender. Amsterdam: John Benjamins.
- 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.
- Hafez, O. M. (1996). *Phonological and morphological integration of loanwords into Egyptian Arabic*. Égypte/Monde Arabe: Les langues en Égypte.
- Hamdi, S. (2017). Number and gender assignment to loanwords in Arabic: Implications from varieties. *International Journal of Linguistics*, 20, 79-88.
- 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.
- Hockett, C. (1958). A Course in modern linguistics. The Macmillan Company: New York.
- Ibrahim, M. H. (1973). Grammatical gender: Its origin and development. The Hague: Mouton.
- 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.
- Loewenstamm, S. E. (1965). The Numerals in Ugaritic. *Proceedings of the International Conference on Semitic Studies*, 19-23 July, Jerusalem, Israel.
- Ibn Mandhour, M. (2007). Lisa:n alarab, Almaktabah altawqifiyah. Cairo. Egypt.
- Madma S Alusat Alarabiyah. (2011). AL-Muadam. AL-Was ^Sit^S. (5th ed.). Cairo. Egypt.
- 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.
- Nichols, J. (1992). Linguistic diversity in space and time. Chicago IL: University of Chicago Press.
- Olstad. J. (2011). On words and sounds: A selection of papers from the 40th PLM. Cambridge: Cambridge Scholars Publishing.
- Poplack, S., Sayahi, L., Mourad, N., & Dion, N. (2015). An exception to the rule? Lone French nouns in Tunisian Arabic. University of Pennsylvania Working Papers in Linguistics, 21(2), 178-186.
- Rice, C. (2006). Optimizing gender. Lingua, 116, 1394-1417.

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

Smeaton, H. (1973). Lexical expansion due to technical change. Bloomington (Indiana): Indiana University Publications.

- 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.

Consonants							
b	Voiced bilabial stop	k	Voiceless velar stop				
m	Bilabial nasal	g	Voiced velar stop				
f	Voiceless labiodental fricative	χ	Voiceless uvular fricative				
θ	Voiceless dental fricative	R	Voiced uvular fricative				
ð	Voiced dental fricative	ħ	Voiceless pharyngeal fricative				
ð٢	Pharyngealized voiced dental fricative	ç	Voiced pharyngeal fricative				
t	Voiceless alveolar stop	h	Voiceless glottal fricative				
t ^٢	Pharyngealized voiceless alveolar stop	5	Voiceless glottal stop				
d	Voiced alveolar stop						
n	Alveolar nasal						
s	Voiceless alveolar fricative	r	Alveolar trill				
z	Voiced alveolar fricative	1	Alveolar lateral approximant				
s٢	Pharyngealized voiceless alveolar fricative						
ſ	Voiceless palato-alveolar fricative	j	Palatal glide				
dz	Voiced palato-alveolar fricative	w	Velar glide				
Vow	Vowels						
i	Short high front unrounded	u	Short high back rounded				
i:	Long high front unrounded	u:	Long high back rounded				
а	Short low front unrounded	e:	Long mid front unrounded				
a:	Long low front unrounded						

Appendix: Transliteration Symbols