

## VARIATION AND RECURRENCE IN THE LEXICAL CHAINS OF ARABIC AND ENGLISH TEXTS

RASOUL AL-KHAFAJI  
*Gothenburg University, Sweden*

### ABSTRACT

The purpose of this paper is to test the validity of two common assumptions as far as the lexis of Arabic and English texts is concerned. The first is that Arabic texts are characterized by an abundance of lexical recurrence; the second is that English texts use more variation than recurrence in their lexis. The paper analyzes a pair of parallel Arabic-English texts and studies their lexical chains for the frequency of occurrence of both variation and recurrence used in them. Moreover, the lexical chains in both texts are closely examined in order to detect, describe, and explain any favorable language-specific tendencies that may be typical of Arabic or English. Both the length of chains as well as the distance of lexical repetition in them is studied for this purpose. One of the above-mentioned assumptions has been supported by the results of data analysis whereas the other has not been substantiated. The paper also arrives at a number of other findings which point to the presence of some significant differences between the linguistic and stylistic preferences of the lexical chains in Arabic and English texts. Such language-specific textual phenomena are believed to be relevant to text-linguistics, translation studies, and language teaching/learning.

### 1. Introduction and research objectives

#### 1.1. Introduction

The use of repetition and the different forms it takes in human language has attracted a lot of attention lately. Suffice it here to refer to the numerous contributions made by text-linguists, sociolinguists, rhetoricians, as well as many others, to the world-conference on 'Repetition in Discourse', which was held at Texas A & M University in May 1990. The comprehensive and varied annotated bibliography prepared by Johnstone and Kirk, which was published in the conference proceedings, presents a good example of the diverse scholarly works and growing interest in this academic field of study (Johnstone 1994). The phenomenon of repetition has been observed and studied both in the language of children and that of adults, in speech as well as

in writing, and in the languages of 'primitive' as well as those of 'advanced' cultures. Repetition can have didactic, playful, emotional, artistic, ritualistic, textual, and rhetorical functions, among others (Johnstone 1994: 6). As for its textual function, repetition is "a central process through which language is created in discourse"; there are even some texts that are "completely organized around patterns of repetition" (Johnstone 1991: 11, 32). In this respect, Winter states that the prime function of repetition in language is its informational value in providing a framework for interpreting what is changed or 'new', by repeating what has already been said in a process which he calls 'repetition and replacement' (Hoey 1991: 20). Similarly, Hoey also maintains that the real significance of the various strategies of repetition in language "lies in their availability as means of connecting sentences, both close to and far off" (Hoey 2001: 41).

The most direct form of repetition in language is repeating a word that has already been used, either exactly in the same form or with some changes. This is called *lexical repetition*. But repetition in text can also be realized in other ways as, for example, by repeating a structure while filling it with new elements. This is called 'parallelism'. Moreover, content, but not form or structure, can be repeated, as in 'paraphrase'. There also exist other means in language for repeating the content or the structure, or both, like 'pro-forms' or 'ellipsis' (Beaugrande and Dressler 1986: 49). The cohesive function of repetition in discourse has been extensively studied in English by Halliday and Hasan (1976: 13) under the categories of 'reference', 'substitution', 'ellipsis', and 'lexical cohesion'. While acknowledging the above textual cohesive categories, Hoey (1991: 10) argues that the relative abundance and high frequency of lexical cohesion, as well as its ability to form long repetition chains, make it "the dominant mode of creating texture". He then concludes that "the study of cohesion in text is to a considerable degree the study of patterns of lexis in text". Consequently, Hoey (1991: 26) asserts that "to a great extent, cohesion is the product of lexical relations (rather than grammatical ones)". Lexical repetition involves using the same or a related *form* of a given lexical item more than once in a given text, and can either be 'simple' or 'complex' lexical repetition as will be defined below.<sup>1</sup>

Though common in human language, lexical repetition may vary both in its frequency as well as its distribution. Certain languages and cultures seem to tolerate lexical repetition, as well as other types of repetition in text, more than others (Johnstone 1994: 16). Arabic is often referred to as a language that makes much more use of repetition, including lexical, than many other languages, like English for example (Johnstone 1994: 11). Repeating the same lexical item many times (i.e., recurrence), whether across sentence boundaries or within the same sentence, has often been singled out by many text-linguists as being a typical characteristic of text in

<sup>1</sup> The terms 'simple' and 'complex' lexical repetition are also used in Hoey (1983) and later Hoey (1991) and Hoey (2001).

Arabic (among others, Williams 1982; Al-Batal 1985; Al-Jabr 1987; Al-Jubouri 1983; Fareh 1988; Johnstone 1991; Hatim and Mason 1997; Al-Khafaji forthcoming). In this respect, Hatim and Mason (1997: 32) point out that "the Arabic text purposefully establishes lexical cohesion via recurrence of the lexical item" and adds that "while recurrence is an option available to users of both Arabic and English, the latter generally see it as a heavily marked form which, to be sustainable, must have some special motivation". By the same token, al-Jubouri (1983: 102) writes that "English discourse rules, codified in rhetoric textbooks under 'variety in word choice', encourage writers to avoid repetition of this sort [i.e., recurrence]. The converse is true in Arabic". Likewise, Stotsky (in Hoey 1991: 243) maintains that in English, "an increase in the use of complex lexical repetition [viz. lexical variation] rather than of simple lexical repetition [lexical recurrence] 'may be an important index of growth'". The assumption made above is then that while both Arabic and English employ lexical repetition in their texts, Arabic favors the 'recurrence' of lexical items whereas English usually opts for lexical 'variation'.

However, the above-mentioned widespread reference to the marked tendency in Arabic to use lexical repetition more frequently, as well as to favor recurrence over variation in its use, has not sufficiently motivated a closer investigation of the chains of lexical repetition so as to examine the *frequency* of recurrence and variation there. Nor have the conditioning factors governing the *distribution* in such chains of 'simple' and 'complex' lexical repetitions, as defined in Section 2 below, been given due attention. Moreover, lexical chains, at least in Arabic, have not been closely inspected in an attempt to detect the favorable and dominant textual trends that distinguish them from those of English. Such are some of the questions that the present study tries to address. It is important, however, to remember before proceeding any further that the purpose of this research paper is *not* to study the frequency of lexical repetition in general, whether in Arabic or in English texts or across the two languages. This has already been dealt with and documented by the many research works just mentioned above. Rather, the specific objective of the present paper is to study the type, frequency and distribution of the *constituents* of lexical repetition represented by simple repetition (recurrence) and complex repetition (variation) in the lexical chains of Arabic and English.

## 1.2. Research objectives

More specifically, the research objectives of the present paper can be formulated as follows:

- (A) To test the validity of the following two assumptions often articulated in the literature:



Assumption 1: Lexical repetition chains in Arabic are characterized by more recurrence than variation.

Assumption 2: Lexical repetition chains in English are marked by more variation than recurrence.

- (B) To discover some distinctive features which are typical of the lexical chains of Arabic or English.

## 2. Description of data and definition of terms

In order to test the validity of the two assumptions in Research Objective (A) above, a quantitative analysis of the chains of lexical repetition has been conducted in a pair of Arabic-English parallel texts. It is believed that since cohesion in general is realized by overt textual devices that are objectively detectable, any of its aspects – including lexical repetition – lends itself to quantitative analysis. (For a similar view, see Blum-Kulka (2000: 304).) However, it is also our conviction that a quantitative description based on frequency counts is not to be taken as an ultimate goal in itself. Rather, such a description is to be considered a useful stepping-stone for the discovery of textual ‘rules’ that, though not predictive, represent preferential trends and useful generalizations in the discourse of a specific text-type or genre. Consequently, and in order to cater for Research Objective (B) above, the lexical chains of repetition in both texts have been scrutinized in an attempt to detect any characteristic textual tendencies, other than those related to the overall frequency of recurrence and variation in the text as a whole.

A pair of parallel argumentative texts has been selected for analysis. Parallel texts are “original texts in two languages that are matched in terms of genre or text type” (Aijmer and Altenberg 1994: 13). Argumentative texts have been selected for investigation because such texts are known to encourage the use of repetition, since repetition is a universal ‘persuasive device’ in argument (Johnstone 1994:6). The Arabic text is an excerpt comprising the first 22 orthographic sentences from an article entitled “naḥwa tarbiya isla:miyya ṣa:liḥa lizama:nina:” (“Towards an Islamic education suitable for our age”) written by Mohammed Fadhil Al-Jamali.<sup>2</sup> The English text consists of the first 16 sentences from Chapter 1 of the first volume of a three-volume work entitled *Masters of political thought* by Michael B. Foster.<sup>3</sup>

Each of the two texts has been thoroughly examined for all the instances of lexical items which enter into repetition chains. The notion of ‘chain’ was first introduced by Halliday and Hasan (1976). A chain is said to be formed when “a cohesive

<sup>2</sup> See Appendix A for a copy of the Arabic text, as well as for its facts of publication.

<sup>3</sup> A copy of the text, together with its facts of publication, is reproduced in Appendix B.

element refers back to an element that is itself cohesive with a still earlier element, and so forth” (Hoey 1991: 14). The description of lexical repetition in terms of ‘chains’ seems to be quite appropriate since “lexical cohesion typically operates through lexical chains that run through a text and are linked to each other in various ways” (Baker 1992: 204). For the purposes of the present research paper, a ‘lexical repetition chain’ (henceforth, LRC) is consequently made up of any one lexical item which is used at least twice in a given text either intra- or inter-sententially. It is to be noted also that only open-set lexical items enter into these LRCs. Besides, each of the constituting forms of a given lexical item in an LRC is said here to represent an instance of either ‘simple’ or ‘complex’ lexical repetition, as defined below.

Hoey (1991: 53) defines ‘simple’ lexical repetition as an instance where “a lexical item that has already occurred in a text is repeated with no greater alteration than is entirely explicable in terms of a closed grammatical paradigm”. The above definition may be adequate enough for the purposes of a language like English, but since the Arabic morphological system is much more complicated, a more explicit definition may be needed. Consequently, I suggest the following as a working definition for simple lexical repetition in both Arabic and English:

Simple lexical repetition occurs when an open-set lexical item is repeated in a given text either without any formal changes at all or with a minimum change *which does not alter its word class*, viz. by adding or deleting at least one *inflectional morpheme*.

Arabic is a highly inflectional language in which inflectional morphemes can occur initially as prefixes in lexical items or finally as suffixes. Conversely, inflection plays a relatively minor role in English: the total number of inflectional morphemes is limited to seven, and all of them occur word-finally as suffixes. These are the plural, the possessive, the tense, the past- and present-participle, and the comparative or superlative morphemes (Hoey 1991: 33). Similar to Arabic, however, inflectional morphemes in English do *not* change the class of words they are added to. Consequently, instances of ‘simple’ lexical repetition are here seen as representative of ‘recurrence’ in the LRCs of both Arabic and English texts since the ‘same’ words are basically repeated. The following are examples taken from the two analyzed texts of some lexical items that are linked by ‘simple’ repetition:

Arabic	English
<i>alhadaf – alhadaf</i>	<i>actual – actual</i>
<i>al?isla:mi – isla:miyya</i>	<i>parts – part</i>
<i>almufakkir – almufakkri:n</i>	<i>point (n) – point (n)</i>
<i>alnazra – alnazratayn</i>	<i>tried – try</i>
<i>mustaqbal – almustaqbal</i>	<i>Aristotle – Aristotle</i>

Complex lexical repetition, on the other hand, is said to occur when “two lexical items share a lexical morpheme, but are not formally identical [...], or when they are formally identical, but have different grammatical functions” (Hoey 1991: 55). This definition by Hoey is found adequate as a working definition for the purposes of the present study. It may be necessary, though, to point out that it is derivational morphemes that are involved in complex lexical repetition. Such morphemes are usually responsible for creating new words. Such ‘new’ words are still, however, related, since they are considered to be derivatives of the same lexical item due to the fact that they share either the same ‘root’ as in Arabic or the same ‘base’ in the case of English. Instances of ‘complex’ lexical repetition, therefore, represent ‘variation’ in the LRCs of Arabic and English since such instances involve the use of ‘new’, though morphologically related, words. In English, “derivation involves affixation, ablaut and compounding, whereas Arabic derivation involves affixation, ablaut and root-and-pattern formation” (Khalil 1996: 60). Noun-, verb-, adjective-, and adverb-forming suffixes are among the most productive derivational morphemes in English. Arabic, on the other hand, uses not only suffixes as derivational morphemes but prefixes and infixes as well. Since both Arabic and English employ many derivational processes in which a wide variety of derivational morphemes are involved, it may be more economical to identify such morphemes by a process of elimination rather than by listing all of them. It can, therefore, be simply said that derivational morphemes in Arabic and English are those that do not appear among the inflectional morphemes already listed under ‘simple’ lexical repetition. Consequently, any kind of formal change involved in the repetition of a given lexical item, other than those listed earlier for simple repetition, makes that repeated item belong to ‘complex’ lexical repetition. The following pairs of words found in the Arabic and English data are thus examples of lexical variation, i.e. items linked by the relation of ‘complex’ repetition:

Arabic	English
<i>istaqbal</i> – <i>almustaqbal</i>	<i>selected</i> – <i>selection</i>
<i>9a:mma</i> – <i>ya9umm</i>	<i>theorists</i> – <i>theory</i>
<i>altarbiya</i> – <i>altarbawiy</i>	<i>importance</i> – <i>important</i>
<i>9ulama:?</i> – <i>al9il</i>	<i>correct</i> (v) – <i>correct</i> (adj)
<i>alwa9an</i> – <i>muwa:9ini:n</i>	<i>students</i> – <i>study</i>

After having arrived at working definitions for simple and complex lexical repetition, each single instance of these repetitions in the LRCs can now be identified and classified. Before embarking on this, however, one important question remains to be settled first: Are both inter- and intra-sentential instances of lexical repetition to be taken into consideration and counted, or only the former? This has been a bone of contention among text-linguists. Following Halliday and Hasan (1976), Hoey (1991)

as well as many others is of the opinion that the only significant type of cohesion is the inter-sentential. On the other hand, many other text-linguists take issue with this and maintain that sentence boundaries have nothing to do with the role of cohesive devices, including lexical repetition, in creating texture (among others, Baker 1992; Korpimies 1978; Al-Batal 1985; Al-Jabr 1987). In the present study, both inter- and intra-sentential instances of lexical repetition are considered since I find the distinction quite arbitrary. Besides, the distinction is not even feasible at times, especially in the case of a language like Arabic where the notion of sentence boundaries is highly elusive; this is because the use of punctuation marks is not fixed and paragraph-long sentences are not uncommon in text (Al-Khafaji 2001).

Now, with the data of the study described, the concept of ‘chain’ in lexical repetition made clear, the definitions of simple and complex lexical repetitions specified, and the question of inter- or intra-sentential repetition settled, the crucial stage of data analysis can be embarked on.

### 3. Data analysis and interpretation

#### 3.1. Frequency of recurrence and variation

On the basis of the working definitions arrived at in Section 2 above for simple and complex lexical definitions, the research data that consists of the two Arabic-English parallel texts described earlier is analyzed. The objective is to identify all the lexical repetition chains in each text as well as to classify each instance of lexical repetition into either ‘simple’ or ‘complex’. Instances of simple lexical repetition, as was argued above, would represent ‘recurrence’ whereas ‘variation’ is realized by complex repetition. The full results of the total number of lexical repetitions in each text, as well as the number and percentage of each of the two types of repetition, are reported below.

Before reporting the results of analysis, however, a brief note on the analytical procedure may be in order. In practice, analysis of the data means that each ‘lexical’ word, viz., ‘content’ word, as opposed to ‘grammatical’ or ‘structural’ word, is to be checked against all the other words which appear subsequent to it in the text. This has to be done starting with the first word in the first sentence of each text. When a minimum of one subsequent word is found to enter into a relation of lexical repetition with a previous word, either of the simple or complex repetition type, a lexical chain has then been detected and noted down. It may also be worth pointing out that the decision to classify a given repetition as ‘simple’ or ‘complex’ depends on the type of morphological or grammatical relation, as was explained above, which links that word with the word *immediately preceding* it in the lexical chain which they share together. A comprehensive table is then worked out for each of the two analyzed texts, as can be seen from Table 1 and Table 2 below. These two compre-



hensive tables list, as well as classify, all the lexical chains found in the respective text. Each of the two tables is basically made up of two columns. In the first column all the lexical chains appear, each represented by the initial word in that chain. The figure, preceded by the letter 'S', which appears directly after the headword in every chain, indicates the specific sentence in which that lexical item occurs in the analyzed text. For the sake of economy, the forms of the other members of a given chain do not appear in the tables, though the sentence numbers in which they occur are indicated in the second column of each of these tables. The second column in each of the two comprehensive tables below is thus taxonomy for *each and every* lexical repetition.<sup>4</sup>

Table 1. Lexical Repetition Chains of the Arabic text

Serial No.	Lexical Chain (represented by initial word) and number of Sentence it occurs in	Type of lexical repetition	
		Simple	Complex
1	<i>istaqbala (S1); ...</i>	(S12)	(S2)
2	<i>al9a:lam (S1); ...</i>	(SS 2; 10)	-
3	<i>al?isla:mi (S1); ...</i>	(SS 2; 2; 4; 8; 8; 8; 9; 10; 10; 14; 15; 20)	-
4	<i>khallafa (S1); ...</i>	-	(S1)
5	<i>masha:kil (S1); ...</i>	(S9)	-
6	<i>akhla:qiyya (S1); ...</i>	(S17)	-
7	<i>yaqza (S2); ...</i>	(S2)	-
8	<i>9a:mma (S2); ...</i>	-	(S10)
9	<i>tahqi:q (S2); ...</i>	(S3)	-
10	<i>altarbawiyya (S3); ...</i>	(SS 9; 13; 15; 22)	(SS 10; 13; 19)
11	<i>yaqifu (S4); ...</i>	-	(S4)
12	<i>almufakkir (S4); ...</i>	(S22)	(SS 9; 16; 18)
13	<i>mutasa:?ilan (S4); ...</i>	-	(S8)
14	<i>alhadaf (S4); ...</i>	(S7)	-
15	<i>sharqan (S6); ...</i>	-	(S13)
16	<i>gharban (S6); ...</i>	(SS 14; 15)	(S13)
17	<i>balad (S8); ...</i>	(SS 8; 14; 16)	-
18	<i>a:ra:? (S8); ...</i>	(S13)	(S10)
19	<i>9ulama:? (S11); ...</i>	-	(SS 14; 22)
20	<i>altaqli:diyya (S10); ...</i>	(S13)	-
21	<i>bimuhtawa:ha: (S10); ...</i>	(S13)	-
22	<i>asa:li:baha: (S10); ...</i>	(S13)	-
23	<i>alnuzum (S11); ...</i>	(SS13; 15; 16; 17; 20)	-

<sup>4</sup> It is to be noted that the initial lexical item in every chain is not classified into either 'simple' or 'complex' since it does not itself constitute an instance of 'repetition' for a preceding word in the text; being the first word in its chain.

Serial No.	Lexical Chain (represented by initial word) and number of Sentence it occurs in	Type of lexical repetition	
		Simple	Complex
24	<i>takhri:j (S11); ...</i>	(S20)	(S11)
25	<i>muwa:fini:n (S11); ...</i>	-	(S12)
26	<i>shaha:da:t (S11); ...</i>	(S18)	-
27	<i>alba:liya (S13); ...</i>	(S16)	-
28	<i>taşluh (S13); ...</i>	(S16)	(S15)
29	<i>lizama:nina: (S13); ...</i>	(S16)	(S16)
30	<i>igtiba:s (S13); ...</i>	(SS14; 15; 16; 17)	(SS15; 17)
31	<i>al9adi:d (S14); ...</i>	-	(SS18; 20)
32	<i>alnazratayn (S14); ...</i>	(S14)	-
33	<i>algha:lib (S14); ...</i>	(S 22)	-
34	<i>alduwal (S15); ...</i>	(SS 21; 22)	-
35	<i>9ina:yatihi (S17); ...</i>	(S 17)	(SS 19; 19)
36	<i>ka:fiya (S17); ...</i>	(S19)	-
37	<i>ya9mal (S18); ...</i>	-	(SS 20; 22)
38	<i>alta:lib (S18); ...</i>	-	(SS 18; 20)
39	<i>alhaya: (S19); ...</i>	(S20)	-

Table 2. Lexical Repetition Chains of the English text

Serial No.	Lexical Chain (represented by initial word) and number of Sentence it occurs in	Type of Lexical Repetition	
		Simple	Complex
1	<i>following (S1); ...</i>	(S12)	-
2	<i>reader (S1); ...</i>	(SS 7; 14)	(SS 10; 12; 14; 16)
3	<i>actual (S1); ...</i>	(S4)	-
4	<i>writings (S1); ...</i>	(S10)	(SS 4; 10; 12)
5	<i>political (S1); ...</i>	(S2)	(S9)
6	<i>theorists (S1); ...</i>	-	(S2)
7	<i>selected (S1); ...</i>	(SS 6; 6)	(S8)
8	<i>parts (S1); ...</i>	(SS 7; 8; 9)	-
9	<i>author's (S1); ...</i>	(SS 4; 8)	-
10	<i>historical (S1); ...</i>	(S2)	(S2)
11	<i>introductory (S1); ...</i>	(S11)	(S14)
12	<i>comments (S1); ...</i>	(SS 5; 11)	(SS 3; 6; 7)
13	<i>understanding (S1); ...</i>	(S12)	-
14	<i>importance (S1); ...</i>	(SS 8; 12)	(S6)
15	<i>quoted (S1); ...</i>	-	(SS 2; 6)
16	<i>book (S2); ...</i>	(SS 4; 8)	-
17	<i>texts (S3); ...</i>	(S11)	-
18	<i>tried (S4); ...</i>	(SS 7; 14; 14)	-
19	<i>work (S4); ...</i>	(SS 4; 4; 8; 10; 10; 15; 16; 16)	-
20	<i>Aristotle (S4); ...</i>	(S9)	-
21	<i>Augustine (S4); ...</i>	(S9)	-

Serial No.	Lexical Chain (represented by initial word) and number of Sentence it occurs in	Type of lexical repetition	
		Simple	Complex
22	<i>student (S4); ...</i>	(SS 12; 14)	(SS 4; 16; 16)
23	<i>point (S6); ...</i>	(S6)	-
24	<i>say (S6); ...</i>	(S14)	-
25	<i>reasons (S8); ...</i>	(SS 8; 16)	(S14)
26	<i>attention (S8); ...</i>	(S12)	-
27	<i>remarks (S11); ...</i>	(S16)	-
28	<i>make (S11); ...</i>	(S13)	-
29	<i>assertion (S12); ...</i>	(S13)	-
30	<i>correct (S12); ...</i>	-	(S12)
31	<i>judgment (12); ...</i>	(S16)	-
32	<i>enlighten (S12); ...</i>	(S12)	(S14)
33	<i>proof (S13); ...</i>	(S13)	-
34	<i>critic (S15); ...</i>	(SS 16; 16)	-

The facts and figures in the above two tables represent the raw results of the data analysis for the two texts analyzed. Table 3 below presents a numerical summary of these results.

Table 3. Summary of the results of data analysis

Text	Type and frequency of lexical repetition	
	Simple repetition Total and percentage	Complex repetition Total and percentage
Arabic	55 = (65%)	29 = (35%)
English	51 = (68%)	24 = (32%)

As can be seen from the above table, the total number of instances of lexical repetition is 84 in the Arabic text and 75 in the English one. These frequency figures are not, however, comparable and do not have any statistical value in themselves since the two texts are not equal in terms of length. Conversely, the 'percentage' figures reported above for each type of lexical repetition are of significance since they are calculated separately for each text *relative to* the total number of instances of lexical repetition in that text. The percentage figures above show that instances of simple lexical repetition are almost twice as frequent in Arabic as complex ones. Hence, it becomes evident that the common view held and expressed by many text-linguists, viz., that Arabic texts tend to employ lexical recurrence abundantly, is statistically justified and well-founded. However, the above percentage figures also show that simple lexical repetitions are similarly predominant in the English text as well. This rather unexpected finding seems to cast doubt on the widely-held conviction, stated in Assumption 2 above, viz. that English texts are typically characterized by more

lexical variation than recurrence. The above percentages across the two texts analyzed show that both texts exhibit a very similar pattern of distribution, as far as the ratio of simple to complex lexical repetition is concerned.

It can therefore be primarily concluded from the above results of data analysis that Assumption 1 in Section 1.2 has been found to be valid while Assumption 2 has not. Many vital questions, however, remain unanswered concerning both assumptions. Why, for instance, is it so widely held then that Arabic texts are predominantly marked by lexical recurrence when more than one third of its repetition chains, as has just been shown, consist of complex repetition? Why, on the other hand, has English lexis in text been often described as typically marked with variation when it only has one third of complex repetition in its lexical chains and, similar to Arabic, about two thirds of recurrence as realized by simple repetition? Linguistically objective explanations must be sought for these questions. Can an interpretation lie elsewhere, other than in the frequency counts? Or, is the size of the data analyzed too limited to allow for conclusions like the above? But data can never be fully representative, no matter how extensive it is. Besides, statistical evidence is, at best, rarely sufficient to fully explain complex socio-cultural phenomena as found in the human language. Bearing this in mind, a closer inspection of the LRCs in the two texts analyzed has been conducted in order to find out what could have possibly given rise to both of Assumption 1 and Assumption 2 in 1.2 above. The objective then is to detect any language-specific probabilistic 'rules', in the sense of de Beaugrande (1980: 30), which are peculiar to Arabic or English. Reported below are some of the relevant findings in this respect.

### 3.2. Length of chains of lexical repetition

The length of LRCs in both texts has been examined as a potential source of difference that might have some explanatory value to the issue at hand. The total length of all lexical chains in each text is calculated in terms of the number of constituent words involved in making up these chains. The average length for all the lexical chains is then worked out for each text by dividing the total length of lexical chains by the number of these chains in every text. The results are as reported in Table 4 below.

Table 4. Average length of Lexical Repetition Chains

Lexical chains	Arabic	English
Total length of lexical chains	123 words	109 words
Total number of lexical chains	39 chains	34 chains
Average length of lexical chains	3.1 words	3.2 words



As can be seen from the figures in Table 4 above, the average length of LRCs in the two texts analyzed has been found to be almost the same. Consequently, it seems that no recourse can be made to the length of lexical chains for detecting any language-specific textual tendencies concerning recurrence or variation in the LRCs of Arabic or English. However, when we focus our attention specifically on the *long* LRCs in the two texts analyzed, we then discover a clear tendency for such chains to be longer in Arabic than they are in English. The average length of LRCs comprising four repetition words and above in both texts is calculated for this purpose, using the same method as in Table 4 above. Table 5 below reports the findings.

Table 5. Average length of long Lexical Repetition Chains

Chain length	Total no. of words		Number of chains		Average length	
	Arabic	English	Arabic	English	Arabic	English
4 words and above	51	53	8	10	6.3	5.3
5 words and above	39	33	5	5	7.8	6.6
6 words and above	34	28	4	4	8.5	7.0

A quick look at the above figures is enough to reveal two noteworthy phenomena concerning LRCs in Arabic and English. The first is that long LRCs tend, on average, to be longer in Arabic than in English. The second is that the longer the lexical chains become, the larger the difference will be between their average lengths in Arabic and English.

Finding that Arabic tends in general to have longer LRCs can provide a plausible, though perhaps neither conclusive nor exclusive, explanation for the above-mentioned assumptions that claim that Arabic favors 'recurrence' whereas English opts for variation. This can probably be better appreciated when we remember that Arabic does in fact use more recurrence as represented by simple repetition than variation in its lexical chains, as was shown in Table 3 earlier. Although the same table also shows that English as well employs the same proportion of simple lexical repetition, i.e. recurrence. However, the fact just revealed by Table Five above that Arabic texts usually use longer lexical chains would tend to give rise to the impression that 'recurrence' in Arabic is more frequent than it is in English. It is these long lexical chains of repetition which attract more attention and which linger more in memory. It is thus not a question of difference in the relative frequency of occurrence but of a difference in the length of certain lexical chains which can at least *partially* explain the basis for the 'valid' first assumption as well as for the 'unsubstantiated' second assumption, as stated in the Research Objectives above. Other differences between the LRCs of Arabic and English may also be involved, however.

### 3.3. The ratio of simple to complex repetition in long LRCs

Another phenomenon that has been discovered to be exclusively characteristic of long LRCs in Arabic is the exceptionally high proportion of simple lexical repetitions in such chains. It is to be remembered that the total numbers of simple and complex repetitions in the two texts analyzed, as was reported in Table 3, are 55 and 29 for the Arabic text and 51 and 24 for the English one. The overall simple-to-complex average ratio is thus roughly about 2:1 in the lexical chains of both Arabic and English. However, when the same ratio of simple-to-complex lexical repetition is calculated in the four longest lexical chains in the two texts analyzed, the results would be found to be markedly different. The four chains concerned in the Arabic text carry serial numbers 3, 10, 23, and 30 in Table 1 above. As for the English text, the four longest chains are 2, 12, 19, and 22 in Table 2. Table 6 below reports the number of instances of each of simple and complex repetitions, as well as the overall ratios, found in the above-mentioned lexical chains.

Table 6. Simple and complex lexical repetitions in longest LRCs

Arabic text			English Text		
Lexical chain (Serial No.)	Simple	Complex	Lexical chain (Serial No.)	Simple	Complex
3	12	—	2	2	4
10	4	3	12	2	3
23	5	—	19	8	—
30	4	2	22	2	3
Total	25	5	Total	14	10
Ratio	5	: 1	Ratio	1.4	: 1

As can be seen from the above table, the average ratio of the frequency of occurrence of simple-to-complex lexical repetition in the four longest chains of the Arabic text, viz. 5:1, is more than double the overall ratio for the Arabic text as a whole; which is about 2:1 as reported in Table 3 above. Conversely, the average ratio of simple-to-complex repetition reported in Table Six above for the longest chains in the English text is less than the overall ratio for the text as a whole. The situation, as far as the frequency of occurrence of simple repetitions is concerned, is thus drastically different in the long lexical chains of Arabic and English. The use of almost an equal blending of simple and complex lexical repetitions in these chains in an English text would help to mitigate the feeling of repetitiveness in the text as a whole. Conversely, however, the feeling of lexical recurrence is only promoted and highlighted in Arabic by the use of longer lexical chains that are unusually rife with instances of simple repetitions, where the same lexical forms are used over and over again.

In order to further ascertain this tendency for long LRCs in Arabic to be marked by more-than-double the average overall frequency of simple lexical repetition, a comparison with short LRCs may be helpful. Thus, the first ten short lexical chains in the Arabic text, each with two to three lexical items only, have been examined for their simple-to-complex repetition ratios. These specific chains appear in Table 1 with the following serial numbers: 1, 2, 4, 5, 6, 7, 8, 9, 11, and 13. For ease of reference, the frequency of occurrence of instances of both simple and complex repetitions in these chains is reproduced below in Table 7. For contrastive purposes, Table 7 also reports the frequency figures for the first ten short lexical chains in the English text, as already listed in Table 2.

Table 7. Simple and complex lexical repetitions in short LRCs

Arabic text			English text		
Lexical chain (Serial No.)	Simple	Complex	Lexical chain (Serial No.)	Simple	Complex
1	1	1	1	1	–
2	2	–	3	1	–
4	–	1	5	1	1
5	1	–	6	–	1
6	1	–	9	2	–
7	1	–	10	1	1
8	–	1	11	1	1
9	1	–	13	1	–
11	–	1	15	–	2
13	–	1	16	2	–
Total	7	5	Total	10	6
Ratio	1.4	: 1	Ratio	1.7	: 1

The above figures show even more clearly the big discrepancy in the ratio of simple-to-complex repetition between long RLCs in Arabic, as reported in Table 6 earlier, and the short ones, as just seen from Table 7. The ratio is 5:1 for the former whereas it is only 1.4:1 for the latter. It thus becomes more evident that it is specifically the distribution ratio of lexical recurrence and variation in the long lexical chains in Arabic, and neither in the short ones nor in the text as a whole, which seems to lend credence to Assumption 1 concerning the predominance of recurrence in the lexis of the Arabic text. On the other hand, Table 7 also shows that the ratio of simple-to-complex lexical repetition is 1.7:1 in the English short LRCs compared to a ratio of 1.4:1 in the long LRCs. This means that, contrary to the situation in Arabic, long lexical chains in English exhibit even less lexical recurrence than shorter ones do; this would help foster a feeling of more variation in the lexis of the English text as a whole, as claimed by the unsubstantiated Assumption 2 in the Research Objectives above.

### 3.4. Distance of lexical repetition

Another phenomenon worthy of further investigation as far as lexical repetition is concerned is the lexical distance which separates the members of a lexical chain. This distance is measured by the number of all words intervening between each lexical item in a chain, except the first word of course, and the one immediately preceding it in that chain.<sup>5</sup> The same four longest lexical chains in both the Arabic and English texts analyzed, as specified in Table 6 above, have also been examined for the phenomenon of the proximity of lexical repetition. In Lexical Chain 3 of the Arabic text, for example, the number of words intervening between the second and the first lexical items which make up this chain is 28, while the number is 15 between the third and the second lexical items, and so on. The total number of words intervening among all the members of this chain comes to 317. When this sum is divided by the total number of all the lexical items which make up Chain 3, viz. 13 as can be seen from Table 1, the average distance of lexical repetition for this chain turns out to be about 24 words. Below is a contrastive table of the average distance of lexical repetition for the longest four chains in each of the Arabic and English texts.

Table 8. Average lexical distance for longest LRCs

Arabic Text		English Text	
Lexical chain	Average distance	Lexical chain	Average distance
3	24	2	47
10	38	12	75
23	10	19	71
30	29	22	49
Overall average distance = about 25 words		Overall average distance = about 60 words	

It can thus be clearly seen from the above figures that, on average, long LRCs are more than twice as dense in Arabic as they are in English. In other words, the lexical 'nodes', viz. the words making up a lexical chain, are much more widely spread out from each other in English than they are in Arabic.<sup>6</sup>

To sum up, this high density of repetition in long Arabic LRCs, viz. repeating same or related words at shorter intervals, is yet another potential source for making

<sup>5</sup> For the purposes of the present study, the word is defined as an 'orthographic' unit, rather than grammatical or semantic for example. The visual entity of the orthographic word has been adopted here for delimiting the word boundary since measuring the 'distance' of lexical repetition has mainly to do with a 'visual', psycholinguistic dimension as well.

<sup>6</sup> This is also partly due to the morphologically composite structure of the Arabic words as orthographic units since Arabic is highly inflectional in comparison to English. Yet, whatever the underlying reason may be, it remains 'visually' true that the distance of intervals within LRCs in Arabic is shorter.



the Arabic lexis in text both look and sound highly repetitive and overwhelmingly dominated by 'recurrence', compared to English. The fact of the matter, however, as argued throughout the paper and as demonstrated by all the discussion and eight tables above, points to differences in the preferential ratios of distribution between simple and complex repetition in the lexical chains of Arabic and English texts. Lexical chains in both the Arabic and English texts have turned out to use almost the same overall relative ratio of 'recurrence' to 'variation'. But, long LRCs in Arabic are characterized by a higher lexical repetition density as well as by an above-average abundance of instances of simple lexical repetition. Besides, long LRCs in Arabic have also been found to have a marked tendency to be longer in Arabic texts than in English. It thus seems plausible to conclude that it is the combination of the above language-specific features of the long LRCs in Arabic and English texts which can help us to understand and explain the linguistic basis and background of both the 'substantiated' Assumption 1 and the 'unsubstantiated' Assumption 2 in the Research Objectives in 1.2 above.

#### 4. Conclusions and suggestions

I have tried, through the analysis of two parallel texts, to examine the frequency and distribution of simple and complex lexical repetition in order to test the validity of two commonly-held assumptions concerning variation and recurrence in Arabic and English texts. Although the research corpus is quite limited in size, certain conclusions can be drawn. These conclusions are to be considered only as tentative at this stage however. At best, they are to be taken to represent indicators, not of rules but of favorable tendencies, which are conventional in Arabic and English texts. If, however, similar findings are arrived at through the analysis of more texts of varied genres, such conclusions would begin to have more value as *probabilistic* rules of a higher explanatory power. In brief, the most significant conclusions arrived at in the present study concerning the lexical repetition chains of Arabic and English texts are:

- 4.1 The overall frequency of 'recurrence' in the lexical chains of Arabic, as represented by simple lexical repetition has been found to be markedly higher than that of 'variation', which is realized by complex lexical repetition. This finding lends credence to Assumption 1 in the Research Objectives; it also indicates a noteworthy trend in Arabic texts which can have significant implications to Arabic applied linguistics, especially in the fields of translation and language teaching.
- 4.2 Similar to Arabic, the lexical chains in the English text analyzed have also been found to be characterized by a frequency-of-occurrence ratio of 2:1 of 'recurrence' to 'variation' in lexical repetition. This finding is not in line

with Assumption 2 in the Research Objectives in that it runs counter to the claim commonly made about variation being peculiar to lexical repetition in English texts. The above claim is not supported by evidence drawn from the statistical analyses of the relative ratios, which have been counted and reported throughout the study.

- 4.3 The lexical chains in the analyzed Arabic and English texts have been further scrutinized in an attempt to discover any typical dominant trends which may distinguish lexical repetition in Arabic from that of English. Such language-specific distinctive features, if detected, might help shed light on the basis of the above-mentioned claims concerning recurrence and variation in the lexis of texts in the two languages. This close inspection of the lexical repetition chains in Arabic and English has led to the following findings which are believed to be relevant to the issues at hand:

- 4.3.1 Long lexical chains tend to be longer in Arabic than they are in English, although the average length of lexical chains in the text as a whole is roughly the same in the two texts analyzed.

- 4.3.2 Long lexical repetition chains in the Arabic text have been found to exhibit a markedly higher percentage of lexical recurrence than that found in the text overall average. Conversely, long chains in English have less-than-average percentage of the lexical recurrence found in the text as a whole, and consequently more of variation.

- 4.3.3 The distance of lexical repetition has been found to be much shorter in the long lexical chains of the Arabic text than it is in English. This would lead to a higher repetition density in Arabic texts.

The three above-mentioned characteristics of long lexical chains detected in the two texts analyzed are candidates for language-specific preferential tendencies in Arabic and English. It seems plausible to conclude, therefore, that it is these distinctive features of long lexical chains, rather than lexical repetition in the text as a whole, which are responsible for (a) the seemingly 'excessive' frequency of lexical recurrence in Arabic and (b) the presumed predominance of variation in English.

- 4.4 The phenomena cited above about the tendency to use longer lexical chains in Arabic texts in comparison to English, as well as for those chains to use instances of simple lexical repetition more frequently and to repeat them at shorter intervals, are worthy of further investigation and interpretation. Various linguistic, cognitive, and rhetorical questions need to be raised and

examined in this respect. Does the more frequent use of simple lexical repetition in long lexical chains in Arabic, for example, tend to beget even more of such use? Can this be psycholinguistically interpreted in terms of a chain effect in which the use of one simple lexical repetition triggers the use of yet another one? Or, is it because argumentative texts in Arabic tend to drive home their arguments by a strategy of 'presentation' (using Johnstone's terminology) which is based on repetition? Hence, long lexical chains come to serve both as carriers and promoters of the argument, on the one hand, and as by-products, on the other, when these chains are themselves motivated and sustained by that very argument.

4.5 The main general conclusion which can be drawn from all of the above discussion is that while repetition of lexical items seems to be a universal strategy in human language, the relative frequency, length, density, and distribution of its chains and their constituting elements are language-dependent. Thus, although both Arabic and English exhibit very similar overall proportional ratios in the frequency of lexical 'recurrence' and 'variation', some specific lexical chains in the texts of the two languages have been shown to favor different distributional patterns. It is these different favorable patterns of distribution which are believed to underlie the two assumptions in Section 1.2 concerning lexical repetition in Arabic and English.

## REFERENCES

- Aijmer, K. and B. Altenberg. 1996. "Introduction." In Aijmer, K., B. Altenberg and M. Johansson (eds.), *Languages in contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*. Lund: Lund University Press. 11-16.
- Al-Batal, M.M. 1985. The cohesive role of connectives in a modern expository Arabic Text. Unpublished Ph.D. dissertation, University of Michigan.
- Al-Jabr, A.M. 1987. Cohesion in text differentiation: A study of English and Arabic. Unpublished Ph.D. dissertation, University of Aston, Birmingham.
- Al-Jubouri, A.J.R. 1983. "The role of repetition in Arabic argumentative discourse." In Swales, J. and H. Mustafa (eds.). 99-117.
- Al-Khafaji, R. 2001. "Punctuation marks in original Arabic texts". *Zeitschrift für arabische Linguistik* 4. 7-24.
- Al-Khafaji, R. 2003. "Lexical cohesion in Arabic texts". *Orientalia Suecana*. Forthcoming.
- Baker, M. 1992. *In other words*. London: Routledge.
- Blum-Kulka, S. 2000. "Shifts of cohesion and coherence in translation." In Venuti, L. and M. Baker (eds.), *The translation studies reader*. London: Routledge. 298-313.
- Beaugrande, R. de. 1980. *Text, discourse, and process*. London: Longman.
- Beaugrande, R. de and W.U. Dressler. 1986. *Introduction to Text Linguistics*. London: Longman.
- Fareh, S.I. 1988. *Paragraph structure in Arabic and English expository discourse*. Unpublished Ph.D. dissertation, University of Kansas.
- Halliday, M.A.K. and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Hatim, B. and I. Mason. 1997. *The translator as communicator*. London: Routledge.
- Hoey, M. 1983. *On the surface of discourse*. London: George Allen/Unwin.
- Hoey, M. 1991. *Patterns of lexis in text*. Oxford: Oxford University Press.
- Hoey, M. 2001. *Textual interaction: An introduction to written discourse analysis*. London: Routledge.
- Johnstone, B. 1991. *Repetition in Arabic discourse: Paradigms, syntagms, and the ecology of language*. Amsterdam: John Benjamins.
- Johnstone, B. (ed.). 1994. *Repetition in discourse: Interdisciplinary perspectives*. New Jersey: Alex Publishing.
- Khalil, A.M. 1996. *A contrastive grammar of English and Arabic*. Jerusalem: Al-Isra' Press.
- Korpimies, L. 1978. "Some aspects of discourse and cohesion." In Östman, J.-O. (ed.), *Cohesion and semantics: Reports on text linguistics*. Åbo: Åbo Akademi. 145-163.
- Swales, J. and H. Mustafa (eds.). 1983. *English for specific purposes in the Arab world*. Birmingham: University of Aston Language Studies Unit.
- Williams, M.P. 1982. *A contrastive analysis of text cohesion and development in Arabic and English*. Unpublished M.A. thesis, University of Leeds.
- Williams, M.P. 1983. "A problem of cohesion." In Swales, J. and H. Mustafa (eds.). 118-128.



APPENDIX A  
THE ARABIC TEXT<sup>7</sup>

(1) استقبل العالم الإسلامي القرن الخامس عشر الهجري وهو متقل بما خلفته عهود التخلف والاستعمار من مشاكل سياسية واقتصادية وأخلاقية وعقائدية. (2) وفي الوقت نفسه نشاهد يقظة عامة تشمل كل أجزاء العالم الإسلامي تقريبا ترافقها عزيمة أكيدة للنهوض والحق بركب الحضارة العصرية وتحقيق المستقبل الأفضل على أسس إسلامية قويمه وبروح متحفزة حركية. (3) ان تحقيق ذلك يتوقف على اختياراتنا التربوية. (4) يقف المفكر الإسلامي في مجابهة هذا الموقف الجديد متسائلا في حيرة وتامل ما هو الهدف؟ (5) وفي أي اتجاه نسير؟ (6) شرقا أم غربا أم على الصراط المستقيم؟ (7) ما هي الوسيلة لبلوغ الهدف؟ (8) من الطبيعي أن تكون الأجوبة على هذه الأسئلة مختلفة من بلد إسلامي إلى آخر إذ هناك آراء ونظريات متضاربة تسيطر على هذا البلد الإسلامي أو ذاك، فمن تمسك بأهداب الشريعة الإسلامية وتعاليمها إلى العلمانية إلى الأحاد. (9) وهذا مما يزيد في بلبلة الفكر التربوي الإسلامي ويعقد مشاكله. (10) ففي العالم الإسلامي اليوم من يرى أن التربية الإسلامية بصيغتها التقليدية وكما ورثناها بمحتواها وأساليبها هي ما يجب أن نعمل. (11) وأن النظم الحديثة قد لا تضمن لنا تخريج علماء نقاة [ sic ] مواطنين شرفاء، انها تخرج أناسا يحملون شهادات تعطى بعد دراسات قد "تمزج السم بالدم". (12) وقد تكون وبالا على مستقبل الأمة والوطن. (13) وهناك من يقول: أن التربية التقليدية بأساليبها البالية وبمحتواها المتحجر لا تصلح لزماننا فهو يرى اقتباس النظم التربوية من الشرق الشيوعي أو الغرب اللبرالي بعجزها وبجرها. (14) والتعليم الرسمي في العديد من البلاد الإسلامية يتأرجح بين النظرتين وفي الغالب تطغى عليه النظرة الثانية نظرة الاقتباس من الغرب. (15) فمعظم الدول الإسلامية اقتبست نظمها التربوية من الغرب وفيما اقتبسته الصالح والضار. (16) فيه ما هو صالح لزماننا وفيه ما هو بال وسقيم باعتراف الباحثين والمفكرين في البلد الذي اقتبست منه النظم المستوردة. (17) فما يقتبس من الخارج قد لا يلائم احتياجاتنا وظروفنا ومما يؤاخذ عليه النظام المقتبس هو عدم عنايته كافية بالتكوين الديني والأخلاقي لناشئتنا. (18) كما أنه في التطبيق عندنا قد يعمل على "قولبة" الطالب ويحجر حريته وافكاره كما أنه قد يرهق الطالب بتعدد المواد النظرية الجافة المطلوب حفظها ببغائنا لغرض اجتياز الامتحان ونيل الشهادة. (19) كما أن التربية كما تمارس عندنا لا تعنى العناية الكافية بالاتصال بالحياة والواقع. (20) النتيجة هي أن الانظمة المعمول بها في العديد على الأقطار الإسلامية قد تخرج طلابا "انكاليين لم يهيأوا للخوض في معترك الحياة. (21) انهم عالة على الدولة ينتظرون من الدولة أن تطعمهم وتكفل معاشهم بأقل ما يمكن من الجهد والتعب من قبلهم. (22) فتربيتهم المدرسية في الغالب تفصل بين العلم والعمل وبين الفكر واليد.

APPENDIX B  
THE ENGLISH TEXT<sup>8</sup>

(1) What is attempted in the following volume is to present to the reader a series of actual excerpts from the writings of the greatest political theorists of the past; selected and arranged so as to show the mutual coherence of various parts of an author's thought and his historical relation to his predecessors or successors; and accompanied by introductory notes and intervening comments designed to assist the understanding of the meaning and importance of the doctrine quoted. (2) The book does not purport to be a history of political theory, with quotations interspersed to illustrate the history. (3) It is rather a collection of texts, to which I have endeavored to supply a commentary. (4) I have tried rather to render the work of Aristotle, Augustine, and the rest accessible to the student, than to write a book about them; and the main object of this work will have been achieved if it serves not as a substitute for a further study of the actual works of these authors, but as an incentive to undertake it. (5) Nor does the commentary make any pretension of being exhaustive. (6) Very often after a long passage has been quoted a single point only has been selected for comment; and sometimes this point has been selected not because it was the most important, but because it was one on which I had something to say. (7) I have not tried to cover all the ground, and shall have done my part if the reader is stimulated, by the samples which I have offered, to complete a commentary of his own. (8) The selection has been confined to a few authors, for reasons not only of space, or of limitations of my own knowledge (though either of these reasons would have been sufficient), but because it is part of the plan of the book to concentrate attention upon the most important works. (9) A knowledge of Plato's *Republic*, of Aristotle's *Politics*, of parts of Augustine's *City of God*, belongs to a general education. (10) The works of lesser writers, or the lesser works of these writers, are doubtless worth reading; but a man who is not a specialist may ignore them without reproach. (11) If the commentary is secondary to the text, still more so must be any introductory remarks which I make here. (12) In commending the writings which follow to the reader's attention, I will indeed stake my credit on the assertion that the study of them will correct the judgment and enlighten the understanding upon matters in which it is important to be enlightened and correct. (13) But if a proof of this assertion is demanded, there is no proof except that of asking the inquirer to make the experiment. (14) The introducer may suggest lines of reasoning, he may try to convey certain lights which he has himself derived from the study, but in doing this he must be tentative and not dogmatic, and in the last resort he must say to the reader, "Go and read for yourself, and try whether this is confirmed by your experience." (15) In this respect his position is like that of the critic of a work of art. (16) However useful the critic's remarks may be in preparing an approach to the work, they can never dispense the reader from the necessity of studying the work itself, nor deprive him of the right, on the basis of this study, of turning critic himself and standing in judgment on the reasonings by which he was led to it in the first place.

<sup>7</sup> This article is by M.F. Al-Jamali and was published in the Tunisian journal *Al-Fikr* 26, 4.

<sup>8</sup> The text is by M. B. Foster and it is quoted by Hoey (1991: 78-79).