

Relativisation across varieties: A corpus analysis of Arabic texts

Abstract

Relative clauses are among the main structures that are used frequently in written texts and everyday conversations. Different studies have been conducted to investigate how relative clauses are used and distributed in corpora. Some studies support the claim that accessibility to relativisation, represented by the Noun Phrase Accessibility Hierarchy (NPAH) which is proposed by KEENAN and COMRIE (1977), predict the distribution of relative clauses in corpora. Other studies found out that discourse functions of relative clauses have an important role in distributing relative clauses in corpora (FOX, 1987). However, little focus has been given to the role of the variety in which relative clauses are written in the distribution of relative clauses in written texts. This study investigates relativisation in Arabic written texts in three varieties: Classical Arabic, Modern Standard Arabic and Iraqi Arabic. A statistical analysis of the results shows that relativisation patterns differ significantly across varieties of the Arabic language and cannot be predicted by one accessibility hierarchy.

1 Introduction

Different studies have been conducted to investigate how relative clauses are used in written and spoken corpora. One of the significant cross-linguistic studies that investigate relativisation in different languages is KEENAN and COMRIE (1977), described as “one of the most influential works in the language universals literature” (Fox, 1987, p. 856). Based on the data of around fifty languages, KEENAN and COMRIE (1977, 1979) state that some grammatical positions are more accessible to relativisation than others, and that accessibility to relativisation follows an implicational hierarchy, the Noun Phrase Accessibility Hierarchy (NPAH¹), which is as follows:

SU > DO > IO > OBL > GEN > OCOMP, where (>) indicates more accessible.

The following English examples are provided to demonstrate the grammatical positions of the NPAH:

SU (subject relative clauses), e.g. the man who bought the book...

DO (direct object relative clause), e.g. the man whom I met...

IO (indirect object relative clause), e.g. the man whom I gave the book to...

OBL (oblique relative clause), e.g. the house which I live in...

GEN (genitive relative clause), e.g. the man whose car is red...

OCOMP (object of comparison), e.g. the man whom I am taller than...

KEENAN and COMRIE presented the NPAH first in 1972, and then, in 1977, they presented the NPAH with a full account of methodological problems and counter examples. According to the NPAH, if a language can relativise only one grammatical position, then that position must be the SU position because it is the most accessible position on the NPAH.

Moving down the NPAH, the difficulty of relativisation increases: IO is more difficult to relativise than DO, and OBL is more difficult to relativise than IO and so on. The OCOMP is the most difficult position to relativise. For example, according to the NPAH, to identify a man, a speaker would rather use SU relative clause “the man who bought the book” than IO relative clause “the man whom I gave the book to” or the OCOMP relative clause “the man whom I am taller than”.

KEENAN (1975) conducted a study to verify the validity of the NPAH by analysing data from written English texts. Two major findings were suggested by that study: First, the NPAH, which is founded on a cross-linguistic basis, “can determine performance constraints within languages” (KEENAN, 1975, p. 147). Second, the SU position is used more frequently in written texts because they are psychologically more accessible to relativisation than the other grammatical positions on the NPAH; accordingly, SU relative clauses are more accessible to comprehension, acquisition and production than other relative clauses. However, the second conclusion has been challenged by FOX (1987), who proposed the Absolute Hypothesis (AH) instead.

FOX (1987), on the basis of the data from English discourse, suggested that accessibility to relativisation in discourse is not determined by the grammatical position of the head noun phrase (henceforth head NP) in the relative clause, rather it is determined by the functions of relative clauses in the text. Instead of what she called ‘the Subject Primacy hypothesis’ (henceforth SPH), which means the subject is more accessible than other grammatical positions to relativisation, she proposed the AH. The AH states that absolute relative clauses: Intransitive Subjects (henceforth ISU, e.g. the man who looks handsome) and DOs (e.g. the man whom I met), are more accessible to relativisation than Transitive Subject (henceforth TSU, e.g. the man who bought the book) relative clauses, because of the discourse functions of the relative clauses “rather than a special cognitive status” (FOX, 1987, p. 869). FOX explains that “Relative clauses serve to situate the referent that is being introduced as a relevant part of ongoing discourse; in a sense they justify the introduction of the referent in the first place” (FOX, 1987, p. 861).

Situating a referent in discourse can be achieved through two strategies: first, providing a static description of the referent; second, linking or anchoring² the referent into discourse through another referent which is well known to the addressee. The function of ISU relative clauses is to provide a description or characterization of the referent, for example, “she is married to this guy who is really quiet” (FOX, 1987, p. 859). On the other hand, the function of TSU and DO relative clauses is to anchor the referent into the text, as in “I know somebody who has her now” and “This man who I have for linguistics is really too much”, respectively (FOX, 1987, p. 859). The DO relative clause links or anchors the head of the relative clause into the context using the SU in the relative clause, as is shown in the example above where the anchor is ‘I’ (in bold type). The TSU relative clause, on the other hand, links the head into the context using the DO of the relative clause, which is ‘her’ in FOX’s example. Noun phrases in the SU position mostly carry given information and tend to be pronominal, so they perform the anchoring function better than noun phrases in the DO position, which usually carry new information.

A number of studies were conducted on the role of the NPAH and the discourse functions of relative clauses in predicting accessibility to relativisation such as JENSEN (1999), GORDON and HENDRICK (2005) and HOGBIN and SONG (2007). However, these studies did not yield similar results. While GORDON and HENDRICK (2005) supported the NPAH,

HOGBIN and SONG (2007) supported the AH and JENSEN (1999) showed that the genre in which the text is written plays a significant role in supporting the NPAH or the AH. Moreover, previous studies were based on the standard varieties of languages, ignoring the differences that might exist between the standard variety and other dialects of a language. In fact, data collected from the standard variety of a language have been found “fairly unrepresentative if compared to the overall picture” (FLEISCHER, 2004, p. 236). This is found to be true in German (FLEISCHER, 2004) and English (KORTMANN, HERRMANN, PIETSCH, & WAGNER, 2005).

Studying the NPAH across varieties of the same language is particularly important in the Arabic language due to the diglossic nature of Arabic. The word diglossia was used by FERGUSON to refer to

“...a relatively stable language situation in which, in addition to the primary dialects of the language (which may include standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety” (FERGUSON, 1959, p. 336).

The coexistence of different varieties, standard and colloquial, of the same language in a community is not enough to result in a diglossic situation; there should be a great gap between formal/ written and colloquial/ spoken (HAMAD, 1992). This gap is found in Arabic.

This study investigates relativisation in three varieties of the Arabic language, Classical Arabic (CA), Modern Standard Arabic (MSA) and Iraqi Arabic (IA), and it compares the data of these varieties to the predictions of the NPAH and the AH. This study aims at answering the following research questions:

1. Which hypothesis, the NPAH or the AH, better predicts the distribution of the relative clauses in Arabic texts?
2. Does the distribution of relative clauses differ from one variety of Arabic into another for the three studied varieties (CA, MSA, IA)?

The rest of this article is organised as follows: in 1.1 an introduction about relativisation in Arabic is presented. Then, the method used in this study is described in section 2. In section 3, results and analysis are produced, which is followed by the discussion and conclusions in sections 4 and 5, respectively.

1.1 Relativisation in Arabic

The relative clause, in Arabic, is a post-nominal clause that is used to modify an item in a way structurally similar to an attributive adjective. Relative markers are used to introduce relative clauses that modify definite heads only (BASHIR, 1982; RYDING, 2005). Hence, when the modified noun is indefinite, no relative marker is used (ABDELGHANY, 2010; SUAIEH, 1980). Relative markers usually shows gender and number agreement with the head of the relative clause as is shown in Table 1.

Table 1: The Relative Markers in Arabic

Number	NOM	ACC	NOM/FEM	ACC/FEM
Singular	allaḍī	allaḍī	allatī	allatī
Dual	allaḍān	allaḍain	allatān	allatain
Plural	allaḍīn	allaḍīn	allawātī	allawātī

As can be observed in Table 1, relative markers are inflected for gender and number. A distinction between nominative case and accusative case only appears with the dual relative marker. However, the agreement indicated by the relative markers in Arabic is different from that in English relative clauses as the relative marker in Arabic agrees with the head's grammatical function in the main clause and not with its grammatical function in the relative clause³.

The relative clauses in the Arabic dialects are similar to those in the standard varieties, MSA and CA, in being post-nominal. However, Arabic dialects differ from CA and MSA in terms of the relative markers they use in relativisation. Relative markers in Arabic dialects are not inflected for gender and number (HOLES, 2004, p. 284). All varieties use the invariable relative pronoun *illi*, or its variants such as (*halli* or *yalli* for Syrian Arabic or sometimes the short form *ill* in the Iraqi dialects⁴) for relativisation in all positions (ALTOMA, 1969; BRUSTAD, 2000; HOLES, 1990, 2004).

2 Methods

Relative clauses are collected from fifteen books, which are written in three different varieties. These books are listed in Table 2. Six CA books are included in this study; the selection of these texts has been done by referring to books that discuss Arabic literary texts such as (JAYYUSI, 2010; ALLEN 1998) in which these texts are discussed as classical works. The second variety from which the other group of texts is collected is MSA or as it is referred to as the "contemporary variant" of CA (CUVALAY-HAAK 1997). Six MSA books are included in this study; these texts are from dates more recent than the CA (1996-2008).

The third variety is IA. Data on IA is collected from three books; all of these books belong to the twentieth century (1972-1988). The reason that only 3 books are included for this variety is that Iraqi Arabic is considered as a spoken variety; therefore, up to the researcher's knowledge, there are no other books that are written in Iraqi Arabic. Furthermore, in these texts, only the conversations between the characters are written in the IA dialect, while the rest is found in MSA, so relative clauses from conversations only are included in this study. That might result in a significantly fewer number of relative clauses in comparison with the other two varieties, yet the statistical method that is used in this study helps in avoiding the consequences of such a difference.

After finishing the data collection, the data are analysed statistically using multi-level Poisson regression analysis. This method of data analysis has been proven to be a good way of analysing textual frequencies (BAAYEN, 2008). By using this method, the effect that differences among texts might have on the results is controlled since texts are considered as a random factor.

Table 2: Texts Included in the Data Collection for this Study

Variety	texts
CA	<ol style="list-style-type: none"> 1. <i>Alf laila wa laila</i> 2. <i>Hayy Ibn Yaḡḡān</i> 3. <i>Maḡāmāt Al-ḡarīb</i> 4. <i>Tārīḡ Al-ḡibarī</i> 5. <i>Tārīḡ Ibn Al-aḡīr</i> 6. <i>Tārīḡ Ibn khalḡūn</i>
MSA	<ol style="list-style-type: none"> 1. <i>Al-ḡarīq 'ilā tall al-muḡrān</i> 2. <i>ḡaḡrīdat al-baḡa 'ah</i> 3. <i>Al-manbūḡ</i> 4. <i>Tārīḡ Al-'arab wa ḡaḡāratihum fi al-Andalus</i> 5. <i>Al-saif wa al-siyāsah fi al-Islam.</i> 6. <i>Tārīḡ Al-'Aarab al-mu'āsir</i>
IA	<ol style="list-style-type: none"> 1. <i>Al-raḡ' al-ba'id</i> 2. <i>Al-naḡlah wa al-ḡīrān</i> 3. <i>Ruba 'iyāt Abu ḡāḡi'</i>

It has been found that CA books, in particular non-fiction books, are quite longer than books written in other varieties. Therefore, to maintain consistency among books in different varieties, only 200 pages are included from each book. This number has been chosen because preliminary results showed that using a lower number of pages did not provide accurate results, where the order of relative clauses is changed dramatically from 20 into 50 pages.

Because the CA is an older variety than the other two varieties, there is a diachronic dimension in the study. However, this study does not focus on the development of the language over time. This study considers CA and MSA as two varieties, as has been done by other linguists such as RYDING (2005), PASHOVA (2002) and VERSTEEGH (2001).

2.1 Relative clauses

The semantic definition that is used by KEENAN and COMRIE (1977) to identify relative clauses will also be used as a basis in this study because. This definition is as follows:

“We consider any syntactic object to be an RC if it specifies a set of objects (perhaps a one-member set) in two steps: a larger set is specified, called the domain of relativization, and then restricted to some subset of which a certain sentence, the restricting sentence, is true. The domain of relativization is expressed in surface structure by the head NP, and the restricting sentence by the restricting clause, which may look more or less like a surface sentence depending on the language” (KEENAN and COMRIE 1977).

The data of this study will include restrictive relative clauses only. Furthermore, structures of relative clauses that are included in this study should have at least one of the following characteristics: first, a relative clause should contain a relative marker, this is only true if the relative clause is definite (see section 1.1); second, a relative clause should contain a verb. Although the first criteria cannot be used to detect indefinite relative clauses, the second one can as is shown in example 1. In the mentioned example, *yukabbiluna* ‘tie-us-up’, is considered a relative clause *yukabbil* for two reasons: first, the clause identifies the noun phrase *hilman* ‘dream’; second, it has a verb ‘tie’.

Counting⁵ relative clauses is the principal method used in this study. Relative clauses in the sample texts are counted and then classified according to the positions of the NPAH⁶. Then the percentage of relative clauses formed on each position is worked out depending on the number of relative clauses found in the texts. This method is chosen because it has been proven to be effective in previous studies, including the two major ones (FOX, 1987; KEENAN, 1975) where the frequency of relative clauses in each position is implemented as a measure of the accessibility of that position to relativisation.

The original hierarchy proposed in KEENAN and COMRIE (1977) is as follows:

1. SU > DO > IO > OBL > GEN > OCOMP

According to KEENAN and COMRIE (1977), the SU position is the most accessible position followed by the DO and the other positions going down the hierarchy, IO, OBL, GEN, OCOMP.

FOX (1987), on the other hand, claims that intransitive subject and direct object are more accessible to relativisation than the transitive subject (refer to section 1). Therefore, the assumption that the SU position is the most accessible position cannot be taken for granted especially in light of other studies that agree with Fox’s claims (e.g. GORDON & HENDRICK, 2005; HOGBIN & SONG, 2007; ROLAND, DICK, & ELMAN, 2007). Thus, in this study, in order to test both the SPH and the AH, SU relative clauses will be further classified into transitive subject relative clauses (TSU) and intransitive subject relative clauses (ISU). Accordingly, hierarchy (1) will be tested as follows:

2. ISU + TSU > DO > IO > OBL > GEN > OCOMP

According to the AH presented by Fox (1987), the predicted hierarchy is:

3. ISU + DO > TSU > IO > OBL > GEN > OCOMP

In this study, both hierarchies (2) and (3) will be considered and the data of this study will show which of these hierarchies is reflected in the distribution of relative clauses in Arabic texts. Examples of relative clauses formed on grammatical positions in hierarchies (2) and (3) are presented in examples (1-6):

TSU

1. la nurīdu ḥilm-an yukabbilu-nā
 not we.want dream-ACC tie up-us
 We do not want a dream that ties us up. [MSA, 2: 249]⁷

ISU

2. šifi-t al-youm wāhid chān yištuḡul wiy-yay bi-l-bank
 saw-I the-day one was work with-me in-the-bank
 I saw today one who was working with me in the bank. [IA, 1: 80]

DO

3. rafaḍ siḡāra-tī allatī qaddam-tu- hā ilai-hi
 refused cigarette-my REL(3.SG.FEM) presented-I-it to-him
 He refused my cigarette that I have presented to him. [MSA, 2: 10]

IO

4. kān awal wāli faraḍ la-hu ra'iatu-hu nafaqāt-hu
 was first governor assigned to-him people-his salary-his
 He was the first governor to whom his citizens assigned a salary. [CA, 5: 306]

OBL

5. fa-ḥaraḡ-tu anā min al-makān allaḡī
 then-went out-I I from the-place REL(3.SG.MAS)
 kun-tu fī-hi sirra
 was-I in-it secretly
 Then I went out secretly from the place in which I was. [MSA, 1: 301]

GEN

6. tawaqaf-tu amām al-manzil allaḡī
 stopped-I in front of the-house REL(3.SG.MAS)
 aš'ala-t Ilene al-nūr fī sālat- i-hi
 turned on- FEM Ilene the-light in lounge-GEN-its
 I stopped in front of the house whose lounge Ilene turned on the light in.
 [MSA, 1: 246]

OCOMP

7. qad nazala bi-nā qawm lam narā
 already came down in-us people not we.see
 qaum qaṭ aḥsana min-hum

people never better than-them

People who we have never seen people better than came to our house...[CA. 4: 103]

4 Results and analysis

The Arabic texts included in this study yielded 2785 relative clauses. ISU (789), DO (749), TSU (601), OBL (475), GEN (161), OCOMP (8), IO (2). TSU+ ISU (1390) are significantly higher than relative clauses in other grammatical positions in the NPAH, which follows the NPAH's predictions, and supports the SPH. However, the IO position is the least frequent position (as is shown in Figure 1); this is counter to the NPAH's order, which is as follows: SU >DO> IO> OBL> GEN> OCOMP.

Thus, the results suggest that there is a gap shown in the IO position since it occurs only twice, although it is the third position in the NPAH. HOGBIN and SONG (2007) revealed similar results and offered two explanations among which the following is found true in the case of Arabic. IO is infrequently used as head NPs in the main clauses in discourse. The infrequent use of indirect object in main clauses in discourse might lead us to the expectation that IO relative clauses would occur infrequently if at all, and this is reflected in the results of this study. The number of relative clauses which have IO heads is only 6 out of 2785. This can be attributed to the fact that indirect object is restricted to the role of beneficiary or recipient and it is also connected with human or animate referents (PALMER, 1994). Because of the small numbers of IO relative clauses (2) and OCOMP (8), these two positions are excluded from the statistical models in this study, as is shown in Table 3.

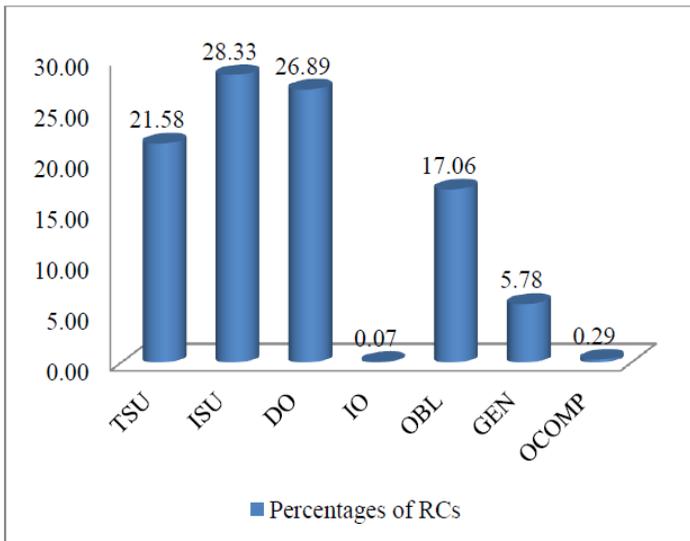


Figure 1: The distribution of relative clauses in Arabic written texts

Table 3: Model 1: Mixed Effect Poisson Regression for the Distribution of Relative Clauses in Arabic Texts Estimate

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.266	0.133	9.500	<0.001 *
TSU	-0.362	0.185	-1.953	0.050 .
DO	-0.036	0.131	-0.279	0.780
OBL	-0.547	0.127	-4.300	< 0.001 *
GEN	-1.573	0.164	-9.577	< 0.001 *
IA	-1.273	0.223	-5.698	< 0.001 *
MSA	-0.279	0.127	-2.206	0.027 *

Model 1 tests two among other predictors: (1) relative clause types with the SU position split into ISU and TSU, (2) variety; these predictors appear in the first column of the table. The dependent variable is the count of relative clauses of the relevant category. Rows 2-5 of Table 3 show a comparison between ISU and other relative clauses on the NPAH. Levels of each predictor are coded by alphabetical order; for example, in the case of the models in this study, all relative clauses on the grammatical positions of the NPAH would be compared to the DO relative clauses since DO comes first in alphabetical order. In this and the following models, alphabetic characters (a, b, c, etc.) are joined to the names of the grammatical positions for ordering purposes. For example, ISU becomes a. ISU, and TSU becomes b. TSU and so on down the NPAH to make the results appear in the order of the positions in the NPAH, (as is shown in Figure 2), which makes the analysis of the results easier. The asterisk (*) in the table indicates that the value is significant, while the dot (.) indicates that the value is approaching significance. Therefore, Model 1 compares relative clauses in all grammatical positions to ISU, as is shown in Figure 2.

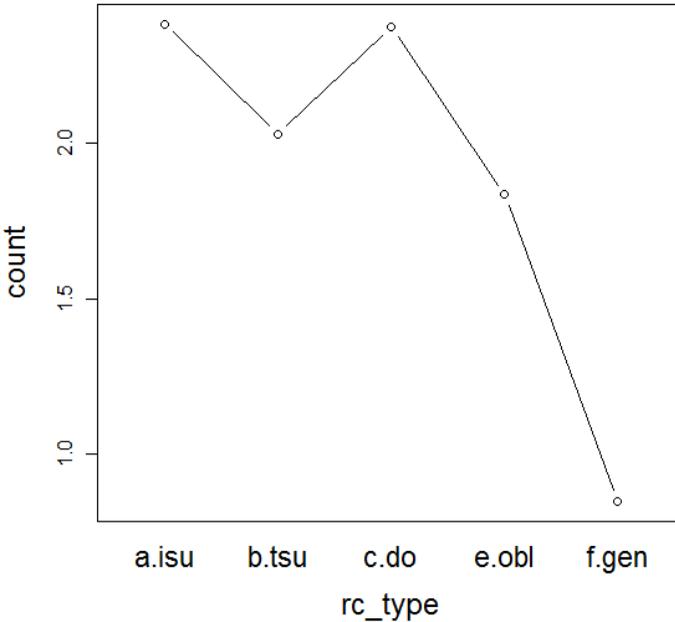


Figure 2: The distribution of relative clauses in Arabic texts

Figure 2 has five points that show the grammatical positions tested in Model 1, ISU, TSU, DO, OBL, GEN. The count axis shows how frequently the relative clauses are used on different positions within the regression model. As is shown in Figure 2, the DO position is close to the ISU position, which is the most frequently used. On the other hand, the DO position is higher than the TSU position with a significant difference (Wald's $z= 3.117$, $p<0.001$)⁸, and there is no statistical difference between TSU and OBL (Wald's $z= 1.128$, $p=0.259$), the difference between the OBL and GEN is significant (Wald's $z= -7.562$, $p<0.001$).

The overall results of this study indicate that ABS relative clauses (ISU+DO) (55.22%) are used more frequently than other relative clauses on the NPAH (TSU, IO, OBL, GEN, and OCOMP). At this stage the results of this study conform to both the SPH and the AH. Therefore, to determine which of these hypotheses the results support more, a comparison is made between the SU category, which includes TSU + ISU and the absolute category, which includes ISU +DO. A model is created in which SU and absolute are treated as two different categories. The results suggest that the number of ABS relative clauses is significantly higher than the number of relative clauses in the SU category (Wald's $z=2.805$, $p=0.005$). This result gives some support to Fox's assertion that "it seems to be the category

ABSOLUTE, rather than SUBJECT, which occupies the leftmost position on the accessibility hierarchy” (FOX, 1987, p. 869) .

4.1 Relativisation across varieties

Three varieties of Arabic are included in this study, CA, MSA, IA. The CA texts reveal 1166 relative clauses, which makes up to 41.86% of all relative clauses in the Arabic corpus, the MSA texts yielded 1451 relative clauses, which make up to 52% of all relative clauses found in Arabic texts, and the IA counts yielded 168 relative clauses, which make up only 6.03% of the data of this study.

To study the influence of variety on the distribution of relative clauses in the text, a model is created to test the interaction between variety and the distribution of relative clauses, as is shown in Table 4. Model 2 tests the interaction between types of relative clauses and variety. In relation to varieties, there is a significant difference between IA and CA (Wald’s $z=-5.279$, <0.001). On the other hand, the difference between MSA and CA does not appear to be significant. The final eight rows show the results of the interaction between variety and relative clauses. There are significant interactions between TSU and IA (Wald’s $z=2.369$, $p=0.018$), and TSU and MSA (Wald’s $z=4.183$, <0.001). There are also significant interactions between DO and IA (Wald’s $z=2.387$, $p=0.017$), DO and MSA (Wald’s $z=2.588$, $p=0.010$). The results suggest that the variety in which relative clauses are written plays an important role in deciding the order of the frequency of relative clauses in the upper grammatical positions (ISU, TSU, DO). The interactions are better shown in Figure 3.

Table 4: Model 2: Mixed Effect Poisson Regression for the Interaction between Relative Clauses and Varieties

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.197	0.141	8.501	<0.001 *
TSU	-0.959	0.188	-5.103	<0.001 *
DO	-0.395	0.159	-2.489	0.013 *
OBL	-0.628	0.192	-3.263	0.001*
GEN	-1.678	0.244	-6.885	<0.001 *
IA	-1.364	0.258	-5.279	<0.001 *
MSA	-0.085	0.165	-0.512	0.609
TSU:IA	0.894	0.377	2.369	0.018 *
DO:IA	0.784	0.328	2.387	0.017 *
OBL:IA	0.077	0.411	0.187	0.852
GEN:IA	0.559	0.502	1.114	0.265
TSU:MSA	1.080	0.258	4.183	<0.001*
DO:MSA	0.574	0.222	2.588	0.010 *
OBL:MSA	0.148	0.272	0.545	0.586

GEN:MSA

0.148

0.345

0.141

0.888

There are three graphs in Figure 3. Each line represents a variety; for each line, there are five points which represent the five grammatical positions tested in this model (ISU, TSU, DO, OBL, GEN). The count axis shows the frequency of relative clauses in each variety. IA (blue) relative clauses are less than the CA (black) and MSA (red). CA and MSA are very close in the ISU, OBL and GEN positions and only differ significantly in the TSU and DO positions. Similarly, IA differs significantly from CA in the three upper positions, ISU, TSU and DO, but there is a slight difference between IA and MSA in the TSU position.

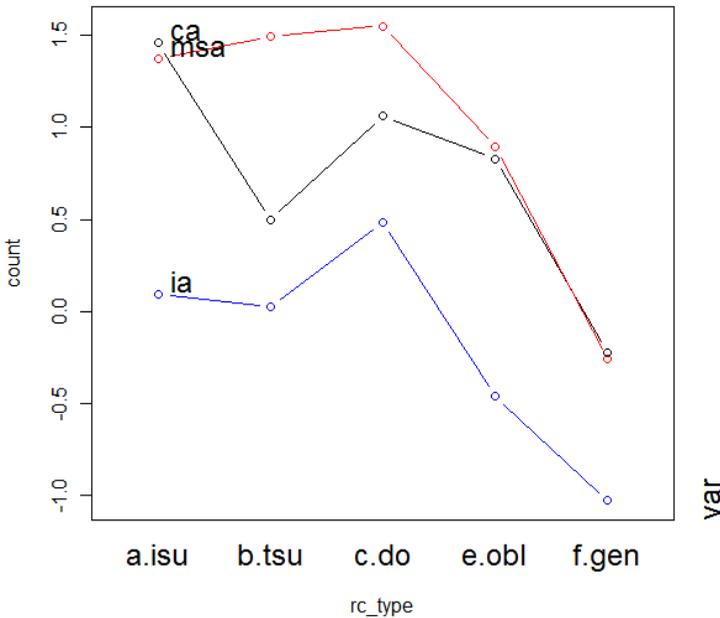


Figure 3: The interaction between relative clauses and varieties

As is shown in Figure 3, the CA line (in black) has the highest point in the ISU position followed by DO, OBL, where there is a slight non-significant difference between the two positions, and then comes the TSU position, which is followed by GEN. To test the difference between SU relative clauses (TSU+ISU) and ABS relative clauses (DO+ISU), a model was created in which TSU and ISU relative clauses were put under SU, and ABS was in-

cluded as a type of relative clauses. The results show that the ABS is significantly more frequent than the SU (Wald's $z = 3.180$, $p = 0.002$) in CA texts.

As is shown in Figure 3, the MSA line (in red) shows that the DO position appears at the highest point which indicates that it has the highest frequency. There are slight differences between ISU, TSU and DO. The OBL and GEN positions are significantly lower than the three upper positions. The difference between ABS and SU relative clauses in MSA is not significant.

Relative clauses in IA texts are found in the following descending order, DO, ISU, TSU, OBL, GEN, as is shown in Figure 3. The differences among relative clauses is found significant only between ISU and GEN (Wald's $z = -2.998$, $p = 0.003$). The ABS relative clauses are used more than SU relative clauses, yet the difference between these two categories is not significant. The differences among the three varieties are better shown in Figure 4, where (\gg)⁹ indicates that the difference is significant, ($>$) indicates that the difference is approaching significance, and ($,$) indicates that the difference is not significant.

The NPAH is not reflected in any of the three varieties considered separately, especially in CA where the frequency of OBL relative clauses is significantly higher than TSU relative clauses. The frequency of SU relative clauses has not been found higher than ABS relative clauses in any of the three varieties. Therefore, the SPH is not supported in the three varieties. For this reason and the infrequent use of IO relative clauses in the three varieties, which is considered as a violation to the hierarchical order of the NPAH, the results do not confirm to KEENAN's (1975) claim that the frequency distribution of relative clauses in texts follows the order of the NPAH.

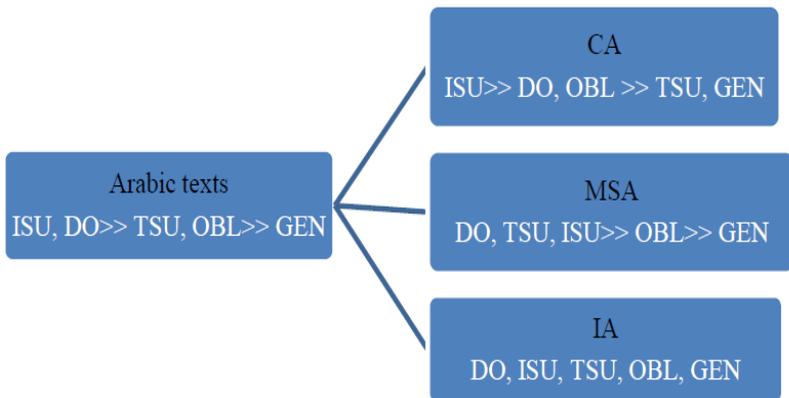


Figure 4: The distribution of relative clauses in the three varieties of Arabic

The results show that there are significant differences among the three varieties of Arabic. ABS relative clauses are used significantly more than SU relative clauses only in CA. Thus, the AH is manifested in the results of CA, which confirms FOX's (1987) claim that it is "the category ABSOLUTE, rather than SUBJECT, which occupies the left-most position on the accessibility hierarchy" (p. 869). The difference between CA and the other two varieties might be attributed to the stylistic changes and the linguistic structures that are used in CA but not in MSA and IA. These stylistic changes take place due to the differences between the chronological periods of CA on one hand and the other two varieties on the other. An example of these linguistic structures is the use of *yuqāl li-* 'said to-', which is always found in the passive form, to give the meaning of 'called' as in example (8). The use of this verb in passive contributes to the high frequency of ISU in CA.

8. kān fī maṣḡid yuqāl la-hu maṣḡid ṣāliḥ
 was in mosque say(PASS) to-it mosque Salih
 He was in a mosque which is called Salih's mosque. [CA, 4: 80]

The distribution of relative clauses differs from one variety to another as is shown in Figure 4. Whereas CA relative clauses appear in the following descending order ISU>> DO, OBL>> TSU, GEN, relative clauses used in MSA have the following descending order DO, TSU, ISU>> OBL>> GEN; and IA relative clauses appear in the following descending order DO, ISU, TSU, OBL, GEN. That is, the order of relative clauses in either CA, MSA or IA is different from the overall order of the relative clauses in the data of this study, which is ISU, DO>> TSU, OBL>> GEN.

CA is the only variety of Arabic that KEENAN and COMRIE (1977) included in their study to represent the Arabic language (p. 76). However, the results of this study show that neither of the varieties can represent the Arabic language because the overall distribution of relative clauses in the data of this study with the distributions of relative clauses in each variety does not reveal similar results. Therefore, CA does not sufficiently represent the Arabic language, which conforms with FLEISCHER's claim that the standard variety is "fairly unrepresentative if compared to the overall picture" (2004, p. 236). Thus, the variety in which relative clauses are written might not contribute to whether relative clauses follow the NPAH or the AH, yet it is an important factor that influences the general distribution of relative clauses in the texts.

5 Conclusions

The overall distribution of relative clauses in Arabic texts conforms with the AH more than the NPAH's predictions. However, looking at varieties of Arabic individually, I have found that each variety of Arabic has revealed a different pattern of relativisation. The distribution of MSA is closer to IA than to CA. IA is different from MSA in two positions only, and both of these varieties differ from CA in four positions. The overall distribution of the whole number of relative clauses in Arabic written texts does not match with any of the distributions revealed by the varieties. Moreover, the AH is reflected in the overall distribution of relative clauses as well as in CA, but not in MSA and IA.

These results seem to suggest two conclusions; first, patterns of relativisation are influenced by the variety in which it occurs. Second, the Arabic language cannot be represented by any single variety; that is, a sample of CA relative clauses is not enough for studying relativisation in Arabic. In general, therefore, these results suggest that it is important to consider different varieties of the same language in deciding accessibility to relativisation in that language. Results also bring up a question of what accessibility in a diglossic situation is. In other words, whether an individual who speaks the three varieties of Arabic has different accessibility hierarchies in his mind, which he uses according to the variety he speaks with. This question can be investigated in future research.

¹ Other abbreviations used in previous studies to refer to the Accessibility Hierarchy is the AH in Song (2001) and NP accessibility hierarchy in Croft (2003)

² Fox adapts the term ‘anchor’ from Prince (1981); “A discourse entity [= ‘referent’ in Fox’s terminology] is anchored if the NP representing it is LINKED, by means of another NP, or ‘anchor’, properly contained in it, to some other discourse entity”

³ There are two clauses in the relative clause sentence: the main clause and the dependent clause, which is the relative clause. For example, the sentence ‘the girls I gave the books to are my friends’ consists of the two clauses: “the girls are my friends” and “I gave the books to the girls”. As a result, the head noun phrase ‘the girls’ has two functions, it is the SU of the main clause, “the girls are my friends”, and at the same time it is the IO of the restrictive clause or dependent clause “I gave the books to the girls”.

⁴ There is more than one dialect in Iraq as the spoken dialect in Baghdad is different from the one spoken in the south of Iraq. However, relative clauses in all dialects of Iraq have the same structure. Therefore, no attempt is made in this study to distinguish among Iraqi dialects.

⁵ Counting is done manually because finding relative clauses in online corpora depends on putting the exact word in the search engine; this can be done with definite relative clauses by putting the relative marker, for example ‘alladī’. However, this is not possible in the case of indefinite relative clauses. Therefore, finding indefinite relative clauses requires reading the whole text.

⁶ In this paper, the way Arabic relative clauses are identified and classified according to the grammatical positions of the NPAH is adopted from (Al- Zagher, 2014).

⁷ Reference to any of the texts is made using the variety abbreviation (e.g. MSA) and the number of the book assigned in Table 2. For example, the reference to Tağridat al-baġa’ah is to be made by using the symbol MSA, 2, this is followed by the page number such as MSA, 2:249

⁸ Since Model 1 compares the grammatical positions to ISU, other models have been created to test whether the difference between other grammatical positions is significant.

⁹ These symbols are used in this figure for the purpose of illustrating the differences among the values as far as the statistical significance, and should not be confused with (>), which is used in the original NPAH.

References

- ABDELGHANY, H. (2010). Prosodic phrasing and modifier attachment in standard Arabic sentence processing. Unpublished doctoral dissertation, City University of New York.
- ALLEN, R. (1998). *The Arabic literary heritage: the development of its genres and criticism*. Cambridge: Cambridge University Press Cambridge.
- AL-ZAGHIR, Z. M. (2014). *Relativisation and Accessibility: A Corpus Analysis of Relative Clauses in Arabic Written Texts*. Doctoral dissertation, University of Otago, Dunedin, New Zealand.
- ALTOMA, S. (1969). *The problem of diglossia in Arabic: A comparative study of classical and Iraqi Arabic*. Cambridge: Harvard University Press.
- BAAYEN, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- BASHIR, S. M. (1982). *A functional approach to Arabic relative clauses with implications for foreign language teaching*. Unpublished Ed.D dissertation, Columbia University Teacher College, New York.
- BRUSTAD, K. (2000). *The syntax of spoken Arabic: A comparative study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects*. Washington D.C.: Georgetown University Press.
- CROFT, W. (2003). *Typology and universals* (2nd ed.). Cambridge: Cambridge University Press.
- EVIATAR, Z., & IBRAHIM, R. (2012). Multilingualism among Israeli Arabs, and the neuropsychology of reading in different languages. In M. LEIKIN, M. SCHWARTZ & Y. TOBIN (Eds.), *Current Issues in Bilingualism: Cognitive and socio-linguistic perspectives* (pp. 57-74). Dordrecht: Springer.
- FERGUSON, C. (1959). Diglossia. *Word- Journal of the International Linguistic Association*, 15(2), 325-340.
- FLEISCHER, J. (2004). A typology of relative clauses in German dialects. In B. Kortmann (Ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective* (pp. 211–243). Berlin: Mouton de Gruyter.
- FOX, B. A. (1987). The noun phrase accessibility hierarchy reinterpreted: Subject primacy or the absolutive hypothesis? *Language*, 63(4), 856-870.
- GAS, S. (1979). Language transfer and universal grammatical relations. *Language Learning*, 29(2), 327-344.
- GORDON, P., & HENDRICK, R. (2005). Relativization, ergativity, and corpus frequency. *Linguistic Inquiry*, 36(3), 456-463.

- HAMAD, A. (1992). Diglossia in Arabic: the beginning and the end. *Islamic Studies*, 339-353.
- HAWKINS, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- HOGBIN, E., & SONG, J. (2007). The accessibility hierarchy in relativisation: The case of eighteenth- and twentieth-century written English narrative. *SKY Journal of Linguistics*, 20, 203-233.
- HOLES, C. (1990). *Gulf Arabic*. London: Routledge.
- HOLES, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Washington, D.C.: Georgetown University Press.
- HYLTENSTAM, K. (1984). The use of typological markedness conditions as predictors in second language acquisition: The case of pronominal copies in relative clauses. In R. ANDERSEN (Ed.), *Second languages: A cross-linguistic perspective* (pp. 39-58). Rowley, Mass: Newbury House.
- JAYYUSI, S. K. (2010). *Classical Arabic stories: An anthology*. New York: Columbia University Press.
- JENSEN, H. B. (1999). *Reduced relative clauses used in eighteenth-century diaries and letters: A sociohistorical perspective*. Unpublished doctoral dissertation, Indiana University, Bloomington.
- JOHNSON, D. E. (1977). On Keenan's definition of "subject of". *Linguistic Inquiry*, 8(4), 673-692.
- KEENAN, E. (1975). Variation in universal grammar. In R. W. Fasold & R. W. Shuy (Eds.), *Analysing variation in language* (pp. 136-148). Washington DC: Georgetown University Press.
- KEENAN, E., & COMRIE, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8(1), 63-99.
- KEENAN, E., & COMRIE, B. (1979). Data on the noun phrase accessibility hierarchy. *Language*, 55(2), 333-351.
- KORTMANN, B., HERRMANN, T., PIETSCH, L., & WAGNER, S. (2005). *Agreement, gender, relative clauses*. Berlin: De Gruyter Mouton.
- MAXWELL, D. N. (1979). Strategies of relativization and NP accessibility. *Language*, 55(2), 352-371.
- PALMER, F. R. (1994). *Grammatical roles and relations*. Cambridge: Cambridge University Press.
- PAVESI, M. (1986). Markedness, discursual modes, and relative clause formation in a formal and an informal context. *Studies in Second Language Acquisition*, 8(01), 38-55.
- PRINCE, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics*. New York: Academic Press.
- ROLAND, D., DICK, F., & ELMAN, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3), 348-379.

- RYDING, K. (2005). *A reference grammar of modern standard Arabic*. Cambridge: Cambridge University Press.
- SAIEGH-HADDAD, E. (2012). Literacy reflexes of Arabic diglossia. In M. Leikin, M. SCHWARTZ & Y. TOBIN (Eds.), *Current issues in bilingualism: Cognitive and socio-linguistic perspectives* (pp. 43-55). Dordrecht: Springer.
- SONG, J. J. (2001). *Linguistic typology: Morphology and syntax*. Harlow: Pearson Education
- SUAIEH, S. I. (1980). *Aspect of Arabic relative clauses: A study of the structure of relative clauses in modern written Arabic*. Doctoral dissertation, Indiana University, Bloomington.
- ZUGHOUL, M. R. (1980). Diglossia in Arabic: investigating solutions. *Anthropological Linguistics*, 201-217.

List of the texts used for the data collection:

- Alf laila wa laila 'Arabian nights'. (1740). Beirut: Al-maktabah Al-sha3biah.
- AL-HARIRI, A. A. (1873). *Maqaamaat Al-hariri*. Beirut: Matba3at Al-ma3aarif
- AL-SAMARRAA'I, K., TAAHA, A.-W., & MATLOUB, N. (2000). *Tariikh al-Aarab wa hadharatihum fi Al-andalus 'History of Arabs and their civilaizations in Andalus'*. Beirut: Dar Al-kitab Al-jadiid.
- AL-SHAYBANI, I. (1231). *Al-kamil fi al-tariikh: Tariikh Ibn Al-athiiir 'The complete history: Ibn Al-athiiir's history'*: International Ideas Home.
- AL-SHAYKH, R. (1996). *Tariikh al-Aarab al-mu'asir 'Contemporary Arab history'*. Cairo: Ein for Human and Social Studies.
- AL-TAKARLI, F. (1980). *Al-raj' al-ba'iid 'The far echo'*. Beirut: Dar Al-mada Lilthaqafah wa Al-finuun.
- AL-TIBARI, M. (838-923). *Tariikh al-umam wa al-muluuk tariikh Al-tibari 'History of nations and kings Al-tibari's history'*: International Ideas Home.
- AL-WARDAANI, S. (1999). *Al-saif wa al-siyaseh fi al-Islam: al-siraa' bain al-Islam al-Nabawi wa al-Islam al-Amawi 'The sword and politics in Islam: The conflict between the Prophet's Islam and the Amawi's Islam'* Beirut: Dar Al-Ra'i.
- AL-YASIRI, S. (1972). *Ruba'iyaat Abu Gaati' 'Abu Gaati's Quadruplet'*. Baghdad: Matba3at Al-sha3b.
- BADR, A. (2005). *Al-tariiq ila tall al-muttraan 'The road to tal al-mutran'*. Beirut: Riad Al-Rayyes Books.
- FURMAAN, G. (1988). *Al-nakhlah wa al-jiraan 'The palm tree and neighbours'*. Baghdad: Dar Al-faraabi/ Dar Babil.
- SA'IID, M. (2008). *Taghriidat al-baja'ah 'The tweet of the swan'*. Beirut: Dar Al-adaab
- TUFAIL, A. (1900). *Hayy Ibn Yaqdhaan Algeria: Fontanah wa Shraka'eh*.
- ZAAAYID, A. (2006). *Al-manbuud 'The castaway'*. Beirut: Al-daar Al-Aarabia lil'uluum-Nashiruun.