

How do Saudi Females Contextualize Male Ancestors? A Corpus-based Investigation

Hisham Alkadi¹, Mohammad Almoaily^{2*}

¹Arabic Linguistics Institute, King Saud University, Riyadh, Saudi Arabia

²Department of English Language and Literature, College of Arts, King Saud University, Riyadh, Saudi Arabia

Corresponding Author: Mohammad Almoaily, E-mail: malmoaily@ksu.edu.sa

ARTICLE INFO

Article history

Received: November 17, 2018

Accepted: February 02, 2019

Published: April 30, 2019

Volume: 10 Issue: 2

Advance access: March 2019

Conflicts of interest: None

Funding: None

Key words:

Corpus,

Newspapers,

Gender,

Females,

Males,

Saudi,

Discourse Analysis

ABSTRACT

The current study attempts to explore the representation of male ancestors in Arabic texts written by Saudi female journalists. These texts appeared in Saudi daily newspapers as daily/weekly columns, which were subsequently put together in a one-million-word corpus. Sketch Engine was used to extract the contexts of eleven Arabic entries for male ancestors, that is, father(s) and grandfather(s), in different linguistic forms (e.g., definite, indefinite, singular, and plural). The total number of tokens for these targeted entries within the corpus was 361, and in each context of these tokens, we determined whether the context was positive, neutral, or negative. Additionally, the contexts were categorized into six groups: upbringing, family relations, guardianship, politics, daily issues, or other. The data show that both the categories of the context and the type of ancestor (father vs. grandfather) can be a determiner for the positive or negative representation of male ancestors within Arabic texts. Overall, there seems to be a negativity towards fathers and positivity towards grandfathers within the corpus. One implication of the study is that the contextualization process employed by female writers tends to be practice-specific instead of portraying a stereotypical picture of male ancestors. A second implication is that the discourse topic seems to be a determiner of the positivity and negativity towards male ancestors. Future research is needed to compare these outcomes with the representation of female ancestors within the same corpus.

INTRODUCTION

Writers can create an image of an individual or a group of people through the way in which these individuals are represented in their writings. The image created by writers can be a negative one, in which the writers demonize people, or a positive one. The representation of people in the media can even shape the way in which certain members or groups within a society are perceived. Therefore, the linguistic representation of society members—especially in the media—is seen as the key to constructing and understanding cultural values (c.f. Hall, 1997; Oumlil, 2017). With this in mind, the aim of this study was to investigate the ways in which male ancestors (namely fathers and grandfathers) are represented in the writings of Saudi female journalists. There are a number of contexts in which the roles of male ancestors are discussed. These contexts include, but are not limited to, upbringing, politics, and guardianship. In each occurrence of the entries, the context can be positive, neutral, or negative. The word *father*, for instance, can be used in the corpus to represent a kindhearted male family member who takes care of his offspring. Contrary to this positive representation, the same

word can be used to denote a disgraceful parent. Conducting a comparison between positive and negative contexts for such entries can be informative in identifying the image of fathers and grandfathers within the writings of Saudi female columnists. This comparison can also be used as an indicator for the representation of male ancestors within Saudi culture.

The one-million-word corpus that was used for data analysis contains texts written by several Saudi female columnists in Saudi daily newspapers. These texts were mostly written in Standard Arabic, and the audience of the columns was the local community (Saudis and foreigners living in Saudi Arabia). The wide range of topics discussed in these articles and the various ideologies of the writers offer great potential for creating a panoramic image of the representation of male ancestors in the corpus.

LITERATURE REVIEW

Seeking Evidence from Corpora

To gain a better understanding of a language and its usage within a society, researchers may analyze the actual applica-

tion of that language (i.e., spoken and written instances) by its speakers. This method of analysis proves to be more informative than relying on one's intuitions (see Baker, 2006; Mooney & Evans, 2015). Hence, research on the characteristics of spontaneous linguistic output—whether written or spoken—has been a subject of interest for researchers since the late 1960s. For example, the project conducted by Kučera and Francis (1967) was one of the earliest attempts at machine-aided linguistic data analysis. In their project, Kučera and Francis analyzed the syntactic differences between the various genres of writing that were found in a one-million-word corpus. Research on “real-world” language flourished in the 1970s, and much of this was owed to the extensive work by sociolinguists such as William Labov (Labov, 1972), Penelope Eckert (1978), Shana Poplack and Sali Tagliamonte (Poplack & Tagliamonte, 2001), and many others. Nevertheless, to devise an accurate analysis of authentic data, researchers need to examine as much data as possible. Thus, without the aid of computer applications specifically designed for the analysis of large texts (potentially exceeding one million words in length), the task of “seeking evidence from corpora” seems to be somewhat unachievable for individual researchers.

According to Sinclair (1991), a lack of the aforementioned tools of computational analysis meant that a thorough analysis of corpora was not achievable until the late 1980s. Corpora are a valuable source for researchers, not just the ones interested in studying spontaneous linguistic data but also those interested in analyzing more sophisticated linguistic output such as formal texts, essays, and journal articles. For instance, researchers can use corpora analysis software to find out the most frequently used words in a text, which may provide ample and insightful information. Examples of this include Shi (2015), who investigated the correlation between word frequency in corpora and bilinguals' performance in the recognition of clinical terms, as well as Okamoto (2015), who investigated whether word frequency in corpora can be a good measure of vocabulary selection in classrooms. Moreover, word frequency may provide useful information concerning the nature of human linguistic interaction and its implications for language teaching.

As noted by Lijffijt et al. (2016), however, words in a corpus are not independent of one another; this raises the need to look at the context in which the frequent words, or word(s) in question, occur. To overcome this issue, corpora analysis software can provide their users with collocations (i.e., co-occurrence of word sequences; c.f. Brown & Palmer, 2015; Kaminski, 2016). Collocation in corpora can answer numerous questions, particularly those related to positivity and negativity. For instance, a researcher might investigate whether a word such as “war” is considered positive or negative in a political corpus by analyzing the sequence that this word collocates with. In addition to collocations, evidence from corpora can be sought using concordances (i.e., the words occurring before and after the word under investigation in a corpus; cf. (Olivar, 2000; Phoocharoensil, 2012).

By using concordance, researchers can investigate various uses of a certain word within different contexts. In the

current study, concordances are used to help identify how Saudi female columnists represent male ancestors in their writing, which were gathered in an approximately one-million-word corpus (refer to Section 3 for more details on the corpus used in this study). What follows is an investigation of the use of corpora in language and gender studies, particularly in determining the a) frequency of certain words, b) what other words they often coexist with, and c) in which contexts they appear.

Word Frequency, Collocation, and Concordance in Language and Gender Studies

The previous section discussed the use of corpora in answering general questions about language and its use. This section conducts a literature review of studies on language and gender, in which various researchers exploited word frequencies, collocations, and concordances to examine the representation of men and women in certain written texts.

In their study concerning the frequency of occurrence of the words “man, men, woman, and women” in three corpora (Brown and LOB, Wellington corpus of New Zealand English, and Freiburg-Brown Corpus of American English), Sigley and Holems (2002) found that the occurrence of the words “woman” and “women” had increased almost twice as much between the years 1960 and 1990. This gradual increase in the use of words for the feminine gender correlated with a gradual decrease in the use of the words for the masculine gender (i.e., “man” and “men”). Further, they found that references made to adult males often tended to be as individuals (using the word man), whereas adult females were more often referred to as a group (i.e., women). One may therefore postulate that this increase in texts can be perceived as an indication of the increased awareness of the need to address issues related to women. Another potential reason for this upturn could be due to the increased discussions about feminism and the feminist movement (see Lewis, Mendenhall, Harwood, & Hunt, 2016; Zimmerman, 2017).

As mentioned in Section 02.1, collocations in corpora can inform readers about positive or negative attitudes that writers intend behind a certain word and about the group of people that they describe by that word. Collocations and concordance can be more precise methods for investigating the representations of both genders in corpora. That is, they say more about the context in which the word appears, which is undoubtedly more informative than just counting the frequency of the target words in a given corpus. Hence, Romaine (2000) examined how collocations can reflect on sexism in language by surveying the words that collocated with the word “spinster” in the British National Corpus; she noticed that this word mostly collocates with negative adjectives such as “gossipy,” “jealous,” and “frustrated.” Moreover, Romaine also found out that words with negative connotations collocate with the words “woman” and “girl” more often than with the words “man” and “boy.” Similarly, Pearse (2008) examined the collocations, as well as the grammatical functions, of the words “man” and “woman” in the British National Corpus, which provided a robust analysis of how men and women were represented in the

corpora. For instance, Pearse found that the word “man” appeared more in the corpus as the subject of verbs that require hard physical activity (e.g., “dig,” “hammer,” “chase,” and “climb”), whereas the word “woman”—as the object of the verb—was mostly used as a recipient of verbs denoting power (e.g., “rape,” “violate,” “limit,” and “marginalize”).

Investigating collocating adjectives is indeed a fascinating method for researchers interested in determining the portrayal of a group of people in a given corpus (e.g., teenagers, men, women, etc.). However, the examination of collocations normally takes into account only two adjacent words. This raises the importance of embracing concordances in the analysis because they enable researchers to analyze the entire context in which the target word(s) appear. This would certainly lead to a more accurate analysis in discourse analysis studies, where the focus would not only be on the coexistence of certain words side by side but also on the entire context in which the word in question appeared. For instance, Carroll and Kowitz (1994) investigated the adjectives used to describe men and women in ESL and EFL textbooks using concordances. They found that adjectives used to describe women were mostly related to their appearance (“beautiful,” “pretty,” and “tall”), whereas adjectives used to describe men included “poor,” “rich,” and “young,” in addition to other adjectives that described their appearance, such as “tall,” “big,” and “fat.”

Arabic texts have received very little attention by corpus linguistics researchers interested in language and gender studies. The only study we are aware of is that of Almujaivel (2016) on feminist and antifeminist discourse in Arabic newspapers. In his study, Almujaivel used a corpus compiled by King Abdulaziz City for Science and Technology that comprised of more than 3,048,447 words appearing in daily Arabic newspapers between the years 1900 and 2013. To take advantage of computer assisted discourse analysis for this large amount of data, Almujaivel used the keywords *المرأة* ‘the woman’ and *النساء* ‘the women’ along with 31n-grams (i.e., 15 words before and 15 words after the target keyword). The 31n-grams of each occurrence from the two aforementioned keywords (i.e., *المرأة* and *النساء*) were compiled in a separate file to assist the researcher in analyzing the context in which they appear (e.g., marital status, domestic violence, education, discrimination against women, niqab/hijab, sexual harassment, etc.). The analysis revealed the existence of both feminist and antifeminist voices in the writings of Arab columnists. For instance, on the topic of marital status, some feminist writers defended the rights of Arab women to manage their own affairs without any constraints, while other writers defended the rights of divorced women. In contrast, there were a number of antifeminist voices that described polygamy as an option for men and argued against the right of women to request divorce. In addition to the concordance analysis, Almujaivel (2016) also listed the collocations (2 n-grams) for the words *المرأة* ‘the woman’ and *الرجل* ‘the man’ in the corpus. The list revealed that the most frequently used word with the word *المرأة* ‘woman’ was “rights,” which appeared 2591 times. Other recurrent collocations in the data were *ضد المرأة* ‘against the

woman’ 2298 times, *دور المرأة* ‘the woman’s role’ 2088 times, and *مشاركة المرأة* ‘participation of the woman’ 1209 times. As for the plural form, the word *النساء* ‘women’ collocates with the words *الرجال* ‘men’ 711 times, *الأطفال* ‘children’ 709 times, and *ضد* ‘against’ 505 times. One may therefore deem this research as highly significant and necessary in this field of study because it seems to be a pioneering attempt to provide a corpus-based analysis of gender-related topics in Arabic newspapers.

Because of the scarcity of research on Arabic corpora, particularly on media texts, the current study attempts to analyze the representations of male ancestors found in the writings of Saudi female columnists. It is important to note that the current study is different from that of Almujaivel (2016), not only in the research question that it seeks to answer (see Section 4 below) but also in data and methodology. First, the current study takes into account the texts written by Saudi female columnists only, unlike Almujaivel, who investigated texts written by female Arab writers in general. As a result, the sample of the current study is more specific. This narrowed down focus is essential, as issues raised and discussed in Arabic newspapers may vary greatly due to different political, social, and financial conditions across the Arab world. Second, the current paper investigates the representation of male ancestors in the corpus using multiple keywords that refer to male adults such as *الأب* ‘the father’ and *الجد* ‘the grandfather’; this is distinct from Almujaivel’s study, which only accounted for the words corresponding to man and woman in Arabic, namely *الرجل* and *المرأة* respectively. Considering such words is essential for a more accurate analysis of the representation of male ancestors in the corpus. More analyses of corpora-based language sexism studies can be found in the literature (de la Ossa, 2016; Fuertes-Olivera, 2007; Goutsos & Fragaki, 2009; Holmes, 2001; Piits, 2016).

THE CORPUS AND DATA ANALYSIS

This study explores the contexts of entries for male ancestors in the writings of Saudi female journalists. The term context here refers to the discourse in which the Arabic words for “father” and “grandfather” appear. Because concordances were utilized within the analysis, we were able to extract and investigate the text in which the entry word(s) appeared. We included eleven entries in our analysis: *والد* and *أب* ‘father’, *الوالد* and *الأب* ‘the father’, *أبي* ‘my father’, *آباء* ‘male parents’, *الآباء* ‘the male parents’, *جد* ‘grandfather’, *الجد* ‘the grandfather’, and *الأجداد* ‘the grandfathers’. For the sake of brevity, any of these entries with affixes attached to them were considered a variation of the same entry¹. For example, *والدها* ‘her father’ is regarded as a variant of the entry *والد* ‘father’. The context of each token of these entries was labelled as either positive, neutral, or negative. Categorizing the context into any of these three categories passed through two stages. First, the authors assessed the context to determine whether the ancestor in question (e.g., the father) was praised, blamed, or neither. Adjectives used to describe the ancestor were also determinants of the positivity, neutrality, or negativity of the text. Next, the categorization accuracy was peer validated. The authors specified the peer-as-

assessment criteria according to which peer assessors should be a) Arabic speakers, b) Saudi Citizens, c) born and living in Saudi, and d) specialized in linguistics. According to these criteria, two fellow Saudi linguists were carefully chosen to examine a randomly selected sample of the authors' analysis (i.e., classification of positivity, neutrality, or negativity). This two-stage process likely ensures that the categorizations are unbiased. Moreover, each context was categorized into one of the six following categories: upbringing, family relations, guardianship, politics, daily issues, or other. In regard to these categories, the difference between upbringing and family relations lies in the type of relation between family members. That is, while any downward interaction within families is described as upbringing, such as the interactions typically between fathers and children, grandfathers and fathers or even grandchildren, any other events in a different direction, whether vertically upward to fathers and grandfathers, or horizontally towards siblings, are all to be labelled and referred to as family relations. As for guardianship, this deals specifically with parents' authority as guardians of their families. Politics deals with contexts in which male ancestors are mentioned in political contexts, whereas daily issues are mentioned in everyday contexts. It should be noted that some instances of the above entries were removed from the analysis, particularly those in which the word did not refer to the genetic father or grandfather. For example, the words *الوالد* and *الأب* 'the father' appeared a number of times as an honorary title for the former Saudi King.

The corpus used in the current study is a collection of Saudi newspaper columns. It includes just over a million words (1,083,396), out of which around 10% (175,073) are unique words. Furthermore, the columns were written by 89 Saudi female columnists in 12 Saudi daily newspapers between 2013 and 2015. The corpus offers great potential for discourse and data analysis, particularly in areas pertaining to the characteristics of the discourse in these texts as well as to how the views of the writers are manifested in their writings. For these reasons, the corpus-based approach seems feasible and viable for fulfilling the aim of this study.

Because Arabic text has its own script features (e.g., letterforms, diacritics, and alhamza²) it was essential that we rely on a capable corpus analyzing tool to handle right-to-left texts with various orthographic characteristics. Alfaifi and Atwell (2016) conducted a comparative evaluation of tools for Arabic corpora search and analysis and found that three tools (namely, Sketch Engine, Khawas [later known as Gawwas], and aConCorde) were the best for supporting Arabic text. Gawwas is a Java program developed in 2012 by Al-Thubaity and his colleagues at different Saudi universities and computer centers (Althubaity & Al-Mazrua, 2014). Though the former has a friendlier interface, we opted for Sketch Engine because of its treatment of Arabic texts in such a large corpus (text processing, which requires high computing ability, is done remotely). Additionally, it offers various features such as word frequency, collocation, and N-gram extraction.

The corpus, as described above, coupled with the invaluable linguistic analytical tool Sketch Engine, allowed us to

address the representation of male ancestors in the corpus by implementing two analytical techniques: keyword list and concordance.

RESULTS AND DISCUSSION

The total number of tokens for male ancestors in the corpus was 361. These tokens occurred in various contexts: upbringing (99 times), family relations (58 times), politics (50 times), guardianship (57 times), and daily issues (70 times). The remaining 27 tokens did not fall under any of the previous labels, so they were subsequently labelled as "other." Out of the eleven entries that were investigated in the current study, the most frequent of them within the corpus was *الأب* 'the father' (106) times, whereas the least frequent was *الجد* 'the grandfather', which appeared only four times. A general trend we found in the data is that singular words for male ancestors were more frequent than plural words (220 for the former and only 141 for the latter). This is evident from Table 1, for instance, where the singular words for male ancestors exist 62 times in the upbringing contexts. Plural words, on the other hand, exist in only 32 contexts. This is in line with Sigley and Holes's (2002) findings discussed in the literature review section above, where males were mostly referred to as individuals.

Table 1 below summarizes the findings of the current study for the first category, "upbringing."

Table 1 reveals a balance between positive and negative contexts related to upbringing (both at around 42%). The occurrence of the entries in neutral contexts, however, is noticeably the lowest, at only 16%. A potential explanation for this is that Saudi columnists have opposing views on the upbringing roles of male ancestors. In turn, their opposing views lead to either discussing positive and virtuous practices (such as proper education, care of children, passing on good morals to younger generations) or negative practices (such as carelessness and bad parenting). Focusing on such issues is more likely to be at the cost of neutral mentions of male ancestors in upbringing contexts. The

Table 1. Male ancestors' upbringing

Entry	Positive context	Neutral context	Negative context
والد 'father'	0	0	0
الوالد 'the father'	3	0	0
أب 'father'	13	3	4
الأب 'the father'	7	6	13
أبي 'my father'	7	1	3
الآباء 'the fathers'	6	3	15
آباء 'fathers'	2	0	5
الجد 'the grandfather'	0	0	0
جد 'grandfather'	1	1	0
أجداد 'grandfathers'	2	1	1
الأجداد 'the grandfathers'	1	1	0
Total (99 tokens)	42 (42.4%)	16 (16.1%)	41 (41.4%)

following examples illustrate the use of ancestors' entries in positive contexts to praise positive practices *انظر إلى علاقة* / انظر إلى علاقة الأب مع أبنائهم يحبهم and in negative contexts to condemn negative parental practices *نشأت البنت في أحضان أمها بعيدا عن الأب*, respectively:

- (1) Inthur ila alaqat al-ab
 ma'a abna'-ih yu-hibbu-hum³
 Look at relation DEF-father
 with children-his 3SG-love-them
 'look at the relationship between the father and his children, he loves them...'
- (2) Nasha'at al-bintu fi ahdani
 ummi-ha ba'idan 'an al-ab
 Grew.up-FEM the-daughter in cradles
 mother-her away from DEF-father
 'the daughter grew up with her mother, away from her father'...

Table 1 also illustrates a noticeable variation in the positivity and negativity across entries. For example, the words *أب* 'father', *الوالد* 'the father', and *أبي* 'my father' are used more in positive contexts, whereas other entries such as *الأب* 'the father' and *الآباء* 'the fathers' are mostly used in negative contexts. The positivity for the word *أبي* 'my father' can be explained by the fact that the female writers who were investigated in the current study are expected to have positive attitudes towards their own fathers. Noticeably, there is a sense of attachment and possession, which is also associated with the possessive pronoun "my". The variation between other entries remains unexplained, possibly due to the relatively low number of tokens per entry.

Another visible trend in Table 1 is that grandfather(s) entries are rarely used in negative contexts (only once). Though we are dealing with small numbers, this could support the claim that was raised earlier that the negativity of upbringing contexts can be due to criticizing modern-day parenthood.

Within the category of family relations, negative contexts for male ancestors seem to be dominant, as illustrated in Table 2, where they number approximately twice as much as in positive contexts (18 vs 34). Neutral contexts, however, number considerably less than the previous two. The argument raised above for the low number of neutral contexts can also be raised here.

When discussing family relations, one is expected to be either positive or negative about the practices of family members. Because the ties between family members does not seem to be as strong as they used to be in the past (c.f. OECD, 2011), the attested negativity in Table 2 is not surprising. For instance, members of extended families in Saudi Arabia used to live together in one house, but this has changed in the last thirty years, and nowadays, most nuclear families live in houses or apartments of their own. In turn, it seems that this social change has led to less compassion for older members of the extended families and created more burden for nuclear families—parents in particular. Other negative practices pertinent to family relations are of fathers mistreating their spouses or children. Hence, discussing such issues has possibly resulted in more negativity for the contexts of the words for male ancestors, as shown in example

(3) below:

- (3) al-ab ya-jidu mut'at-ah ma'a
 Asdiqa-ih
 DEF-father 3SGM-find joy-his with
 friends-his
 'The father only enjoys sitting with his friends.'

Similar to the findings in the upbringing category, Table 3 shows that there is a sense of balance in the use of male ancestors' words in relation to politics. A possible explanation for this is that the writers seem to have opposing views on the political roles of fathers and grandfathers. Table 3 shows that the father(s) entries are mostly used in negative contexts, while the words for grandfathers are mostly used in positive contexts. That is, the total number of positive contexts for fathers' entries is only 8, while that for negative contexts is twice as much (16); however, the positive con-

Table 2. Male ancestors and family relations

Entry	Positive context	Neutral context	Negative context
والد 'father'	1	1	2
الوالد 'the father'	0	0	0
أب 'father'	2	0	6
الأب 'the father'	7	0	14
أبي 'my father'	2	0	1
الآباء 'the fathers'	1	2	6
آباء 'fathers'	1	0	2
الجد 'the grandfather'	1	1	0
جد 'grandfather'	2	2	0
أجداد 'grandfathers'	0	0	3
الأجداد 'the grandfathers'	1	0	0
Total (58 tokens)	18 (31%)	6 (10.3%)	34 (58.6%)

Table 3. Male ancestors and politics

Entry	Positive context	Neutral context	Negative context
والد 'father'	0	0	2
الوالد 'the father'	0	0	0
أب 'father'	2	1	3
الأب 'the father'	0	0	5
أبي 'my father'	1	0	0
الآباء 'the fathers'	3	1	3
آباء 'fathers'	2	0	3
الجد 'the grandfather'	0	0	0
جد 'grandfather'	3	2	0
أجداد 'grandfathers'	9	1	6
الأجداد 'the grandfathers'	3	0	0
Total (50 tokens)	23 (46%)	5 (10%)	22 (44%)

texts of grandfathers' entries are 15, and negative contexts are only 6. The word *الأب* 'the father', for instance, is used five times in negative contexts and has never appeared in positive or neutral contexts, while other words such as *جد* 'grandfather', *أجداد* 'grandfathers', and *الأجداد* 'the grandfathers' are also mostly used in positive contexts. This imbalance of negativity across fathers' and grandfathers' contexts could be due to the positive political roles that grandfathers have played in the establishment of Saudi Arabia, as shown in example (4) below:

(4) *al-athariyat allati shayyad-a-ha al-ajdad*
 DEF-artefacts which create-PST-it DEF.grandfathers
 'Artefacts built by the grandfathers'

A comparison between the percentage of positive guardianship contexts (only 7%) and negative contexts (77%) in Table 4 reveals that the use of ancestors' words in guardianship is more negative than any other category in the current study. This can be explained by the fact that the issue of guardianship in Saudi Arabia is mostly associated with violations of women's or children's rights. Indeed, negative practices such as preventing one's daughter from completing her education or preventing a divorced wife from seeing her children regularly will yield more negativity for male ancestors. This is apparent in the word *الأب* 'the father', which was used 22 times in negative contexts, compared to 4 times in neutral contexts and only once in a positive context. Similarly, the word *الوالد* 'the father' appears 4 times in negative contexts and never appears in neutral or positive contexts. This could be due to the father being mostly blamed for negative guardianship practices. Hence, it is not surprising that none of the words for grandfather appear in the corpus of guardianship contexts. Example (5) below demonstrates how *الأب* 'the father' improper guardianship practices are used in negative contexts:

(5) *al-ab allathi yu-zawwiju*
bint-ah qabl al-sin
 DEF-father who 3SGM-marry
 daughter-his before DEF-age
 'the father, who arranges for the marriage of his under-age daughter'

Table 5 reveals that ancestors' words in relation to daily issues are mostly discussed in negative contexts within discourse, with the entries appearing more than 60% of the time in such contexts. Conversely, they appear as positive and neutral in fewer contexts, both at around 28% of the time. This predominant negativity may be linked to the negative daily practices by male ancestors against vulnerable family members (e.g., children). This is clearly noticed in the entry *الأب* 'the father', which appears 16 times in negative contexts and only twice in positive contexts. Example (6) shows the use of *الأب* 'the father' in negative daily issues' contexts:

(6) *dawr al-ab wa dawr al-amel*
 role DEF-father and role DEF-worker
 '[the house worker performs] the role of the father in addition to his job'

Table 6 summarizes the positivity and negativity for each entry across all categories. That is, the table seeks to compare the positive, neutral, and negative contexts of each en-

Table 4. Male ancestors and guardianships

Entry	Positive context	Neutral context	Negative context
والد 'father'	0	1	0
الوالد 'the father'	0	1	4
أب 'father'	0	0	1
الأب 'the father'	1	4	22
أبي 'my father'	0	0	7
الآباء 'the fathers'	3	2	6
آباء 'fathers'	0	1	4
الجد 'the grandfather'	0	0	0
جد 'grandfather'	0	0	0
أجداد 'grandfathers'	0	0	0
الأجداد 'the grandfathers'	0	0	0
Total (57 tokens)	4 (7%)	9 (16%)	44 (77%)

Table 5. Male ancestors and daily issues

Entry	Positive context	Neutral context	Negative context
والد 'father'	0	0	1
الوالد 'the father'	0	6	1
أب 'father'	1	1	2
الأب 'the father'	2	6	16
أبي 'my father'	2	2	2
الآباء 'the fathers'	3	0	3
آباء 'fathers'	0	1	1
الجد 'the grandfather'	0	0	0
جد 'grandfather'	0	0	0
أجداد 'grandfathers'	6	3	5
الأجداد 'the grandfathers'	6	0	0
Total (70 tokens)	20 (28.5%)	19 (27.15)	31 (61.2%)

try in the entire corpus, regardless of their categories. Note that some percentages are for a very low number of tokens (e.g., *الجد* 'the grandfather' and *والد* 'father').

It is clear from the table above that the fathers' entries mostly appear in negative contexts and that the grandfathers' entries mostly appear in positive contexts. For instance, the word *الأب* 'the father' is used 72 times (67%) in negative contexts, and only 34 times in neutral and positive contexts. Similarly, the word *آباء* 'fathers' appears 16 times (69%) in negative contexts and only 5 times (22%) in positive contexts. In contrast, the word *الأجداد* 'the grandfathers', appears more than 60% of the time in positive contexts and never appears in negative contexts.

Table 6. The positivity and negativity of each entry across the entire corpus

	Positive		Neutral		Negative		Total
	No.	%	No.	%	No.	%	
والد 'father'	3	27	3	27	5	45	11
الوالد 'the father'	3	19	8	50	5	31	16
أب 'father'	18	45	5	12	17	42	40
الأب 'the father'	17	16	17	16	72	67	106
أبي 'my father'	12	40	5	17	13	43	30
الآباء 'the fathers'	16	26	8	13	37	61	61
آباء 'fathers'	5	22	2	9	16	69	23
الجد 'the grandfather'	1	25	2	50	1	25	4
جد 'grandfather'	6	46	5	38	2	15	13
أجداد 'grandfathers'	18	46	6	15	15	38	39
الأجداد 'the grandfathers'	11	61	7	39	0	0	18

The total number of tokens for fathers' entries was 287. Out these tokens, 25% were used in positive contexts, 17% in neutral contexts, and 57% in negative contexts. The contexts of grandfathers' entries were noticeably different, with 49% positive, 27% neutral, and 24% negative. This trend can be linked to Almujaivel's (2016) study, reviewed above, in that feminist voices in the texts written by female authors could have led to negative representations of fathers' roles as family members (see Table 2) and as "guardians" of their female relatives (see Table 4), as well as to negative discussions of fathers' roles in daily issues (see Table 5).

CONCLUSION

The current paper attempted to investigate the representation of male ancestors in the writings of female Saudi columnists in daily newspapers. More specifically, the study aimed to conduct a quantitative analysis of the positivity, neutrality, and negativity of the contexts of the words for male ancestors in a one-million-word corpus. The data analysis revealed that negativity or positivity towards male ancestors mainly stems from the topic discussed by the writer. In upbringing and political contexts, for instance, there was a similarity in the number of occurrences of negative and positive contexts. In other categories, however, male ancestors' entries mostly appeared in negative contexts. The positivity and negativity can also be predicted by the ancestor (i.e., the father or the grandfather). With the exception of أب 'father', the 'father(s)' contexts were mostly negative, whereas the contexts of the 'grandfather' entries were different in the sense that none of them had predominantly negative contexts.

There are many implications for this study. The discourse about fathers in the media can provide insights into their roles as members of society. Negative discourse about fathers' social relations as well as their guardianship of their family members, for instance, is an indication of the fallacy in family relations that requires attention from local social reformers. Indeed, negative discourse surrounding fathers calls for more public and government attention to issues such as domestic violence and bad parenting. Positive discourse

about grandparents, on the other hand, can be indicative of nostalgic discourse about grandparents' era stemming from the writers' aforementioned criticism of fathers. Another implication is that such discourse about male ancestors can be used by social activists, as well as governments, to convince the public of social change. Many of the positive changes in the Saudi context, such as lifting the ban on women driving, could have been more controversial if the public was not fully aware of the wrong practices by some male parents.

We conclude this paper with a recommendation to compare male and female ancestors (i.e., mother and grandmother) in the same corpus. This could reveal different degrees of positivity and negativity across these two groups. Another comparison can also be made between male ancestors and male descendants (i.e., sons and grandsons) because the images of new generations may be different from those created for older generations.

ACKNOWLEDGEMENT

We deeply thank Dr. Abdulmuhsin Althobaity from King Abdulaziz City for Science and Technology and his students for providing us with a copy of their offline corpus, which was used in the current study.

END NOTES

- 1 Words with the definite prefix and the possessive suffix were reported as separate entries because they occur more frequently than other lemmas.
- 2 Alhamza (ء), representing the glottal stop, is arguably regarded as a letter because, unlike diacritics, it is obligatory to write Alhamza in most of its word positions. Contrasting with the letters that are represented by two or three letterforms, Alhamza is represented by more than nine different letterforms as it joins the letters <وا>, <وا>, or <وا>. This status of Alhamza makes it quite difficult sometimes to write the correct letterform and therefore to computationally recognize it.
- 3 Leipzig glossing rules were applied (<https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>).

REFERENCES

- Alfaifi, A., & Atwell, E. (2016). Comparative evaluation of tools for Arabic corpora search and analysis. *International Journal of Speech Technology*, 19(2), 347-357. <https://doi.org/10.1007/s10772-015-9285-5>
- Almujaiwel, S. (2016). المعالجة الآلية للصحف العربية: تحليل الأنماط باستخدام BCU [Arabic Newspaper Processing: Analysing Discourse Patterns Using BCU]. In S. Almujaiwel (Ed.), لغويات المدونة الحاسوبية: تطبيقات تحليلية على العربية الطبيعية [Corpus Linguistics: Analytical Applications on Natural Arabic Language]. (pp. 11-56). Riyadh: KAICA.
- Althubaity, A., & Al-Mazrua, M. (2014). *Khawas: Arabic Corpora Processing Tool USER GUIDE*. Retrieved from <https://sourceforge.net/projects/kacst-acptool/files/?source=navbar>
- Baker, P. (2006). *Using corpora in discourse analysis*. London; New York: Continuum.
- Brown, D. W., & Palmer, C. C. (2015). The phrasal verb in American English: Using corpora to track down historical trends in particle distribution, register variation, and noun collocations. *Studies in the History of the English Language VI: Evidence and Method in Histories of English*, 85, 71-97.
- Carroll, D., & Kowitz, J. (1994). Using concordance techniques to study gender stereotyping in ELT textbooks. In J. Sunderland (ed.) *Exploring Gender Questions and Implications for English Language Education*, 73-83.
- de la Ossa, A. C. (2016). Using corpora to analyze gender. *Discourse Studies*, 18(1), 109-111. <https://doi.org/10.1177/1461445615612668b>
- Fuertes-Olivera, P. A. (2007). A corpus-based view of lexical gender in written business English. *English for Specific Purposes*, 26(2), 219-234. <https://doi.org/10.1016/j.esp.2006.07.001>
- Goutsos, D., & Fragaki, G. (2009). Lexical choices of gender identity in Greek genres: The view from corpora. *Pragmatics*, 19(3), 317-340. <https://doi.org/10.1075/prag.19.3.02gou>
- Hall, S. (1997). *Representation: Cultural representations and signifying practices*. London: SAGE.
- Holmes, J. (2001). A corpus-based view of gender in New Zealand English. In H. B. Marlis Hellinger (Ed.), *Gender Across Languages: The linguistic representation of women and men* (Vol. 1, pp. 115-136). Amsterdam: John Benjamin's Publishing Company. <https://doi.org/10.1075/impact.9.10hol>
- Kamiński, M. (2016). Corpus-based Extraction of Collocations for Near-Synonym Discrimination. *Proceedings of the XVII Euralex International Congress*, 367-374. Lexicography and Linguistic Diversity. Tbilisi: Ivane Javakhishvili Tbilisi State University.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.
- Labov, W. (1972). *Language in the inner city: Studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.
- Lewis, J. A., Mendenhall, R., Harwood, S. A., & Hunt, M. B. (2016). "Ain't I a woman?": Perceived gendered racial microaggressions experienced by black women. *Counseling Psychologist*, 44(5), 758-780. <https://doi.org/10.1177/0011000016641193>
- Lijffijt, J., Nevalainen, T., Saily, T., Papapetrou, P., Puolamaki, K., & Mannila, H. (2016). Significance testing of word frequencies in corpora. *Digital Scholarship in the Humanities*, 31(2), 374-397. <https://doi.org/10.1093/llc/fqu064>
- Mooney, A., & Evans, B. (2015). *Language, society and power: An introduction* (Fourth edition). Abingdon, Oxon, England; New York: Routledge.
- OECD. (2011). Families are changing. *Doing Better for Families*. OECD Publishing. <https://dx.doi.org/10.1787/9789264098732-en>
- Okamoto, M. (2015). Is corpus word frequency a good yardstick for selecting words to teach? Threshold levels for vocabulary selection. *System*, 51, 1-10. <https://doi.org/10.1016/j.system.2015.03.004>
- Olivar, A. (2000). Concordance of the 'Corpus regularum' by Benedict of Aniano. *Studia Monastica*, 42(2), 403-404.
- Oumlil, K. (2017). The representation of women in Moroccan television talk shows. *The Journal of North African Studies*, 23(3), 23. <https://doi.org/10.1080/13629387.2017.1307901>
- Pearse, M. (2008). Investigating the collocational behaviour of man and woman in the BNC using Sketch Engine. *Corpora*, 1(3), 29.
- Phoocharoensil, S. (2012). Language corpora for EFL teachers: An exploration of English grammar through concordance lines. *12th International Educational Technology Conference - IETC 2012*, 64, 507-514. <https://doi.org/10.1016/j.sbspro.2012.11.060>
- Piits, L. (2016). Man's gender and age as based on the collocations of the Estonian word mees 'man'. *Folklore-Electronic Journal of Folklore* 64, 143-158. <https://doi.org/10.7592/FEJF2016.64.piits>
- Poplack, S., & Tagliamonte, S. (2001). *African American English in the diaspora*. Malden, MA: Blackwell.
- Romaine, S. (2000). *Language in society: An introduction to sociolinguistics* (2nd ed.). Oxford; New York: Oxford University Press.
- Shi, L. F. (2015). English word frequency and recognition in bilinguals: Inter-corpus comparison and error analysis. *International Journal of Audiology*, 54(10), 674-681. <https://doi.org/10.3109/14992027.2015.1030509>
- Zimmerman, T. (2017). Intersectionality: The fourth wave feminist Twitter community. *Atlantis-Critical Studies in Gender Culture & Social Justice*, 38(1), 54-70.